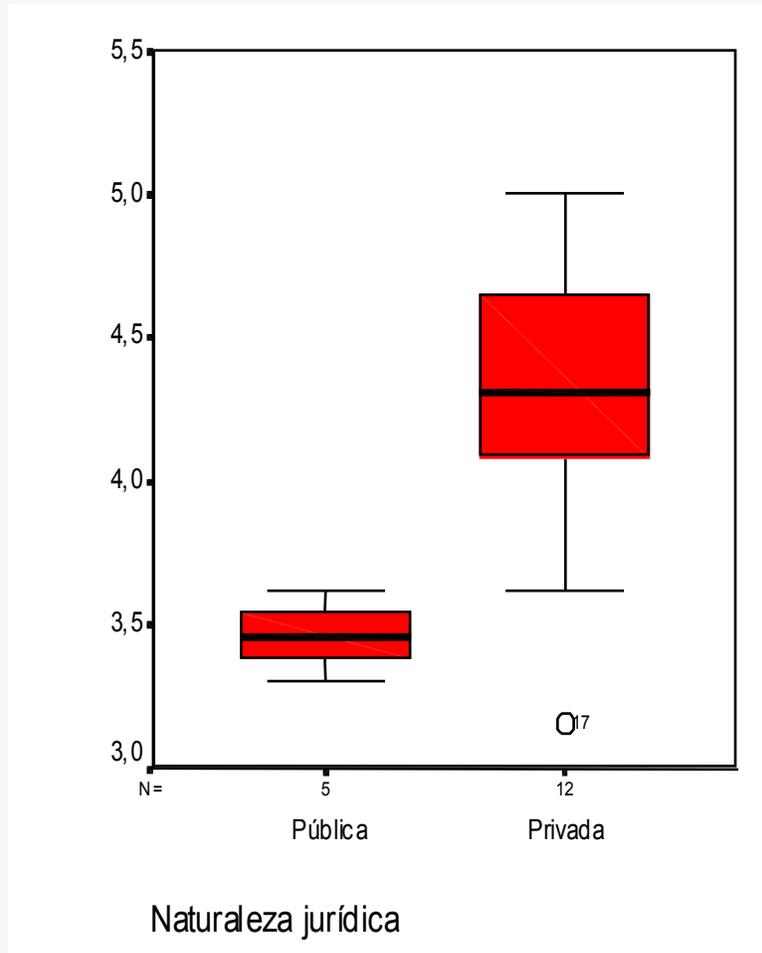


ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



Primera parte edición electrónica: Descriptiva y Distribuciones de Probabilidad.

León Darío Bello Parías

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Aviso Legal

Este texto no puede ser distribuido por ningún medio impreso, electrónico o cualquier otra forma sin permiso escrito del autor. Para adquirir esta primera parte sobre **Estadística descriptiva y distribuciones de probabilidad** debe visitar uno de los sitios que le menciono a continuación o <http://www.leondariobello.com/ebook-estadistica/>, allí encontrará la manera de realizar los pedidos, además, diversa información sobre temas de estadística y de mercadeo on line.

En caso de requerir alguna consulta puntual sobre los temas expuestos en el ebook, pueden expresarla escribiendo al correo ldbello753@gmail.com o solicitar una charla on line en mi sala de conferencias a la cual podrá acceder en la dirección www.gvowebcast.com/conference.leondario.

Tengo la convicción de que éste libro digital le será de gran ayuda para mejorar y fortalecer sus conocimientos en el área de la estadística y comprenderá adecuadamente los conceptos de la estadística descriptiva, probabilidades y distribuciones de probabilidad, además, de manejar con soltura diversos software estadísticos como: Excel, STATG, y SPSS para la parte descriptiva y Excel para los modelos probabilísticos como son: Poisson, Binominal, Normal, T, F, chi cuadrado.

Con la dedicación del caso, confío que las páginas que siguen le servirán para mejorar sus habilidades y conocimientos en la parte de estadística descriptiva, incluyendo parte del análisis exploratorio de datos y el manejo de las probabilidades con aplicación en el área de la salud (especificidad y

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

sensibilidad).

La obra “Estadística como apoyo a la investigación” en su versión digital, está compuesta por 6 ebooks, todos ellos con enlaces que nos llevan a objetos de aprendizaje (OA), los cuales nos facilitan como su nombre lo indica el aprendizaje, como muestra de uno de ellos, lo invito a que acceda al enlace <http://www.leondariobello.com/OA/Promedioaritmético/>

Con la compra de uno sólo de los ebooks referenciados en nuestros sitios web, tiene derecho a la membresía del club estadístico y marketing por un año a partir de la fecha de su primera compra, dicha membresía se valora en 100 dólares anuales, adquiriendo el derecho de participar en algunos seminarios cortos de reforzamiento en temas afines a nuestro interés, además, de información útil, propia y actualizada.

Adelante y muchos éxitos.



León Darío Bello Parías

www.leondariobello.com/e-books-estadistica

www.ciemonline.info/blog/libros-electronicos-estadistica

ldbello@leondariobello.com

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CONTENIDO

INTRODUCCIÓN

PRESENTACIÓN

Página

CAPÍTULO 1

CONCEPTOS BASICOS

1.1 ESTADÍSTICA	19
1.2. ESTADÍSTICA DESCRIPTIVA	20
1.3 ESTADÍSTICA INFERENCIAL	21
1.4 POBLACIÓN ESTADÍSTICA O UNIVERSO	21
1.5 MUESTRA	22
1.6 MUESTREO	22
1.7 CENSO	24
1.8 PARÁMETRO	24
1.9 ESTADIGRAFO	24
1.10 ESTADÍSTICA PARÁMETRICA	25
1.11 ESTADÍSTICA NO PARÁMETRICA	25
1.12 VARIABLES CUANTITATIVAS Y CUALITATIVAS	26
1.13 MEDICION	28
1.13.1 Dato	28
1.13.2 Elementos de la Medición	29
1.13.3 Escalas de Medición	32

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

1.13.3.1 Medida Nominal	32
1.13.3.2 Medida Ordinal	33
1.13.3.3 Medida de Escala	33
1.13.4 Fuentes de Datos	34
1.14 MEDIDAS DE FRECUENCIA EN EPIDEMIOLOGÍA	36
1.15 ALGUNAS APLICACIONES DE LA ESTADÍSTICA	39
Ejercicios capítulo	43

CAPÍTULO 2

LA INVESTIGACIÓN Y LA ESTADÍSTICA

2.1. ETAPA CONCEPTUAL	49
2.3 DISEÑO Y PLANEACIÓN	50
2.3 ETAPA EMPIRICA	54
2.4 ETAPA ANALITICA	56
2.5 DIFUSIÓN	57
Ejercicios	58

CAPÍTULO 3

PROCESAMIENTO DE DATOS

3.1 Presentación en forma de tablas o distribuciones de frecuencia	61
3.1.1 Partes de una tabla estadística	61
3.1.2 Tipos de Tablas Estadísticas según el número de variable	63

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

3.1.2.1 Tablas simples o univariadas	63
3.1.2.2. Tablas bivariadas o de doble entrada o de contingencia	63
3.1.2.3. Tablas compuestas o tipo banner	65
3.1.3 Tablas Estadísticas según el nivel de medición de las variables	68
3.1.3.1. Variables Nominales u Ordinales	68
3.1.3.2 Una variable cuantitativa con pocas categorías	71
3.1.3.3. Una variable cuantitativa continua	77
3.2 MUESTRA AGRUPADA	79
3.2.1 PROCEDIMIENTOS PARA AGRUPAR EN INTERVALOS	80
3.2.2 CONSTRUCCIÓN DE INTERVALOS CON G-stat	82
Ejercicios	85
3.3 CONSTRUCCIÓN DE TABLAS E HISTOGRAMA CON EXCEL 2010	88
3.4 REPRESENTACIÓN GRÁFICA	96
3.4.1 Gráficos simples	98
3.4.1.1 Gráfico de Barras	98
3.4.1.2 Gráfico Circular o de Sectores	99
3.4.1.3 Gráfico de Segmentos o Siluetas	100
3.4.1.4 Histograma de Frecuencias	102
3.4.1.5 Polígono de Frecuencia	102
3.4.1.6 Diagrama de tallo y hoja	103
3.4.1.7 Diagrama de Pareto	105
3.4.1.8 Diagrama de causa efecto o Espina de Pescado	106

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

3.4.2 GRAFICOS COMPUESTOS	108
3.4.2.1 Barras Comparativas	108
3.4.2.2 Pirámide poblacional	109
3.4.2.3 Diagrama de Dispersión o Nube De Puntos	111
3.4.2.4 Caja y Sesgo	112
3.5 IMPORTANCIA DE LA ESCALA	115
Ejercicios	117

CAPÍTULO 4

ESTADISTICA DESCRIPTIVA Con Excel y SPSS V.18

4.1 MEDIDAS DE TENDENCIA CENTRAL	122
4.1.1 Medida Aritmética	123
4.1.1.1 Interpretación Media Aritmética	127
4.1.1.2 Propiedades Útiles de la Media Aritmética	127
4.1.2 Mediana (Me)	129
4.1.3 Media Armónica (Ma)	132
4.1.4 Media Geométrica (Mg)	133
4.1.5. Promedio recortado	134
4.2 MEDIDAS DE POSICIÓN	135
4.2.1 Moda (Mo)	135
4.2.2 Cuartiles y Percentiles	137
4.3 MEDIDAS DE DISPERSIÓN O VARIACIÓN	138
4.3.1 Rango	139
4.3.2 Rango Intercuartilico	139
4.3.3 Varianza y Desviación Típica o Estándar	140

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

4.3.3.1 Propiedades de la Varianza	141
4.3.3.2 Interpretación	143
4.3.4 Coeficiente de Variación	144
4.3.5 ¿Qué dicen las medidas de dispersión?	145
4.4 MEDIDAS DE FORMA	145
4.4.1 Distribuciones Simétricas	146
4.4.2 Distribuciones Asimétricas	146
4.4.3 Curtosis	148
4.5 ORDENES ESTADÍSTICA DESCRIPTIVA CON EXCEL	149
4.6 ORDENES ESTADÍSTICA DESCRIPTIVA CON SPSS 19	155
Ejercicios capítulo	165

CAPÍTULO 5 INTRODUCCION A LAS PROBABILIDADES

5.1 Experimento Aleatorio	173
5.2 Espacio Muestral	174
5.3 Suceso o Evento	174
5.4 Suceso Contrario	175
5.5 Definición Clásica de Probabilidad	176
5.6 Definición Frecuencial o Empírica	179
5.7 Definición Subjetivista	179
5.8 Axiomas y Teoremas de Probabilidad	180
Ejercicios	186
5.9 TEOREMA DE PROBABILIDAD TOTAL	188
5.10 TEOREMA DE BAYES	192
5.11 APLICACIÓN DE LAS PROBABILIDADES A LA MEDICINA	196

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

5.12 TESTS DIAGNÓSTICOS USANDO BAYES	203
Ejercicios	205

CAPÍTULO 6

DISTRIBUCIONES DE PROBABILIDAD Con Excel

6.1 VARIABLES ALEATORIAS (v.a)	214
6.1.1 Variables Aleatorias Discretas	215
6.1.2 Función de Probabilidad Discreta	215
6.1.3 Función de Probabilidad Acumulada Discreta (F.D.A)	215
6.1.4 Valor Esperado o Esperanza Matemática Discreta	218
6.1.5 Desviación Estándar	222
Ejercicios	223
6.2 MODELOS PROBABILÍSTICOS	225
6.2.1 Modelo Binomial	226
6.2.2 Modelo Poisson	233
6.2.3 Modelo Normal	242
Ejercicios	254
6.2.4 Distribución T de Students	259
6.2.5 Distribución Chi cuadrada - X^2 -	268
6.2.6 Distribución F	276
Ejercicios Capítulo	281

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

INTRODUCCIÓN

El siglo XXI se conoce como el siglo del conocimiento y hoy luego de una década se habla de “Gestión del Conocimiento”, todo ello enmarcado dentro de un proceso investigativo que logre construir conocimiento en las diversas áreas del saber, por ello, en todas las profesiones y a cualquier nivel de formación, se ofrecen cursos de estadística, la situación está en identificar los procedimientos y métodos adecuados para lograr los objetivos que se buscan con el aprendizaje de ésta disciplina, que soporta gran parte del trabajo investigativo, desde la construcción de los instrumentos de recolección de datos, pasando por el diseño metodológico, procesamiento de datos, y la producción de información con fines de apoyar la toma de decisiones a todos los niveles.

En esta primera edición digital, y primera parte, se continúan realizando esfuerzos para convertir en fácil lo tradicionalmente complicado y mostrar las bondades de la estadística como apoyo a la investigación científica, haciendo uso para ello de programas de cómputo de fácil manejo como el Excel de uso cotidiano en muchos ambientes de estudio y trabajo, el Statg, de uso gratuito y que lo puede descargar desde el sitio www.leondariobello.com en la sesión programas estadísticos y el IBM SPSS, el cual servirá para introducirlo en el manejo de un software de reconocido valor mundial y supremamente amigable, lo que nos facilitará el camino para que en procedimientos más avanzados se tenga el manejo de este poderoso software, una versión libre por 30 días lo podrán obtener dando clic en el enlace siguiente: www.ibm-spss-statistics.sofnic.com

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

El libro electrónico en su primera parte, pretende como uno de sus objetivos crear cultura estadística, que no es más que entender y saber leer informes e investigaciones que utilicen técnicas estadísticas descriptivas, de otro lado, busca que los lectores adquieran las destrezas necesarias para manejar software estadístico y además se apropien de las competencias para interpretar las correspondientes salidas del computador.

Como un apoyo importante, se tiene el diseño de dos cursos colocados en la red Internet e implementados en la plataforma educativa Moodle, las direcciones en las cuales encontrará la posibilidad de reforzar lo leído en éste texto digital y en otros ebooks de temas más avanzados, accediendo a las presentaciones en power point y a diversos documentos y enlaces, son:

<http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=402>

<http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=322>

En los cursos, encontrará foros con opiniones y enlaces colocados por más de 100 participantes a lo largo de los últimos 3 años.

Adicionalmente, a lo largo del ebook encontrará enlaces que lo llevarán a Objetos de Aprendizaje(OA) que le ayudarán a mejorar sus competencias en los diversos temas tratados en el texto digital.

Otra situación que se privilegia es lograr una mayor comprensión de los temas, resaltando la parte interpretativa y de ayuda a la toma de decisiones, además, de presentar los temas que son realmente usuales en la

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

investigación y en la cotidianidad de las personas que les toca entregar informes numéricos en su trabajo diario. Por lo tanto, son usuarios de éste texto los estudiantes, profesore, trabajadores de la información y/o investigadores, teniendo mayor ejemplificación en las áreas sociales, salud y Administración.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

PRESENTACIÓN

Tomando la experiencia del autor acumulada en 25 años de docencia universitaria en diversas instituciones en el ámbito nacional (Colombia), actualmente docente de la Universidad de Antioquia y aunado lo anterior con el trabajo investigativo, originado por su carácter de asesor estadístico en múltiples trabajos y áreas del conocimiento, el presente texto busca cerrar la brecha presentada entre la enseñanza de la estadística y su aplicación en las diferentes áreas del conocimiento, de una manera comprensible y amigable para los lectores.

El objetivo entonces, es desarrollar los temas de la estadística descriptiva y los modelos probabilísticos en esta primera parte. Siguiendo metodologías validadas y que garantizan resultados eficientes. Se aborda el tema desde el análisis exploratorio de datos, de tal manera que los resultados obtenidos sean confiables. El usuario de este e book va, desde personas de diferentes pregrados, en los cuales la investigación es un soporte importante en el curriculum, hasta aspirantes a Magister, cuyo producto final, es generalmente una investigación científica, pasando por los especialistas. Se pretende dar mayor énfasis al concepto, a la interpretación que al manejo operativo, presentando, no obstante, las fórmulas con ejemplos, procurando entender y comprender cual es el sentido de la mismas. Claro está que mientras más alto es el nivel de formación, se requerirá complementar con los partes más avanzadas del texto, es decir, complementar con otros e books: parte 2, Muestreo e Inferencia Estadística Paramétrica, 3 Estadística No Paramétrica, 4 Modelos de Regresión, 5 Diseño de Experimentos y 6 Aplicación de Series de Tiempo en el Área de la Salud.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

SÍNTESIS DE LOS CAPITULOS.

Capítulo 1. CONCEPTOS BASICOS: La sensibilización, la actitud positiva y el saber para qué sirve algo, es la base de la primera disertación sobre: La Estadística como apoyo a la Investigación, valga decir, se presentan los conceptos básicos de la estadística, enmarcados dentro de cual es su utilidad en el manejo de un proyecto de investigación o en la labor cotidiana de entrega de informes periódicos, además, de enfatizar sobre las escalas de medición, siendo estas el soporte de cualquier trabajo estadístico, ya que de ahí se derivan los diferentes procedimientos que se puedan aplicar para apoyar la toma de decisiones.

Capítulo 2. LA INVESTIGACIÓN Y LA ESTADÍSTICA: Es uno de los pocos textos de estadística en el cual se presentan los pasos sugeridos para una investigación cuantitativa y se muestra el aporte de los procedimientos estadísticos en ella, es decir, se liga la labor del estadístico con el trabajo investigativo, es decir, se menciona cuál es el aporte de la estadística en cada uno de los pasos sugeridos para llevar a cabo una investigación científica, útil para cualquier área del conocimiento.

Capítulo 3. PROCESAMIENTO DE DATOS: Hace referencia a la importancia del procesamiento de datos y el cómo se deben de manejar estos para lograr resultados confiables. Es habitual hablar en términos de frecuencias absolutas o relativas cuando de tablas estadísticas se trata, en esta oportunidad, se pretende dar mayor relevancia a como se presenta la tabla de tal manera que la información presentada sea de fácil comprensión. Mostrando tablas y gráficos de estudios asesorados por el autor, logrando

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

con ello, detectar dificultades prácticas, además, dando mayor énfasis a la interpretación de las mismas con un lenguaje comprensible para cualquier usuario.

La temática se desarrolla diferenciando la presentación según las escalas de medición definidas anteriormente. Otro hecho a destacar, es que se colocan los gráficos mostrando casos prácticos y su verdadera aplicación en las investigaciones, por ejemplo, el gráfico de caja y sesgo, es muy dicente en el análisis exploratorio de datos de una variable cualitativa y una cuantitativa, mientras que el gráfico de Pareto, tiene su fuerza en la presentación de fortalezas y debilidades, además, de su aplicación en Control de Calidad. La inserción del diagrama de espina de pescado, es importante, ya que los problemas de investigación en gran medida se inician con tormentas de ideas que se pueden sintetizar en éste diagrama. Se muestra como se construye en Excel el histograma y la pirámide poblacional, ambas de uso frecuente en estudios descriptivos y epidemiológicos. Adicionalmente, se muestra la forma de construir los histogramas con el programa Gstat, el cual es de uso gratuito y se descarga del sitio www.leondariobello.com

Capítulo 4. MEDIDAS RESUMEN O DESCRIPTIVAS Con Excel, Gstat y SPSS: Definitivamente todo trabajo investigativo, tiene una parte descriptiva, la cual se hace dependiendo de la naturaleza y escala de medición de las variables objeto de estudio, para ello se requiere de las medidas resumen, las cuales son abordadas desde el punto de vista de la comprensión de los resultados y su interpretabilidad, por ello, se utiliza el Excel como herramienta de apoyo en los diferentes cálculos, sin detrimento,

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

de razonar sobre la parte de la formulación. Se complementa con el manejo del SPSS en la parte descriptiva, dado el auge y popularidad que adquiere este software, además, que servirá para sentar bases para manejar otros procedimientos avanzados analizados en los libros electrónicos siguientes, que podrá adquirir según sus necesidades.

El énfasis está dado pues, por el soporte de los programas de cómputo y la interpretación de los resultados, discutiendo la pertinencia de los resultados entregados por el computador. Una prueba de ello, es el video que puede encontrar en la dirección <http://www.youtube.com/watch?v=IKWxUySThpAy> que hace parte del OA sobre el promedio y desviación estándar, lo invito a que lo observe y opine sobre él.

Capítulo 5. INTRODUCCION A LAS PROBABILIDADES: Entrega las bases para la teoría de riesgos e incertidumbre, la cual es fundamental para la toma de decisiones, situación a la cual estamos abocados permanentemente. Se presentan los temas de una manera simple, sin ejemplarizar con juegos, que según la experiencia de participantes en mis múltiples cursos, tiende a confundir en lugar de aclarar conceptos, a parte de que no se ve una aplicación real. Se destaca la forma de presentar los temas, buscando sencillez y aplicabilidad. Se adicionan los temas atinentes a la medicina, como son, las pruebas diagnósticas (especificidad y sensibilidad). También y como es usual en los diferentes capítulos se tienen enlaces a OA y a videos complementarios. Entre al enlace www.leondariobello.com/OA/pruebasdiagnosticas y aprecie la estrategia de reforzamiento.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Capítulo 6. DISTRIBUCIONES DE PROBABILIDAD Con Excel: Luego de introducir el lenguaje probabilístico de una manera simple, se avanza con los modelos de probabilidad, relacionándolos con las aplicaciones detectadas en temas como tamaño de muestra e inferencia estadística. En todos los casos se presentan ejemplos de la cotidianidad. De nuevo aparece el soporte del programa Excel como ayuda a resolver la parte operativa, que en ningún caso será el objetivo relevante. Se adiciona un mayor número de ejemplos para las variables discretas como la binomial y Poisson, además de las continuas: Z, T, Chi-cuadrado y F.

Visite nuestros sitios www.leondariobello.com/e-books-estadistica o www.ciemonline.info/blog/libros-electronicos-estadistica y adquiera éste ebook, sin duda, le será de gran ayuda en el avance de su dominio de la estadística descriptiva y los modelos probabilísticos con un enfoque práctico y que realmente le sirva para mejorar en su labor cotidiana, minimamente, adquiere una cultura estadística adecuada a los tiempos modernos, donde la información abunda. Recuerde que nos puede contactar para dudas en vivo en la sala www.gvowebcast.com/conference.leondario con cita previa y desde cualquier parte del mundo.

Continúe avanzando en su formación profesional, eso le permitirá optimizar su tiempo y eficiencia en la entrega de informes cuantitativos.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



León Darío Bello Parias
Docente Universidad de Antioquia

<http://www.leondariobello.com/ebook-estadistica/>
www.ciemonline.info/blog/libros-electronicos-estadistica

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CAPITULO 1

Quien hace puede equivocarse, quien no hace ya está equivocado.

DANIEL KON

CONCEPTOS BÁSICOS

Para iniciar, es importante y necesario informar que se tiene apoyo conceptual (lecturas, presentaciones) en el curso “Estadística Descriptiva” colocado en la plataforma educativa Moodle en un servidor de la Universidad de Antioquia, Medellín, Colombia, el cual viene activo desde el año 2007 y elaborado por el autor de este e book, la dirección para acceder a él es:

<http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=402>. Los usuarios del libro, podrán interactuar con el autor en su sala de video conferencia www.gvowebcast.com/conference.leondario para expresar sus inquietudes y comentarios que consideren pertinentes, incluso para sugerir la inclusión o eliminación de temas.

1.1 ESTADÍSTICA.

Al hacer referencia a la palabra ESTADÍSTICA, la mayoría de las personas poco versadas en el tema, la asocian con una serie de datos o columnas de números; de sucesos pasados o de hechos históricos, como las llamadas comúnmente “Estadísticas deportivas”, o una lista de defunciones, nacimientos, causas de muerte en una región en un determinado tiempo, listados de precios de un sin número de artículos, los salarios netos en los

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

últimos años, el precio de las acciones de una determinada empresa etc.

La estadística ha sido definida por muchos autores de diferentes maneras, dependiendo del perfil y formación de quien la define, manteniendo eso sí el sentido de manejar datos. Lo anterior, se debe a que los métodos estadísticos se aplican en todas las áreas del conocimiento, valga decir, la Medicina, la Psicología, Sociología, Administración, Contabilidad, Economía, las Ingeniarías, las Humanidades y su área de desempeño. De una manera general, la estadística se puede definir como:

"Manejo adecuado de datos con el fin de producir información útil, de tal forma que se minimicen los riesgos en la toma de decisiones y se produzca o fortalezca el conocimiento."

La estadística se divide en dos grandes ramas según el nivel de generalización: Estadística descriptiva y Estadística Inferencial.

Según la forma de los datos, se clasifica en estadística Paramétrica y No Paramétrica.

1.2. ESTADÍSTICA DESCRIPTIVA.

El objetivo central de la estadística descriptiva es presentar información en forma conveniente, útil y comprensible; de tal forma que se puedan hacer comparaciones entre una o varias informaciones de un mismo suceso. Además, de describir el comportamiento de un fenómeno objeto de estudio, siempre con el objetivo de apoyar la toma de decisiones. Se puede afirmar, que la estadística descriptiva hace parte del trabajo preliminar que luego

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

complementa la Inferencia Estadística, aunque por si sola tiene una importancia capital.

1.3 ESTADÍSTICA INFERENCIAL.

Se ocupa de generalizar la información obtenida por medio de los métodos de la estadística descriptiva, más específicamente hace inferencias (estimaciones) acerca de poblaciones, a partir de muestras extraídas de estas. Eso sí, las muestras deben ser representativas de la población de la cual fueron extraídas.

Para una mejor comprensión es necesario aclarar algunos conceptos de la Estadística.

1.4 POBLACIÓN ESTADÍSTICA O UNIVERSO.

Son todas y cada una de las posibles observaciones o medidas que se estén considerando en una situación dada. De acuerdo a su tamaño, la población puede considerarse finita o infinita. Las observaciones en estudio pueden ser de cualquier naturaleza; por ejemplo; la estatura de todos los niños en un preescolar (finita), el puntaje obtenido en un test sobre estrés para todos los empleados del estado (finita), el total de vehículos que llegan a un peaje durante un mes (infinita), el número de computadores existentes en Colombia (finita), etc. Es útil aclarar que en estadística cuando la población es muy grande, se considera infinita.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

1.5 MUESTRA.

Es un subconjunto de la población estadística, es decir, es una parte de ella y por lo tanto tiene que poseer las mismas características de la población objeto de estudio.

Ejemplos: La estatura de 45 niños de un preescolar, el puntaje obtenido de 50 empleados del estado en la prueba sobre el estrés, los primeros 100 vehículos que llegan a un peaje durante un mes, el número de computadores vendidos por EPM en Medellín – Colombia - en el año 2010 etc.

1.6 MUESTREO.

Es claro que todos los pasos de una investigación son relevantes y deben tener la mayor rigurosidad posible, no obstante, mucho de este esfuerzo se puede perder si no se seleccionan las unidades de análisis adecuadas y de manera coherente con los objetivos del estudio. Muestreo se puede definir como:

"El método estadístico por medio del cual se definen los criterios y técnicas que deben orientar el proceso de recolección u obtención de datos con el fin de producir información útil."

Existen diversas técnicas de muestreo, tales como:

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

NO PROBABILISTICOS	PROBABILISTICOS
Conveniencia	Aleatorio Simple
Por Juicio u Opinión.	Estratificado
Cuotas o Prorratoe.	Conglomerados
Con Fines Especiales.	Sistemático en Fases.
Bola de Nieve	Conglomerados Monoetápico
	Conglomerados Bietápico
	Conglomerados Bietápico Estratificado

Lo relevante según mi experiencia, es no perder de vista "nunca" los objetivos de la investigación. Para refrendar este comentario, he encontrado que muchas personas piensan que es mejor el muestreo probabilístico debido a que dentro de sus características está el poder generalizar los resultados a la población objeto de estudio, sin embargo, parece que no se tiene en cuenta en muchos casos, que no siempre el interés es generalizar resultados, es más, hay casos donde lo que se requiere es obtener información de expertos y por lo tanto es mejor el muestreo no probabilístico, sin contar que en otras ocasiones no es posible usar un muestreo probabilístico.

El muestreo es una de las herramientas fundamentales de la estadística, por lo tanto, será el tema del segundo ebook de esta serie de libros electrónicos. Le sugerimos entrar a la dirección <http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=322>

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

donde podrá encontrar presentaciones y participaciones de usuarios sobre el tema de muestreo, esto, más la recomendación de leer el ebook "[Muestreo e Inferencia Estadística](#)", nos dará una mayor y mejor visión sobre los muestreos descriptivos.

1.7 CENSO.

Es el recuento que se hace de una población finita; es decir, debe ser numerable y contable. Es claro que si la población es grande, los costos y el tiempo serán mucho mayores que cuando se utiliza muestreo, no obstante, dado la naturaleza de la variable en estudio y de los objetivos de la investigación, es posible que en algunos casos se recomiende éste procedimiento. De cualquier manera, el censo no garantiza una confiabilidad en los resultados del 100%, precisamente por que existen otros factores que inciden en las diferentes mediciones en una investigación.

1.8 PARÁMETRO.

Son medidas que describen con un sólo valor las características de una población. Una población puede tener varios parámetros, por ejemplo, promedio, varianza poblacional, proporción, ejemplo, promedio de edad de todos los estudiantes de su universidad, la desviación o variabilidad de los salarios de todos los empleados de su organización, la proporción de personas que se matricularon en cursos virtuales en su ciudad.

1.9 ESTADIGRAFO.

Es el número que resulta de la manipulación de ciertos datos iniciales de

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

acuerdo con determinados procedimientos específicos, comúnmente, usamos un estadígrafo que se calcula a partir de una muestra para estimar el parámetro de una población. Por ejemplo, promedio de edad de una muestra de 400 estudiantes de pregrado de su universidad, desviación estándar del salario de una muestra de 50 empleados de su organización, la proporción de personas matriculadas en cursos virtuales en una muestra de 1000 estudiantes en su ciudad.

Como vé, La diferencia radica en que el parámetro se calcula con todos los datos de la población, mientras que el estadígrafo usa los datos de una muestra para estimar lo real.

1.10 ESTADÍSTICA PARAMÉTRICA.

Hace referencia a aquella que utiliza como su nombre lo indica parámetros, es decir, aquella que requiere para su análisis que los datos tengan una distribución de probabilidad determinada, de ahí, que es vital identificar la forma de los datos antes de definir que procedimiento seguir. Los métodos gráficos son importantes para determinar esta situación. No sobra comentar que los métodos presentados en las estadísticas iniciales en casi todas las carreras pertenecen a ésta categoría, no obstante, hay que estar atento a que sigue cuando los datos no cumplen este supuesto. El ebook "[Muestreo e Inferencia Estadística](#)", profundiza sobre los procedimientos más usuales de esta categoría.

1.11 ESTADÍSTICA NO PARAMÉTRICA.

Se utiliza precisamente cuando los datos no se ajustan a la forma de

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

distribuciones de probabilidad conocidas. Es muy utilizada, por esta misma razón, cuando los tamaños de muestra son pequeños. Es útil mencionar que algunas pruebas paramétricas tienen su análoga en la no paramétrica. Se caracteriza, por no requerir estimar parámetros, utiliza en lo general pruebas que utilizan rangos (diferencias entre las posiciones de las mediciones), por lo tanto, suelen ser más fáciles de desarrollar. En cualquier caso, los cálculos no serán el problema, ya que en lo general siempre se podrá acceder a software especializado.

Entre las pruebas no paramétricas se encuentran las siguientes: Prueba χ^2 de Pearson, Prueba de Cochran, Prueba de Cohen kappa, Prueba de Fisher, Prueba de Friedman, Prueba de Kendall, Prueba de Kolmogórov-Smirnov, Prueba de Kruskal-Wallis, Prueba de Mann-Whitney, prueba de Wilcoxon, Prueba de McNemar.

Todas ellas, están implementadas en diferentes software estadístico, en el ebook "[Estadística No Paramétrica](#)", podrá conocer su aplicación y en que casos se recomienda cada una de ellas. Recuerde todos los ebook los puede adquirir entrando a uno de los sitios: <http://www.leondariobello.com/ebook-estadistica/> 
www.ciemonline.info/blog/libros-electronicos-estadistica

1.12 VARIABLES CUANTITATIVAS Y CUALITATIVAS O ATRIBUTOS.

Los datos estadísticos pueden clasificarse en variables cuantitativas (números) y cualitativas (categorías). Al ser variables pueden tomar diferentes valores o diferentes categorías. Por ejemplo; la edad, el cociente intelectual, el número de alumnos por grupo en una institución de educación

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

superior en Colombia, los días de estancia en un Hospital, el nivel de colesterol basal de un grupo de personas, el precio del dólar en un periodo determinado de tiempo, el salario de un grupo de profesionales de la administración, el color de moda, los tipos de usuarios según el tipo de vinculación a una EPS, etc. La variable se diferencia de una constante, ya que esta última nunca puede cambiar de valor; por ejemplo $P=3.14115\dots$, $e=2.7182\dots$ (Base de los logaritmos naturales).

Las variables cuantitativas se dividen en variables **discretas y continuas**. Las variables discretas, en la práctica se pueden definir como aquellas que surgen de un proceso de conteo y por lo tanto toman valores enteros, por ejemplo, número de accidentes de tránsito por día en Envigado, número de nacimientos en un determinado día en el I.S.S, el número de pacientes con estrés, el número de empresas que solicitan créditos mensualmente, el número de clientes que son reportados como deudores morosos, etc. Nótese que todas ellas empiezan como número de..... No obstante, pueden existir variables discretas que no necesariamente sean valores enteros, eso sí, se podrán numerar los valores que toma la variable.

Una variable continua es aquella que puede tomar un número no contable de valores dentro de cualquier intervalo dado, surgen de un proceso de medición, tales como: el precio del dólar día a día, el precio de las acciones de una empresa cada día, la estatura de un grupo de personas, el tiempo que se demora en realizar una determinada actividad (operación, viaje, examen), el nivel de colesterol basal, el nivel de triglicéridos, etc.

Las variables cualitativas son atributos o categorías que pueden tomar los

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

sujetos u objetos en una investigación. Por ejemplo; Profesión, Estado civil, credo religioso, estrato, color, tamaño (grande, mediano y pequeño), tipo de institución, categoría de municipio, tipo de enfermedad, etc.

Los métodos estadísticos para analizar variables cuantitativas son distintos de los utilizados para analizar atributos o cualidades. Cuando los datos están en forma cuantitativa se puede obtener mayor información.

En cualquier caso se requiere obtener datos bien sean cualitativos o cuantitativos y para ello es necesario saber como se miden las variables, de ahí, la importancia de tener claridad sobre estos conceptos.

1.13 MEDICIÓN.

Es la asignación de números a las características objeto de estudio de acuerdo con ciertas reglas, también se conoce como el proceso sistemático de asignación de números a individuos, de forma que se representen características de ellos. Lo anterior aplica más para las variables cuantitativas. En las categóricas no se requiere tener estándares ni siempre se utilizan números.

1.13.1 Dato: Es un número o valoración que surge de un proceso de medición o de una categorización, puede ser cuantitativo o cualitativo.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejemplo:

Edad	Sexo	Salario	Opinión (1-5)
28	Femenino	1589354	4
35	Masculino	2159357	3

Cada número, categoría o valoración es un dato. Ahora, si cada fila corresponde a información de una persona o producto, toma el nombre de registro.

1.13.2 Elementos de la Medición: En la investigación se requiere entre otros aspectos que los datos obtenidos sean lo suficientemente confiables y válidos para garantizar resultados pertinentes y útiles. Los aspectos claves a tener en cuenta en la medición de variables son:

Conceptualización: Es una definición de la variable que procura uniformar criterios entre los investigadores, no todo el equipo investigativo tiene que estar de acuerdo, el objetivo es estandarizar el lenguaje para todas las personas involucradas en un estudio.

En un estudio sobre adolescentes, que se puede entender como “adolescente”, ¿jóvenes entre 10 y 15 años?, ¿entre 8 y 12?, ¿según lo definido en otros estudios? Hay que dar claridad sobre como se abordará para nuestro trabajo. Un caso que generalmente presento en mis charlas, es el caso de: Qué se entiende por mujer adulta?, las respuestas han sido tan diversas, que le transmito esta pregunta a usted apreciad@ lector. Definitivamente, las respuestas son tan diversas como inesperadas, por ello,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

es tan importante tomar una definición para todos. Para el caso que expongo, mujer adulta, puede ser mayor de 16 o 18 años, que tenga hijos, que sea casada, que tenga madurez mental?, en fin sería una serie de posibilidades que confirman la necesidad de definir conceptualmente las variables.

Es prudente mencionar que no todas las variables requieren de una definición conceptual, esto debido a que la variable es conocida universalmente de la misma manera o al menos para una gran mayoría. Tal es el caso de: Sexo, estrato social, edad, entre otras, a no ser que tenga un sentido diferente del común para un estudio en particular.

Operacionalización: Se trata de volver tangible la definición conceptual, es decir, como voy a obtener datos sobre la variable en estudio. En algunos casos una variable se puede operacionalizar de varias maneras, cuando esto ocurre, es recomendable seleccionar aquella que entregue mayor información, sea más precisa y tenga en cuenta el contexto en el cual se utilizará.

También se da el caso de que algunas variables son de fácil medición y por diversas fuentes, tal es el caso de la edad, la cual se puede valorar, con la fecha de nacimiento, por encuesta, preguntándola, pidiendo la cédula, la estatura, se hace generalmente por observación, pero también se puede utilizar el metro, y así múltiples casos. Sin embargo, hay muchas otras, y más en el caso de las ciencias sociales, donde no es nada fácil hacerlo, por ejemplo, ¿cómo se mide la ansiedad de una persona?, ¿cómo se mide la calidad de vida?. Ahí se tiene un reto importante y que denotará la habilidad

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

de los investigadores.

Otro aspecto importante a tener en cuenta es que la variable debe ser observable, medible y cuantificable.

Representatividad: Es el poder de generalización de resultados, hace referencia a que tanto se puede extrapolar las conclusiones encontradas, si es del caso, es decir, no siempre se requiere que la medición sea representativa, en algunos casos, sólo se requiere que refleje la realidad de la variable en un momento dado.

Dos conceptos adicionales de los elementos de la medición, pero que hacen referencia a la parte interna de los instrumentos, son:

Confiabilidad: Hace referencia a la consistencia de resultados en diferentes mediciones en varios momentos en el tiempo o en partes de un mismo cuestionario. No es coherente una mujer con 20 años y que tenga 8 hijos, para medir la consistencia interna de los instrumentos existen técnicas estadísticas, siendo la más popular el “Alfa de Conbrach”, el cual usa las correlaciones de los elementos de un cuestionario. También se utiliza el concepto de correlaciones en dos mitades en una misma muestra, es decir, los resultados obtenidos en una parte de la muestra, deberían ser similares a los resultados de la otra parte.

Validez: Se refiere a que se mida lo que realmente se quiere. Si se quiere medir el ingreso familiar, no es adecuado obtener datos sobre el salario formal del jefe de hogar por razones obvias. Parece trivial el tema de validez, no obstante, es común encontrar dificultades en los estudios, por no

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

tener claridad sobre lo que se pretende medir. Una de las maneras de evitar el problema, es realizar una prueba piloto, es decir, tomar información de una pequeña parte de la población y verificar si con esos datos se logra cumplir con los objetivos planteados en la investigación.

1.13.3 Escalas de Medición: Una debilidad sentida en los que se inician en la investigación y por ende en algunos procedimientos estadísticos es no tener claridad sobre las escalas de medición de las variables, lo que impide tener claridad sobre lo que se pretende medir. Estas se emplean fundamentalmente para definir el plan de análisis en una investigación, además, son de gran utilidad en la codificación de instrumentos de recolección de información. Se clasifican en: nominal, ordinal, de intervalo y de razón. Las escalas nominal y ordinal se utilizan exclusivamente en variables cualitativas o categóricas, mientras que las de intervalo y de razón se aplican a las variables cuantitativas. **El programa estadístico SPSS**, las identifica como variables de escala, lo que sugiere que se pueden clasificar en sólo tres categorías. Les recuerdo que en el curso de estadística descriptiva ubicado en:

<http://aprendeonline.udea.edu.co/lms/moodle/course/view.php?id=402> podrá obtener información complementaria sobre el tema.

1.13.3.1 Medida Nominal: Se aplica a las variables cualitativas o categóricas que no tienen un orden predeterminado, generalmente son las variables de identificación, valga decir, profesión, estado civil, sexo, tipo de institución, procedencia, tipo de crédito, ocupación, etc. La escala nominal es la forma más débil de medición, debido a que se puede calcular un

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

menor número de estimadores que al usar las otras. Como indicadores se utilizan las frecuencias (número de veces que se presenta la variable), porcentajes, tasas y razones, como gráficos, se usa el circular y el de barras preferentemente.

1.13.3.2 Medida Ordinal: A diferencia de la anterior, la variable tiene un orden preestablecido, por ejemplo, estrato, nivel de escolaridad, grado de quemadura, nivel de alcohol, grado de sordera, nivel de estrés, etc. La escala ordinal es una forma de medición más fuerte, ya que permite calcular un mayor número de estadígrafos que la escala nominal. Aparte de los indicadores utilizados en la escala nominal, se usan los porcentajes acumulados, valga decir, la mediana (50%).

1.13.3.3 Medida de Escala: Son variables cuantitativas, bien sean discretas o continuas, por lo tanto, se pueden realizar operaciones matemáticas. No obstante, existe una diferencia importante según el cero tenga significado de ausencia de la medida o no. Si el cero no indica ausencia de medida y además, no se permite hacer equivalencias, por ejemplo, la temperatura en cero grados no significa ausencia de calor, además, la diferencia entre 5 y 10 grados centígrados no equivale a la diferencia entre 25 y 30 grados centígrados, se conoce como escala de intervalo. Si por el contrario el cero significa ausencia de la medida, y se pueden realizar equivalencias entonces se habla de una escala de razón. Por ejemplo, cero revistas, significa que no hay revistas; la diferencia entre \$1500 a \$2.000 es la misma diferencia entre \$80.000 y \$80.500.

La razón de considerar las variables cuantitativas en una sola categoría es

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

que con ellas se pueden realizar operaciones matemáticas, teniendo en cuenta, eso sí, que la redacción de algunos valores calculados con fórmulas estadísticas dependen de la escala de medición. Dado la importancia de este tema en el desarrollo adecuado de una investigación, y las fallas reiteradas al definir las variables se invita a que se enlace al siguiente OA (objeto de aprendizaje) www.leondariobello.com/OA/escalasdemedicion

1.13.4 Fuentes de Datos: Son las que se utilizan para obtener información sobre aspectos a investigar, pueden ser primarias y secundarias.

Son datos **primarios** aquellos que los investigadores recogen con el fin de satisfacer las necesidades propias de su estudio. Por ello, se prevé un contacto directo con los objetos y/o sujetos de los cuales se obtiene información, los investigadores construyen los instrumentos de recolección de información, logrando obtener la información pertinente.

Para obtener los datos se utiliza la observación directa del fenómeno bajo estudio; en otros casos, se utiliza el cuestionario como estrategia de recolección de información.

Los datos **secundarios** son aquellos que fueron obtenidos por otras entidades y/o personas y por lo tanto, fueron diseñados para resolver problemas diferentes al estudiado, no obstante, pueden ser de gran utilidad. No es aventurado afirmar, que el análisis de los datos secundarios es un requisito para recopilar los datos primarios.

Algunas instituciones que proveen éste tipo de datos en Colombia son: el Dane, Banco de la República, los Ministerios, Secretarías de Salud,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Camacol, Fenalco entre muchas otras. Las universidades e instituciones tienen bases de datos, las cuales se puede acceder de diversas maneras, por parte de sus usuarios.

Recopilando algunos elementos tratados anteriormente y con el fin de dar una aplicación útil en los proyectos se sugiere el siguiente formato para concretar la direccionalidad de los estudios.

Variable	Definición Conceptual	Definición Operacional	Nivel de Medición	Fuente de Datos
Sexo		Observación	Nominal	Primaria
Edad	Tiempo transcurrido desde el nacimiento, en años cumplidos.	Fecha de nacimiento.	Escala.	Primaria.
Antigüedad	Tiempo transcurrido desde su vinculación a la empresa en meses.	Registro de la oficina de laborales.	Escala.	Secundaria
Estrato	Nivel socio económico	Se le consulta por medio de encuesta.	Ordinal.	Primaria.
Salario	Devengado semanal en su empresa	Se le consulta por medio de encuesta.	Escala.	Primaria.
Número de hijos		Encuesta	Escala	Primaria
Tipo de Sangre		Encuesta	Nominal	Primaria

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

1.14 Medidas de frecuencia en Epidemiología.

Si entendemos la epidemiología como una disciplina que se encarga de estudiar la frecuencia de las enfermedades en las poblaciones humanas, así como los factores que definen su expansión y gravedad, además, dentro de las ramas de la epidemiología, se destaca la de **epidemiología descriptiva**, esta describe los fenómenos epidemiológicos en tiempo, lugar y persona, cuantificando la frecuencia y distribución mediante medidas de incidencia, prevalencia y mortalidad. Por ello, es importante abordar el tema de la medición de la morbilidad y mortalidad, para ello, se utiliza las medidas de frecuencia. La construcción de estas medidas se realiza calculando razones, proporciones y tasas.

Estas medidas son importantes en epidemiología, en la medida que permiten realizar comparaciones entre distintas poblaciones o en la misma población a través del tiempo, evitando, la influencia del tamaño de la población y/o los cambios demográficos presentados por diversas causas y que no permitirían una adecuada comparabilidad.

La situación es fácil de entender, si en Medellín -Colombia-, se presentan 20 casos de la influenza (enfermedad respiratoria) H1N1 y en Bogotá los casos son 40, no se puede afirmar que en Bogotá el problema es el doble que en Medellín. Esto debido a que en la capital Bogotá el número de expuestos es mucho mayor.

Lo anterior llevó a construir medidas que no estuvieran influenciadas por el tamaño de la población en la que se realiza la medición. Las medidas de frecuencia más usadas en epidemiología se refieren a la medición de la

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

mortalidad o la morbilidad en una población.

Proporción es el cociente de dos variables, el numerador está contenido en el denominador, por lo tanto, siempre será un valor entre 0 y 1. Si se requiere presentar el indicador porcentualmente, se multiplica el valor de la proporción por 100.

La expresión general se puede colocar así: $a/(a+b)$

Los ejemplos son múltiples, la proporción de mujeres en edad de procrear.

Proporción: Número de mujeres en edad de procrear/ Número de mujeres

Proporción de hombres con cáncer de próstata

Proporción: Número de hombres con cáncer de próstata/ Número de hombres.

Dando un poco mayor de contexto, se puede presentar la siguiente situación: un municipio determinado, tiene una población de 25.000 habitantes y se presentan 1.500 casos con influenza, la proporción de personas con esta enfermedad es de $1.500/25.000 = 0.06$, que expresada en porcentaje es del 6 %.

Razón es el cociente de dos variables, el numerador y el denominador son independientes, por lo tanto, ninguno está contenido en el otro, sin embargo, el denominador se toma como referencia, como la unidad. Generalmente son magnitudes que expresan una relación entre dos

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

sucesos en una misma población, o un solo evento en dos poblaciones.

La expresión general es: a/b .

Si en la Facultad Nacional de Salud Pública de la Universidad de Antioquia, se estima que de 300 fumadores, 200 son hombres, la razón por sexo sería:

$$\text{Razón sexo} = 300/200 = 3/2 = 1.5.$$

Lo que quiere decir, que por cada mujer fumadora, existe 1.5 hombres con ese hábito; de otra manera, por cada dos mujeres fumadoras, se encuentran 3 hombres fumadores.

Otros ejemplos comunes en epidemiología:

Razón de masculinidad en Colombia según censo 2005

$$= \text{Total hombres} / \text{total mujeres} = 20336117 / 21132267 = 0.9633.$$

Razón de masculinidad en Colombia según censo 2005 y para el grupo de edad de 0 a 4 años = $2106179 / 2002682 = 1.0516$

Otra aplicación podría ser: En una institución de salud, se tienen 150 camas disponibles y tienen 50 enfermeras, la razón es:

$$\text{Razón camas} / \text{enfermera} = 150/50 = 3 \text{ cada enfermera puede atender 3 camas.}$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Tasa es una proporción cuyo principal referente es la variable tiempo, mide la magnitud del cambio de la variable que mide un evento por cada periodo de tiempo. Las tasas se expresan multiplicando el resultado obtenido por una potencia de 10, con el fin de permitir rápidamente su comparación con otras tasas, el operador multiplicador, busca facilitar la interpretación de la tasa.

Las tasas brutas son calculadas para toda la población. Las tasas específicas se calculan para un subgrupo específico que está en riesgo de presentar el evento. Las más usuales se dan por edad, sexo, raza y ocupación.

Ejemplo: tasa general de fecundidad relaciona el número de nacimientos con el de mujeres de 15 a 44 años de edad (edad fértil), dando una visión de cuantos niños están naciendo por cada mil mujeres capaces de procrear

Tasa General de Mortalidad por una causa= N° de muertes por una causa, en un lugar y tiempo determinado $\times 1000 /$ Población promedio en ese lugar y en ese tiempo

Tasas de Mortalidad General por sexo Turbo año 2000.

Tasa mortalidad hombres por 1000 habitantes = $(311/45400) \times 1000 = 6.85$ por cada mil hombres.

1.15 ALGUNAS APLICACIONES DE LA ESTADÍSTICA.

Uno de los objetivos de la estadística es suministrar información en forma

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

útil y comprensible con el fin de poder observar las particularidades más importantes y sobresalientes de una serie de datos, es decir, describir hechos y situaciones de la realidad. Sin embargo, el objetivo principal de la estadística es hacer predicciones y estimaciones acerca de una población, sobre la base de una información contenida en una muestra, es el soporte para apoyar diversos departamentos en las empresas, por ejemplo:

DEPARTAMENTO DE CONTROL DE CALIDAD (PRODUCCIÓN).

Por medio de cartas estadísticas de control se vigila y mejora la calidad, tanto de los productos como de los servicios, es una manera de control procesos y de identificar sobresaltos en los mismos, con el fin de prevenirlos en el futuro. Es un procedimiento, como su nombre lo indica, de control, no de prevención. Existen diversos gráficos para controlar la precisión y la variabilidad de los procesos, se llaman, cartas \bar{X} media, S^2 de varianzas, de rangos, para prevenir defectuosos y defectos.

DEPARTAMENTO DE MERCADOS. E book adicional.

Tanto la diversificación de los productos como el aumento de la competencia, lleva a las empresas a estar enteradas constantemente de las necesidades y preferencias de los consumidores; para poder satisfacer sus necesidades y permanecer en el mercado. La estadística nos dice como hacer los diseños para capturar información y su posterior análisis. Es vital en los estudios de posicionamiento e imagen, así como para medir percepciones.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

DEPARTAMENTO DE PERSONAL

Es necesario tener un control sobre el número de horas trabajadas, accidentes de trabajo, rendimiento de operarios, reprocesos, información de quejas y reclamos, etc. La estadística nos ayuda a llevar dicho control y a identificar los aspectos más relevantes de las variables manejadas en dicho departamento con el fin de tomar acciones de mejoramiento.

DEPARTAMENTO DE COMPRAS

La estadística se utiliza para conocer la calidad y cumplimientos de los proveedores y/o usuarios de un servicio o producto. Sirve además, para hacer estudios de inventarios con el fin de optimizar los pedidos, para pronosticar y estimar ventas futuras.

DEPARTAMENTO DE CONTABILIDAD Y FINANZAS

Con el fin de realizar pronósticos de presupuestos, gastos e inversión. Las series de tiempo tienen una importancia relevante en estos temas, así como los procedimientos de regresión que se analizarán más adelante.

LA ESTADÍSTICA Y LA SALUD PÚBLICA

El hecho de que la estadística o bioestadística manejen las teorías sobre grandes números, variabilidad y la incertidumbre, aporta suficientemente a la Salud Pública, ya que se estudia el comportamiento de las comunidades y poblaciones con el fin de diagnosticarlas y con ello, definir e implementar políticas públicas en bien de la sociedad.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

El salubrista, requiere de sólidos conocimientos en bioestadística para participar en investigaciones que aporten a la solución de problemas de salud pública y que les permita entender el manejo de los datos sanitarios en un país.

Con respecto a las áreas de aplicación de la estadística en salud pública, en la dirección <http://www.respyn.uanl.mx/iv/1/ensayos/bioestadistica.html>, se encuentra el boletín “EL VALOR DE LA ESTADÍSTICA PARA LA SALUD PÚBLICA”, dirigido por Pedro César Cantú y Luis Gerardo Gómez Gúzman, de la Universidad Autónoma de Nuevo León, se plantean las siguientes: estudios de cambios en características objetos de estudio, se dan en estudios de impacto. Diagnóstico de enfermedades y de la salud de la comunidad, predicción del resultado probable de un programa de intervención, elección apropiada de intervención en paciente o comunidad, administración sanitaria y planificación, realización y análisis en la investigación en salud pública.

Con mi experiencia de 10 años en la Facultad Nacional de Salud Pública de la Universidad de Antioquia, donde se ofrecen las Maestrías y Doctorados en Salud Pública y Epidemiología, no queda duda, que los procedimientos estadísticos son necesarios en los trabajos investigativos y por ello, se abordarán algunos temas usuales en la solución de problemas de la Salud Pública.

INVESTIGACION CASO APARTE.

Es válido pensar que en los departamentos antes mencionados y en los demás que poseen las organizaciones se realizan investigaciones, no

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

obstante, la investigación científica está más arraigada en la comunidad educativa, que de alguna manera es la que impulsa el trabajo investigativo.

Por lo tanto, los procedimientos estadísticos son el soporte de gran parte de las investigaciones realizadas en cualquier área del conocimiento, valga decir, Salud, Ciencias Naturales, Ciencias Sociales, Humanidades, Áreas administrativas, Contables, Financieras, Ciencias del Comportamiento etc. Una prueba de lo anterior se sustenta en que los medios encargados de divulgar las investigaciones de los diferentes frentes del conocimiento, presentan artículos en los cuales sobresalen los métodos estadísticos utilizados. Se va desde la descripción inicial de las variables objeto de estudio, hasta la validación de hipótesis planteadas por los investigadores. Luego de validar el supuesto de normalidad (forma acampanada de los datos), se selecciona el procedimiento apropiado (Paramétrica o no Paramétrica).

Ejercicios

1. ¿Cuál es la diferencia fundamental entre estadística descriptiva e inferencial?
2. ¿Cuál es la diferencia fundamental entre estadística Paramétrica y la no Paramétrica?
3. ¿Cuál es la diferencia fundamental entre las variables cualitativas y las cuantitativas?, de tres ejemplos de cada una de ellas.
4. ¿Cuál es la diferencia fundamental entre un indicador de tasa y otro de razón?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

5. ¿Cuál es la diferencia fundamental entre un indicador de proporción y otro de razón?. Mencione un ejemplo de cada caso.
6. ¿Cuál es la diferencia fundamental entre la escala nominal y la ordinal?. Cite dos casos de cada escala.
7. Cite cinco ejemplos de variables cualitativas y cinco cuantitativas relacionadas con su que hacer diario, defínalas tanto conceptual como operacionalmente. Diga que nivel de medición tiene cada una de ellas.
- ¿Qué fuente de datos usaría para obtener información de las variables definidas en el punto anterior?.
 - ¿Qué escala (s) de medición le corresponde a las variables cualitativas y cual a las cuantitativas, además, que indicadores (estadísticos) le puede calcular (investigue al respecto).?
8. En la solicitud de ingreso a la Universidad, le solicitan la siguiente información:
- Nombre, edad, estatura, sexo, estrato, Estado civil, N° hijos, N° de hermanos, tipo de sangre, ingresos.
- Clasifique las variables anteriores, utilizando el formato propuesto en el apartado 1.13.4.
9. Con el fin de realizar un estudio sobre enfermedades profesionales, se diseño un instrumento donde se solicita entre otros aspectos los siguientes:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Variable	Definición Conceptual	Definición Operacional	Nivel de Medición	Fuente de Datos
Nombre				
Edad				
Estatura				
Sexo				
Nº consultas				
Estado civil				
EPS				
Diagnóstico				
Valor del Copago				
Valor del servicio				

Complete el cuadro. Justifique su respuesta para las dos últimas columnas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CAPITULO 2

"Más vale conocer cosas inútiles que no saber nada" L.A. Seneca.

LA INVESTIGACIÓN Y LA ESTADÍSTICA

Los procedimientos estadísticos surgen como una necesidad de aumentar la capacidad explicativa de los investigadores, máxime que estos se enfrentan al problema de manejar en la mayoría de casos, grandes cantidades de datos sin organizar, no obstante, aún, en aquellos casos donde la producción de datos no es alta, la estadística aporta importantes procedimientos para direccionar adecuadamente las investigaciones de corte cuantitativo fundamentalmente. Por esta razón, se inserta este capítulo referente a los pasos sugeridos a seguir en una investigación cuantitativa, haciendo énfasis en el aporte de la estadística en dicho proceso investigativo. Una investigación científica debe ser llevada a cabo de una manera rigurosa y sistemática, no obstante, si bien se tiene una estructura predeterminada, ésta puede tener variaciones dependiendo del perfil del investigador principal. El siguiente enlace nos lleva a otro OA www.leondariobello.com/OA/Procesoinvestigativo que nos permite ilustrar de otra forma una de las maneras de abordar el proceso investigativo, lo cual redundará en mejorar nuestra producción académica y laboral.

Se observa como el proceso empieza con una idea, originada generalmente por el deseo de mejorar algo y/o de solucionar algún problema, esa idea se tendrá que sustentar en razones y argumentos válidos para desarrollarla,

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

como estrategia se redacta un número amplio de preguntas que luego de una segunda lectura se irán consolidando; de ahí, se identifica el tema de investigación, para continuar con la formulación del problema, es decir, organizar las preguntas de investigación de tal manera que sustenten el porqué de la utilidad y necesidad del proyecto investigativo. De este razonamiento deben salir los objetivos, con los cuales se puede definir el marco teórico y con él seleccionar el diseño de la investigación, que como se verá a continuación, es un punto clave para lograr calidad en el estudio y por lo tanto credibilidad, en este punto la estadística aporta herramientas vitales para garantizar la realización de un buen estudio. A partir de ahí, se recogen los datos, se procesan y se ejecuta el plan de análisis propuesto para cumplir con los objetivos.

Es útil comentar que las etapas de una investigación científica, toma diferentes nombres según el enfoque que se quiera dar, sin embargo, existen unas etapas generales que se pueden sintetizar de la siguiente manera:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

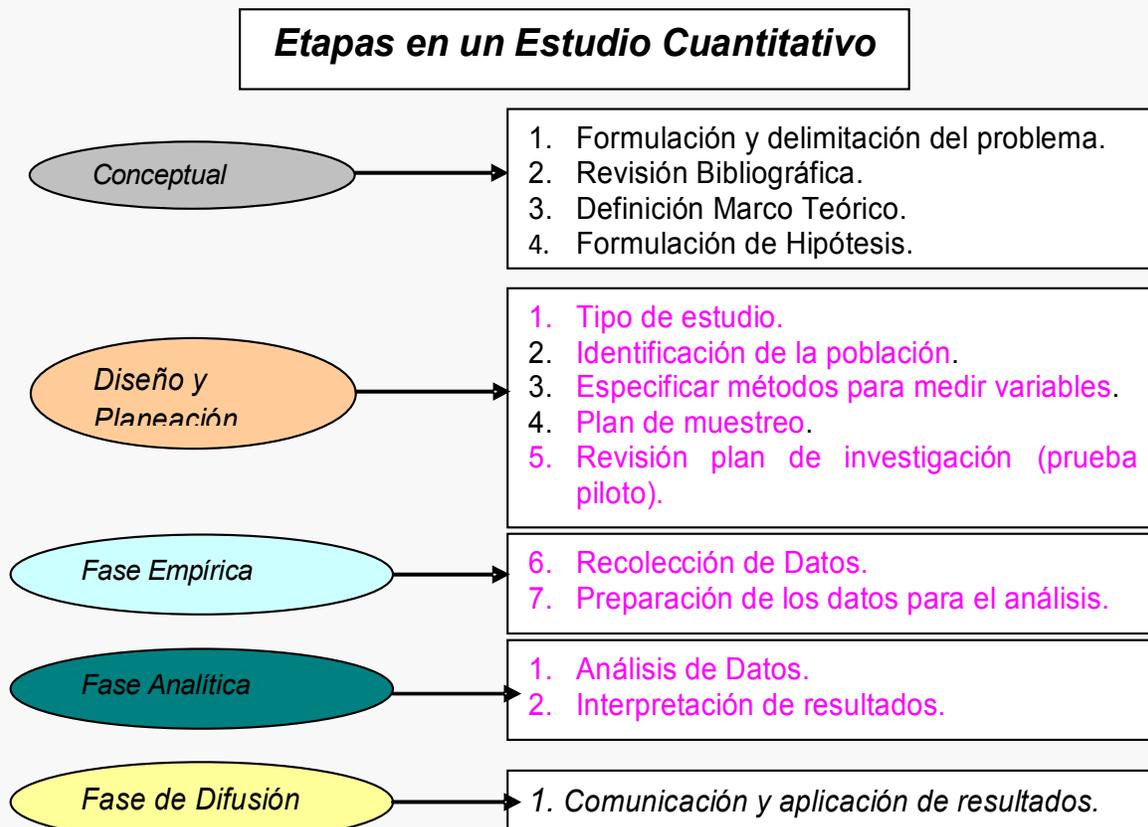


Figura 2.2: Etapas en un estudio cuantitativo.

La estadística apoya todo el proceso de investigación, sin embargo, tiene su mayor aporte en: el diseño, la fase empírica y la fase analítica. En aras de puntualizar cual es el aporte, se presentan algunos aspectos en cada una de las etapas del proceso de investigación, haciendo mayor énfasis donde la estadística es más evidente, sin querer decir que en las otras etapas no se necesita.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

2.1. ETAPA CONCEPTUAL.

Es aquella etapa, donde se presenta la formulación y delimitación del problema, se da la revisión bibliográfica, se desarrolla el marco teórico y se plantean las Hipótesis si fueran necesarias, es pues el soporte teórico de la investigación. Si se quiere garantizar un buen diseño de la investigación, el estadístico o quien haga sus veces, deberá estar informado sobre los aspectos teóricos del problema en estudio.

Todo el proceso de investigación, empezando por el diseño teórico tendrá como base los objetivos, estos serán la luz que alumbrarán a lo largo del desarrollo del estudio. Para ir definiéndolos, se sugiere plantear un número de preguntas que se espera serán resueltas al finalizar el trabajo. Las preguntas deben ser construidas por el equipo investigativo. Esta estrategia dará un insumo importante para argumentar la necesidad e importancia del estudio, es decir, hay que justificar el trabajo, donde se debe tener en cuenta el aporte que la investigación dará a la comunidad, o sea, cual será el impacto esperado a una población objeto de estudio, si bien es cierto, lo económico es relevante, también lo es el lograr mejorar el bienestar en la población. La justificación también debe mostrar razones institucionales e incluso personales para llevar a cabo un trabajo que en lo general requiere de una inversión en tiempo y en recursos de diversa índole. Esto lleva a sugerir que en esta etapa del proyecto, se debe analizar la viabilidad y factibilidad del estudio; para ello, se debe interrogar mínimo sobre: ¿La investigación se puede llevar a cabo?, ¿se tienen los recursos disponibles para ello?, ¿se tiene la voluntad de los entes financiadores y/o de las personas para ejecutar el proyecto?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

En cuanto al marco teórico, es importante destacar que no por lo extenso es garantía de un buen diseño teórico de la investigación. En nuestra experiencia, es importante desagregar el “Borrador del título de la investigación” y documentar al lector sobre estos temas, mencionó “Borrador” en el sentido que en esta etapa del proyecto, no necesariamente se tiene un título definido. Como ejemplo, si se va a realizar el estudio “Comportamiento de las tasas de mortalidad por cáncer de mama en mujeres mayores de 50 años en Medellín entre los años 2000 al 2008”, se sugiere que en el marco teórico, se hable de manera concreta y con la visión del objetivo del estudio sobre: generalidades de la ciudad de Medellín, tasas de mortalidad, cáncer de mama, características de mujeres de más de 50 años, procedimientos a utilizar para determinar el comportamiento; para el caso, podría colocarse métodos de series de tiempo. Por ello, es importante hacer una revisión actualizada de la posible bibliografía existente sobre el tema; esto podría minimizar los posibles errores en el diseño de la investigación.

Los anteriores, son algunos aspectos a tener en cuenta y que se deben abordar con buen criterio para lograr avanzar en el estudio con posibilidades de éxito en las etapas subsiguientes.

2.2 DISEÑO Y PLANEACIÓN.

Es una parte clave para el desarrollo y posterior utilidad de la investigación, ya que es aquí donde se lleva a cabo la planeación del estudio; para ello, se requiere definir el tipo de estudio (exploratorio, descriptivo o analítico). Además, se requiere identificar la población objetivo, especificar los métodos para medir las variables de investigación, diseñar el plan de muestreo o por

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

el contrario realizar censo y revisar el plan de la investigación (Prueba Piloto). Como se puede observar, la **estadística es el soporte de esta etapa crucial de la investigación**. Es importante en este punto empezar a diferenciar cuando se prefiere realizar muestreo y cuando censo. El siguiente cuadro muestra las desventajas del censo con respecto al muestreo.

MUESTREO	CENSO
Económico	Costoso
Oportuno	Requiere de mayor tiempo.
Puede ser más exacto	Mayor error humano
Ensayos destructivos (control de calidad)	Imposible de realizar

Pese a lo anterior, el censo se prefiere cuando la población es pequeña, con posibilidad de acceder a ella y/o cuando la variable clave del estudio presenta mucha variabilidad. En cualquier caso no se puede pretender la confiabilidad total usando censo, ya que como se ha notado no sólo es el número de observaciones a medir y cómo se seleccionan, sino, que la investigación requiere de una serie de pasos todos ellos con factores aleatorios que pueden alterar en mayor o menor medida los resultados obtenidos. La labor de los investigadores será precisamente minimizar esos factores. No sobra mencionar que en algunos casos la población objetivo no coincide con la población estudiada, esto ocurre por imposibilidad de poder tomar mediciones a algunas unidades de análisis, el caso más fácil de ilustrar, es el no tener acceso a determinados lugares geográficos por el riesgo físico que pueden correr los investigadores y o recolectores de datos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

A manera de reforzar los comentarios presentados en el sentido de la importancia de la definición de la población objeto de estudio, se comenta el siguiente caso, donde la falla en la identificación de la población ocasionó un error fatal en las conclusiones del estudio. Uno de los casos típicos es el llevado a cabo en 1936 en los Estados Unidos, donde la revista Literary Digest predijo la victoria del candidato Landon en perjuicio de Roosevelt de cerca del 15% con una muestra de más de 2 millones de personas. El resultado fue catastrófico, Roosevelt ganó con buen margen. Luego de analizar las posibles causas del error, se llegó a la conclusión que el problema radica en la mala definición de la población objetivo (marco muestral), ya que se buscó información sólo de las personas que figuraban en los directorios telefónicos, listados de suscriptores de revistas, poseedores de vehículos, entre otros, dejando por fuera gran parte de la población en posibilidades de votar. Los resultados fueron acertados para esa parte de la población, más no para todos los posibles votantes en los Estados Unidos.

En cuanto a los métodos para medir las variables de investigación de tal manera que se logren los objetivos del estudio, se lleva a cabo lo que se conoce como el plan de análisis, el cual tiene una relación directa con las escalas de medición y el tipo de variable (capítulo 1).

Los análisis aparte de los gráficos, se hacen bien sea univariados (poco), bivariados y multivariantes (lo sugerido). El siguiente esquema muestra una guía para llevar a cabo un plan hasta con dos variables.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CLASIFICACION DE LAS TECNICAS ESTADISTICAS



Figura 2.3: Técnicas estadísticas generales univariadas y bivariadas.

La síntesis anterior muestra aspectos muy generales cuando el análisis es univariado o bivariado. No obstante, el trabajo con una sola variable y como se verá más adelante es útil en la mayoría de los casos sólo para hacer control de calidad a los datos. Para las variables cuantitativas, se utilizan procedimientos exploratorios, que conllevan a calcular las medidas descriptivas como el promedio, mediana, percentiles, desviación típica entre otras y para las variables cualitativas, la moda, y en general frecuencias.

Para el caso bivariado, como su nombre lo indica, se relacionan dos

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

variables, es el caso de regresión simple, de procedimientos de correlación que pretenden identificar relación entre las variables objeto de estudio. Entre otras pruebas y dependiendo de la forma de los datos y de los objetivos del estudio, se mencionan las siguientes: la normal, la t, pruebas de Mann Whitney, Wilcoxon, rangos de Spearman entre otras. En las cualitativas, sobresale la prueba Ji cuadrado, teniendo en el menú estadístico muchas otras que se abordarán poco a poco en diversos ebooks.

Con respecto a pruebas multivariantes, ese es objeto de otro ebook, el cual podrá adquirir con descuento al haber adquirido el que esta leyendo. No, obstante, se pueden mencionar procedimientos comunes como: regresiones múltiples (lineal, cuadráticas, de cox, logística etc.), diseños experimentales, análisis de sobrevida, análisis de conglomerados, análisis de factor, análisis de árbol (tree), minería de datos, entre muchos otros procedimientos.

2.3 ETAPA EMPIRICA.

De nuevo la **estadística** es el soporte para desarrollar este punto, ya que tiene que ver con la manera de recolectar la información y de cómo se preparan los datos para el análisis. Esta es la parte que requiere por lo general mayor cantidad de tiempo. El esquema sugerido para procesar los datos es el siguiente:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Figura 2.4: Pasos sugeridos en el procesamiento de datos.



Teniendo en cuenta lo anterior, se puede afirmar que: la estadística provee a los investigadores de técnicas adecuadas para la recolección de datos, tales como la encuesta, la entrevista, la medición en el laboratorio o lugar de experimentación, así como de la depuración de los datos, valga decir, aporta técnicas de exploración de datos con el fin de identificar diferentes tipos de error. Otra situación importante en la que la estadística no se puede dejar de lado, es en la transformación de datos con el fin de lograr los objetivos del

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

estudio. Estas transformaciones se hacen cuando se requieren cambios en las escalas de medición y/o cuando hay que realizar cálculos para estimar indicadores, todo guiado por los objetivos del estudio, los cuales serán la guía para orientar las demás acciones.

Se presume que la recolección de los datos en el campo ya se realizó y que además, el proceso se hizo de manera adecuada. La primera parte hace referencia a la edición de los instrumentos de recolección de información, es decir, a una revisión de los mismos y a la corrección si hubiera lugar. La estadística, también apoya el proceso de codificación, teniendo en cuenta las escalas de medición y los objetivos del estudio. El tercer momento donde la teoría estadística apoya el procesamiento de datos es en la digitación y posterior transformación de datos si así se requiriera, además, de proveer las técnicas para realizar el análisis exploratorio, fundamental para optimizar el recurso humano y garantizar resultados confiables. Por último, la estadística, colabora con los investigadores en la construcción del plan de análisis, el cual define que procedimientos se deben desarrollar y en que momento, valga decir, si se requiere de análisis estadístico univariado y/o bivariado y/o multivariante. Sí se requiere de gráficos y/o de tablas, si el informe es global o desagregado según algunas variables de identificación. Se define cuales pruebas son las apropiadas para cada caso. Con lo realizado hasta ahora, se puede producir información y con ella, construir conocimiento para luego gestionarlo en favor de la comunidad.

2.4. ETAPA ANALITICA.

Tiene que ver con el análisis propiamente dicho y con la interpretación de

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

resultados, es decir, luego de definir y desarrollar los procedimientos **estadísticos** adecuados para cumplir con los objetivos del estudio, se hace la lectura a los cuadros y gráficos elaborados y se contextualizan teniendo en cuenta el marco teórico. En este punto es imprescindible tener el apoyo de personal experto en el área objeto de estudio.

Es importante mencionar que la lectura no es simplemente “volver” a decir que lo que ya se muestra en las tablas y/o gráficos, es “conversar” sobre lo que se quiere describir y/o informar, es importante avanzar en conclusiones según los hallazgos encontrados. Esta habilidad se adquiere leyendo informes de investigación y aplicándolos a nuestro que hacer.

2.5. DIFUSIÓN.

Hace referencia a la comunicación de los resultados y su aplicación. En la actualidad es primordial socializar los resultados en diferentes ámbitos y en lo posible publicarlos. Por ello, es imprescindible tener un conocimiento claro sobre los tópicos estadísticos y poder, dependiendo de las escalas de medición realizar los cálculos adecuados. Este punto tiene que ver con la siguiente frase: “Lo que no se escribe no existe”, esto da la importancia de que las investigaciones trasciendan de lo personal a lo más general que es la comunidad y no sólo la académica.

Como se puede intuir, la estadística es el brazo derecho y algo más para llevar a cabo investigaciones de carácter científico.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejercicio

1. Ubique tres investigaciones de tipo descriptivo, por cualquier medio (Internet, biblioteca, etc.) y analice si el marco teórico cumple con la teoría vista, además, reflexione sobre la justificación de ellas y valórela en términos de si realmente fue necesaria llevarla a cabo.
2. De manera sucinta, redacte una idea que considere importante para desarrollar una investigación, defina algunas preguntas y justifique el por qué es importante llevarla a cabo.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CAPITULO 3

"Nada nos engaña tanto como nuestro propio juicio" Leonardo Da Vinci"

PROCESAMIENTO DE DATOS

Para recoger información estadística útil, es necesario definir con exactitud (valor medio) y precisión (variabilidad) la clase de datos que se desea, y definir con anterioridad el proceso de obtención de los datos; en caso contrario el resultado final del estudio puede llevar a conclusiones completamente erróneas. Por ejemplo, si se desea obtener información acerca de los ingresos de una entidad bancaria durante los dos últimos años, hay que aclarar, que comprende el término "**INGRESOS**", ya que puede variar de acuerdo a los propósitos de la investigación; si se va a tener en cuenta los ingresos familiares, o son ingresos del jefe del hogar, si se tendrá en cuenta el ingreso mensual o diario, etc. Definiendo bien estos aspectos y otros más se llega a tener una definición de ingreso que sea adecuada para los fines del estudio. Otro ejemplo, puede ser: se desea llevar a cabo una investigación acerca de las defunciones de una determinada región durante los dos últimos años, hay que aclarar, que comprende el término "**DEFUNCIÓN**", ya que puede variar de acuerdo a los propósitos de la investigación; si se va a tener en cuenta las defunciones por muerte natural, o si por el contrario se tendrá en cuenta las muertes accidentales; si se tendrá en cuenta las muertes de los adolescentes, o si únicamente se tendrá en cuenta las personas adultas, etc. Definiendo bien estos aspectos y otros más se llega a tener una definición de defunción que sea adecuada para los fines de la investigación. Esto tiene que ver

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

directamente con lo ya definido, la **conceptualización de las variables**.

Lo importante y lo esencial es que la información que se recoja sea toda homogénea en su naturaleza, que se refiera a una misma cosa y que las categorías en que se agrupan sean también internamente homogéneas. Luego de haber obtenido la información correctamente, se nos plantea el problema de organizarlos y presentarlos de tal forma que se noten sus particularidades. La presentación de la información es importante por dos razones básicas:

- a) Facilita el análisis de los datos.
- b) Coloca la información al alcance de otras personas, que de otra manera implicaría mayor costo y tiempo entenderla.

La presentación de los datos se puede hacer por medio de texto, de tablas o cuadros y de gráficos o figuras o una combinación de ellas. Cuando se habla de texto, se realiza una disertación sobre el tema en estudio y se usa cuando tanto las tablas como los gráficos no son buenas ilustraciones de lo que se quiere presentar. Es una técnica más utilizada para variables cualitativas o en aquellos casos donde el número de categorías o su composición no ameritan una tabla o un gráfico.

Como complemento de este capítulo, puede abrir el enlace www.leondariobello.com/OA/frecuenciassimples

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

3.1 Presentación en forma de tablas o distribuciones de frecuencia.

Son tablas resumen que presentan los datos en bruto (bases de datos) en una forma ordenada y coherente.

3.1.1 **Partes de una tabla estadística:** Independiente del tipo de tabla, es decir, si es para una sola variable o para múltiples variables, se requiere que ésta sea lo más explícita posible y para ello se recomienda el siguiente esquema:

Numeración

Título

Tabla 2. Distribución porcentual de tutelas en salud interpuestas por usuarios. 2002-2007.

Año	Total	Porcentaje
2002	10	2,6
2003	34	8,9
2004	116	30,2
2005	163	42,4
2006	56	14,6
2007	5	1,3
Total	384	100

Encabezados

Matriz de datos
Cuero de la información

Fuente: fallos de tutela Dirección Seccional de Juzgados de Antioquia. 2002-2007. La información del año 2007 correspondía a los 2 primeros meses del año.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Título: Expresa la información que contiene, se coloca en la parte superior de la tabla. Debe ser breve, concreto y completo. Además, requiere de un número consecutivo.

Un título completo debe contener lo siguiente:

-La circunscripción espacial, es decir, debe indicar institución o área geográfica a la que pertenecen los datos. Responde a la pregunta **¿A DÓNDE** pertenece la información?

- El criterio de clasificación de los datos. Responde a **¿CÓMO** se presenta el contenido de la tabla?

- El espacio temporal o período que abarca la información que se presenta. Responde a **¿A QUÉ TIEMPO** pertenecen los datos de la tabla?

A veces es necesario indicar debajo del título, indicar las unidades de medida que se han utilizado para indicar la variable y si se necesitó realizar un cambio de escala para calcular los porcentajes.

Encabezados: Está formado por la primera fila de la parte superior y la primera columna de la izquierda, nos indica las características (variables) del fenómeno estudiado.

Cuerpo: Es el contenido mismo de la información de la tabla, se presenta en celdas o casillas de intersección, ordenadas por los títulos de las filas y columnas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Fuente: Indica la fuente de recolección de los datos presentados en la tabla, se coloca siempre en la parte inferior de la tabla y hace referencia a quien realizó el estudio.

Las tablas también se conocen como distribuciones de frecuencias y están formadas por variables y/o atributos, que por definición toman diferentes valores o categorías. La columna donde se encuentran los valores de la variable se encabeza con su respectivo nombre. No obstante, la mayoría de textos colocan la letra X_i (que se lee X sub i), a cada uno de los valores de las variable los determinamos en forma general $x_1, x_2, x_3, \dots, x_n$ (que se lee X sub 1, x sub 2, x sub 3, ..., x sub i), donde el subíndice toma tantos valores como categorías tenga la variable.

3.1.2 Tipos de Tablas Estadísticas según el número de variables.

3.1.2.1 Tablas simples o univariadas: Se denominan así por el hecho de que representan una sola variable o característica. Tal es el caso de la tabla anterior. En la actualidad que se utilizan programas de cómputo, su principal utilidad radica en explorar las variables cualitativas, es decir, en detectar errores en la digitación e identificar la importancia de posibles cruces de variables.

3.1.2.2. Tablas bivariadas o de contingencia: Son tablas en las que se presentan dos variables con el fin de determinar la relación entre ellas. Se sugiere que la variable independiente se coloque en las filas, sin embargo, prima la parte estética, es decir, se ubica en las filas la variable que tenga más categorías. Lo que es primordial es la forma de calcular los porcentajes. Se

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

debe de calcular en el sentido de la variable independiente, que en la salud, será en la dirección del factor causal.

La construcción de estas tablas busca identificar posibles causas de problemas de salud, y sus factores asociados, aún sin ser causales. En definitiva, se pretende determinar asociación o no entre dos variables. Las tablas de 2×2 , tienen un manejo especial en el área de la salud, su diseño general es:

	Enfermo		Total
Factor de riesgo	Sí	No	Total
Expuesto			
No expuesto			
Total			

Es claro que puede tener otro diseño en otras áreas del conocimiento. Adicionalmente, también se pueden construir otras tablas de contingencia con mayor dimensión. Una estructura puede ser:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Municipio-Provincia	Causas de muerte			Total
	Accidentes de Tránsito	Muertes Violentas	Suicidio	
Medellín				
Bello				
Envigado				
Itagui				
Total				

En definitiva lo que se pretende identificar es la relación entre las variables analizadas, por lo tanto, es importante en la lectura de la tabla, tratar de visualizar dicha relación o su negación. Este procedimiento se puede desagregar aún más, entrando otra variable, lo que origina un cruce de 3 variables, útil para detectar relaciones “Espureas”.

3.1.2.3. **Tablas compuestas o tipo banner:** son aquellas que presentan en forma simultánea tres o más variables o características de la realidad en estudio, útil para presentar información sintética y práctica.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

	Sexo		Hábito de Fumar		Hábito de Beber	
Medellín						
Bello						
Envigado						
Itagui						
Total						
Porcentaje						

En esta categoría se puede mencionar las tablas que presentan medidas resumen de más de tres variables cuantitativas, como ilustración, se tomó una tabla del trabajo realizado por la empresa Incomat y de la cual fui el asesor estadístico. Se adiciona, una posible manera de leer los datos, teniendo en cuenta, eso sí, que un Ingeniero de suelos sería útil para complementar la redacción.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Tabla No 21: Medidas resumen de las variables descriptoras del suelo en los puntos de exploración en Medellín Colombia. 2007.

	Estadísticos						
	Media	Coef. Variación	Mínimo	Máximo	Percentiles		
					25	50	75
Resistividad promedio	7666,62	76,14	1654,5	28500,0	3610,350	5553,150	10166,250
Resistividad diferencia	1052,98	102,13	,0	4600,0	278,800	740,000	1540,000
PH promedio	7,3612	6,34	5,80	8,00	7,2000	7,5000	7,7000
Promedio Redox	442,968	11,41	348,20	543,20	397,8750	446,2500	482,0750
Promedio sulfato	204,139	80,46	28,200	981,000	101,62500	149,60000	261,50000
Promedio cloruro	1860,57	24,90	863,50	2759,00	1594,1250	1921,5000	2258,6250
Evaluación corrientes parásitas	19,4404	55,49	3,40	52,60	10,3000	18,2000	25,5000
Corrientes con resistividad	3,984	75,02	,6	15,7	1,800	3,550	5,125
Agresividad del suelo -AWWA-	1,520	99,16	,0	5,5	,000	1,000	3,000
Agresividad del suelo -Incomat-	7,1378	28,30	3,00	12,50	5,5000	7,0000	8,0000

La resistividad promedio se calculó promediando los valores encontrados de la resistividad del terreno en contacto con la estructura metálica. El resultado promedio dio 7666,62 Ω /cm, no obstante, se encontró una alta variación en estas mediciones, dado el coeficiente de variación de 76,14%, lo que indica que existen diferencias importantes en los valores de resistividad calculados a los 2 y 3 metros de profundidad. Por ello, es importante analizar otros indicadores. El menor valor encontrado fue de 1654,5, es decir, *****. Es de destacar que el 75% de los valores encontrados están por encima de 3610, es decir no aportan puntos en la agresividad del suelo, *****.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Los cambios significativos encontrados en las mediciones de la resistividad a los 2 y 3 metros son confirmados al analizar los indicadores de la diferencia de la resistividad. Estos datos presentan la mayor variabilidad de todos los indicadores analizados, el coeficiente de variación es de 102,13%, lo que indica que las mediciones tienen una fluctuación alta, encontrando valores desde 0 hasta 4.600. Observando los el percentil 25, se concluye que el 25% de las mediciones están por debajo de 278,8 Ω/cm , mientras que el 50% están por debajo de 740 Ω/cm , es decir, son valores que *****.

3.1.3 Tablas Estadísticas según el nivel de medición de las variables.

En este punto del aprendizaje, sólo se hará mención a la construcción de tablas usando **conteos o frecuencias** con sus respectivos porcentajes si son necesarios. En capítulos posteriores, se calcularán otras medidas resumen de acuerdo al tipo de variable.

3.1.3.1. Variables Nominales u Ordinales: Como ya se mencionó anteriormente, para estas variables es útil calcular fundamentalmente: Porcentajes, proporciones, tasas y razones, además, de valores absolutos. Lo más común es relacionarlas con otras variables. Ejemplos univariados y bivariados son los siguientes:

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Tabla No 31: Sugerencias recibidas por los usuarios de Inter, Municipio de Medellín, diciembre 2009.

GIRARDOTA	FRECUENCIA	PORCENTAJE
Conciencia y educación ciudadana	5	23,81
Limpiar el parque y otras zonas	4	19,05
Establecer horarios fijos para la recolección	3	14,29
Aumentar la frecuencia de recolección	3	14,29
Más botes de basura	2	9,52
Vigilar la recolección en la plaza de mercado	2	9,52
Que el carro pase más despacio	1	4,76
Rebajar los costos	1	4,76
	21	100,00

*Fuente: Estudio de percepción de usuarios de Inter***** S.A E.S.P, diciembre del 2009.*

Es fácil entender que la tabla se construyó contando el número de veces que se repite cada categoría, es decir, al contar en la base de datos, se encontró que 2 personas consideran que se requieren más botes de basura en la zona, sólo una persona solicitó rebajar los costos y otro que el carro pase más despacio. El texto no hace énfasis en la construcción manual, ya que con certeza el lector tendrá la ayuda de algún sistema de cómputo.

Note que es útil presentar los resultados de las categorías en orden descendente, con el fin de apreciar lo más relevante de manera rápida. Dado que la parte de la lectura de las tablas se ha detectado como una debilidad, se presenta un ejemplo de la redacción para la tabla anterior, sin pretender que sea la única posible para el caso.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

En el Municipio de Girardota no se percibe una opinión generalizada sobre el tema de aseo, al menos en cuanto a lo de sugerencias se trata. Cinco personas comentan que las campañas educativas son útiles para mejorar el aseo de la zona, 4 solicitan limpiar el parque. Otros comentarios presentados con menor frecuencia apuntan a establecer horarios fijos de recolección, aumentar la frecuencia de la misma, colocar más botes, vigilar la recolección en la plaza de mercados, procurar que el carro pase más despacio y rebajar los costos. De lo anterior, se deduce la importancia que tienen las campañas educativas para mejorar la limpieza del Municipio.

Tabla *****: Relación entre la opinión sobre si la drogadicción debe ser tratada durante el tratamiento o posterior al abandono (Medellín 2009).

La Drogadicción debe ser tratada	Consumo de droga durante el tratamiento				Consumo de droga posterior al abandono			
	SI		NO		SI		NO	
	Nº	%	Nº	%	Nº	%	Nº	%
SI	9	90.0	26	92.9	26	96.3	9	81.8
NO	---	---	2	7.1	1	3.70	1	9.1
NS/NR	1	10.0	---	---	---	---	1	9.1
TOTAL	10	26.3	28	73.6	27	71.0	11	28.9

Fuente: Estudio realizado por estudiantes de la especialización en Fármaco dependencia en la Universidad ***** 2009

Se destaca que el 81.8% de las personas encuestadas consideran que el problema de la drogadicción debe ser tratada. También se observa que durante el tratamiento, cerca de las $\frac{3}{4}$ partes no consumieron droga, sin

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

embargo, un porcentaje significativo (26.3%) si lo hicieron, indicador que el programa requiere de ajustes para minimizar éste problema. Otra situación que confirma lo anterior, es que el 96.3% recaen, es decir, vuelven a consumir droga. En lo que sí se está de acuerdo, es que el problema debe ser tratado.

3.1.3.2 Una variable cuantitativa con pocas categorías: Posibles valores de la variable no quiere decir, pocos datos. Si bien el enfoque es utilizar algunos de los programas estadísticos existentes en el mercado, es útil conocer cómo se pueden construir las tablas, tanto de variables categóricas como cuantitativas y especialmente continuas.

Ejemplo: Los siguientes datos en bruto, se obtuvieron de un grupo de estudiantes de la especialización en Salud Ocupacional de una prestigiosa universidad Colombiana en el año 2010.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Genero	Profesión	Estado civil	Edad	Nº Hijos
Femenino	Médic@	Solter@	28	0
Masculino	Psicólog@	Divorciad@	40	2
Femenino	Enfermera	Casad@	32	1
Masculino	Sociólogo	Unión libre	38	1
Femenino	Enfermera	Casada	33	3
Femenino	Médic@	Solter@	28	1
Masculino	Sociólogo	Divorciado	45	1
Femenino	Enfermera	Soltera	40	2
Femenino	Médic@	Solter@	32	2
Masculino	Sociólogo	Unión libre	38	0
Femenino	Médic@	Solter@	28	0
Masculino	Estadístico	Casado	33	1
Femenino	Médic@	Solter@	28	0
Femenino	Médic@	Solter@	28	0

Construir las siguientes tablas: a. Univariada para la profesión, b. Univariada número de hijos, c. Bivariada genero vs profesión.

a. Univariada: profesión.

Al ser una variable categórica nominal no se requiere orden, sin embargo, es recomendable colocar las categorías en orden descendente con el fin de identificar aquellas que tienen mayor presencia. Al lado se construyen las columnas referentes al número de veces que está la categoría, seguida de su correspondiente porcentaje, así.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Profesión	Número de Personas	Porcentaje
Médico	6	42.85
Enfermera	3	21.42
Sociólogo	3	21.42
Psicólogo	1	7.14
Estadístico	1	7.14
Total	14	100.0

Es claro que el porcentaje se calcula dividiendo el número de personas sobre el total de ellas, valga decir, $6/14 = 0.4285$. Otro hecho a destacar es que no se calcula el porcentaje acumulado, ya que no tiene sentido.

a. Univariada: número de hijos.

Número de hijos	Número de Personas	Porcentaje	Porcentaje Acumulado
Ninguno	5	35.71	35.71
Uno	5	35.71	71.42
Dos	3	21.42	92.84
Tres	1	7.14	100.0
Total	14		100.0

Al ser una variable con un orden natural, las categorías vienen organizadas y se pueden calcular los porcentajes acumulados, además, de las frecuencias

Es súper recomendable que las frecuencias simples se usen en la mayoría de casos sólo para realizar el análisis exploratorio de datos, es decir, para detectar errores de digitación, más no para el análisis e interpretación de resultados. El interés debe ser buscar relación entre variables, por ello se recomienda realizar cruces de variables.

De manera manual el esquema y la nomenclatura para construir las tablas simples son:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Al total de las observaciones, la denominamos n si es una muestra o N si es la población y es igual a la suma de los f_i ; donde cada f_i está definido por el número de veces que se repite la variable (x_i), y se conoce como **frecuencia absoluta, que se puede rotular como número de**

Frecuencia absoluta acumulada: La denotamos por la letra mayúscula F_i y es igual a la suma parcial de la frecuencia absoluta, se puede colocar como **número acumulado de....**

Porcentaje valido: Es más explícito colocar el nombre **porcentaje**, es igual a la razón entre la frecuencia absoluta y el número de observaciones multiplicado por 100. Dice la participación porcentual de cada valor con respecto al total de observaciones.

$$\text{Porc.} = (f_i / n) * 100$$

Porcentaje acumulado: Es igual a la suma parcial del porcentaje valido o a la razón entre la frecuencia absoluta acumulada y el número de observaciones.

$$\text{Porc. acumulado} = (F_i / n) * 100$$

PROCEDIMIENTO

Para construir una tabla de frecuencias, lo primero que se hace es ordenar la serie de datos en forma ascendente, a continuación colocamos la

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

frecuencia absoluta (f_i); es decir; el número de veces que se repite la variable. Continuamos la tabla con el siguiente orden: frecuencia absoluta acumulada (F_i), porcentaje acumulado y por último el porcentaje acumulado. En forma general la tabla de frecuencias queda de la siguiente forma.

Nombre de la Variable	Número de ...	Número acumulado	Porcentaje	Porcentaje acumulado
x_1	f_1	$F_1 = f_1$	$(f_1 / n) * 100$	$(F_1 / n) * 100$
x_2	f_2	$F_2 = f_1 + f_2$	$(f_2 / n) * 100$	$(F_2 / n) * 100$
x_n	f_n	$F_n = f_1 + f_2 + \dots + f_n$	$(f_n / n) * 100$	$(F_n / n) * 100$

Dicho esquema fue el utilizado en los ejemplos anteriores. No sobra resaltar que dada la escala de medición, se calcula las columnas referentes a valores acumulados o no.

c. Bivariada: genero vs profesión. Se realizan los conteos de las variables a relacionar y se colocan en las márgenes de la tabla, para luego identificar a que subcategoría pertenecen, así:

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Profesión	Genero					
	Femenino		Masculino		Total	
	Nº	Porc.	Nº	Porc.	Nº	Porc.
Médic@	6	66.67	0	0.0	6	42.85
Enfermera	3	33.33	0	0.0	3	21.42
Sociólog@	0	0.0	3	60.0	3	21.42
Psicólog@	0	0.0	1	20.0	1	7.14
Estadístic@	0	0.0	1	20.0	1	7.14
Total	9	64.28	5	35.71	14	100.00

De manera manual, primero se cuentan los totales de cada categoría, es decir, de la misma manera que se hizo la tabla simple, para luego, identificar de ese total, cuantos pertenece a cada subcategoría. Para el ejemplo, de los 6 [médic@s](#), se determina cuantos son hombres y cuantas mujeres, igual para todas las demás profesiones.

Un tema de especial importancia es ¿cómo se deben calcular los porcentajes? Estos se calculan en el sentido de la variable independiente, es decir, la que se presenta primero o la que no está influenciada por la otra variable. El género, no depende en manera absoluta por la profesión, más bien podría pensarse que éste influye en la profesión. De hecho, según los resultados, esto sucede, ya que todas las mujeres son enfermeras o médicas, mientras que los hombres pertenecen a las otras profesiones.

Es usual sugerir que la variable independiente se coloca en las filas, no obstante, parece ser más relevante la parte estética de la tabla. Los invito a ver el siguiente vídeo con el fin de observar como se crean tablas de doble entrada con el software SPSS, es de gran ayuda que mire el vídeo, [Tablas Bivariadas](#).

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

El fin último de los cruces de variables, es determinar si existe relación entre las variables ubicadas en la tabla, y si bien existen pruebas para determinar dicha relación (Chi cuadrada), de manera descriptiva y como se planteó en el recuadro anterior, es posible intuir dicha relación o no.

Ejercicios: Construya las tablas simples para las variables Edad, Estado civil, además, cruce las variables, estado civil y profesión, diga si parece existir relación entre ellas y ¿porqué? Relacione la edad con el número de hijos, que concluye. Use la tabla del ejemplo anterior.

3.1.3.3. Una variable cuantitativa continua: Es usual que este tipo de variables, posea un alto número de posibles respuestas, por lo tanto, la tabla simple, tendría un número muy alto de valores, lo que impediría un análisis adecuado de las características más importante de los datos, por ello, se utiliza la estrategia de agrupar dichos valores en intervalos o rangos. De cualquier manera, cada vez es menos probable que se realice esta tarea de manera manual, por ello, no se pretende ser demasiado exhaustivo en la construcción de estas tablas.

Para familiarizarnos con esta nomenclatura, desarrollemos un ejemplo más, manualmente.

Supongamos que el salario decenal de una muestra de 40 profesionales recién egresados, pertenecientes a una empresa del sector educativo fueron los siguientes:

400000 420000 450000 550000 420000 530000 600000 550000
530000 400000 450000 480000 500000 530000 570000 600000

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

420000 450000 480000 500000 450000 480000 500000 530000
 500000 530000 550000 570000 530000 550000 550000 570000
 600000 480000 570000 500000 480000 530000 530000 500000

Se requiere construir una tabla de frecuencias simple. Siguiendo el procedimiento estudiado anteriormente, lo primero que se hace es ordenar de menor a mayor la variable sueldo (x_i). Hay dos personas que ganan \$400000, luego $f_i = 2$ y así sucesivamente, lo mismo con las demás columnas analizadas anteriormente.

Tabla No **: Distribución de los salarios decanales en instituciones del sector educacional en Medellín, marzo del 2004.

Salario Decanal	Número de profesionales	Número acumulado de profesionales	Porcentaje	Porcentaje acumulado
400000	2	2	5.00	5.00
420000	3	5	7.50	12.50
450000	4	9	10.00	22.50
480000	5	14	12.50	35.00
500000	6	20	15.00	50.00
530000	8	28	20.00	70.00
550000	5	33	12.50	82.50
570000	4	37	10.00	92.50
600000	3	40	7.50	100.00

$n = 40$

De cualquier manera la clave está en la interpretación, se plantea la siguiente: “El salario decanal de los profesionales se concentra en mayor medida en los valores centrales, es decir, la mayoría de ellos devengan entre \$480.000 y \$550.000, siendo el valor más usual el de \$530.000. Además, se observa que 20 profesionales que equivale al 50% devengan \$500.000 decanales o menos. Los menores porcentajes de personas se

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

encuentran en las categorías extremas, valga decir, menos de \$420.000 y más de \$600.000.

3.2 MUESTRA AGRUPADA

En estadística, por lo general, se utiliza un gran número de datos, ya que mientras más datos se tengan de una determinada población, el estudio que se haga será más confiable. Debido a esto, se resumen los datos originales, es decir, se convierte la tabla de frecuencias simple en una tabla de frecuencias por intervalos, de tal manera que sea más comprensible la información entregada.

Existe disparidad de criterios para definir el número de intervalos, sin embargo, el criterio y la experiencia del investigador es lo más relevante al momento de definir dicho número. Esta misma experiencia dice que la dispersión de los datos será un elemento fundamental al momento de definirlo. En cualquier caso, menos de cinco y más de 15 es poco probable que se encuentren, a no ser que los objetivos de la investigación lo requieran, por ejemplo, las edades quinquenales en estudios demográficos, los salarios, si de retenciones se trata.

Un criterio sano al momento de definir el número de intervalos o celdas es que no debe existir ninguno de ellos con puntaje o frecuencia cero, ya que estaríamos alargando la distribución en forma innecesaria. Ahora bien, si agrupamos la información en muy pocas celdas, podemos estar dejando de observar algunas particularidades importantes para el estudio.

Las razones fundamentales para hacer este tipo de agrupaciones son:

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- Es antieconómico y poco práctico tratar con un gran número de casos distribuidos en muchos puntajes o frecuencias.
- Algunos datos tienen asociados una frecuencia tan baja, que no se justifica mantenerlos como entidades distintas y separadas.

3.2.1 PROCEDIMIENTOS PARA AGRUPAR EN INTERVALOS.

Lo primero que debo mencionar, es que definitivamente el procedimiento de manera manual, no lo va a utilizar, o al menos la probabilidad es mínima de que ello ocurra, por lo tanto, no se hará énfasis en ese procedimiento.

Para agrupar una serie de datos en intervalos, se sugiere los siguientes pasos.

- 1) Definir en cuántos intervalos se va a agrupar, como ya se mencionó, esto dependerá de la variabilidad de los datos, se va a denotar con la letra **K**.
- 2) Hallar la variación o dispersión de los datos; es decir, el rango o recorrido de estos, donde el rango es igual al valor máximo menos el valor mínimo de la distribución (datos), es decir,

$$R = X_{\max} - X_{\min}$$

- 3) Calcular la amplitud de los intervalos, donde dicha amplitud deberá ser constante (en algunos casos no lo es). La amplitud es igual a la razón entre el rango y el número de intervalos. A la amplitud la denotamos con la letra minúscula **a**.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$a = R / K$$

Es importante tener en cuenta que la frecuencia absoluta de los intervalos es mutuamente excluyente, es decir, un determinado puntaje o frecuencia no puede pertenecer a dos intervalos a la vez.

De nuevo, lo esencial está en la lectura de la tabla (interpretar).

INTERVALOS LI – LS	Número de profesionales	Número acumulado de profesionales	Porcentaje	Porcentaje acumulado
1) 400000 a 440000	5	5	12.50	12.50
2) 440000 a 480000	9	14	22.50	35.00
3) 480000 a 520000	6	20	15.00	50.00
4) 520000 a 560000	13	33	32.50	82.50
5) 560000 a 600000	7	40	17.50	100.0

$$n = 40$$

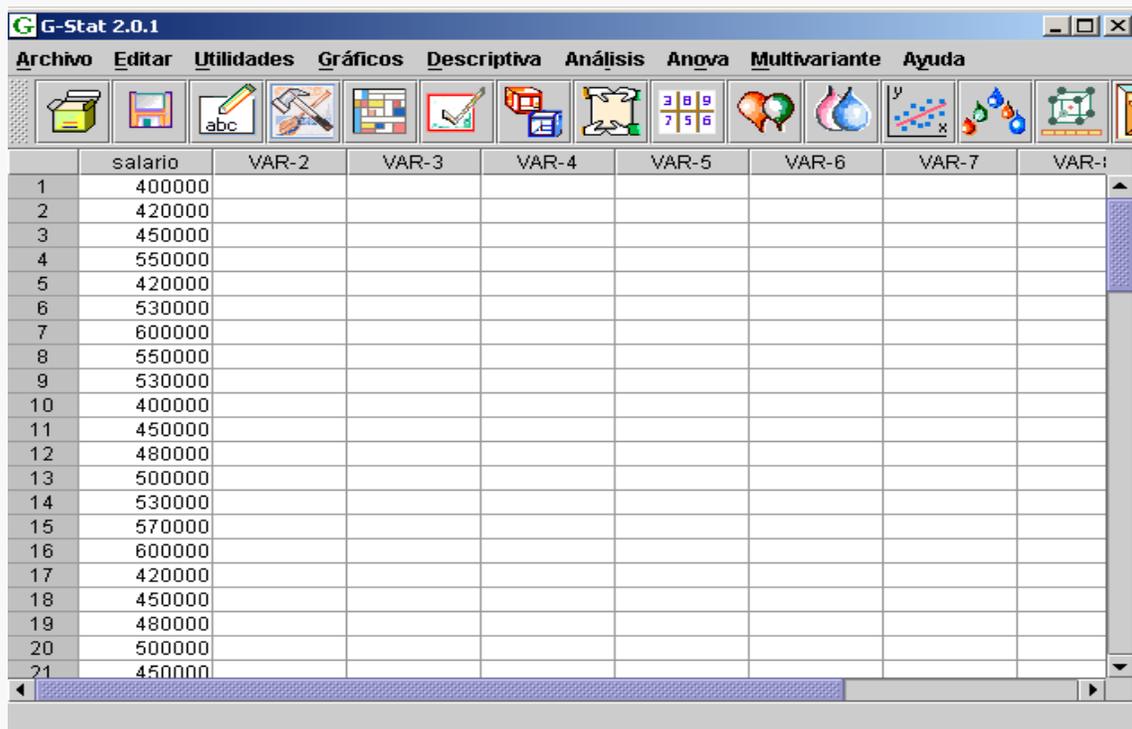
Se aprecia que el mayor número de profesionales del sector educativo devengan salarios quincenales entre \$520.000 y \$560.000, además. El menor número de estos devengan menos de \$440.000. En la mayor categoría salarial se encuentra el 17.50%. También es importante destacar que el 82.50% ganan quincenalmente menos de \$560.000.

Otro comentario útil, es que en algunos casos, se acostumbra adicionar una cifra decimal más de la que tienen los datos, para evitar la ambigüedad de donde puede quedar un valor, para el caso, se pudo definir el primer intervalo como: \$399.999.50 a \$440.000.50 y así para los demás intervalos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

3.2.2 CONSTRUCCIÓN DE INTERVALOS CON G-stat.

En el sitio www.leondariobello.com puede acceder al link donde baja el software gratuito G-stat. Luego digite los datos tal como si fuese una hoja electrónica, los datos se apreciarían así:



The screenshot shows the G-Stat 2.0.1 software interface. The menu bar includes: Archivo, Editar, Utilidades, Gráficos, Descriptiva, Análisis, Anova, Multivariante, Ayuda. The toolbar contains icons for file operations, data entry, and analysis. The main window displays a data entry table with the following data:

	salario	VAR-2	VAR-3	VAR-4	VAR-5	VAR-6	VAR-7	VAR-8
1	400000							
2	420000							
3	450000							
4	550000							
5	420000							
6	530000							
7	600000							
8	550000							
9	530000							
10	400000							
11	450000							
12	480000							
13	500000							
14	530000							
15	570000							
16	600000							
17	420000							
18	450000							
19	480000							
20	500000							
21	450000							

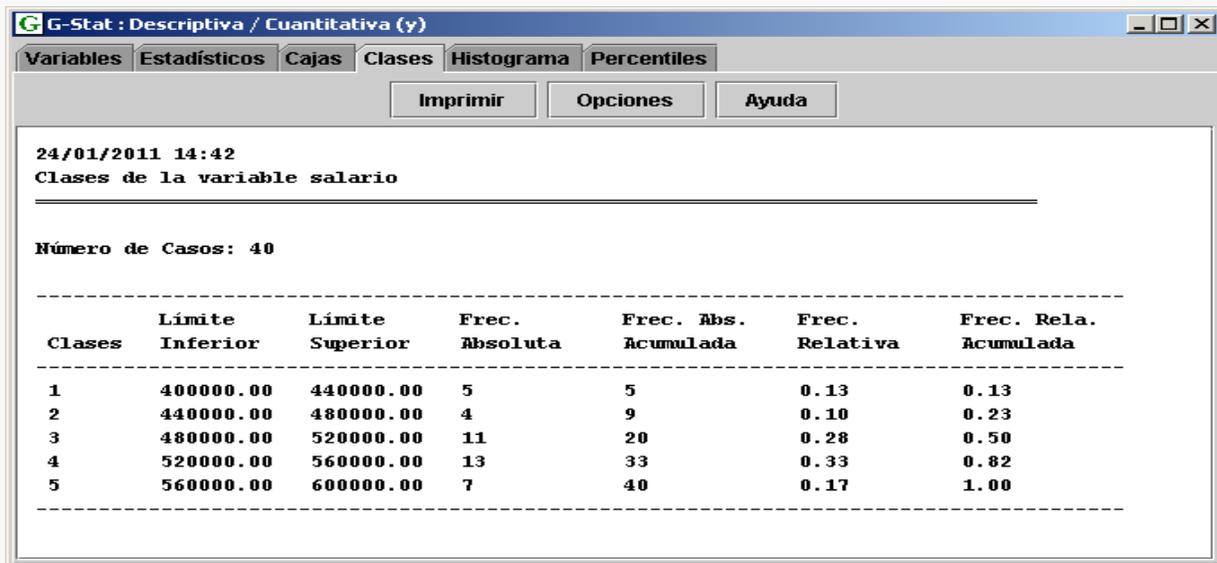
Para colocar el nombre a la variable se requiere dar click en el nombre VAR-1 de la columna, teniéndola sombreada, en la primera opción puede colocar el nuevo nombre, además, de realizar otros ajustes si fuese el caso.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Luego de abrir el archivo, lo que logra, con los siguientes pasos:

1. Descargar el programa del sitio www.leondariobello.com
2. Descomprimir la carpeta G-stat 2.0 a su PC
3. Entrar a la carpeta G-stat 2.0 y solicite ver detalles de los archivos, lo cual se logra dando click en la parte superior de la pantalla en el icono vistas.
4. Dar click en el archivo G-stat 2.0 tipo aplicación y que tiene un peso de 312 KB, existen varios archivos con el mismo nombre.

Luego proceda a entrar por la opción Descriptiva+cuantitativa(y), seleccione la variable cuantitativa a procesar, en este caso salario, obtendrá la siguiente pantalla, al solicitar clases.



The screenshot shows the G-Stat software interface with the 'Clases' tab selected. The window title is 'G-Stat : Descriptiva / Cuantitativa (y)'. The interface includes tabs for 'Variables', 'Estadísticos', 'Cajas', 'Clases', 'Histograma', and 'Percentiles'. Below the tabs are buttons for 'Imprimir', 'Opciones', and 'Ayuda'. The main display area shows the date and time '24/01/2011 14:42' and the variable name 'Clases de la variable salario'. Below this, it indicates 'Número de Casos: 40'. A table displays the class intervals and frequencies for the variable 'salario'.

Clases	Límite Inferior	Límite Superior	Frec. Absoluta	Frec. Abs. Acumulada	Frec. Relativa	Frec. Rela. Acumulada
1	400000.00	440000.00	5	5	0.13	0.13
2	440000.00	480000.00	4	9	0.10	0.23
3	480000.00	520000.00	11	20	0.28	0.50
4	520000.00	560000.00	13	33	0.33	0.82
5	560000.00	600000.00	7	40	0.17	1.00

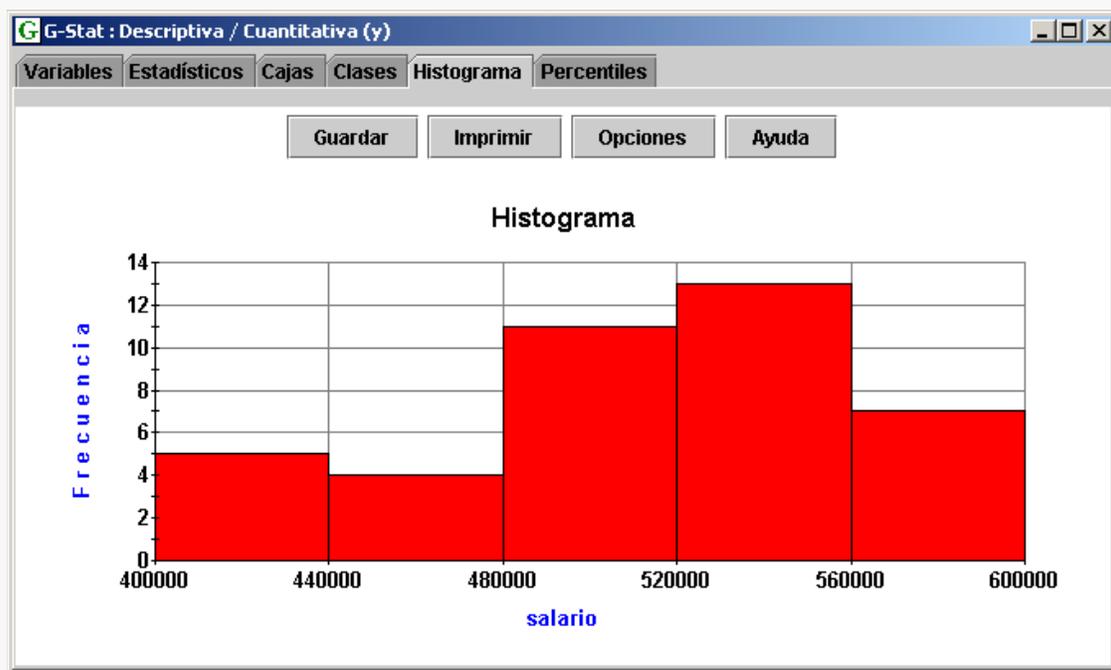
Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Observe que los resultados difieren, esto se debe a que en el G-stat, no se incluye el límite superior, es decir, para el primer intervalo, el 44000 se cuenta para el segundo intervalo. Si su interés es cambiar el número de clases, le da click en Opciones y podrá ensayar otras distribuciones de frecuencias.

Si bien no se ha conversado sobre los gráficos, es útil mencionar que el histograma es el usado para dibujar datos cuando se tienen agrupados en intervalos, para el ejemplo, la presentación es la siguiente:



Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejercicios

1) EL pago de servicios públicos, (en miles de pesos) en un determinado sector de la ciudad de un grupo de 40 familias aparece a continuación.

300 - 320 - 350 - 380 - 370 - 300 - 320 - 360
340 - 320 - 380 - 340 - 320 - 360 - 390 - 380
370 - 400 - 320 - 330 - 350 - 370 - 390 - 420
420 - 440 - 400 - 370 - 390 - 420 - 400 - 440
320 - 340 - 360 - 350 - 370 - 380 - 390 - 360

- Construya una distribución de frecuencia Simple, e interprete un valor de cada columna.
- ¿Cuántas personas pagan menos de \$370.000?
- ¿Qué porcentaje de las personas pagan entre \$320.000 y \$440.000?
- Construya una distribución en cinco intervalos e interprete un valor de cada frecuencia.

2) El tiempo de espera de un grupo de 40 personas (en minutos) para recibir un servicio en una cierta corporación mostró los siguientes resultados:

8 - 10 - 12 - 15 - 13 - 12 - 10 - 6 - 8 - 8
11 - 12 - 14 - 12 - 12 - 10 - 8 - 9 - 14 - 6
8 - 9 - 9 - 14 - 12 - 13 - 13 - 14 - 12 - 10
13 - 16 - 18 - 20 - 14 - 13 - 10 - 11 - 8 - 12

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- a. Construya una distribución de frecuencia simple e interprete lo más relevante.
 - b. ¿Qué porcentaje de las personas tienen un tiempo de espera de 13 minutos?
 - c. ¿Qué porcentaje de personas tuvo un tiempo de espera de 14 minutos o menos?
 - d. ¿Qué porcentaje de personas tuvo un tiempo de espera mayor de 12 minutos pero menor que 15 minutos?
 - e. ¿Cuántas personas esperaron más de 16 minutos para recibir su servicio?
 - f. ¿Cuántas personas tuvieron un tiempo de espera de 13 minutos?
 - g. ¿Cuántas personas esperaron más de 12 minutos pero menos de 15 minutos para recibir el servicio?
 - h. Construya una distribución en seis intervalos e interprete un valor de cada frecuencia, use G-stat.
3. Los siguientes datos representan las notas de un grupo de estudiantes en **Estadística Descriptiva**.

4.5	3.6	2.8	4.3	3.5	4.0
2.3	3.1	3.5	2.5	3.1	4.2
3.5	4.1	4.2	3.6	3.5	2.6

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

2.7	4.0	3.7	2.8	3.5	3.1
3.3	3.8	4.3	3.5	3.1	3.8
4.0	2.7	2.5	2.7	2.8	2.7

- a. Construya la distribución de frecuencia simple.
 - b. ¿Cuántos estudiantes tienen una nota de 4.0?
 - c. Al menos que porcentaje de estudiantes obtuvo una nota de 3.0
 - d. ¿Que porcentaje de estudiantes ganó la materia?
 - e. Si se requiere que las notas estén entre 4.0 y 4.3. ¿Qué porcentaje de ellas no cumple con la regla?
 - f. Si para habilitar, se requiere mínimo una nota de 2.0, ¿qué porcentaje de estudiantes no puede habilitar?
 - g. Agrupe en cinco intervalos e interprete un valor de cada columna, apóyese en el G-stat.
4. Con la siguiente base de datos, construya las tablas de frecuencia simple e interprételas. Además, cruce todas las variables cualitativas con el municipio. Determine si el municipio influye en las otras variables.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Municipio del Sur oeste de Antioquia	Tipo de Institución	Nivel de Complejidad	Facturación (millones)	Quien entrega la información
Andes	Pública	2	550	Gerente
Hispania	Privada	1	50	Director
Jardín	Pública	1	320	Gerente
Andes	Privada	2	420	Director
Jardín	Pública	2	89	Director
Andes	Pública	2	120	Asistente
Hispania	Privada	2	120	Gerente
Andes	Privada	3	50	Director
Andes	Pública	1	80	Asistente
Hispania	Privada	2	300	Director
Jardín	Privada	2	50	Asistente
Jardín	Pública	1	50	Gerente

¿Qué puede concluir al tabular la variable cuantitativa facturación? ¿Es útil cruzar ésta variable según municipio? Justifique su respuesta.

3.3 Construcción de tablas e histograma con Excel 2010.

Cuando se trate de construir tablas simples, el Excel, tiene una función que permite realizar los conteos, la manera para proceder es la siguiente:

1. Digitar los datos en bruto en la hoja de cálculo.
2. Digitar en otras celdas los valores individuales de menor a mayor, sólo un valor por cada categoría, como se observa a continuación.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

The screenshot shows an Excel spreadsheet titled 'Ejerciciomuestraagrupada - Microsoft Excel'. The data is organized as follows:

Salario decenal									
A	B	C	D	E	F	G	H	I	J
1	Salario decenal								Valores individuales
2									
3	400000	420000	450000	550000	420000	530000	600000	550000	400000
4	530000	400000	450000	480000	500000	530000	570000	600000	420000
5	420000	450000	480000	500000	450000	480000	500000	530000	450000
6	500000	530000	550000	570000	530000	550000	550000	570000	480000
7	600000	480000	570000	500000	480000	530000	530000	500000	500000
8									530000
9									550000
10									570000
11									600000

3. Utilizar la función **=frecuencia(rango de datos; rango de valores a contar)**, al ser una orden de matriz, se debe sombrear primero las celdas donde se quiere el resultado, usualmente al lado del rango de valores, para el ejemplo, la función es la siguiente: **=FRECUENCIA(A3:H7;J3:J11)** y se presiona al mismo tiempo las teclas Ctrl + Shift + Enter.

4. Se construye el resto de la tabla usando fórmulas, obteniendo los

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

resultados que se aprecia a continuación.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Salario decanal									Salario	N Personas	Acum	Porcenta	Porc.	Acumulado	
2										Decanal						
3	400000	420000	450000	550000	420000	530000	600000	550000		400000	2	2	5	5		
4	530000	400000	450000	480000	500000	530000	570000	600000		420000	3	5	7,5	12,5		
5	420000	450000	480000	500000	450000	480000	500000	530000		450000	4	9	10	22,5		
6	500000	530000	550000	570000	530000	550000	550000	570000		480000	5	14	12,5	35		
7	600000	480000	570000	500000	480000	530000	530000	500000		500000	6	20	15	50		
8										530000	8	28	20	70		
9										550000	5	33	12,5	82,5		
10										570000	4	37	10	92,5		
11										600000	3	40	7,5	100		
12											40		100			

Para construir tablas agrupadas, se requiere primero activar las opciones estadísticas que trae el programa, para ello, y luego de estar en la hoja de cálculo Excel, se procede a dar clic al botón inicio colocado en la parte superior izquierda de la pantalla, para luego entrar por la pestaña “**opciones de Excel**”, la cual origina otras posibilidades entre las que se encuentra la de “**complementos**”. Por último, se entra por el botón “**ir...**”, el cual permite seleccionar “**Herramientas para análisis**” y aceptar. Este proceso se

Estadística Descriptiva y Distribuciones de Probabilidad.

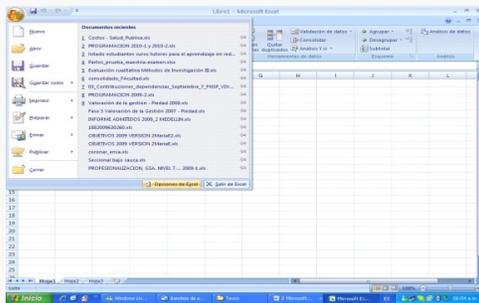
©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

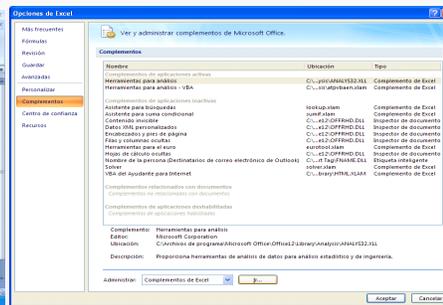
realiza una sola vez y activa seleccionando **“Datos”** del menú principal, la opción **Análisis de Datos**, la cual nos lleva a los procedimientos estadísticos diseñados en el Programa Excel versión 2010.

De manera gráfica se presenta la siguiente secuencia buscando mayor facilidad a los lectores para activar estas valiosas posibilidades estadísticas.

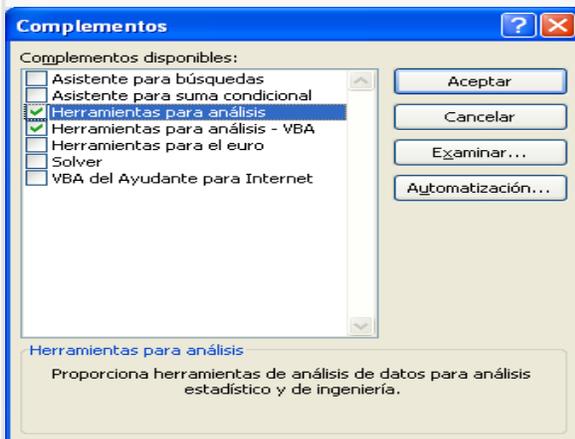
Paso 1



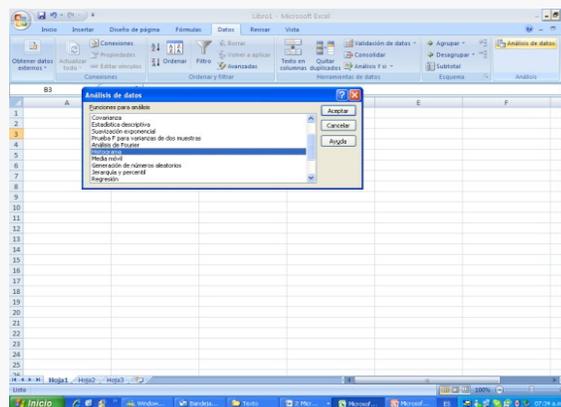
Paso 2



Paso 3



Paso 4



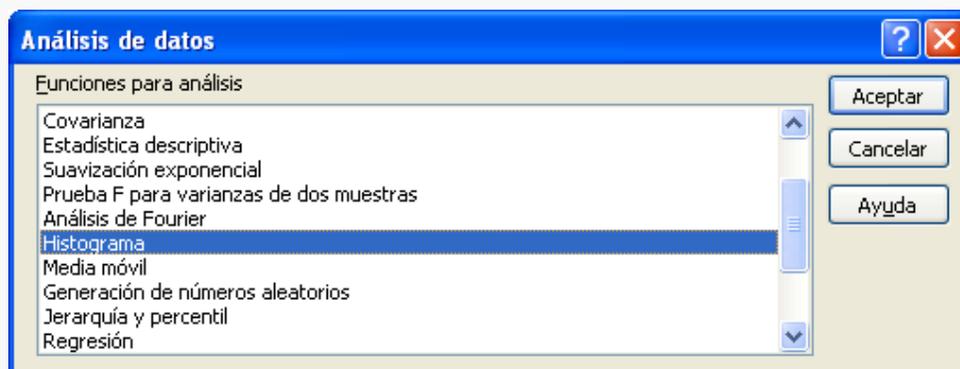
Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Con la opción Análisis de Datos, se puede calcular, además, de las frecuencias e histogramas, las estadísticas resumen (media, mediana, moda, desviación) y otros procedimientos más avanzados (Intervalos de confianza, Regresión, ANOVAS, Series de tiempo entre otros) que sin embargo, requieren de algunos supuestos que no están implementados en la hoja de calculo.

Para proceder, se activa la opción Análisis de Datos, del menú Datos, presentando las siguientes opciones.



Se aprecia que se pueden realizar procedimientos avanzados como: Análisis de Covarianza, Series de Tiempo (suavizado exponencial), Regresión y Correlación, además, de pruebas Paramétricas de diferencias de medias. Para nuestro objetivo inmediato, se utilizará la opción **Histograma** que permite realizar los conteos de una manera ágil y confiable. La ruta es:

Datos + Análisis de Datos + Histograma

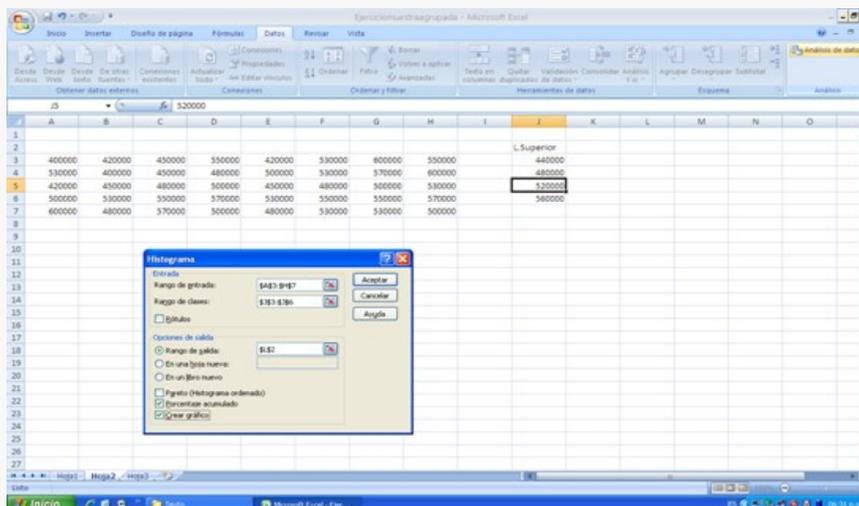
Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Se entiende como **rango de entrada** las celdas donde se encuentran los datos sin tabular, en el **rango de celdas** se colocan los límites superiores de cada intervalo, **rango de salida** es la celda donde se quieren los resultados. Si se selecciona la opción **Crear gráficos**, se construye el Histograma con barras separadas, lo cual implica unir las utilizando opciones luego de seleccionar las barras.

La siguiente figura ilustra como se procede para construir las tablas agrupadas. Reiterando que el rango de entrada hace referencia a los datos originales, rango de clases a los límites superiores de cada intervalo y el rango de salida al lugar donde se quiere que se visualice los resultados.



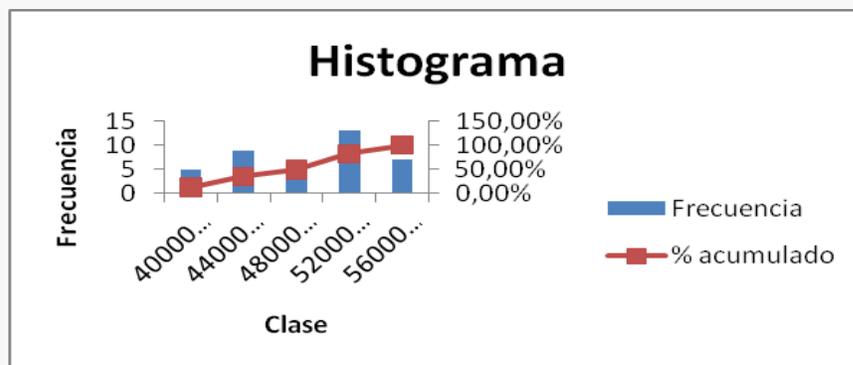
Para el ejemplo, se seleccionó los porcentajes acumulados y el gráfico. La respuesta del programa es la siguiente.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Clase	Frecuencia	% acumulado
440000	5	12,50%
480000	9	35,00%
520000	6	50,00%
560000	13	82,50%
y mayor...	7	100,00%



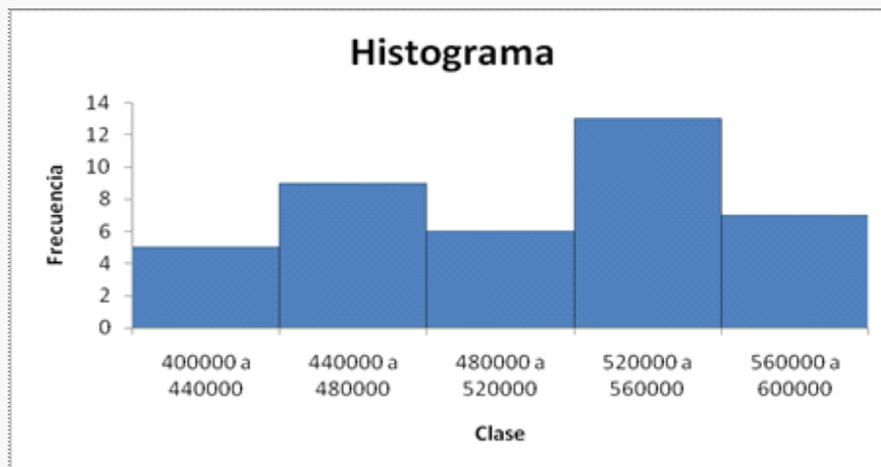
Es claro que no sólo es poco práctica sino inadecuada. Por lo tanto, se sugiere de manera manual, ajustar la forma de presentación, es decir, se construyen las dos columnas faltantes utilizando fórmulas en Excel y copiando los intervalos teniendo en cuenta que sea coherente con lo visto anteriormente, en la forma de construir los intervalos. El resultado es evidentemente más claro y entendible.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Intervalos salario	Frecuencia	Fre. Acumulada	Porcentaje	Porc. acumulado
400000 a 440000	5	5	12,5	12,50
440000 a 480000	9	14	22,5	35,00
480000 a 520000	6	20	15	50,00
520000 a 560000	13	33	32,5	82,50
560000 a 600000	7	40	17,5	100,00



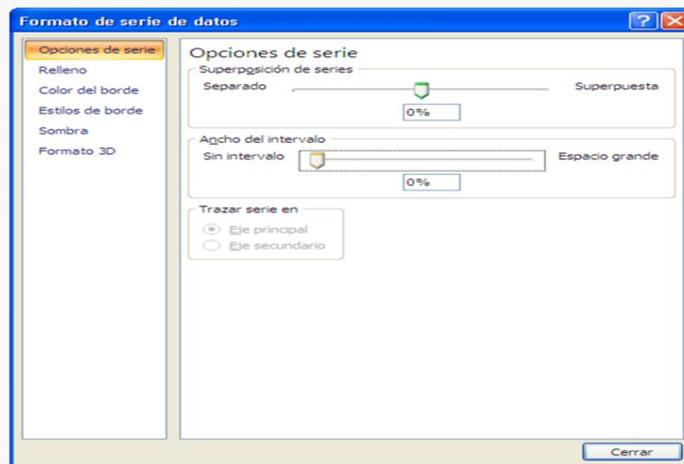
El histograma requiere de ajustes, dado que una de las características es que las barras sean unidas, ya que se utiliza para variables cuantitativas continuas.

Como se requiere de modificaciones para construir el histograma usando Excel, se colocan los pasos para adecuarlo.

1. Luego de tener la figura que el programa entrega y que se muestra en la parte superior, se eliminan las leyendas y la línea que une las barras separadas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

2. Se da clic en una de las barras, de tal manera que todas queden seleccionadas.
3. De clic en botón derecho del Mouse y coloque cero en la opción intervalo tal como se muestra a continuación.



3.4 REPRESENTACIÓN GRÁFICA.

Se conoce con el nombre de gráficos la representación geométrica que se hace de la magnitud de los datos de una serie estadística, de forma que estos datos queden representados cuantitativamente mediante el empleo de una escala adecuada de comparación entre unos y otros.

Los gráficos, constituyen uno de los medios de uso más difundidos para la presentación y el análisis de información estadística. Esto se debe a que las ideas presentadas gráficamente son entendidas con mayor rapidez y comodidad que las explicaciones numéricas y/o verbales.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Un gráfico para que sea de utilidad real debe ser sencillo y hacer énfasis en los rasgos más significativos de los datos, por tal razón no puede incluirse tanta información como en los cuadros. Podemos afirmar que la presentación tabular y la gráfica son un buen complemento para la presentación de datos. Es útil tener en cuenta que algunos gráficos requieren de una convención.

Existen diversos gráficos para representar datos estadísticos, tales como gráficos de barras (vertical u horizontal), segmentos o líneas, circular o pastel, de dispersión, diagrama de Pareto, pictograma, pirámides, cartograma etc. Sin embargo, nuestro objetivo es colocar a los lectores en capacidad de interpretar y elaborar gráficos de utilidad diaria y que además, son usuales en las diferentes investigaciones.

Se deben numerar y tener las mismas consideraciones que las tablas en cuanto a los títulos se refieren, es decir, deben ser claros e ilustrativos de donde se hizo el estudio y en que fecha. Los ejemplos presentados hacen referencia a trabajos asesorados por el autor.

Los gráficos se utilizan para:

- Explorar los datos (caja y sesgo, tallo y hojas).
- Visualizar la forma de la distribución de los datos (caja y sesgo, Histograma).
- Observar patrones o tendencias (líneas).
- Agrupar información por factores (caja y sesgo).
- Observar relaciones (nube de puntos).
- Comparar distribuciones (caja y sesgo).

Estadística Descriptiva y Distribuciones de Probabilidad.

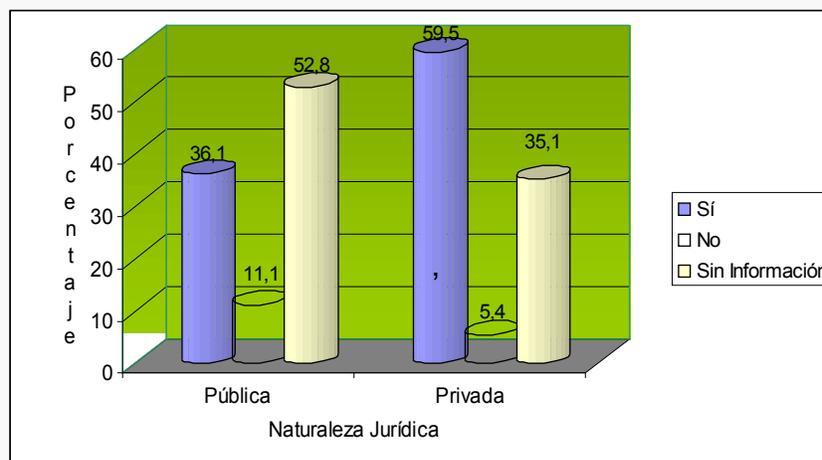
©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- Comparar medidas estadísticas (caja y sesgo)
- Identificar particularidades (barras, circular).
- Identificar fortalezas y debilidades en procesos (Espina de pescado, Pareto).

3.4.1 Gráficos simples: Se utilizan para presentar información de una sola variable, bien sea cualitativa o cuantitativa.

3.4.1.1 Gráfico De Barras: Es uno de los gráficos más simples para su elaboración y es usado con mayor frecuencia, se utiliza principalmente en la presentación de series cuantitativas si la variable toma pocos valores o cualitativas, si se tienen pocas categorías. En este tipo de gráficos se trata de construir barras paralelas dejando espacio entre ellas, colocadas en forma vertical u horizontal de tal forma que se puedan hacer comparaciones entre categorías de una misma variable.



Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

En la actualidad prima la estética, es decir, se tiene en cuenta en primera instancia que la figura sea entendible y por lo tanto, no debe existir una camisa de fuerza para colocar las barras de manera horizontal o vertical. La amplitud de las barras es constante y se deja un espacio entre ellas; la altura de cada barra corresponde a la respectiva frecuencia. Es importante anotar que las barras construidas de variables cualitativas se deben ordenar de acuerdo a su magnitud, mientras que en las barras verticales se mantiene el orden cronológico o cuantitativo de la serie. Gráfico 1: Distribución de las Instituciones según naturaleza jurídica y tenencia de Grupo de Salud Ocupacional, Antioquia mayo 2003.

3.4.1.2 Gráfico Circular o de Sectores: Esta clase de gráficos es muy utilizada por ser una forma clara y sencilla de representar diversas cantidades que han de compararse rápidamente. Además es útil, si lo que se quiere es destacar la importancia relativa de cada una de las categorías dentro del total, es decir, las porciones se presentan de manera porcentual.

Este gráfico consiste como su nombre lo indica en un círculo, el cual se divide en tantas partes como categorías se tienen, de manera que el área sea proporcional a la importancia relativa de cada categoría, no obstante, cuando se tienen muchas categorías (más de ocho) es mejor realizar el gráfico de barras.

Es importante tener en cuenta que en aquellos gráficos diseñados en tres dimensiones, se puede tener un efecto visual contraproducente, es decir, se puede confundir al lector en lugar de aclarar la información que se quiere entregar.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

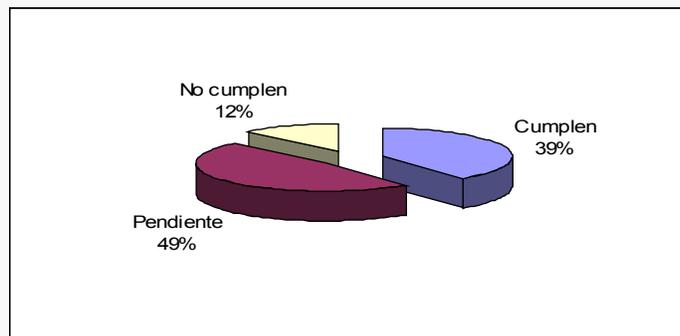


Gráfico 1: Distribución porcentual sobre el cumplimiento de los Grupos Interdisciplinarios de Salud Ocupacional, Antioquia Mayo de 2003.

Dada la variable cualitativa y pocas categorías, es la estética y la facilidad de entender la información lo que debe primar, a manera de ejemplo se presentan dos gráficos presentando lo mismo.

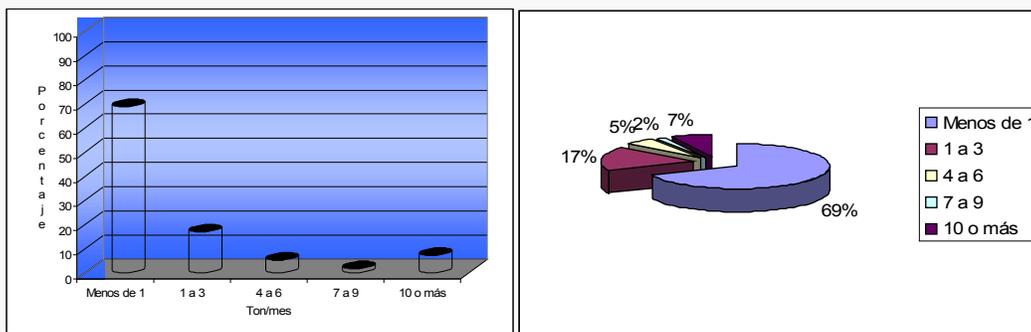


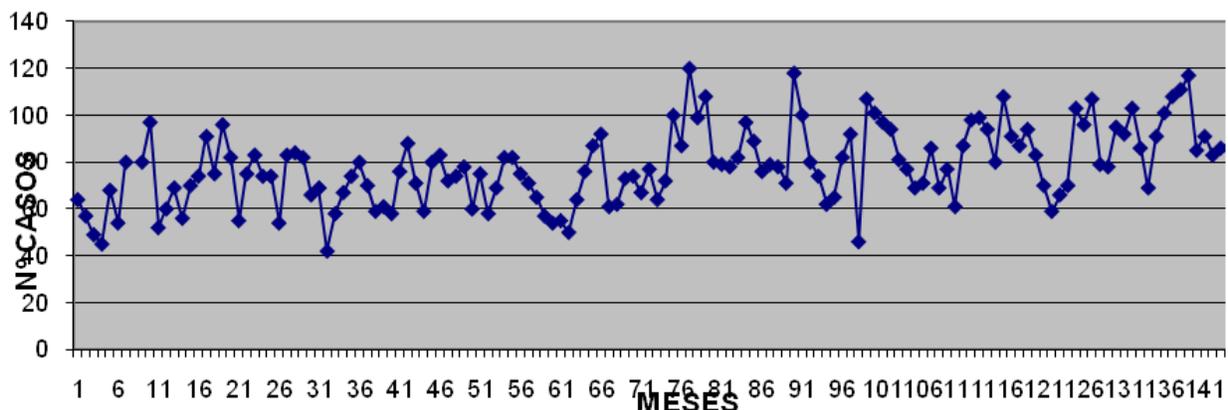
Gráfico 4: Distribución de la cantidad de carga transportada en Colombia vía aérea en promedio mensual por tonelada. Junio de 2004.

3.4.1.3 Gráfico de Segmentos o Siluetas: Este tipo de gráfico es utilizado principalmente para representar series de tiempo, permite representar varias

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

series en un mismo gráfico. La técnica para su construcción es similar a la utilizada por el gráfico de barras verticales, con la diferencia de que en lugar de representar el fenómeno con barras, se utiliza una línea formada uniendo puntos con segmentos de recta. Cada punto de origen de los segmentos de recta representa un par de valores ordenados de la tabla correspondiente. Es útil para comparar perfiles, por lo tanto, tienen una gran aplicación en investigaciones de mercados, donde es usual realizar comparaciones.

COMPORTAMIENTO DE LA INFECCION RESPIRATORIA AGUDA EN SANTA FE DE BOGOTA (1982-1992)



Fuente: Secretaría de Salud Bogotá D.C.

El gráfico presenta el comportamiento del número de casos de IRA en Santa Fe de Bogotá durante los años 1982 y 1992. Se aprecia un leve incremento, originado posiblemente por el crecimiento natural de la población, por esta razón, los Epidemiólogos prefieren trabajar con tasas, las cuales tienen en el denominador el número de personas en una determinada región y en un

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

tiempo dado.

3.4.1.4 Histograma de Frecuencias: El histograma es la manera de representar gráficamente las diferentes frecuencias (absoluta y relativa), de tal manera que se noten las características principales, considerando que la variable es de tipo cuantitativo y esta dada en intervalos. Sin embargo, el histograma más utilizado en la práctica es el porcentual.

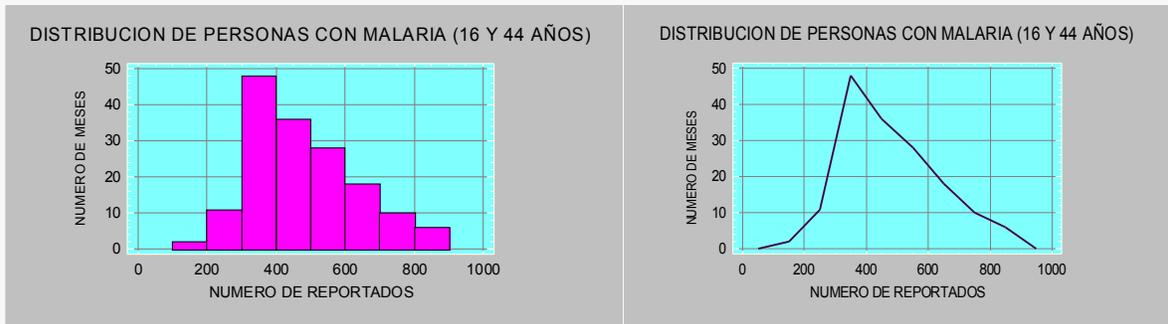
Un histograma es un gráfico de barras verticales, sin dejar espacio entre ellas, donde la altura de cada barra representa el valor de la frecuencia o porcentaje correspondiente, y la amplitud de la barra equivale a la amplitud del intervalo.

3.4.1.5 Polígono de Frecuencia: Un polígono de frecuencia es un gráfico de líneas, que se construye uniendo los puntos medios de cada intervalo con segmentos de recta.

El punto medio de cada intervalo se llama MARCA DE CLASE (M_c), y es el punto que representa a todos los de su clase; la marca de clase es igual a la semi-suma del límite superior y del límite inferior del intervalo o celda correspondiente.

A continuación se presentan el histograma, el polígono de frecuencia y la tabla correspondientes al número de personas con edades entre 14 y 44 años reportadas con malaria en Medellín durante el periodo 1986-1998 en periodos epidemiológicos (Un año es considerado con 13 periodos).

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.



Es claro que ambos gráficos proporcionan la misma información, será el investigador, dependiendo de sus objetivos o incluso de la parte estética la que se defina cual será el que se presente en el informe. Lo anterior implica, que sólo se requiere de una sola figura. Si se requiere comparar varias distribuciones, se debe construir el polígono.

En la práctica se prefiere construir el histograma con distribuciones discretas y el polígono de frecuencia con distribuciones continuas. Cuando dos o más distribuciones de frecuencias se comparan entre si, el polígono proporciona una visión mas clara. Se recuerda que la marca de clase es la semisuma de los límites de cada intervalo.

3.4.1.6 Diagrama de tallo y hoja: Tiene su principal aplicación en el análisis exploratorio de datos, ya que su forma es un gráfico de barras horizontal, que además, entrega información sobre como es el comportamiento de los datos al interior de cada intervalo. Viene implementado en los programas estadísticos más usuales en nuestro medio (SPSS, Statgraphics.). Toma su nombre, debido a que al dígito(s) significativos de la variable se le asigna el

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

nombre de tallo, y a partir de ahí, salen las hojas, es decir, las otras cifras significativas de la variable. El software determina e informa si encuentra valores extremos.

Los siguientes datos pertenecen a un índice de conocimiento evaluado sobre 100 puntos.

78 93 61 100 70 83 88 74 97 72
66 73 76 81 83 64 91 70 77 86

Para efectos del procedimiento, se define que las decenas hacen parte del tallo y las unidades serán las hojas, es decir, el número 78 se verá como 7 | 8. De esta manera las decenas se pondrán en una columna, en forma vertical, y las unidades a su derecha:

6	1 4 6
7	0 0 2 3 4 6 7 8
8	1 3 3 6 8
9	1 3 7
10	0

En los casos reales donde se tiene una gran cantidad de datos, es usual asignar una escala a cada hoja, es decir, si la escala es de 10, quiere decir que cada hoja equivale a 10 casos, dicho de otra manera, al encontrar dos veces cero, en realidad representan 20 casos, 10 cada uno.

Es importante resaltar, que la representación da una idea clara de que forma tienen los datos, valga decir, si la mayoría de ellos se encuentran al principio o al final de los valores, pero además, dentro de cada intervalo nos damos

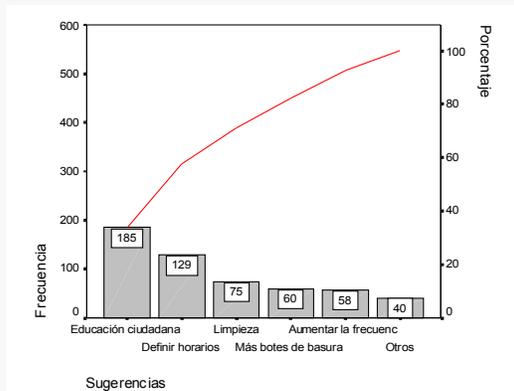
ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

cuenta que valores tienen mayor frecuencia. Para el ejemplo, los mayores puntajes están entre 70 y 79 puntos.

Este tipo de gráficos es práctico siempre y cuando se cuente con el software adecuado para construirlo.

3.4.1.7 Diagrama de Pareto: Su origen se remonta a 1987 debido al economista V. Pareto, el cual planteó que la mayoría de dificultades son originadas por muy pocas causas, no obstante, hoy en día, éste gráfico se aplica en innumerables situaciones, manteniendo eso sí, que la variable en lo común es cualitativa nominal o categórica, lo que permite ordenar descendientemente las diferentes categorías de la variable, tal como se construye el gráfico de barras, la diferencia radica en que adicionalmente a esa información, se calcula el porcentaje acumulado y se coloca en la margen derecha de la figura, de tal manera, que se pueda identificar que porcentaje se presenta por dos o más categorías de la variable. Es útil al identificar fortalezas y debilidades en los procesos de autoevaluación de procesos, además, en los procesos de control de calidad en las empresas.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

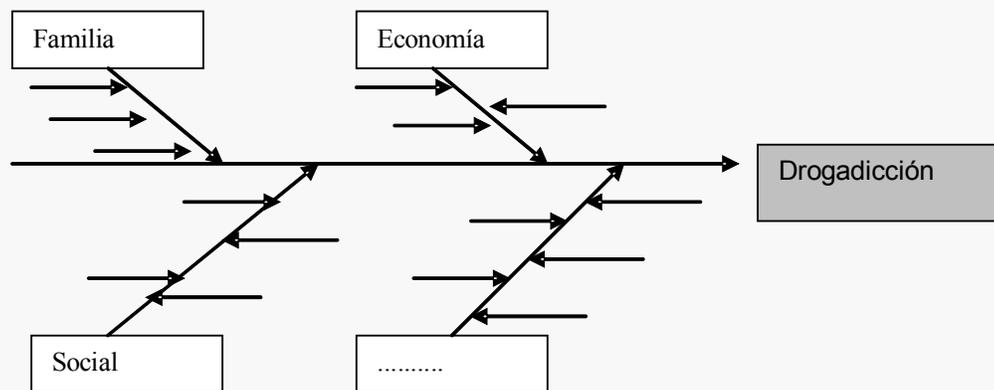


El gráfico muestra como la educación ciudadana y la definición de horarios es lo más mencionado como sugerencias para mejorar el servicio de recolección de información, pero además, el otro eje permite ver como aproximadamente el 73% de las sugerencias se concentran en las tres primeras categorías.

3.4.1.8 Diagrama de causa efecto o Espina de Pescado: Se usa como paso previo al gráfico de Pareto, ya que permite identificar de entre múltiples razones cuales pueden ser las más críticas. Es útil seleccionar una parte del proceso bajo estudio, para ir determinando las causas relevantes, aplicarles Pareto y de esta manera lograr avanzar a todo un proceso. Su forma toma la figura de la espina de pescado, de ahí su nombre.

Se quiere determinar cuáles son las causas del incremento de la adicción a sustancias Psicoactivas, un diagrama incipiente podría ser.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



De cada flecha surgirán otras causas y de estas otras, hasta agotar todas las posibles razones para luego determinar cual es la prioritaria y enlazar el análisis con el diagrama de Pareto.

Este diagrama requiere de actualización permanente.

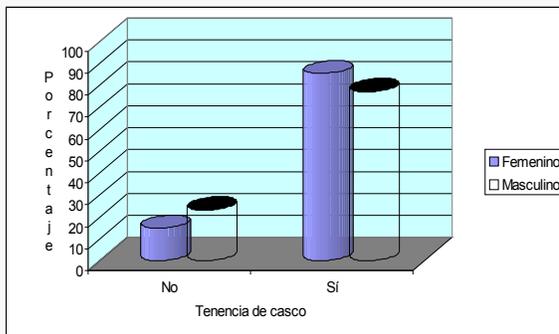
Para construir un gráfico de Causa-Efecto se debe tener en cuenta:

1. Decidir cual es el factor que se desea analizar. Trazamos un flecha gruesa que representa el proceso y a la derecha se coloca el factor:
2. Se indica que origina o influye en el factor objeto de estudio, colocando los más generales y dibujando flechas secundarias con otras posibles razones de cambio en el factor principal.
3. Se incorporan más causas y con más detalle. Así seguimos ampliando el Diagrama de Causa-Efecto hasta que contenga todas las causas posibles de dispersión.

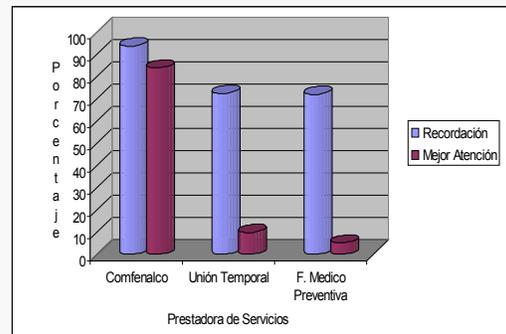
ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

3.4.2 GRAFICOS COMPUESTOS: Presentan información de varias variables, bien sean cualitativas o cuantitativas o una combinación entre ellas.

3.4.2.1 Barras Comparativas: Este gráfico es de gran utilidad para la ilustración de los fenómenos demográficos y económicos y en general en las diversas áreas del conocimiento, por cuanto permite estudiar una serie de detalles que a la vista de los valores numéricos (tablas) sería difícil darnos cuenta, permite observar de una manera simple la relación existente entre dos o más valores del mismo fenómeno, de tal forma que toda diferencia o similitud entre ellos quede plenamente establecida, permitiendo de una simple mirada efectuar la comparación.



Relación entre el genero del conductor y la tenencia de casco. Año 2001, Municipio de Neiva



Comparación porcentual de recordación y atención al usuario. Año 2004, Antioquia

Los gráficos se construyeron con Excel y utilizando datos de diversas investigaciones, de ahí, que la numeración y los títulos no tengan necesariamente secuencialidad. Se destaca además, que las variables

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

involucradas en la presentación para este caso son cualitativas.

3.4.2.2 Pirámide Poblacional: Se utiliza para comparar el comportamiento de dos variables en un momento dado, es muy usual en demografía, ya que permite ver el comportamiento de la población según género y edad. Si bien es cierto que el sexo masculino se usa en la izquierda, no es necesario hacerlo. Además, este tipo de gráficos se puede utilizar en otras áreas del saber, siempre y cuando el objetivo sea comparativo.

Observe como se construye la pirámide usando el SPSS, entre al enlace [Piramide SPSS](#).

Para construir el ejemplo, se tomaron los datos de la información del DANE (Departamento Nacional de Estadística) que en Colombia es quien se encarga de la información censal. Los datos corresponden al censo 2005.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Tabla No **: Población total censada, por sexo y grupos de edad, total nacional 2005.

Tabla ***: Población total censada, por sexo y grupos de edad, total nacional 2005.

Edad	Hombre	Mujer
0 a 4	2106179	2002682
5 a 6	2197689	2098224
10 a 14	2214464	2124582
15 a 19	1975856	1957898
20 a 24	1783320	1858519
25 a 29	1590993	1689774
30 a 34	1404139	1516151
35 a 39	1392512	1526649
40 a 44	1304948	1427556
45 a 49	1088238	1203070
50 a 54	876301	959039
55 a 59	692733	757925
60 a 64	524576	580157
65 a 69	428876	492178
70 a 74	321765	380753
75 a 79	228608	275830
80 a 84	121846	157029
85 y más	86074	124251

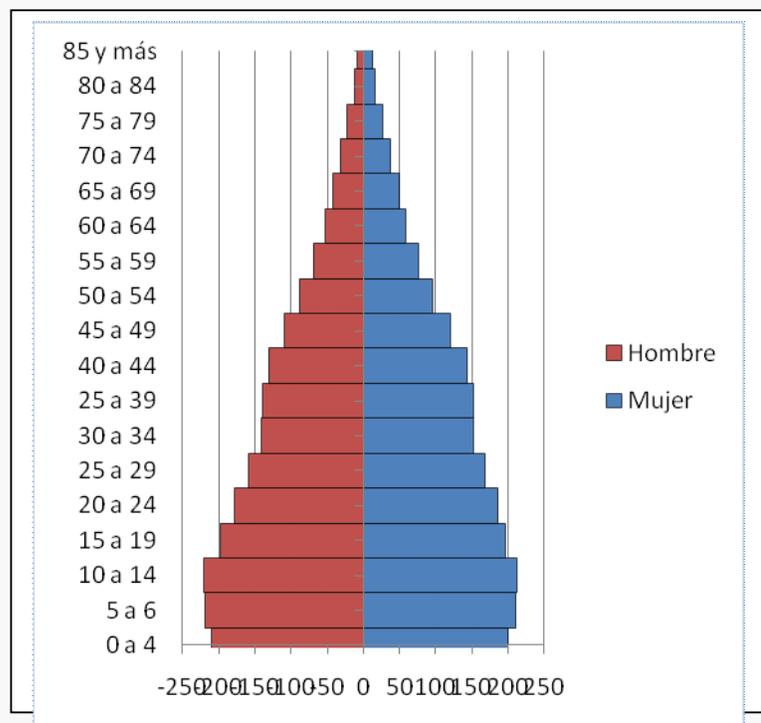


Gráfico ***: Comparación del total nacional 2005 de personas por 10.000 según sexo y grupos de edad total.

Se aprecia que hay un mayor número de hombres hasta los 19 años, para equilibrarse el número de personas por sexo en la mitad de la pirámide y finalizar con más mujeres en las edades mayores.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

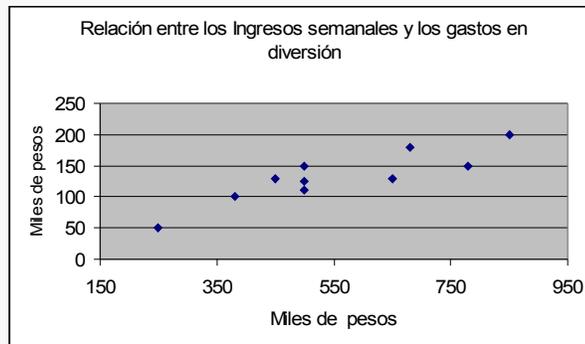
Como se requiere de alguna modificación para construir el gráfico de pirámide usando Excel, se colocan los pasos para elaborarla.

1. Se digita los datos de: edad, hombres y mujeres, en ese orden.
2. Para que los datos de los hombres queden a la izquierda, se digitan y se coloca formato de moneda, seleccionando la opción rojo, se requiere digitar los números precedidos del signo menos. Todos los números de esa columna quedan en rojo.
3. Se seleccionan las tres columnas y se da clic en gráficos de la barra de herramientas, escogiendo el de barras horizontales. Se seleccionan las barras de uno de los lados.
4. Clic en botón derecho y se coloca cero en la opción intervalo y 100 en superpuestas.
5. Si se quiere algún diseño especial, se puede dar formato a la figura creada.

3.4.2.3 Diagrama de Dispersión o Nube De Puntos: Se utiliza cuando se quiere determinar la relación entre dos variables cuantitativas, es un paso obligado para llevar a cabo el procedimiento de regresión que se verá más adelante. Por ahora, es importante decir, que se construye uniendo parejas ordenadas en un plano cartesiano y se pretende ver la existencia de una relación entre las variables objeto de análisis. Por ejemplo, determinar si hay relación lineal entre el ingreso familiar semanal y los gastos en diversión.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ingreso (miles)	Gastos.
500	125
850	200
250	50
500	150
650	130
780	150
500	110
380	100
450	130
680	180

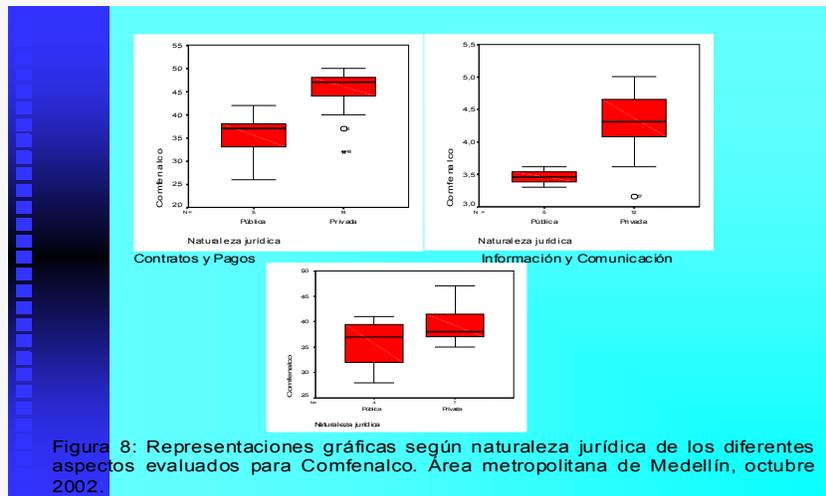


El gráfico muestra una posible relación lineal y directa, es decir, a mayores ingresos, se esperaría mayores gastos en diversión.

3.4.2.4 Caja y Sesgo: Compara los resultados de una variable cuantitativa con una cualitativa, además, de identificar en las variables cuantitativas valores extremos o atípicos, se determina la forma de los datos, es decir, si estos tienen un comportamiento aproximadamente normal o si por el contrario, son asimétricas, éste hecho es importante, en la medida que da pautas para definir cuál o cuáles son los indicadores adecuados para darle sentido a la información.

Se utiliza para revisar los datos de entrada y validar la pertinencia de algunos valores y definir su inclusión o no dentro de la base de datos definitiva, ya que con esa base, se calculan los indicadores definitivos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



Cuando las cajas no están en un mismo nivel, se puede afirmar que hay diferencias en las categorías analizadas. Para el gráfico anterior, es el caso para el recuadro superior izquierdo, mientras que el recuadro inferior, a pesar de que no están exactamente en el mismo horizonte, se puede afirmar que los cambios no son significativos. Los puntos que se encuentran fuera de los bigotes son llamados valores atípicos y por lo tanto, requieren de un análisis especial. El recuadro superior derecho, muestra como los datos representados por la caja de la izquierda es muy homogénea y con forma normal, mientras que la segunda caja permite inferir que los datos son aproximadamente normales pero con una variabilidad alta.

Para ilustrar el criterio técnico para identificar los valores atípicos y extremos, se utiliza un ejemplo de una investigación realizada para describir algunas variables de los suelos en Medellín, Colombia, dentro de las cuales se utilizó la variable Redox. los criterios son: si un valor se encuentra más alejado de $1.5 \cdot$ Rango intercuartílico del tercer cuartil o del primero,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

entonces el valor es poco común. Si el valor es más alejado de 3 * Rango intercuartilico del tercer cuartil o del primero, el valor se considera extremo. El cálculo de los cuartiles se verá en el capítulo siguiente:

Rango intercuartilico = Cuartil 3 – Cuartil 1

Además, el cuartil 3 equivale al 75% de los datos y el cuartil 1 al 25%

A modo de ilustración, para el promedio Redox, se encontró lo siguiente en el estudio ya mencionado:

$$RI = Q_3 - Q_1 \quad RI = 482,07 - 397,87 = 84,2$$

$$\text{Valor atípico por encima: } 482,07 + (1.5 * 84,2) = 482,07 + 126,3 = 608,37$$

$$\text{Valor atípico por debajo: } 397,87 - (1.5 * 84,2) = 397,87 - 126,3 = 271,57$$

$$\text{Valor extremo por encima: } 482,07 + (3.0 * 84,2) = 482,07 + 252.6 = 734,67$$

$$\text{Valor extremo por debajo: } 397,87 - (3.0 * 84,2) = 397,87 - 252.6 = 145.27$$

Cualquier valor por encima de 608.37 pero menor de 734.67 es considerado atípico por encima. Si es mayor de 734.67 es considerado extremo, igual reflexión se realiza para el lado inferior.

Otra muestra de estos gráficos es:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

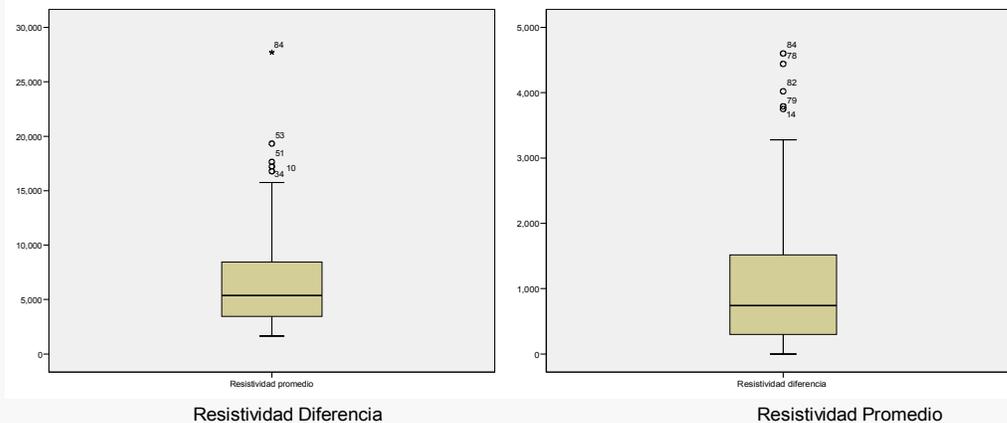


Figura 1: Distribución de la resistividad promedio y la diferencia a 2 y 3 metros de profundidad.

El comportamiento en ambos casos, muestra una asimetría positiva, es decir, hay una mayor concentración de mediciones en los valores bajos de la escala, sin embargo, en ambos casos, se presentan valores atípicos por la parte superior, es decir, hay pocas mediciones con valores muy altos.

3.5 IMPORTANCIA DE LA ESCALA.

Para obtener una buena unidad de medida, debemos tener en cuenta que no debe ser muy pequeña, para poder ver con claridad y exactitud sus características principales. Las escalas elegidas para los ejes suelen ser diferentes, pero teniendo en cuenta que no existe mucha desproporción entre ellas, ya que se podría desvirtuar la representación.

Los mismos datos pueden dar la impresión de dos comportamientos diferentes, debido al cambio de escala, por ejemplo, obsérvese los gráficos

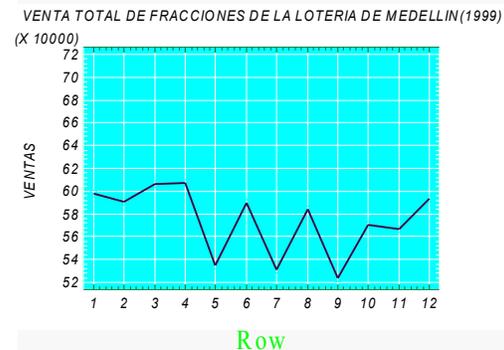
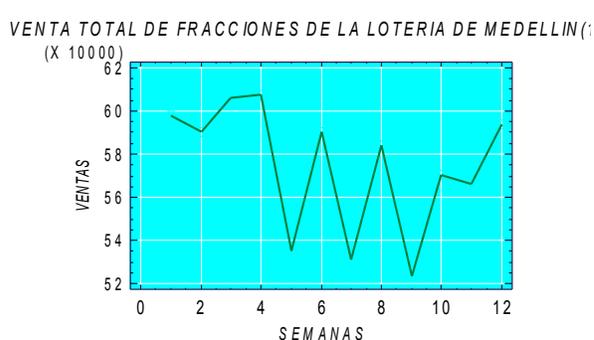
ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

de la página siguiente tomados a diferente escala.

Como se puede apreciar en ambos gráficos se representa las ventas de fracciones de lotería en 1999 por semana, sin embargo, a primera vista nos da la idea de que son dos representaciones diferentes; donde en el gráfico b) las ventas de fracciones de la lotería muestran una mayor estabilidad que en gráfico a), lo que nos lleva a pensar, que los mismos datos pueden ser utilizados para mostrar situaciones de beneficio particular.

Otro tipo de escalas son los utilizados para los llamados gráficos logarítmicos y semi - logarítmicos. Se llaman gráficos logarítmicos cuando la escala de los dos ejes es logarítmica, y se utiliza para representar una función exponencial.

Ahora, cuando uno de los ejes tiene escala natural y el otro escala logarítmica se llama gráfico con escala semi - logarítmica.



Las figuras anteriores, se construyeron con los mismos datos, no obstante,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

presentan una visión diferente de la temática en estudio. El de la derecha muestra un comportamiento más estable, mientras que el de la izquierda, da a entender que la venta de fracciones de la lotería de Medellín presenta muchos altibajos.

Luego de haber leído sobre gráficos, te invito a redondear las ideas viendo este vídeo sobre el tema, considero te aclararán las posibles dudas que tengas, enlace: [Gráficos](#).

Ejercicios.

1. Una muestra aleatoria de personas empleadas, nos muestra cuánto tiempo en minutos transcurre hasta llegar a su lugar de trabajo.

29.9	27.0	27.0	20.0	23.2	29.9	25.3	23.2	25.3	20.0
22.5	25.3	20.1	22.5	22.5	28.2	22.5	28.2	27.0	29.9
20.0	25.7	24.0	20.1	20.0	23.2	29.9	22.5	28.2	28.2
29.9	27.0	27.0	20.0	23.2	29.9	25.3	23.2	25.3	20.0

Agrupe en 5 intervalos utilizando G-stat o Excel y responda los siguientes interrogantes:

- Al menos cuantos carros recorren 24 millas/galón.
- ¿Qué porcentaje de vehículos recorren 27 millas/galón?
- Construya el histograma de frecuencias e interpretelo.

2. Los siguientes datos corresponden al número de camas disponibles en un hospital por día.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

67	75	63	62	65	65	65	62
75	72	61	61	61	65	75	69
70	65	61	72	65	60	70	69
61	78	62	67	65	60		

Usando un software estadístico responda los siguientes interrogantes:

- El gerente decide colocar solamente 68 camas disponibles por día. ¿Cuál será el porcentaje en que no se pueda satisfacer la demanda?
- ¿Está de acuerdo con la decisión del gerente? ¿Por qué?
- Construya un gráfico adecuado.

3. Los aspirantes a ocupar la dirección en el hospital San Juan de Mitú requieren tener como mínimo 270 puntos en las pruebas exigidas para tal cargo. La información obtenida se presenta a continuación. Se quiere agrupar en seis intervalos y calcular las diferentes medidas resumen.

250	260	278	356	209	265	250	230	240	245	260	250	211	270	280
215	286	312	287	324	321	212	294	275	200	188	314	250	260	198
275	210	256	235	287	298	265	250	230	240	245	285	291	287	265

- ¿Cuántas personas no pueden aspirar al cargo? Realice el conteo de manera manual y usando software.
- ¿Qué porcentaje de aspirantes sacaron menos de 213 y más de 300?
- Agrupe en 5 intervalos, interprete y construya el gráfico adecuado.

4. Almacenes ÉXITO, con el fin de realizar un estudio sobre las secciones que presentan mayor movimiento, realizó un conteo del número de

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

transacciones por día y dependencia, encontrando lo siguiente:

ROPA HOMBRE	300
ROPA MUJER	1500
ROPA BEBE	600
ROPA DE HOGAR	200
FARMACIA Y MISCELANEA	350
VARIOS Y ELECTRODOMÉSTICOS	50

Dibuje dos gráficos que sean adecuados a los datos anteriores y explique por qué esos son los que se deben presentar. Coloque los títulos adecuados para cada gráfico.

5. Al aplicar tres tipos de drogas a 18 pacientes con el fin de determinar cuál de ellas actúa con mayor rapidez, se obtuvieron los siguientes resultados, en minutos:

Paciente	Medicamento A	Medicamento B	Medicamento C
1	20	15	10
2	15	15	12
3	16	13	10
4	14	12	8
5	15	13	10
6	18	15	12

a) Entregue un informe gráfico de los datos anteriores, ¿cuál presentaría usted? ¿Por qué?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

b) ¿Cómo haría un gráfico que muestre la importancia relativa de los tiempos de cada droga?

6. Durante un mes se obtuvo información sobre el pago de consultas en un consultorio médico por día, en miles de pesos, los resultados en bruto son los siguientes:

15	8	10	15	12	20	14	14	16	8	9	10	12	8	10
15	14	10	12	8	10	12	15	14	10	12	14	12	10	10

- Identifique el tipo de variable y su nivel de medición.
- Ordene los datos en forma de frecuencia simple e interprete un valor de cada columna. Coloque el título acorde con la información presentada.
- Lea la tabla, es decir, interprete los resultados encontrados.
- Agrupe en cuatro y cinco intervalos, interprete un valor de cada columna.
- Construya el gráfico adecuado según la presentación de los datos en la parte c. Tanto para 4 como para 5 intervalos.

Interprete el gráfico en términos del enunciado.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

CAPÍTULO 4

MEDIDAS RESUMEN O DESCRIPTIVAS CON EXCEL y SPSS

"El Pensar requiere tiempo, y lo pensado ejecución"
Quevedo

En los capítulos anteriores, se ha podido observar la forma de la presentación de datos a través de tablas y gráficos; los cuales nos dan una buena idea sobre el comportamiento y la forma de la(s) variable(s) en estudio, sin embargo, es de vital importancia resaltar algunas características cuantitativas y/o cualitativas de las distribuciones de frecuencia, ya que serán de gran utilidad para la descripción de la población o muestra, además de facilitar la interpretación y análisis de los datos.

Para resaltar tales características y observar sus particularidades se utilizan las llamadas medidas resumen o estadísticas descriptivas, estas se clasifican así:



Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Si bien es cierto, las medidas de tendencia central están inmersas dentro de las de posición, para efectos prácticos, se presentan en forma separada.

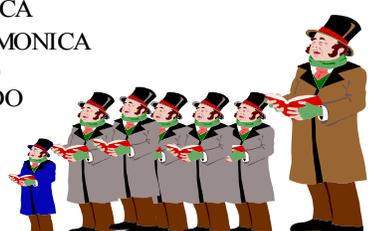
Como un complemento importante se coloca el enlace a un vídeo de 1 hora y 43 minutos donde se conversa sobre los temas introductorios de la estadística descriptiva, fortaleciendo algunos que ya se explicaron antes y otros que se verán a continuación, además, de una pequeña introducción al SPSS, el enlace es [Conceptos Generales Descriptiva](#).

4.1 MEDIDAS DE TENDENCIA CENTRAL.

Para que sean de utilidad real, deben ser representativas de un grupo homogéneo en el cual los miembros se puedan reconocer como similares, las más usadas son:

MEDIDAS DE TENDENCIA CENTRAL

- En general se busca eliminar el problema de valores extremos.
- **MEDIA ARITMETICA**
- MEDIA GEOMETRICA
- MEDIA ARMONICA
- PROMEDIO RECORTADO
- **MEDIANA**



Materia preparada por:
Profesor León Darío Bello Pumas

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

La presentación en Power Point se encuentra en el curso de estadística descriptiva al cual puede acceder desde la página www.leondariobello.com.

Adicional a las anteriores, existen otras, sin embargo, no han ganado el suficiente respaldo práctico para tratarlas en este texto. El criterio básico para definir cuál o cuáles son las medidas a calcular tiene que ver con el problema de minimizar los valores atípicos y/o extremos. Otro comentario importante, es que en general los promedios son de tendencia central, si los datos se comportan aproximadamente acampanados.

4.1.1 Media Aritmética: Es la medida de tendencia central más utilizada y generalizada, también se conoce con el nombre de promedio aritmético. No obstante, en algunos casos no se recomienda, valga decir, cuando los datos presentan mucha variabilidad, en estos casos se sugiere segmentar o estratificar los datos de acuerdo a alguna variable de identificación (estrato, profesión, sexo, escolaridad etc.). Por lo anterior, siempre que se trabaje la media aritmética, se requiere de calcular e interpretar una medida de variabilidad, las cuales se analizarán más adelante.

La media aritmética se denota \bar{x} (se lee x- barra), se calcula de la siguiente forma.

$$\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$$

pero como $x_1 + x_2 + \dots + x_n = \sum x_i$ entonces.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Ejemplo 1.

Supóngase que el número de reclamos recibidos por el defensor del cliente en una entidad financiera durante una semana fue el siguiente:

20, 20, 19, 21, 22, 21

Nos piden calcular la medida aritmética.

$\bar{x} = 20 + 20 + 19 + 21 + 22 + 21 / 6 = 20.5$ En promedio se reciben 21 reclamos por semana en esa entidad.

En Excel la función es: =Promedio(Rango de los datos)
Para el caso sería: =Promedio(A3:A8). Sí los datos se encuentran en esas celdas.

Cuando se tiene una distribución de frecuencia simple, no unitaria, la media aritmética, o media ponderada el cálculo es el siguiente:

Primero se multiplica cada uno de los datos (x_i) por su respectiva frecuencia (f_i), luego se suman dichos productos y la suma se divide por el número de observaciones.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{n}$$

Donde m es el número de categorías de la variable.

Ejemplo 2.

Dada la siguiente distribución, cuya variable x_i representa el número de personas reportadas a Data crédito en una semana, calcular la media aritmética.

x_i	f_i	$x_i f_i$
6	8	48
7	3	21
8	6	48
10	4	40
12	3	36
14	3	42
15	2	30
n = 29		265

$$\bar{x} = \frac{\sum_{i=1}^7 x_i f_i}{n} = \frac{265}{29} = 9.14$$

El número promedio de personas reportadas por semana a Data crédito es de aproximadamente 9. A continuación se debe hacer referencia a la variabilidad de los datos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Cuando se tiene una muestra agrupada por intervalos, para calcular la media aritmética, se busca un valor que represente a todos los valores de la celda o intervalo. Este valor va a ser siempre el que se definió anteriormente como marca de clase y lo llamamos Mc , donde:

$$Mc_i = \frac{ls_i + li_i}{2}$$

$$\bar{x} = \frac{\sum Mc_i f_i}{n}$$

Ejemplo 3.

Determine el promedio aritmético del número de niños con dificultades en el aprendizaje de una segunda lengua, para ellos se obtuvo la siguiente información:

Nº niños					
Li	La	Mc	f _(i)	Mc _(i)	f _(i)
5	a	10	7.5	4	30
10	a	15	12.5	6	75
15	a	20	17.5	8	140
20	a	25	22.5	6	135
25	a	30	27.5	4	110
			n = 28	490	

$$\bar{x} = 490 / 28 = 17.5$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

4.1.1.1 Interpretación Media Aritmética. Hasta ahora se ha hecho énfasis en mostrar cómo se encuentra la media aritmética, según la entrada de los datos. Sin embargo, más que hallar el valor numérico de la media es necesario saber interpretar el valor representativo de una serie de datos, o el valor promedio de estos.

Hay que saber cuando se puede utilizar la media aritmética; ya que esta medida está afectada por valores extremos, es decir, si existe algún dato muy diferente a los demás, nos puede dar un resultado totalmente erróneo para la labor estadística. Por ejemplo, supongamos que las edades de cinco personas que componen una familia son las siguientes; 60, 20, 24, 19, 27 al calcular la media aritmética nos da:

$$\bar{x} = 60 + 20 + 24 + 19 + 17 / 5 = 28$$

En este caso es un error estadístico afirmar que las personas que componen esta familia tienen en promedio una edad de 28 años, ya que esta es afectada por un valor muy grande (extremo) 60. Este problema se detecta, además, calculando la desviación estándar que se presenta posteriormente y se debió percibir al realizar el análisis exploratorio de datos.

4.1.1.2 Propiedades Útiles de la Media Aritmética.

- La media aritmética de una constante es igual a la constante.
- La media aritmética del producto de una constante por una variable es igual al producto de la constante por la media de la variable.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- La media aritmética de una variable más o menos una constante es igual a la media de la variable más o menos la constante.
- Cuando una población o muestra se subdivide en sub muestras, entonces, la media total es igual a la media ponderada de las medias de las sub muestras.

$$\bar{x} = (\bar{x}_1 f_1 + \bar{x}_2 f_2 + \dots + \bar{x}_m f_m) / n$$

Una síntesis de lo que se pretende con la media aritmética se presenta en la diapositiva siguiente, donde se pretende mostrar como un valor extremo afecta de manera radical el valor del promedio.



El primer promedio se calculó teniendo en cuenta que el segundo valor del dólar era de \$2771, mientras que para el segundo promedio se trabajó con

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

\$6571, presentándose un cambio grande en el resultado, siendo por lo demás erróneo, ya que exceptuando un día todos los valores del dólar estuvieron por debajo del promedio, situación que lleva a concluir que el promedio aritmético no es el apropiado en este caso.

4.1.2 Mediana (Me): Esta medida depende únicamente de la posición que tenga en la tabla; por lo tanto, no está afectada por valores extremos. La mediana, que se denota Me, va a ser siempre un valor central, es decir, va a superar no más del 50% de la información y a su vez será superada por el otro 50% restante. Debe tenerse presente que los valores estén ordenados de menor a mayor.

Independiente del número de observaciones y de la forma de presentar los datos (frecuencia simple o agrupada), se identifica en que lugar se encuentra la mediana, para ello se utiliza la siguiente expresión:

$$Me = X_{((n+1)/2)}$$

Ejemplo 1. Hallar la mediana del número de solicitudes de créditos rechazados mensualmente, si los datos son:

20, 24, 30, 26, 18

Primero se ordenan los datos ascendentemente. Se utilizan los subíndices entre parentesis para indicar que los valores de la variable están ordenados.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$X_{(1)}$ $X_{(2)}$ $X_{(3)}$ $X_{(4)}$ $X_{(5)}$
18, 20, 24, 26, 30

$Me = X_{((n+1)/2)} = X_{((5+1)/2)} = X_{(3)} = 24$. Lo que quiere decir que, el 50% de los créditos rechazados mensualmente equivale a 24 o menos.

Si se cambia el valor de 20 por 2, la mediana no sufre alteraciones, ya que el 24 se mantiene en la mitad de la distribución, lo que quiere decir, que esta medida no está afectada por valores extremos y por lo tanto se recomienda en esos casos.

En Excel la función es: =Mediana(Rango de los datos)

Para el caso sería: =Mediana(B3:B7). Si los datos se encuentran en esas

Ejemplo 2. Los salarios de los empleados en una multinacional son los siguientes en un mes.

Salario		Frecuencia
(mínimos)	Frecuencia	Acumulada
2	25	25
3	30	55
4	24	79
5	26	105
6	15	120
7	10	130
8	5	135
9	2	137
	137	

Cuando se tienen los datos en forma de frecuencias, se requiere de la frecuencia absoluta acumulada, lugar donde se ubica la posición de la mediana, utilizando la misma fórmula, así:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$Me = X_{((n+1)/2)} = X_{((137+1)/2)} = X_{(69)} = 4.$$

Como el número 69 no se encuentra en la frecuencia acumulada, se aproxima por exceso, correspondiéndole el 79. Luego la mediana es igual a 4 y se interpreta así: Aproximadamente el 50% de los empleados de la multinacional tienen un salario de 4 mínimos o menos, es claro, que también se puede redactar: Aproximadamente el 50% de los empleados de la multinacional tienen un salario de 4 mínimos o más.

Para los casos en el que el tamaño de muestra es par, se puede utilizar la misma fórmula, el cambio radica en que se promedian los dos valores centrales.

Ejemplo 3. La facturación realizada por ocho instituciones del sector salud, en millones de pesos, fueron los siguientes:

$$\begin{array}{cccccccc} X_{(2)} & X_{(4)} & X_{(3)} & X_{(5)} & X_{(7)} & X_{(1)} & X_{(6)} & X_{(8)} \\ 135, & 155, & 135, & 200, & 450, & 120, & 250, & 600 \end{array}$$

Si se aprecian los datos, estos presentan alta variabilidad, ya que cambian de manera importante de institución a institución. Por lo tanto, se recomienda calcular la mediana.

$$Me = X_{((n+1)/2)} = X_{((8+1)/2)} = X_{(4.5)} = (155+200)/2 = 177.5$$

Como se observa cuando el número de datos es par, la mediana es el

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

promedio de los dos valores centrales, para el caso los ubicados en las posiciones 4 y 5.

Una manera de disminuir la variabilidad, es agrupar por tipos de institución, valga decir, pública y privada o nivel I, II y III.

El cálculo de la mediana cuando los datos están agrupados por intervalos, se trabaja igual, solamente que se requiere calcular la marca de clase y dicho valor será el representativo del intervalo que contiene la mediana.

4.1.3 Media Armónica (Ma): La media armónica es la inversa de la media aritmética de los valores recíprocos. Se recomienda cuando se requiere utilizar el tamaño promedio de un grupo cuando los tamaños de las muestras de los grupos no son iguales, además, cuando se tiene una relación inversa, valga decir, tiempo y velocidad. La fórmula viene dada por:

$$M_a = \frac{1}{\sum_{i=1}^n \frac{1}{x_i} * f_i}$$

El Excel la calcula usando la siguiente expresión:

MEDIA.ARMO (número1;número2; ...)

MEDIA.ARMO (A10:A30) Calcula la Ma de los datos ubicados entre las filas A10 y A30.

La media armónica es siempre inferior a la media geométrica, que a su vez

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

es siempre inferior a la media aritmética. Tiene su principal aplicación en el área de las ingenierías.

4.1.4 Media Geométrica (Mg): Suaviza un poco el efecto de los valores extremos, por lo tanto, siempre será menor o igual a la media aritmética. Su principal aplicación se da: 1) Para promediar porcentajes, índices y cifras relativas, 2) Para determinar el incremento porcentual promedio en ventas, producción u otras actividades o series económicas de un periodo a otro y 3) Para calcular la tasa de crecimiento promedio, dado un interés compuesto por tasas variables. Claro está, que en los anteriores casos, se supone que el comportamiento de los datos sigue un comportamiento geométrico. En todos los casos éste indicador es menor o igual al promedio aritmético.

Su cálculo se puede hacer utilizando la raíz de orden n del producto de los valores de la variable elevados a su respectiva frecuencia. La ecuación para la media geométrica es:

$$Mg = \sqrt[n]{X_1 * X_2 * X_3 * \dots * X_n}$$

No obstante, no se podría calcular cuando se tienen valores negativo o un valor de cero. Por lo tanto, se tiene una fórmula alternativa, en función de logaritmos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejemplo.

Las utilidades obtenidas por una compañía constructora en cuatro proyectos fueron: 3, 2,4 y 6% respectivamente. Calcule los promedios aritmético y geométrico. Compare los resultados.

$$\bar{x} = (3+2+4+6)/4 = 3.75\%.$$

$$Mg = \sqrt[4]{3*2*4*6} = 3.46\%$$

Si bien el valor de 6 no es extremo, si influye para que el promedio se incline hacia los valores mayores de la serie. La Mg de 3.46% da una cifra más conservadora para las utilidades.

Utilizando el Excel la sintaxis para calcularla es la siguiente:

MEDIA.GEOM (número1; número2;...)

MEDIA.GEOM (A10:A30) Calcula la Mg de los datos ubicados entre las filas A10 y A30.

4.1.5. Promedio recortado: Su principal aplicación se da en el análisis exploratorio de datos. Consiste en eliminar después de ordenar la serie de datos, algunos valores de los extremos, usualmente el 2.5% en cada extremo. Si es similar a la media aritmética clásica, se puede pensar que no existen valores extremos o que estos se compensan. Por lo tanto, permite inducir conclusiones sobre la existencia o no de valores atípicos. La

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

incorporación de este indicador en los programas estadísticos, obliga su interpretación en éste texto.

Utilizando el Excel la sintaxis para calcularla es la siguiente:

```
=MEDIA.ACOTADA (número1; número2;..., porcentaje.)  
=MEDIA.ACOTADA (A3:A12; 0.2) Calcula el promedio recortado de los  
datos ubicados entre las filas A3 y A12, eliminado el 10 de los datos a cada  
lado, después de ordenar.
```

4.2 MEDIDAS DE POSICIÓN.

Se caracterizan como su nombre lo dice para dar información sobre la ubicación o concentración de los datos en cualquier punto de la distribución. Sobresalen la moda y los percentiles.

4.2.1 Moda (Mo): Es usual considerarla como medida de tendencia central, sin embargo, ésta puede estar ubicada en cualquier punto de la distribución, es más, cuando se encuentra al principio de los datos la distribución toma el nombre de asimétrica positiva, si se encuentra al final, la distribución se llama asimétrica negativa. En el caso de que este en el centro y su valor sea similar a la mediana y la media aritmética, la distribución se considera aproximadamente normal, siempre y cuando no exista más de una moda.

La moda es simplemente el valor o la característica que más se repite, es el valor más común de la distribución y se denota Mo. En algunas distribuciones sólo existe una moda (unimodal), si tiene dos puntos máximos

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

se conoce como bimodal y en general se puede decir que hay distribuciones multimodales (más de una) o amodal (ninguna).

La M_o se obtiene por simple inspección, ya que es el valor de la variable que le corresponde la mayor frecuencia absoluta (f_i). Si los datos están agrupados por intervalos, la M_o es la marca de clase del intervalo con mayor frecuencia.

En Excel la función es: =Moda(Rango de los datos)

Para el caso sería: =Moda(C3:C7). Sí los datos se encuentran en esas celdas y se repite algún valor.

Ejemplo.

Para las siguientes distribuciones, encontrar la M_o e interprétela en términos del problema.

Salarios (miles)	Nº Empleados
500	25
525	20
530	20
600	15
750	12
800	10
900	5

Rangos de Edad	Nº Personas
20 a 25	10
26 a 30	12
30 a 35	18
35 a 40	25

Para el salario, la Moda es 500.000, ya que es el valor que más se presenta, no quiere decir, que es la mayoría, es lo que está de **moda**. Se puede decir, entonces, que el salario que mas empleados ganan es de \$500.000.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Para las edades, la moda es de 32.5 años, debido a que es el valor que se encuentra en la mitad del intervalo que posee la mayor frecuencia, no sobra recordar, que existen fórmulas de interpolación para calcular este valor, sin embargo, considero que no amerita el esfuerzo en el cálculo con respecto a lo que se gana en precisión. La edad más usual en dicha población es de 32.5 años.

Ejercicio: Tanto para la tabla simple como para la tabla agrupada, calcule las diferentes medidas e interprete sus resultados. Cuál promedio considera usted que es el adecuado según cada caso.

4.2.2 Cuartiles y Percentiles: Se utilizan especialmente en variables continuas y dividen la información en cuatro partes (25, 50 y 75%) si son los cuartiles o en 100 si son los percentiles. Con el fin de simplificar los cálculos, se plantea una sola fórmula, tanto para datos agrupados como para frecuencias simples, además, de si es un cuartil o un percentil.

$$Q_r (K) = n * \%deseado.$$

Donde r es el orden del cuartil o percentil, es decir, si se quiere calcular el 3 cuartil, la expresión sería: $Q_3 (4) = n * 0.75$.

Si lo que se requiere es el 80%, se utiliza el percentil 80.

$$Q_{80} (100) = n * 0.80.$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

El número obtenido con la fórmula anterior, se ubica en la frecuencia absoluta acumulada (F_i). Veamos el ejemplo en la siguiente diapositiva.

De igual manera como se calcularon otras medidas de tendencia central, el Excel permite calcular tanto los cuartiles como los percentiles. La sintaxis es la siguiente:

=CUARTIL (A1:A5, 3) calcula el tercer cuartil, teniendo los datos en las celdas A1 a A5.

=PERCENTIL (A1:A5, 0.9) Calcula el percentil 90, la diferencia está en que para calcular los percentiles se requiere digitar después del rango de datos el porcentaje deseado en forma de proporción (número entre 0 y 1).

Ejercicio Excel: los siguientes datos muestran el número de accidentes de tránsito ocurridos en Cucuruba – Huila – durante 12 meses.

8 15 12 7 6 8 8 12 7 6 8 10

Usando las órdenes de Excel, calcule las medidas de posición y de tendencia central analizadas anteriormente, interprételas.

4.3 MEDIDAS DE VARIACIÓN O DISPERSIÓN.

En todos los casos se va a necesitar más información que la obtenida con las medidas de posición. Debido a esto se necesita encontrar las medidas de variación, las cuales sirven para medir el grado en que los datos de una

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

muestra o población tienden a concentrarse o alejarse alrededor de un valor central. También son útiles para comparar la variabilidad de varias muestras o poblaciones.

El despreciar la existencia de la variación puede conducir a decisiones incorrectas en problemas importantes.

La estadística ayuda a analizar, y a sacar conclusiones teniendo en cuenta la existencia de la variación.

4.3.1 Rango: Es la diferencia entre el valor máximo y el valor mínimo, es una medida poco usada puesto que proporciona poca información útil cuando el número de datos es pequeño, por eso, es usado frecuentemente en control de calidad donde se utilizan muestras pequeñas, menores de 20¹.

$$R = X_{\text{máx}} - X_{\text{mín}}$$

4.3.2 Rango Intercuartílico: Mide la variabilidad en el 50% central de los datos, se utiliza para construir el gráfico de caja y sesgo. Es importante comparar la magnitud de esta medida con el rango, ya que se puede determinar si existe concentración o no de los datos alrededor de los valores centrales.

$$R_i = Q_3 - Q_1$$

¹ La generalidad es considerar un tamaño de muestra pequeño cuando $n \leq 30$.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

4.3.3 Varianza y Desviación Típica o Estándar: La varianza y la desviación estándar son las medidas de variabilidad más frecuentemente empleadas.

Se estudiarán simultáneamente, ya que para su cálculo se sigue un procedimiento exactamente igual, diferenciándose únicamente en que para calcular la desviación típica se requiere de un paso más.

Considero prudente insertar acá el OA sobre el cálculo del promedio y la desviación estándar, entre en él, dando click en:

<http://www.leondariobello.com/OA/Promedioaritmético>

Confío que el objeto de aprendizaje le aclarará el tema, continuemos hablando de la varianza. Se denota S^2 (se lee S al cuadrado), se define como la media aritmética de los cuadrados de las variaciones respecto a la media aritmética de dicha distribución. Como se ve, no es muy dicente dicha definición, lo que si se observa es que se construye con valores al cuadrado, lo que impide darle una interpretación a la misma.

La desviación típica (S) más conocida y usada, desempeña un papel muy importante en el análisis de los datos estadísticos. Igualmente permite la comparación de dos o más distribuciones, cuando están en la misma unidad de medida, para determinar cuál de ellas tiene mayor o menor grado de homogeneidad o estabilidad.

Las fórmulas a utilizar son:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$\sigma = \sqrt{\frac{(\sum (x - \bar{x})^2 f_i)}{n}}; S = \sqrt{\frac{(\sum (x - \bar{x})^2 f_i)}{n - 1}}$$

En Excel se usa la siguiente sintaxis:

=VAR (A3:A12) calcula la varianza de la muestra, teniendo los datos en las celdas A3 a A12.

=DESVEST (A3:A12) Calcula la desviación estándar de la muestra, teniendo los datos en las celdas A3 a A12.

4.3.3.1 Propiedades de la Varianza.

1. La varianza de una variable debe ser siempre un valor no negativo. $S^2 > 0$, por definición de la varianza, siempre va a hacer una suma de cuadrados, lo que garantiza el cumplimiento de esta propiedad.
2. La varianza de una constante es igual a cero. $S^2_{(k)} = 0$.
3. La varianza de una variable más o menos una constante, es igual a la varianza de la variable. $S^2_{(x+k)} = S^2_{(x)}$.
4. La varianza de una constante por una variable, es igual al producto de la constante al cuadrado por la varianza de la variable. $S^2_{(kx)} = k^2 * S^2_{(x)}$.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

EJEMPLO MANUAL.

Los siguientes datos hacen referencia a los pagos realizados por un grupo económico durante un mes por concepto de insumos. Al lado se colocan los resultados obtenidos con Excel.

Pagos	Días
100	5
110	7
115	8
118	10
120	12
130	9
133	7
135	3
	61

Pagos * Días	(xi-x) ² fi
500	1986.91
770	690.85
920	194.79
1180	37.42
1440	0.05
1170	911.84
931	1194.96
405	680.91
7316	5697.74

Se quiere calcular la media aritmética y la desviación típica o estándar.

$$S = \sqrt{\frac{(\sum (x - \bar{x})^2 f_i)}{n - 1}} = \sqrt{\frac{5697.74}{60}} = 9.74485$$

El promedio aritmético es igual a 119.934.426.23 (7316/61) y la desviación igual a 9.744.859.93 aplicando la fórmula anterior.

Dichos resultados se pueden calcular utilizando las funciones del Excel, así:

=Promedio (rango de datos)

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

=Desvest (rango de datos)

Interpretación del Ejemplo: el grupo económico paga en promedio por día por concepto de insumos un valor de \$119.93423 con una desviación de \$ 9.74485 lo que indica que los pagos realizados son homogéneos, es decir, poco variables. Lo anterior se deduce al relacionar la desviación con respecto a la media (S_x/\bar{x}).

4.3.3.2 Interpretación.

Como venimos estudiando dos medidas de dispersión conjuntamente, es apenas lógico dar sus interpretaciones en este mismo numeral.

La varianza es muy conocida y usada, sin embargo, carece de interpretación; no obstante, su importancia radica especialmente en que da origen a la desviación típica o estándar (S); si se quieren comparar dos o más distribuciones se puede utilizar una de estas medidas, teniendo en cuenta que deben tener la misma unidad de medida y magnitudes similares; tal comparación nos va a decir cuál de las distribuciones tiene mayor o menor variabilidad o dispersión en los datos.

La desviación estándar por el contrario, sí se puede interpretar por si sola, ya que al sacar la raíz cuadrada, queda en la misma unidad de medida que la variable en estudio; el inconveniente que puede generar, es que se requiere tener buen dominio de la variable en estudio para determinar cuando la desviación estándar es alta o baja, ya que en algunos casos un valor grande no significa variabilidad alta y viceversa. El caso puede ser que

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

una desviación de \$152.000.000 puede ser poca, si se trata de transacciones de multinacionales, por el contrario, \$120.000 puede ser muy alta si se refiere a deudas de comerciantes minoristas por día.

4.3.4 Coeficiente de Variación.

El coeficiente de variación es una medida relativa y presenta el resultado, como un porcentaje que debe emplearse para hacer comparaciones entre poblaciones que tengan diferentes unidades de medidas y/o magnitudes diferentes. A medida que dicho valor se aleja del cero, la distribución se vuelve más heterogénea, mientras que si se aproxima a dicho valor los datos son más homogéneos².

El coeficiente de variación esta definido como la razón existente entre la desviación y la media de los datos multiplicado por 100:

EJEMPLO.

Un grupo de profesionales, tiene un salario mensual promedio de \$700.000 y desviación de \$90.800, otro grupo de empleados profesionales que trabajan en un país diferente, en una actividad similar, tiene un salario promedio de 250 dólares y una varianza de 625 dólares. Se quiere determinar cuál grupo de salarios presenta una menor variabilidad.

$$Cv_A = \left(\frac{90.800}{70.000} \right) * 100 = 12.97\%$$

² Se consideran que un $Cv \leq 20\%$ indica estabilidad en los datos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$Cv_B = \left(\frac{25}{250}\right) * 100 = 10\%$$

25 es la raíz de 625.

Interpretación: En el país donde se paga los salarios con dólares es más estable, ya que dio un coeficiente de variabilidad menor ($10\% < 12.97\%$), no obstante, se puede considerar que en ambos países los salarios son relativamente homogéneos, debido a que sus coeficientes dieron valores no muy alejados del cero.

4.3.5 ¿Qué dicen las medidas de dispersión?.

Al tratar con las medidas de dispersión es importante tener en cuenta que sólo dicen la mitad de la historia, situación similar sucede con la media aritmética. Una interpretación efectiva de lo que dice una medida de dispersión se puede dar en dos tipos de situaciones:

Cuando se quiere comparar la variabilidad de varias distribuciones, estén o no en la misma unidad de medida.

Cuando se complementa la información entregada por la media (o alguna otra medida de localización).

4.4 MEDIDAS DE FORMA.

Miden la curva que forman los valores de la serie y se clasifican en

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

simétricas o asimétricas y en puntiagudas o no. Por sus características requieren que los datos sean cuantitativos y en lo general continuos.

4.4.1 Distribuciones Simétricas: Una distribución simétrica es una distribución en forma de campana que presenta ciertas características en los valores de la variable. En una distribución simétrica la media aritmética, la mediana y la moda tienen valores similares: $x \approx Me \approx Mo$. Un ejemplo de una distribución simétrica es la siguiente:

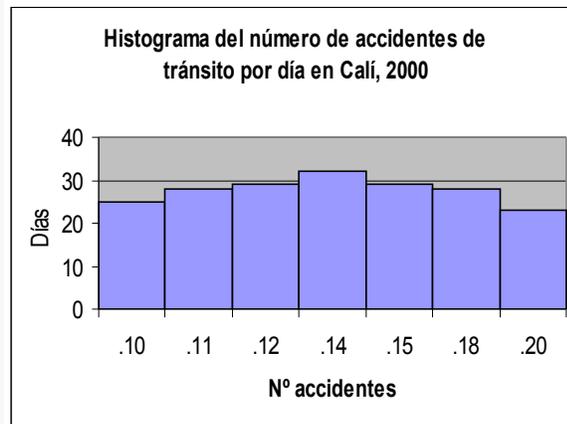
X_i : 2 - 4 - 6 - 8 - 10 - 12 - 14

f_i : 4 - 6 - 8 - 10 - 8 - 6 - 4

Aunque múltiples situaciones conllevan al estudio de las distribuciones simétricas, o supuestamente simétricas, una de ellas desempeña un papel importantísimo, por lo que es frecuentemente utilizada en diferentes análisis estadísticos: **LA DISTRIBUCION NORMAL**

4.4.2 Distribuciones Asimétricas: Presenta un alargamiento bien a la derecha (Asimétrica positiva) o bien a la izquierda (alargamiento a la izquierda). La manera más fácil de detectar la forma de los datos es dibujando el histograma de frecuencia o un gráfico que permita ver la composición de los datos. En la actualidad, el gráfico de caja y sesgo es usado frecuentemente para éste fin. A continuación se presentan tres histogramas construidos en Excel y que muestran los tres tipos de formas analizados antes, no obstante, se pueden presentar múltiples formas, tal es el caso de: Rectangular (uniforme), convexa, cóncava etc.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



Se nota en el gráfico que cuando $M_o < M_e < \bar{x}$. La distribución es asimétrica positiva, mientras que si $M_o > M_e > \bar{x}$ los datos son asimétricos negativos y si estas medidas son similares los datos son acampanados (simétricos).

Algunas distribuciones asimétricas positivas muy utilizadas en la estadística son la distribución " **JI CUADRADO** " (χ^2) y la distribución **F DE FISHER**.

Algunas medidas utilizadas para determinar el grado de asimetría son conocidas como **PRIMERO y SEGUNDO COEFICIENTE DE SESGO PEARSON**, dadas por:

$$A_1 = \frac{\bar{X} - M_o}{S}$$

$$A_2 = \frac{3(\bar{X} - M_e)}{S}$$

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Sí A_1 o $A_2 > 0$ La distribución será asimétrica positiva.

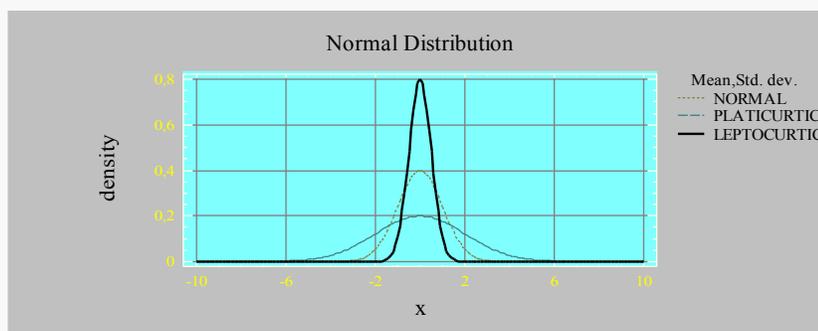
Sí A_1 o $A_2 < 0$ La distribución es asimétrica negativa

SIEMPRE Y CUANDO SE ALEJE DEL CERO DE MANERA SIGNIFICATIVA.

El Excel tiene la siguiente función:

=COEFICIENTE.ASIMETRIA (A3:A12).

4.4.3 Curtosis: La curtosis es el grado de agudeza o apuntamiento que presentan algunas distribuciones simétricas y generalmente es tomada con relación a la distribución normal. Si una curva simétrica es más plana que la normal, se dice que la distribución es **ACHATADA** o **PLATICURTICA**; si es más aguda que la normal se dice que la distribución es **APUNTADA** O **LEPTOCURTICA** y si es normal se dice que la distribución es **MESOCURTICA**.



Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Una medida de curtosis empleada es el **COEFICIENTE DE CURTOSIS PERCENTELICO (K)**

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \quad \text{donde:}$$

Q_3 y Q_1 Son el tercero y primer cuartíl.

P_{90} y p_{10} Son el percentíl 90 y el percentil 10.

Si $K \leq 0,263$ la distribución es normal

Si $K > 0,263$ la distribución es leptocúrtica

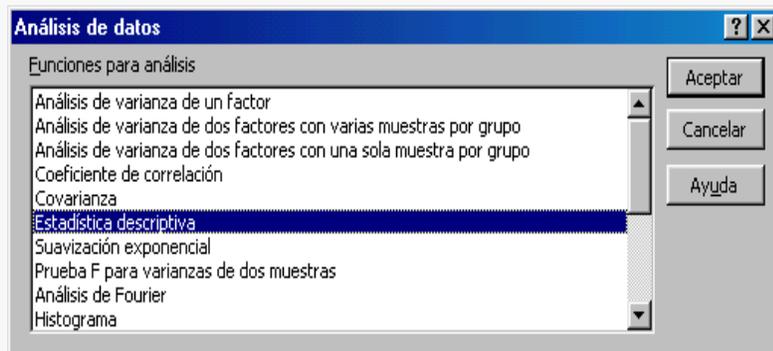
Si $K < 0,263$ la distribución es platicúrtica

=CURTOSIS (A3:A12), para trabajar en Excel, en cualquier caso, se tiene como referente el cero.

4.5 ORDENES ESTADÍSTICA DESCRIPTIVA CON EXCEL.

Como se mencionó en la sesión 3.2 del capítulo anterior, los procedimientos estadísticos más avanzadas no se activan cuando se instala el programa de Microsoft Excel, se requiere activar las subrutinas, para luego dar clic en la opción **Datos + Análisis de datos**, presentando lo siguiente.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



A continuación se selecciona la opción **Estadística Descriptiva**, permitiendo seleccionar el rango de entrada de los datos en bruto y la celda donde se quiere que queden los resultados obtenidos. Además, se puede seleccionar la alternativa rótulo, siempre y cuando se haya seleccionado el título de la serie de datos, en el pantallazo de la izquierda se presenta la manera como se debió seleccionar la información para obtener los resultados mostrados a la derecha.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

The screenshot shows the 'Estadística descriptiva' dialog box in Microsoft Excel. The 'Entrada' (Input) section has 'Rango de entrada' set to '\$A\$4:\$A\$16', 'Agrupado por' set to 'Columnas', and 'Rótulos en la primera fila' checked. The 'Opciones de salida' (Output options) section has 'Rango de salida' set to '\$C\$4', 'Resumen de estadísticas' checked, and 'Nivel de confianza para la media' set to 95%. The background shows the Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K
4	Salario diario			Salario diario							
5	20000										
6	25000			Media	25666,66667						
7	30000			Error típico	1634,538825						
8	18000			Mediana	26500						
9	15000			Moda	25000						
10	28000			Desviación estándar	3662,208885						
11	25000			Varianza de la muestra	33360606,06						
12	30000			Curtosis	-0,061568176						
13	29000			Coefficiente de asimetría	-0,486093528						
14	35000			Rango	20000						
15	25000			Mínimo	15000						
16	28000			Máximo	35000						
17				Suma	308000						
18				Cuenta	12						

Ejemplo 1: Los siguientes son los precios promedios por mes de tres acciones. Calcule las estadísticas descriptivas entregadas por el Excel y que tengan utilidad práctica.

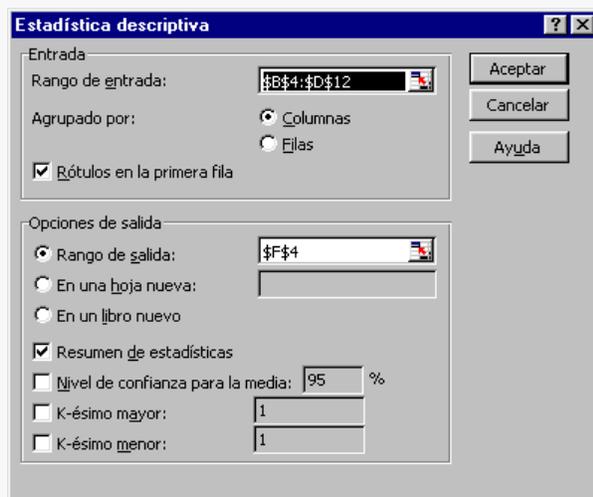
Mes	Acción A	Acción B	Acción C
Enero	2500	150	752
Febrero	2550	148	780
Marzo	2530	140	759
Abril	2485	148	781
Mayo	2500	140	790
Junio	2489	139	849
Julio	2560	142	820
Agosto	2600	140	835

Los resultados fueron los siguientes:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

	Acción A	Acción B	Acción C
Media	2526.75	143.375	795.75
Error típico	14.40579001	1.591691598	12.48391823
Mediana	2515	141	785.5
Moda	2500	140	#N/A
Desviación estándar	40.74572721	4.50198369	35.30985293
Varianza de la muestra	1660.214286	20.26785714	1246.785714
Curtosis	-0.372585717	-1.89473857	-1.29223681
Coefficiente de asimetría	0.781479018	0.59905066	0.367946909
Rango	115	11	97
Mínimo	2485	139	752
Máximo	2600	150	849
Suma	20214	1147	6366
Cuenta	8	8	8

El pantallazo para calcular los valores anteriores es el siguiente:



ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Los estadísticos ubicados en las filas con diferente color no son usualmente interpretados. El error típico hace referencia a la desviación de los promedios con respecto a su valor esperado, lo que no es útil en ésta primera parte, ya que se trabaja con los valores de la variable, no con los promedios. La varianza no tiene interpretación por sí misma como ya se mencionó. Los coeficientes de asimetría y curtosis si bien son comparables con un estándar (-0.27 a 0.27), es más práctico y entendible utilizar el histograma o la caja y sesgos para detectar la forma de los datos. Además de que existen pruebas adicionales para determinar la normalidad de los datos (forma acampanada).

En los demás, se puede interpretar así para la acción A: El promedio de los precios de la acción en los últimos 8 meses fue de \$2.526,75, con una desviación de 40.74, lo que indica que es baja con respecto a su promedio y por lo tanto, se puede pensar que el invertir en dicha acción no conlleva riesgos altos. Se encontró, además, que el 50% de los meses el precio de la acción fue de \$2515 o menos y el valor que más se presentó fue de \$2.500. En algunos casos no es necesario ni lógico, presentar los otros valores. A veces, la distribución no tiene M_o , como ocurre en la acción C.

Interprete los resultados para las acciones B y C. Además, diga cual de ellas es más variable.

Para comparar la variabilidad cuando se tienen magnitudes diferentes es necesario calcular el coeficiente de variabilidad. En Excel se requiere construir la fórmula, así:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

C. Variación = (Desvest/promedio)*100.

Ejemplo 2: Utilizando las funciones de Excel calcule los estadísticos para el precio del dólar durante 12 días:

	Precio Dólar
	2820
	2815
	2826
	2850
	2900
	2915
	2895
	2985
	2895
	2900
	2982
	2986
Promedio Aritmético	2897.42
Promedio Geométrico	2896.80
Promedio Armónico	2896.18
Promedio Recortado	2897.42
Mediana	2897.50
Moda	2900.00
Min	2815.00
Max	2986.00
Desviación Típica	62.63
Percentil 80	2968.60
Cuartil 25	2931.75

Promedio Aritmético	=PROMEDIO(B66:B77)
Promedio Geométrico	=MEDIA.GEOM(B66:B77)
Promedio Armónico	=MEDIA.ARMO(B66:B77)
Promedio Recortado	=MEDIA.ACOTADA(B66:B77,0.1)
Mediana	=MEDIANA(B66:B77)
Moda	=MODA(B66:B77)
Min	=MIN(B66:B77)
Max	=MAX(B66:B77)
Desviación Típica	=DESVEST(B66:B77)
Percentil 80	=PERCENTIL(B66:B77,0.8)
Cuartil 25	=CUARTIL(B66:B77,3)

Los resultados se obtuvieron utilizando las fórmulas de Excel, la sintaxis se coloca en la tabla a la derecha de los datos.

Todos los promedios dieron resultados similares, indicativo de que el precio del dólar en esos días no presentó valores atípicos. Es más, el promedio aritmético y el promedio recortado dieron lo mismo. Tanto la Mg como la Ma no presentan cambios sustanciales, en estos casos es preferible interpretar la media aritmética.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

La Me la Mo y la \bar{x} son aproximadamente iguales lo que lleva a suponer que los datos tienden a ser simétricos. Coherente con lo dicho anteriormente la desviación es baja con respecto a la media, por lo tanto, el precio del dólar ha venido cambiando poco. También se encontró que el 80% de los días el precio del dólar fue de \$2968.60 o menos. Además, el 25% del tiempo el precio del dólar fue de \$2931.75 o menos. No sobra comentar que, cuando la desviación no dé información por si sola, es útil calcular el coeficiente de variabilidad, construyendo la fórmula en Excel $(S_x/\bar{x})*100$.

4.6 ORDENES ESTADÍSTICA DESCRIPTIVA CON SPSS 19.0.

Para acceder al programa SPSS, se procede así: Inicio + Todos los programas + SPSS inc + PASW Statistics 19.

Observe el siguiente vídeo [Generalidades SPSS](#).

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

descriptiva libro.sav [Conjunto_de_datos1] - PASW Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

1: salario 20000 Visible: 6 de 6 variables

	salario	sexo	estrato	experiencia	estudio	edad	var	var	var	var	var	var
1	20000	Hombre	2	4	15	35						
2	25000	Mujer	2	4	18	36						
3	30000	Hombre	3	6	20	38						
4	18000	Hombre	3	2	18	30						
5	15000	Hombre	4	2	15	32						
6	28000	Mujer	2	3	19	35						
7	25000	Hombre	3	4	20	40						
8	30000	Hombre	4	5	22	42						
9	29000	Hombre	2	5	21	40						
10	35000	Mujer	2	8	22	38						
11	25000	Mujer	3	5	20	45						
12	28000	Mujer	3	5	23	40						
13	20000	Hombre	2	4	18	38						
14	25000	Mujer	2	4	22	39						
15	30000	Hombre	3	5	25	44						
16	18000	Hombre	3	2	18	32						
17	15000	Hombre	4	2	15	29						
18	28000	Mujer	2	5	20	35						
19	25000	Hombre	3	4	21	30						
20	30000	Hombre	4	5	25	35						
21	29000	Hombre	2	5	24	33						
22	35000	Mujer	2	6	26	35						
23	25000	Mujer	3	5	25	29						
24	20000	Mujer	2	5	25	29						

Vista de datos Vista de variables

PASW Statistics Processor está listo

Inicio E5 05:04 p.m.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

descriptiva libro.sav [Conjunto_de_datos1] - PASW Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	salario	Numérico	6	0	Salario diario	Ninguna	Ninguna	8	Derecha	Escala	Entrada
2	sexo	Cadena	1	0	Sexo encuesta... (0, Mujer)...	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
3	estrato	Cadena	1	0	Estrato	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
4	experiencia	Numérico	8	0	Años de experi...	Ninguna	Ninguna	8	Derecha	Escala	Entrada
5	estudio	Numérico	8	0	Años de estudio	Ninguna	Ninguna	8	Derecha	Escala	Entrada
6	edad	Numérico	8	0	Años de vida	Ninguna	Ninguna	8	Derecha	Escala	Entrada
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

Vista de datos **Vista de variables**

PASW Statistics Processor está listo

Inicio [Taskbar icons] 05:06 p.m.

Se abre una ventana parecida al Excel en la cual luego de cancelar el primer pantallazo (permite abrir últimos archivos usados), se puede digitar la información, teniendo en cuenta que en la parte izquierda inferior de la pantalla, se puede mover entre vista de datos y vista de variables. Se sugiere entrar en vista de variables para definir las características de ellas, en el pantallazo de la izquierda se observan los datos digitados para el

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

ejemplo, y en la derecha la configuración de las variables. Lo relevante es definir el tipo de variable, cuantitativa o cualitativa, con lo que se lleva a definir automáticamente la escala de medición, que puede ser: Escala (numérica), Ordinal (cualitativa) o Nominal (cualitativa). Es posible abrir archivos con formato Excel.

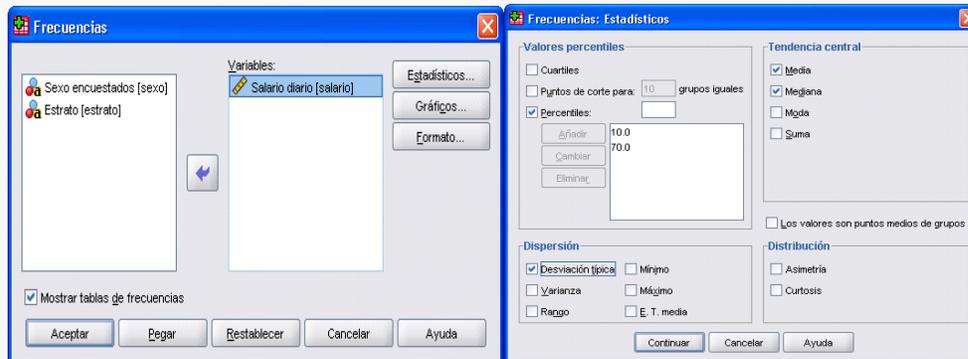
La etiqueta es una aclaración de la variable, ya que ésta, no puede tener espacios ni caracteres especiales, los valores son de vital importancia en las variables cualitativas, ya que ahí, es donde se colocan los nombres que equivalen a códigos, valga decir; Hombre=1, mujer= 2. El objetivo de éste texto no es mostrar el manejo del programa de manera exhaustiva, por lo tanto, con estos insumos, se construye la base de datos y se procede a calcular los estadísticos resumen, tanto para las variables cualitativas como para las cuantitativas. En la figura siguiente, se aprecia que se debe activar del menú principal, la opción **Analizar**, la cual, será la que siempre nos guiará en todos los procedimientos a realizar, sean básicos o avanzados. Para nuestro objetivo, se presentarán los resultados entregados por los sub menús: **Frecuencias, Descriptivos y Tablas de contingencia**.

En la página del curso Estadística Descriptiva ya citada previamente, se puede tener acceso a un archivo donde se profundiza en la manera de empezar a usar el programa SPSS 19.0.

A continuación se presentan los pantallazos adecuados para calcular estadísticos resumen para variables cuantitativas, además, de la construcción de tablas simples y compuestas. La ruta es:

Analizar + Estadísticos descriptivos + Frecuencias.

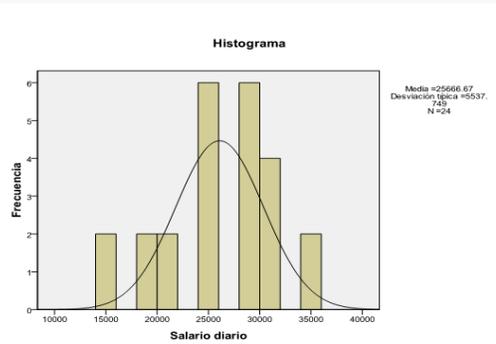
ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



La caja de diálogo de la derecha, permite seleccionar la(s) variable(s) a las cuales se calcularán los estadísticos resumen y/o las tablas de frecuencia simple. Para el ejemplo, se seleccionó la variable cuantitativa salario (se pueden seleccionar varias al tiempo). En la caja de la izquierda, a la cual se accede dando clic en **Estadísticos**, se muestra como hay diferentes estadísticos a calcular, se solicitaron: La media, Mediana, Desviación típica y los percentiles 10 y 70.

La opción de **Gráficos** ofrece otras posibilidades, tales como: gráfico de barras, sectores e histogramas, tal como se aprecia a continuación:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



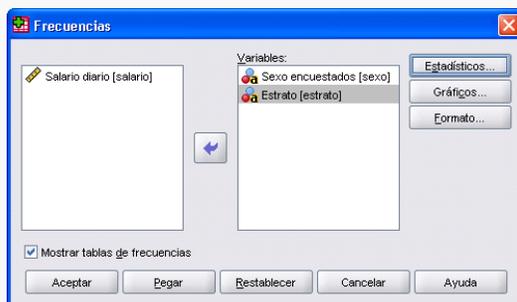
Los resultados arrojados por el programa son:

Salario diario		
N	Válidos	24
	Perdidos	0
Media		25666.67
Mediana		26500.00
Desv. típ.		5537.749
Percentiles	10	16500.00
	70	29000.00

El promedio del salario por día es de \$25.666,67 con una desviación típica de \$5.537,75, es decir, los salarios tienen una variabilidad moderada, además, el 50% de las personas, devengan \$26.500 o menos. También, se encontró que el 10% de las personas ganan \$16.500 o menos y el 70% ganan \$29.000 o menos. Dado el número de datos, el histograma muestra unos vacíos, que en la práctica no debería suceder, dado un número mayor de observaciones.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Esta misma ruta, permite encontrar las frecuencias simples para variables nominales u ordinales, en este caso, se hará con las variables sexo y estrato, tal como se aprecia a continuación.



La opción Formato, permite solicitar la salida según un determinado orden, en algunos casos en las variables nominales, es posible solicitar que las categorías se ordenen según sus frecuencias. Las ordinales, no es útil realizar estos cambios, se mostrará por que.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
Mujer	10	41.7	41.7	41.7
Hombre	14	58.3	58.3	100.0
Total	24	100.0	100.0	

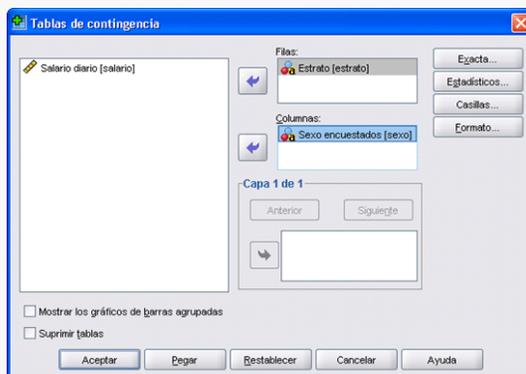
	Frecuencia	Porcentaje	Porcentaje acumulado
Válidos			
4	4	16.7	16.7
2	10	41.7	58.3
3	10	41.7	100.0
Total	24	100.0	

Como se aprecia en las salidas, el SPSS, entrega 4 columnas, frecuencia, porcentaje, las más útiles, además de porcentaje válido y porcentaje acumulado. El primer detalle, es que el acumulado no aplica para las variables nominales, no tiene sentido decir que el 85% de las personas son casadas o menos. En las ordinales sí tiene interpretación, por ejemplo, el 35% son instituciones de segundo nivel o menos. Otro hecho que no se recomienda es el organizar valores según su frecuencia para variables ordinales, en el ejemplo, quedó ubicada primera la categoría de 4, luego 2 y por último 3, en estrato esto no tiene una interpretación lógica.

Para eliminar una columna se da doble clic en la tabla para luego sombrear la o las columnas a eliminar, lo cual se hace con la tecla suprimir del teclado. Una consideración importante es que las tablas simples se usen para identificar valores anómalos, es decir, para detectar posibles códigos o palabras mal digitadas; su interpretación no es relevante, dado que en la

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

práctica es de interés entregar información útil para la toma de decisiones, situación que en muchos casos no se logra con tablas simples, además, cuando se realizan los cruces de variables conocidas también como tablas de contingencia, se obtiene la información de las tablas simples, más las intersecciones entre las categorías, posibilitando análisis e interpretaciones más relevantes. El proceso para construir esas tablas es: **Analizar + Estadísticos descriptivos + Tablas de contingencia**. Llegando a la siguiente caja de diálogo:



El cruce de variables pretende identificar si existe relación entre dos o más de ellas, además, de posibilitar el análisis de estas variables en una sola tabla, en lugar de construir dos cuadros simples. Estas tablas aplican fundamentalmente para variables categóricas o incluso cuantitativa discreta si ésta no tiene muchas categorías de respuestas. Para el ejemplo se relacionan las variables cualitativas: Sexo y Estrato.

Es usual colocar en las filas la variable considerada independiente, sin embargo, en aquellos casos, donde la dependiente tiene muchas categorías, es preferible colocar esta variable en las filas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Dependiendo donde se coloque la variable independiente, se debe pedir que se calculen los porcentajes, es decir, si la variable sexo es la que no depende del estrato, si se coloca en la fila, los porcentajes se piden por fila o viceversa, si la variable esta en la columna, así se deben pedir los porcentajes. Esto es fundamental y llevaría a conclusiones erróneas en algunos casos. La activación de los porcentajes se da por la opción **casillas**, mostrando lo siguiente:



Los resultados encontrados son:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

		Estrato			Total
		2	3	4	
Sexo encuestados	Hombre	4 28.6%	6 42.9%	4 28.6%	14 58.3%
	Mujer	6 60.0%	4 40.0%	0 .0%	10 41.7%
Total		10 41.7%	10 41.7%	4 16.7%	24 100.0%

Es útil recalcular manualmente los porcentajes marginales donde da 100%, con el fin de determinar los porcentajes de las categorías pertenecientes a la variable independiente Sexo. Lo usual es redactar los datos encontrados en las márgenes del cuadro, para luego identificar posibles relaciones entre las variables redactando las celdas interiores de la tabla. Para el caso se tiene (realizando el cambio sugerido del 100%):

Se presenta una diferencia en cuanto al porcentaje de hombres (58.3) y mujeres (41.7) encuestados. Con respecto a los estratos, se tuvo un mismo porcentaje de personas del estrato dos y tres (41.7%), el menor porcentaje de personas es del estrato 4, con un 16.7%.

Se percibe una diferencia en los porcentajes de personas en los estratos según el sexo. Los hombres están distribuidos en mayor porcentaje en los estratos más altos, valga decir, el 28.6% de los hombres están en el estrato 4, mientras que ninguna mujer pertenece a ese nivel. De otro lado, el 60% de las mujeres son del menor estrato evaluado, el dos, mientras que los hombres de ese estrato sólo corresponde el 28.6%. Es importante mencionar que es sólo un ejemplo de muchas maneras de redactar, lo que se pretende es que los analistas, no sólo hablen de lo encontrado en la tabla, sino de lograr llegar a alguna conclusión.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejercicios

1. Una compañía bancaria está estudiando el número de reclamaciones de los usuarios en el último trimestre, tanto para hombres como para mujeres. Los resultados son los siguientes:

Monto de la Reclamación	N° de reclamaciones		Total
	Hombres	Mujeres	
100 a 200	8	15	23
201 a 300	10	20	30
301 a 400	15	8	23
401 a 500	20	10	30
501 a 600	22	6	28

a. Represente gráficamente los datos anteriores de tal manera que se aprecien las diferencias o similitudes según el sexo.

b. Represente el gráfico que compare el total de reclamos según el monto.

c. Represente el gráfico adecuado para mostrar la participación relativa del número de reclamos según sexo.

d. Calcule el promedio aritmético y la desviación estándar para el total de reclamos. Interprete.

e. Diga en cuál de los dos sexos es más homogéneo el número de reclamos.

2. Los precios de cinco acciones del sector confecciones se incrementaron

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

desde 1985 en: 37.1, 1140, 1.5, 2.7 y 840% respectivamente. ¿Cuál es la media geométrica del incremento porcentual en el precio de esas acciones? ¿Por qué utiliza Mg en lugar de la media aritmética? Use la función de Excel.

3. Los salarios de varios grupos de [enfermer@s](#) por día en dos países se registran a continuación:

Colombia		Estados Unidos Dólares	
Pesos (miles)	Personas	Ls – Li	Personas
30	5	10 a 12	10
32	8	12 a 14	8
35	10	14 a 16	6
36	7	16 a 18	4
38	6	18 a 20	4

- Determine en cuál país es más estable el salario ¿Por qué?
- Usando medidas de posición diga que forma tienen los datos en Estados Unidos.
- Suponga que en Colombia los salarios son aproximadamente normales; diga entre qué valores se encuentra el 95% de las observaciones.
- Si el incremento para el año entrante en Colombia es del 10%, cuales son las nuevas medidas (Media, Desviación Típica)?.
- Diga que gráficos son los adecuados para cada país.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

4. El número promedio de accidentes de tránsito en diciembre del año 2010 fue de 12 por día con una desviación típica de 2. Las autoridades estiman que con las medidas tomadas el número de accidentes disminuya en dos. Si esto es cierto, ¿cuáles serán las nuevas medidas?

5. Las siguientes distribuciones muestran el número de defunciones reportadas en dos entidades de Salud. a) Se le solicita identificar cual de ellas presenta mayor homogeneidad y cual es más heterogénea.

EMIII		ISSSS	
Xi	Frec.	Xi	Frec.
10	5	20	10
12	6	21	12
14	7	22	14
16	6	23	15
18	5	24	14

b) Para EMIII diga que forma parecen tener los datos y calcule las medidas de posición y de variabilidad más comunes, use la hoja electrónica Excel.

6. Los siguientes son los precios promedios por mes de tres acciones. Calcule las estadísticas descriptivas entregadas por el Excel y diga en cual de ellas invertiría. Justifique su respuesta.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Mes	Acción A	Acción B	Acción C
Enero	2500	150	752
Febrero	2550	148	780
Marzo	2530	140	759
Abril	2485	148	781
Mayo	2500	140	790
Junio	2489	139	849
Julio	2560	142	820
Agosto	2600	140	835

7. En Excel, Gstat y/o SPSS calcule las estadísticas resumen. Interprete los resultados y tome las decisiones que considere necesarias.

Edad	Salario miles	Nº Moras > 30 Días
40	400	4
38	500	3
51	450	4
44	450	5
41	480	4
40	1250	5
38	400	3
40	450	6
38	450	3
51	250	4
30		5
32	450	4
28	400	3
45	380	3
40	420	5
38	515	5
28	545	6
30	485	4
30	379	4
30	450	5

Usando las funciones, calcule las medidas de tendencia central y las compara. Calcule las medidas de variabilidad que considere y las compara. Encuentre el percentil 10 y el 60, además, del cuartil 2. Interprete todos los

Diga que gráfico haría para cada variable y por qué?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Calcule las medidas descriptivas entregadas con la opción Análisis de Datos del Programa Excel.

8. Los siguientes datos presentan el subsidio otorgado por diferentes Cajas de Compensación y sus respectivos aportes mensuales, además del tiempo de respuesta que estas entidades demoran en responder inquietudes de sus usuarios.

Subsidio	Aportes(millones)	Tiempo Respuesta
15000	5800	15
12500	6250	22
13500	3685	14
14600	5489	10
15000	2103	30
12500	5487	23
13000	56985	25
13500	2598	20
16300	3548	14

Usando las funciones de Excel y los procedimientos de SPSS, calcule las medidas de tendencia central y las compara. Calcule las medidas de variabilidad que considere y las compara. Encuentre el percentil 20 y el 50, además, del cuartil 2. Interprete todos

9. Usando Excel y/o SPSS desarrolle los siguientes puntos con los datos que se muestran a continuación:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

EDAD	SEXO	ESTADO CIV.	PROFESIÓN	SALARIO (miles semanal)	ESTRATO
40	Hombre	Soltero	Estadístic@	400	4
38	Mujer	U .libre	Enfermer@	500	3
51	Mujer	Soltero	Médic@	450	4
44	Mujer	Soltero	Enfermer@	450	5
41	Hombre	Soltero	Médic@	480	4
40	Mujer	U .libre	Bacteriólog@	1250	5
38	Mujer	Soltero	Médic@	400	3
40	Mujer	Casado	Otra	450	6
38	Hombre	Casado	Enfermer@	450	3
51	Hombre	U. libre	Otra	250	4
30	Mujer	Soltero	Médic@		5
32	Mujer	Soltero	Enfermer@	450	4
28	Hombre	Soltero	Bacteriólog@	400	3
45	Hombre	Soltero	Otra	380	3
40	Mujer	U. libre	Enfermer@	420	5
38	Hombre	Casado	Médic@	515	5
28	Mujer	Soltero	Bacteriólog@	545	6
30	Mujer	Casado	Enfermer@	485	4
30	Hombre	Soltero	Bacteriólog@	379	4
30	Mujer	Casado	Bacteriólog@	450	5

- Realice el análisis exploratorio de datos (variables cualitativas y cuantitativas). Realice los cambios que considere necesarios (justifique si es del caso)
- Calcule las medidas descriptivas para las variables cuantitativas y diga cual es más homogénea.
- Con las medidas de posición diga que forma parecen tener los datos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

d) Construya el histograma para las variables cuantitativas. Interpretelo.

e) Calcule los percentiles 20, 40, 60 y 80. Interpretelos.

f) Calcule los cuartiles e interprételos.

10. Digite información para 8 variables que usted maneje diariamente, de las cuales, 4 sean cuantitativas (50 registros o datos por variable). Realice el análisis descriptivo adecuado y concluya según lo encontrado. Use el programa G-stat.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

CAPITULO 5

INTRODUCCIÓN A LAS PROBABILIDADES

Para el cálculo de las probabilidades se requiere tener una información previa, la cual se puede dar por varias vías, una de ellas se refiere a la probabilidad relativa de que ocurrirá algún suceso o evento, en otros casos se apoya en datos históricos para definir el valor probable de que ocurra un hecho, y en otras oportunidades se apoya en la experiencia de las personas e incluso en aspectos emocionales. En cualquier caso es de gran ayuda para la toma de decisiones en casos donde la incertidumbre está presente, tal como en los siguientes casos. ¿Cuál es la posibilidad (probabilidad) de que una persona contraiga una determinada enfermedad, si se sabe que en el último año la han contraído 7 de cada 10000 personas?. ¿Qué tan probable es, que en un parto se presenten gemelos?, ¿qué tan posible es que en un día ocurran tres accidentes aéreos? Los ejemplos sólo están limitados por la imaginación del lector. No obstante, es usual que el tema se estigmatice con los juegos de azar, ya que, de hecho, esta teoría tiene sus orígenes en dichos juegos.

En realidad, la probabilidad es una medida de la incertidumbre, en la medida, que se presentan situaciones en las que no se puede determinar que va a suceder. ¿Quién sabe cuanto se demora para llegar a su casa después de salir del colegio u oficina? ¿Está seguro de que el próximo año le aumentarán el 8% en su salario? Considero que NO, el caso es que se dan algunas situaciones donde la probabilidad es muy alta de que ocurra el suceso, y por el contrario, existen otras donde la probabilidad es muy baja,

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

por ejemplo, de que usted señor lector, gane más de 20 salarios mínimos. Claro está, que eso depende de otros factores, tal es el caso, de que los lectores en su mayoría sean reconocidos ejecutivos.

La teoría de probabilidades nos resuelve estos problemas y otros más. Para ir avanzando en esta temática se requiere definir algunos conceptos de la teoría probabilística, tales como las que se describen a continuación.

5.1 EXPERIMENTO ALEATORIO

Es una prueba o ensayo en la cual no se conoce el resultado, aunque se hayan realizado muchas pruebas anteriores, bajo las mismas condiciones. Es prudente comentar que existen experimentos determinísticos, sin embargo, la estadística no se encarga de estudiarlos.

Ejemplo N° 1: Sea el experimento de aplicar un medicamento con el fin de eliminar un dolor de muela.

Ejemplo N° 2: Determinar el número de accidentes de tránsito en un cruce preferencial durante las horas pico en un mes.

Ejemplo N° 3: Determinar el número de personas que asisten al Congreso de Salud Pública, que se realiza cada dos años.

Ejemplo N° 4: El tiempo de duración de una operación quirúrgica.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

5.2 ESPACIO MUESTRAL (S)

Es el conjunto de todos los posibles resultados de un experimento aleatorio. Puede ser discreto o continuo. Los ejemplos que a continuación se enuncian se refieren a los mencionados en el punto anterior.

Ejemplo N° 1: $S = \{\text{Elimina el dolor, No elimina el dolor}\}$

Ejemplo N° 2: $S = \{0, 1, 2, 3, \dots\}$. Espacio muestral discreto.

Ejemplo N° 3: $S = \{0, 1, 2, \dots, n\}$, donde n es el total de personas que asisten a una reunión. Espacio muestral discreto.

Ejemplo N° 4: $S = \{t / t > 0\}$, donde t es el tiempo y como es una variable continua, se escribe en forma de intervalos. Espacio muestral continuo.

5.3 SUCESO O EVENTO (A).

Es un subconjunto del espacio muestral, el suceso en que estamos interesado que ocurra se llama suceso favorable. Además, el suceso que esté compuesto por un solo elemento se conoce como suceso simple y por el contrario, el que contiene más de un elemento, recibe el nombre de suceso compuesto y por ende se puede subdividir en otros sucesos.

Otra clasificación, hace referencia a que los sucesos pueden ser: Seguros e imposibles. Entendiéndose como suceso seguro, aquel que con toda seguridad pasa y suceso imposible, por el contrario, no tiene ninguna oportunidad de ocurrencia. Valga decir, si se tiene en un comité, médicos y

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

enfermeras, no se puede seleccionar un filósofo, entendiendo que no puede tener dos profesiones, por el contrario, con seguridad el seleccionar una persona de ese comité, tiene que ser médico o enfermera.

Ejemplo N° 1: Suceso que se alivie. $A = \{\text{Elimine}\}$. Suceso Simple.

Ejemplo N° 2: Suceso que ocurra máximo dos accidentes. $A = \{1, 2\}$.

Ejemplo N° 3: Suceso que asistan mínimo 50 personas $A = \{50, 51, \dots, n\}$. Suceso compuesto.

Ejemplo N° 4: Suceso que la operación dure menos de 10 horas. $A = \{t/t < 10\}$

Todos los sucesos continuos son compuestos. ¿Por qué?

¿Existe algún suceso seguro, en los casos anteriores? ¿Por qué?

5.4 SUCESO CONTRARIO (A^c)

Es lo que le hace falta al suceso A para completar el espacio muestral.

Ejemplo N° 1: $A^c = \{\text{No elimina}\}$

Ejemplo N° 2: $A^c = \{3, 4, \dots\}$

Ejemplo N° 3: $A^c = \{0, 1, 2, \dots, 49\}$

Ejemplo N° 4: $A^c = \{t / t \geq 10\}$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Recuerde que al adquirir éste ebook, adquirió el derecho de realizar inquietudes puntuales al autor en su sala de conferencias ubicada en la dirección www.gvowebcast.com/conference,leondario

5.5. DEFINICIÓN CLÁSICA DE PROBABILIDAD.

La probabilidad de ocurrencia de un suceso **A**, es igual al cociente entre el número de casos favorables y el número de casos posibles de un experimento aleatorio.

$$p(A) = \frac{\text{casos}_- \text{favorables}}{\text{casos}_- \text{posibles}}$$

Ejemplo N° 1: De un grupo de 30 personas, de las cuales 20 son mujeres. Cuál es la probabilidad de que si se selecciona una persona al azar, cuál es la probabilidad de que: a) sea hombre, b) no sea hombre.

Solución:

Se define el suceso: A: suceso que sea hombre.
20 mujeres, 10 hombres, total 30 total de la población
a)

$$P(A) = \frac{10}{30} = 0.3333$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$b) P(A^c) = (20/30) = 0.666 \text{ ó } P(A^c) = 1 - 0.333$$

Ejemplo Nº 2: En su grupo de clase se tiene la siguiente composición: 10 hombres, de los cuales, 2 son casados y el resto solteros. 15 mujeres, 6 casadas, 3 unión libre y el resto solteras. Si se selecciona una persona al azar, cuál es la posibilidad de que: a) sea casada, b) sea hombre, c) no sea hombre, d) casada o soltera, e) hombre o mujer, f) mujer y casada, g) hombre y mujer, h) unión libre o soltera, i) viuda, j) hombre o soltera.

Solución: Otra forma es definir los sucesos teniendo en cuenta las iniciales de lo que representan, así:

M: Suceso que sea mujer. H: Suceso que sea hombre.

S: Suceso que sea soltera. C: Suceso que sea casada.

U: Suceso que sea unión libre.

Solución:

a) Probabilidad de que la persona seleccionada sea casada.

$$P(C) = \frac{8}{25} = 0.32$$

b) Probabilidad de que la persona seleccionada sea hombre.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$P(H) = \frac{10}{25} = 0.40$$

c) Probabilidad de que la persona no sea hombre.

$$P(H^c) = \frac{15}{25} = 0.60$$

Otra manera de calcular esta probabilidad, es utilizando el complemento, así:

$$P(H^c) = 1 - P(H) = 1 - 0.40 = 0.60$$

d) Probabilidad de que la persona seleccionada sea casada o soltera es de 0.88

$$P(CuS) = \frac{22}{25} = 0.8888$$

e) $P(H \text{ ó } M) = 1$. Suceso Seguro.

f) Probabilidad de que la persona sea mujer o esté casada.

$$P(MuC) = \frac{15 + 8 - 6}{25} = 0.68$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Resuelva los numerales g, h, i, j. Interprete los valores y diga si es muy probable o poco probable que sucedan los eventos.

5.6 DEFINICIÓN FRECUENCIAL O EMPÍRICA.

Tiene como base el hecho de que los experimentos son repetibles, cuando la población es muy grande, los resultados son similares a los obtenidos con la definición clásica. La frecuencia relativa de un evento A, estima la probabilidad de dicho evento, es decir, si en N repeticiones de un experimento hay n(A) casos favorables del evento A, entonces:

$$P(A) = \frac{n(A)}{N}$$

5.7 DEFINICIÓN SUBJETIVISTA.

Como su nombre lo indica, cambia dependiendo de quien realiza el análisis, valga decir, para el lector, la posibilidad de terminar una investigación sobre el desempleo en la capital del país, puede ser muy pequeña, por aspectos económicos o por situaciones de inseguridad, mientras que para el Departamento Nacional de Estadística –DANE -, el realizar dicho estudio, es muy probable. En la vida real, se escucha a menudo, que es muy probable que llueva, mientras que otra persona considera que esto no es tan posible.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

5.8 AXIOMAS Y TEOREMAS DE PROBABILIDAD.

1) $P(A) \geq 0$

2) $P(S) = 1$

3) $P(\bar{A}) = 1 - P(A)$

4) $0 \leq P(A) \leq 1$

5) $P(A \cup B) = P(A) + P(B)$, si los sucesos son excluyentes.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$, si los sucesos no son excluyentes.

6) $P(A \cap B) = P(A) * P(B)$, si los sucesos son independientes.

$P(A \cap B) = P(A) * P(B/A)$, si los sucesos son dependientes.

7)

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Esta se conoce como probabilidad condicional. Si los sucesos son independientes, la $P(B/A) = P(B)$

Para ilustrar los casos anteriores, se presenta el siguiente ejemplo:

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Número de registros causas externas región Urabá Antioqueño, año 1998

MUNICIPIOS	HOMICIDIOS	SUICIDIOS	ACCIDENTES TRANSITO	TOTAL
Apartado	143	3	26	172
Arboletes	3	1	3	7
Carepa	28	1	11	40
Chigorodo	50		18	68
Mutata	25	1	1	27
San Juan de Urabá	3		2	5
San Pedro de Urabá	4		1	5
Turbo	67	6	30	103
Total	323	12	92	427

Preparado León Darío Bello P.

Fuente: DANE y D.S.S.A. Murindo y Vigía del Fuerte 0 registros

La tabla anterior, presenta el número de registros de muertes por causas externas en el año 1998, en el Urabá Antioqueño-Colombia-.

Esta es otra manera de tener el espacio muestral, se mantienen datos no actualizados, con la intención que el lector realice el ejercicio con datos propios de su interés.

Al seleccionar un registro al azar se quiere determinar cuál es la probabilidad de que:

- Sea de Carepa.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- b) No sea de Carepa
- c) Sea de Medellín
- d) Haya muerto por accidente de tránsito o se haya suicidado o lo mataron.
- e) Sea de Apartado o Murindó
- f) Sea de Apartado o lo hayan matado.
- g) Si se sabe que es de Mutatá, ¿cuál es la probabilidad de que se haya suicidado?.
- h) Si murió por accidente de tránsito, ¿cuál es la probabilidad de que sea de Chigorodo o San Juan de Uraba?.
- i) ¿Son independientes los sucesos, Municipio y causa externa de muerte?. Demuéstrelo estadísticamente.

Solución:

Dado la cantidad de sucesos posibles, es prudente diferenciarlos con las iniciales respectivas.

a) $P(Ca) = 40/427 = 0.0936$.

b) $P(Ca') = 1 - P(Ca) = 0.9063$.

c) $P(\text{Medellín}) = 0$. Es lo que se conoce como suceso imposible. De aquí se sustenta la teoría de que no hay probabilidad negativa, en el peor de los casos la probabilidad es cero.

d) $P(Ac \text{ ó } Su \text{ u } Ho) = P(Ac) + P(Su) + P(Ho) = 1$.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Como se tomaron todos los posibles sucesos y estos son excluyentes, entonces, con toda seguridad se murió por una de esas causas. De éste caso y el anterior, se demuestra que la probabilidad estará siempre entre cero y uno.

e) $P(Ap \text{ ó } Mu) = P(Ap) + P(Mu) = 172/427 + 0$. Los sucesos son excluyentes. ¿Por qué? Interprete el resultado.

f) $P(Ap \text{ u } Ho) = P(Ap) + P(Ho) - P(Ap \wedge Ho) = 172/427 + 323/427 - 143/427$
 $= 0.8243$

Los sucesos son no excluyentes. ¿Por qué?. Interprete el resultado

g) $P(Su/Mu) = 1/27 = 0.037$

h) $P(Ch \text{ ó } Sju/At) = (18+2)/92 = 0.2173$

La probabilidad de que una persona que se sabe era de Chigorodó o de San Juan de Urabá haya muerto por un accidente de tránsito es de 0.2173, lo que quiere decir que es factible que ocurra. Esta última interpretación es subjetiva, ya que pueden existir diversas lecturas sobre lo que es factible, muy factible o poco factible. El contexto será el que ilustre sobre temas particulares.

i) Anteriormente se colocó la fórmula para calcular probabilidades cuando los sucesos son independientes, basados en ella, se puede demostrar estadísticamente que si los resultados encontrados utilizando los valores

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

observados son similares a los resultados teóricos, se puede concluir que los sucesos son independientes. Otro comentario, es que los resultados son coherentes tomando cualquier categoría de las variables en estudio.

Para el efecto se tomó: Ser de Carepa y haber muerto por accidente de tránsito.

Resultados por fórmula.

$$P(\text{Ca} \wedge \text{At}) = (40/427) * (92/427) = 0.936 * 0.2154 = 0.2066$$

Resultados por tabla.

$$P(\text{Ca} \wedge \text{At}) = (40/427) = 0.0936$$

Se aprecia que las probabilidades calculadas por los dos métodos es muy diferente, por lo tanto, se puede afirmar que el municipio sí tiene incidencia en el tipo de muerte reportada. Esto implicaría que los municipios deben de tomar diferentes medidas para disminuir las causas externas de muerte.

Redacte cinco preguntas para resolver con la tabla anterior, además, seleccionando otros dos sucesos, confirme que el municipio sí influye en la forma de morir.

Dos de los conceptos que requieren mayor entendimiento para poder resolver los problemas que se pueden encontrar son: Independencia y Exclusión.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Independencia se requiere analizar cuando se requiere solucionar problemas que involucran intersecciones, mientras que exclusión trata en interrogantes que tratan sobre la regla de la adición, valga decir, en el tema de unión.

Otro concepto clave para definir independencia, hace referencia al muestreo con reposición, entendido éste como el reemplazo de la unidad seleccionada previamente por otra con las mismas características. Mientras que el muestreo sin reposición es aquel donde disminuye el espacio muestral, dado que no se reemplaza la unidad seleccionada.

La siguiente diapositiva puede ayudar a dilucidar la situación.

SUCESOS INDEPENDIENTES Y EXCLUYENTES

✓ Sean A, B dos eventos, se dicen probabilísticamente independientes si:

$$P(A \cap B) = P(A) P(B) \quad \Rightarrow \quad \begin{aligned} P(A | B) &= P(A) \\ P(B | A) &= P(B) \end{aligned}$$

Muestreo con reposición



2. Sean A, B dos eventos, se dicen EXCLUYENTES si:

$$P(A \cap B) = \emptyset$$

Mismo día
Misma hora



Preparado León Darío Bello P.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

En un mismo instante, no se puede ir viajando en avión y estar por carretera, por eso se llaman sucesos excluyentes, la ocurrencia de un hecho imposibilita la ocurrencia del otro suceso. Mientras que si viaja en avión si puede ir vestido de corbata (sucesos no excluyentes).

Reforzando el concepto de independencia, si a una persona se le incrementa el salario, esto no tiene que ver con que a otra persona de otra empresa le suceda lo mismo, siempre y cuando no sea por ley.

Ejercicios

1. Enumere cinco pares de sucesos los cuales sean: a) Independientes b) Dependientes. Defina cinco pares de sucesos los cuales sean: a) Excluyentes b) No excluyentes. Justifique sus respuestas.

2. Se quieren seleccionar dos entrenadores de natación para que dirijan el equipo que nos representará en los Juegos Nacionales, para lo cual se tienen disponibles 4 de ellos. ¿Cuántos elementos forman el espacio muestral y numérelos. ASUMA QUE SON PEDRO, JOSE, MARÍA Y GONZALO.

3. Con el fin de seleccionar un profesional para dirigir una selección departamental, se citaron a tres ex jugadores del deporte, dos Especialistas en la materia y una persona empírica (sólo experiencia). Dado la idoneidad de los aspirantes, el cargo se va a proveer al azar. a) ¿Cuál es la probabilidad que se seleccione un Especialista? b) ¿Se seleccione un ex jugador? c) ¿No se seleccione un Especialista? d) ¿Sea Especialista o ex jugador? e) ¿Sea Estadístico?.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

4. Para el caso anterior, si se seleccionan dos personas, cuál es la probabilidad de que: f) ambos sean ex jugadores, g) todos Especialistas, h) uno ex jugador y el otro Especialista, l) El primero Especialista y el segundo empírico, j) Siquiera uno ex jugador, k) ninguno ex jugador. Si se seleccionan tres aspirantes cual es la probabilidad de: l) Dos sean Especialistas, m) mínimo uno sea ex jugador, n) mínimo dos sean Especialistas, o) Ninguno sea Especialista, p) todos sean Especialistas, q) Redacte otra pregunta que se acomode a los datos del problema.

5. Luego de realizar una encuesta sobre el desempleo en el área deportiva se obtuvo la siguiente tabla.

	EMPLEADO	DESEMPLEADO	TOTAL
HOMBRES	460	40	500
MUJERES	140	260	400
TOTAL	600	300	900

Si se selecciona una persona al azar, cuál es la probabilidad de que:

- Sea empleado o mujer.
- Sea hombre o mujer
- Sea desempleado y hombre.
- Si es desempleado, cuál es la probabilidad de que sea hombre?
- Son los sucesos independientes, porqué?

6. El número de estudiantes admitidos para el primer semestre del 2002, en tres facultades (datos ficticios), se muestra en la tabla siguiente:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

SEXO	ENFERMERIA	MEDICINA	EDUCACIÓN FÍSICA
HOMBRES	30	150	60
MUJERES	170	50	60

Determine, cuál es la probabilidad de que:

- No sea estudiante de Medicina.
- Sea de Medicina o Enfermería.
- Sea hombre o de Enfermería.
- Sea mujer y sea de Enfermería.
- Sea hombre de Educación Física
- Si es de Educación Física. ¿Cuál es la probabilidad de que sea hombre?
- Si es hombre, ¿cuál es la probabilidad de que estudie Medicina o Enfermería?
- ¿Son los sucesos estadísticamente independientes? Demuéstrelo matemáticamente.
- Redacte dos preguntas que se puedan resolver con los datos anteriores.

5.9 TEOREMA DE PROBABILIDAD TOTAL.

Como condición necesaria se tiene que el espacio muestral está conformado por varios sucesos mutuamente excluyentes A_i cuya probabilidad es la unidad. El interés es encontrar la probabilidad de otro suceso B , que tiene parte (intersección) en todos los A_i .

Por lo tanto, éste teorema es una suma de intersecciones, es decir, la

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

probabilidad de ocurrencia de un suceso, es igual a la posibilidad de que se puedan dar varias opciones. Además, estas probabilidades son condicionales. Veamos el **siguiente ejemplo** para ilustrar lo mencionado anteriormente.

De un grupo de personas, algunas están afiliadas a diferentes EPS (Entidades Promotoras de Salud), COOMEVA, otras a SUSALUD y otras a COMFENALCO. También se sabe, que hay algunas cotizantes y otras beneficiarias. Si está interesado en determinar la probabilidad de que una persona seleccionada al azar sea cotizante, es claro que dicha persona puede ser de una de las tres EPS. Por lo tanto, se requiere conocer la probabilidad de que sea de cada EPS y además, saber la probabilidad de ser cotizante de cada una de ellas.

La solución se puede plantear así:

$$P(\text{Cotizante}) = P(\text{Cooomeva y cotizante}) + P(\text{Susalud y cotizante}) + P(\text{Comfenalco y cotizante})$$

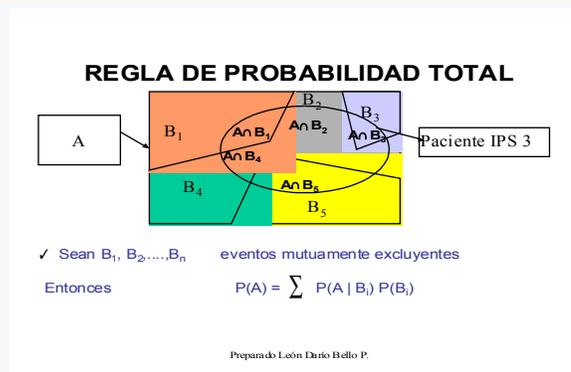
Que al desarrollar las intersecciones se puede escribir de la siguiente manera:

$$P(\text{Cotizante}) = P(\text{Cooomeva}) * P(\text{Cotizante/Cooomeva}) + P(\text{Susalud}) * P(\text{Cotizante/Susalud}) + P(\text{Comfenalco}) * P(\text{Cotizante/Comfenalco}).$$

Lo anterior, se puede generalizar usando la siguiente expresión:

$$P(B) = \sum (A_i) * P(B/A_i) \quad (\text{donde "i" toma valores entre 1 y n})$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



Si se define B como el suceso de que sea cotizante y A_i las diferentes EPS, entonces la probabilidad de que ocurra el suceso B es igual a la suma de multiplicar cada una de las probabilidades condicionadas de este suceso con los diferentes sucesos A por la probabilidad de cada suceso A . La siguiente diapositiva muestra lo planteado anteriormente, presentando al suceso A como el evento a calcular, esto indica, que la letra asignada a cada suceso es definida por el investigador y en nada influye en el resultado final.

Es importante tener en cuenta que los sucesos forman un espacio muestral y que además, son mutuamente excluyentes.

Veamos otro posible caso:

“Usted tiene para invertir en: acciones de empresas, en bonos a largo plazo y en comprar divisas extranjeras de su país. Recibe asesoría de tres empresas de valores. Preguntese ¿qué opciones tiene si decide comprar acciones?”

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

¿Reflexionó?. Pues puede comprar acciones en la empresa E_1 , o E_2 o E_3 . La solución se puede expresar así:

$$P(\text{Acciones}) = (P(E_1nA) + P(E_2nA) + P(E_3nA))$$

Ejemplo 1: La probabilidad de que a una persona le dé infarto al miocardio si tienen más de 50 años es del 0.3% y para 50 o menos del 0.1%. Si la probabilidad de que en esa población una persona tenga más de 50 años es del 25% ¿Cuál es la probabilidad de que se presente un infarto en esa población?

$A_1 = \{\text{más de 50 años}\}$ $A_2 = \{50 \text{ o menos}\}$ estos sucesos constituyen el espacio muestral.

$B = \{\text{padecer infarto}\}$

$$P(B|A_1) = 0.003; p(B|A_2) = 0.001; p(A_1) = 0,25$$

$$P(A_2) = 0.75$$

La probabilidad de que en esa región se presente un infarto al miocardio es de 0.0015, es decir, es poco probable que se presente éste hecho.

$$P(B) = 0.003 \times 0.25 + 0.001 \times 0.75 = 0.0015$$

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejemplo 2: Una población está formada por tres grupos de profesionales: Psicólogos, 30%, Sociólogos 10% y Administradores 60%. Además, se sabe que el porcentaje de personas con maestría en cada una de estas profesiones es, respectivamente, del 20%, 40% y 5%. Se quiere determinar la probabilidad de que un individuo elegido al azar tenga maestría.

Si se definen los sucesos así:

A: Psicólogos, B: Sociólogos, C: Administradores, M: Maestría

Se tiene:

$$P(M) = P(A) \cdot P(M/A) + P(B) \cdot P(M/B) + P(C) \cdot P(M/C) = 0.3 \cdot 0.2 + 0.1 \cdot 0.4 + 0.6 \cdot 0.05 = 0.13.$$

La interpretación será siempre redactar la pregunta, para el caso se tiene: La probabilidad de que una persona seleccionada al azar de esa población haya realizado estudios de Maestría es de 0.13, es decir, que es relativamente baja.

5.10 TEOREMA DE BAYES.

Se utiliza para calcular la probabilidad de las causas a partir de los efectos que han podido ser observado, es una generalización de las probabilidades condicionales, ya que el interrogante a solucionar es una condicional. El nombre lo toma de su desarrollador, Thomas Bayes (1702-1761). Se apoya en el teorema de probabilidad total para su cálculo. La probabilidad $P(A_i/B)$ viene dada por la expresión:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

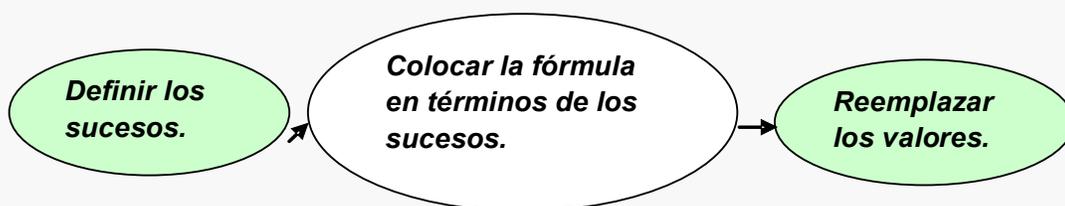
$$P(A_i / B) = \frac{P(A_i) \cdot P(B / A_i)}{P(A_1) \cdot P(B / A_1) + P(A_2) \cdot P(B / A_2) + \dots + P(A_n) \cdot P(B / A_n)}$$

El denominador es igual al resultado de la probabilidad total, por lo tanto, cumple los requisitos que tienen los de la probabilidad total.

$$p(A_i | B) = \frac{p(B|A_i)p(A_i)}{\sum_{j=1}^n p(B|A_j)p(A_j)} \quad \text{para } i = 1, \dots, n$$

Es claro que el denominador es mayor que cero, para evitar la indeterminación.

La **metodología** sugerida para este tipo de problemas que son considerados **complejos**, y luego de releer el enunciado, es la siguiente:



Se sugiere lo anterior, ya que los enunciados al ser en alguna medida más extensos que los que se han trabajado hasta ahora, pueden confundir, además, es más fácil ubicar las diferentes probabilidades sabiendo qué se está buscando.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Ejemplo 1: El editor de un libro, está analizando la posibilidad de editar un nuevo ebook de estadística con énfasis en la investigación de mercados. Dado la poca oferta de estos textos, se determina que la posibilidad de tener éxito es de 40%. Luego de evaluar el texto y el mercado potencial, se encuentra que: El 80% de los textos exitosos tuvieron un informe favorable, y el 30% de los que no tuvieron éxito, también recibieron informes favorables. Encuentre la probabilidad de que si se recibió un informe favorable, el libro sea un éxito.

Sean los sucesos:

A: Suceso que el informe sea favorable.

B: Suceso que sea un éxito.

La diferencia con la regla de probabilidad total, radica en que ésta última no tiene un conocimiento previo de la ocurrencia de otro suceso, mientras que en el teorema de Bayes, siempre se conoce algo, de ahí, que se calcula una probabilidad condicional. Sin embargo, se puede calcular primero el denominador, que no es más que el resultado de la probabilidad total.

$$P(B/A) = P(B) \cdot P(A/B) / P(A)$$

$$P(A) = P(B) \cdot P(A/B) + P(B') \cdot P(A/B') = 0.4 \cdot 0.8 + 0.6 \cdot 0.3 = 0.50.$$

$$P(B/A) = 0.32 / 0.50 = 0.64$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

La probabilidad de que un libro sea exitoso, sabiendo que fue evaluado positivamente es de 0.64, lo que equivale a decir, que es más probable que tenga éxito que fracase, ya que la probabilidad de que ocurra esto último. si recibió informe positivo es el complemento. es decir. 0.36.

Ejemplo 2: Un estudiante sabe que hay 10 preguntas que pueden formularle en un examen. Sin embargo, sólo tiene tiempo para estudiar a fondo tres de ellas, y las selecciona al azar. Supongamos que la probabilidad de que apruebe si le preguntan una de las que estudio, es de 0.95, mientras que la probabilidad de que apruebe si figura una de las otras preguntas es de 0.20. El examen contiene sólo una pregunta, y es una de las diez. A) Calcule la probabilidad de que pierda el examen. B) sí el dijo que ganó el examen. Cuál es la probabilidad de que le hayan preguntado una de las que no estudio?

Definición de sucesos:

A: Suceso que pierda el examen (pudo definirlo que gane el examen).

B: Suceso que le pregunten una de las que estudio.

Note que los sucesos salen de las preguntas, por eso, se recomienda colocar la fórmula a aplicar antes de ubicar los datos del problema.

$$a) P(A) = P(B) \cdot P(A/B) + P(B') \cdot P(A/B')$$

El interrogante cumple las características de la RPT. Se reemplazan las

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

respectivas probabilidades, así:

$$P(A) = 0.30 \cdot 0.05 + 0.70 \cdot 0.80 = 0.575.$$

$$P(A') = 1 - P(A) = 0.425 \quad P(\text{Ganar})$$

En estas circunstancias, se tiene un poco más de probabilidad de perder que de ganar.

b) Interpretar no es más que responder la pregunta. La probabilidad de que una persona que ganó el examen, le hayan preguntado algo que no estudio es de 0.3294. Tiene casi una tercera parte de ganar en esas circunstancias.

$$P(B' / A') = \frac{P(B') \cdot P(A' / B')}{P(A')} = \frac{0.70 \cdot 0.20}{0.425} = 0.3294$$

Considero que es el momento de acceder al OA sobre éste tema, visite el enlace: www.leondariobello.com/OA/teoremadebayes y refuerce los conceptos.

5.11 APLICACIÓN DE LAS PROBABILIDADES A LA MEDICINA.

La medicina dentro de sus características se encuentra la incertidumbre, el riesgo. La incertidumbre va desde actividades preventivas, terapéuticas, predictivas y diagnósticas. En el proceso de diagnóstico, están: la historia clínica, el examen físico y si es del caso, pruebas complementarias. Estas últimas pretenden aclarar las dudas sobre el estado del paciente.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Cuando se realizan varias pruebas, a la vez, se incrementa la probabilidad de diagnosticar a un enfermo, pero también aumenta la probabilidad de considerar como enfermo a un sano.

Por lo tanto, hay que realizar esfuerzos para lograr una buena prueba diagnóstica, es decir, que se encuentren probabilidades muy altas de resultados positivos en enfermos y negativos en sanos.

Las pruebas diagnósticas deben cumplir:

- **Validez:** Es el grado en que un test mide lo que se supone que debe medir. La sensibilidad y la especificidad de un test son medidas de su validez.
- **Reproductividad:** Es la capacidad del test para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares.
- **Seguridad:** Se determina por el valor predictivo de un resultado positivo o negativo.

Acá, se presentan algunos conceptos necesarios para abordar el tema de la validez de las pruebas diagnósticas, veamos.

- **Incidencia:** Porcentaje de nuevos casos de la enfermedad en la población.
- **Prevalencia:** Porcentaje de la población que presenta una enfermedad.
- **Sensibilidad:** Indica la acertividad de una prueba para clasificar

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

adecuadamente a un individuo en términos de identificarlo como enfermo cuando realmente lo es. Se da en proporción o porcentaje, mientras más cerca a 100% es mejor. También se conoce como la probabilidad de encontrar un verdadero positivo.

$$\text{Sensibilidad} = \frac{\text{Enfermos}_{+}}{\text{Enfermos}_{-} + \text{Enfermos}_{+}}$$

$$S = EP / (EP + EN)$$

- **Especificidad:** Indica la acertividad de una prueba para clasificar adecuadamente a un individuo en términos de identificarlo como sano cuando realmente lo es. Como para el caso anterior, se da en proporción o porcentaje, mientras más cerca a 100% es mejor. También se conoce como la probabilidad de encontrar un verdadero negativo.

$$\text{Especificidad} = \frac{\text{Sanos}_{-}}{\text{Sanos}_{-} + \text{Sanos}_{+}}$$

Los índices de calidad de las pruebas anteriores, realmente son insumos para calcular lo realmente importante en la medicina; ahí es donde surge de nuevo el teorema de Bayes.

Otro aspecto importante en las pruebas, es medir la seguridad, para ello, se requiere calcular los valores predictivos positivos y negativos.

- **Valor predictivo positivo:** Es la probabilidad de padecer la

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

enfermedad si se obtiene un resultado positivo en el test.

$$VPP = \frac{VP}{VP + FP}$$

- **Valor predictivo negativo:** Es la probabilidad de que un sujeto con un resultado negativo en la prueba esté realmente sano.

$$VPN = \frac{VN}{FN + VN}$$

Para visualizar de mejor manera se construye la siguiente tabla:

	Diagnóstico	
Resultados Prueba	Enfermo	Sano
Positivo	VP(verdadero +)	FP (Falso +)
Negativo	FN (Falso -)	VN(verdadero -)

Aplicamosla a los siguientes resultados hipotéticos de la prueba de una ecografía de seno y su correspondiente biopsia.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Resultados Ecografía	Resultado Biopsia		Total
	Cancer Seno	Sano	
Anormal	785	265	1050
Normal	490	1960	2450
Total	1275	2225	3500

$$\text{Sensibilidad} = \frac{785}{785 + 490} = 0.6156 = 61.56\%$$

$$\text{Especificidad} = \frac{1960}{2225} = 0.8808 = 88.08\%$$

$$\text{VPP} = \frac{785}{1050} = 0.7476 = 74.76\%$$

$$\text{VPN} = \frac{1960}{2450} = 0.80 = 80\%$$

$$\text{RVP} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}} = \frac{0.6156}{1 - 0.8808} = 0.4964$$

$$\text{RVN} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}} = \frac{1 - 0.6156}{0.8808} = 0.4364$$

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Refuerce lo conversado, entre a la dirección www.leondariobello.com/OA/pruebasdiagnosticas

Razones de probabilidad: Son índices que permiten comparar métodos diagnósticos sin depender de la prevalencia de la enfermedad, algunos programas estadísticos los entregan por defecto, como es el caso del Epidat 3.1, a continuación se presentan dos de ellos.

- **Razón de verosimilitudes positiva:** se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos. Es, el cociente entre la sensibilidad y 1-especificidad.

$$RV+= \text{Sensibilidad}/(1-\text{Especificidad})$$

- **Razón de verosimilitudes negativa:** se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma. Es cociente entre 1-sensibilidad y especificidad.

$$RV-= (1- \text{Sensibilidad})/\text{Especificidad}$$

Se trata de dilucidar el asunto calculando los indicadores anteriores usando la siguiente tabla:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

	Tiene cáncer de próstata		Total
	Sí (E ⁺)	No (E ⁻)	
Prueba de antígeno			
Positivo (P ⁺)	60	120	180
Negativo (P ⁻)	15	180	195
Total	75	300	375

Sensibilidad= Enfermos positivos/ (Enfermos positivos + Enfermos negativos)

$$S = E^+ / (E^+ + E^-) = 60 / 75 = 0.80$$

Especificidad= Sanos negativos/ (Sanos positivos + Sanos negativos)

$$E = S^- / (S^- + S^+) = 180 / 300 = 0.60$$

V. predictivo positivo= Positivos enfermos / (Positivos enfermos + Positivos no enfermos)

$$VPP = 60 / 180 = 0.33$$

V. predictivo negativo= Negativo sano / (Negativo sano + Negativo enfermo)

$$VPN = 180 / 195 = 0.9230$$

RV+= Sensibilidad/(1-Especificidad)

RV+= 0.80/(1-0.60)= 2.0. Indica que la prueba de antígeno es dos veces

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

probable que detecte el problema en personas con cáncer que en los que no lo tienen.

$RV = (1 - \text{Sensibilidad}) / \text{Especificidad}$

$RV = (1 - 0.20) / 0.60 = 0.33$

5.12 TESTS DIAGNÓSTICOS USANDO BAYES.

Las pruebas diagnósticas sirven para estimar de mejor manera la probabilidad de que un individuo presente una enfermedad. Los tests diagnósticos son una aplicación del teorema de Bayes, teniendo como base los conceptos de sensibilidad y especificidad, estos valores son estimados previamente. En la medicina, interesa responder preguntas como: si un paciente resultó positivo, ¿cuál es la probabilidad de que esté verdaderamente enfermo?, o si el sujeto resultó negativo en la prueba ¿cuál es la probabilidad de que realmente esté sano? Estas probabilidades se calculan usando el teorema de Bayes.

Ejemplo 1: Según estudios previos en una IPS (Institución Prestadora de Salud), la diabetes afecta al 15% de los usuarios afiliados. La presencia de glucosuria se usa como indicador de diabetes. Su sensibilidad es de 0,2 y la especificidad de 0,98. Calcular los índices predictivos, verdaderos positivos y verdaderos negativos.

Definición de sucesos:

E: Suceso que el usuario tenga diabetes, esté enfermo.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

V+: Suceso que la prueba dio positivo.

V-: Suceso que la prueba dio negativo.

$$P(E/V_+) = \frac{P(E) * P(V_+ / E)}{P(V_+)} = \frac{0.15 * 0.20}{0.15 * 0.20 + 0.85 * 0.02} = 0.6382$$

$$P(E^c / V_-) = \frac{P(E^c) * P(V_- / E^c)}{P(V_-)} = \frac{0.85 * 0.98}{0.85 * 0.98 + 0.15 * 0.80} = 0.8740$$

Ejemplo 2: Se ha identificado que aproximadamente el 10% de las personas con más de 70 años, sufren de algún tipo de artritis. Se aplica una prueba y efectivamente detecta al 85% de los enfermos, sin embargo, el 4% da falsos positivos. Determine la probabilidad de que si la prueba dio positiva, realmente padezca la enfermedad. Además, sino dio positiva, cuál es la probabilidad de que realmente esté sana la persona?

Sucesos:

- E: Padecer la *enfermedad* (artritis).
- E^c: No padecer la enfermedad.
- T+: El resultado del test es positivo.
- T-: El resultado del test es negativo.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$P(E/T+) = \frac{P(E) * P(T+ / E)}{P(T+)} = \frac{0.10 * 0.85}{0.10 * 0.85 + 0.90 * 0.04} = 0.70$$

Se tiene una probabilidad de 0.70 de que si el test dio positivo, el paciente realmente tenga la enfermedad. Obviamente, mientras más alta sea esta probabilidad, la prueba es mejor.

$$P(E^c / T-) = \frac{P(E^c) * P(T- / E^c)}{P(T-)} = \frac{0.90 * 0.15}{0.90 * 0.15 + 0.10 * 0.96} = 0.5844$$

Se tiene una probabilidad de 0.5844 de que si el test dio negativo, el paciente realmente éste aliviado.

Ejercicios

1. Con el fin de seleccionar un profesional para dirigir una unidad intermedia, se citaron a cuatro Médicos Generales, tres Especialistas en Pediatría y dos Gerentes en Sistemas de Información en Salud. Dado la idoneidad de los aspirantes, el cargo se va a proveer al azar. a) Cuál es la probabilidad de que se seleccione un Especialista? b) De que se seleccione un Médico General? c) No se seleccione un Especialista? d) Sea Gerente de Sistemas de Información en Salud. o Especialista? e) Sea Estadístico. f) Redacte dos interrogantes más y los soluciona.

2. Para el caso anterior, si se seleccionan dos personas, cuál es la probabilidad de que: f) ambos sean Médicos Generales, g) todos

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Especialistas, h) uno Gerente en Sistemas de Información y el otro Especialista, l) El primero Especialista y el segundo Gerente en Sistemas de Información, j) Siquiera uno Médico General, k) ninguno Médico General. Si se seleccionan tres aspirantes cuál es la probabilidad de: l) Dos sean Médicos Generales, m) mínimo uno sea Médico General, n) mínimo dos sean Médicos Generales, o) todos sean Médicos Generales, p) se seleccione uno de cada profesión, q) Redacte otra pregunta que se acomode a los datos del problema.

3. En un experimento para estudiar la dependencia de la hipertensión de los hábitos de fumar, se obtuvieron los siguientes datos:

	No Fumadores	Fumadores Moderados	Fumadores fuertes
Con Hipertensión	150	250	600
Sin Hipertensión	500	150	50

Si se elige una persona al azar, determine, cuál es la probabilidad de:

- Sea fumador fuerte y tenga Hipertensión.
 - No fume o tenga problemas de hipertensión.
 - Si se sabe que no tiene problemas de hipertensión, cual es la probabilidad de que se no fumador o fumador moderado.
 - Si se sabe que es fumador fuerte, cuál es la probabilidad de que tenga hipertensión.
4. Una mujer portadora de hemofilia transmite la enfermedad a la mitad de sus hijos varones y ninguna a sus hijas. Cuál es la probabilidad de que:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- a) El siguiente hijo hombre esté afectado.
- b) De los dos siguientes hijos varones, ambos tengan la enfermedad.
- c) Si no se sabe el sexo, cuál es la probabilidad de que uno tenga la enfermedad.
- d) Si no se sabe el sexo, cuál es la probabilidad de que ninguno tenga la enfermedad.

5. Si la probabilidad de que un vendedor encuentre a un usuario en casa es de 0.8 el día sábado, cuál es la probabilidad de que en dos visitas hechas en un día sábado, a) ambos usuarios estén en casa, b) sólo uno éste en casa, c) todos estén en casa, d) ninguno esté, e) por lo menos uno, f) siquiera uno esté.

6. Cierta intervención quirúrgica tiene un índice de mortalidad de 30%. Si esta operación es realizada de manera independiente en tres pacientes distintos, cual es la probabilidad de que: a) las tres intervenciones sean mortales, b) todos se salven, c) ninguno se muera, d) se salven exactamente dos, e) siquiera uno se salve.

7. El número de profesionales que pretenden ser admitidos en el Censo poblacional del año 2005, para labores de campo se muestra a continuación, (datos ficticios):

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Región	Administradores	Economistas	Estadísticos
Antioquia	350	150	20
Choco	70	35	-----
Córdoba	150	40	5
Total	570	225	25

Determine, cuál es la probabilidad de que:

- Sea Administrador de Córdoba.
- Sea de Córdoba o Antioquia.
- Sea de Choco o estadístico..
- Sea de Antioquia y no sea economista.
- Si es de Choco, cuál es la probabilidad de que sea Estadístico o Administrador.
- ¿Son los sucesos estadísticamente independientes? Demuéstrelo matemáticamente.

8. Después de las 11:00 P.M la Empresa Metrocable dispone de 2 burbujas para transportar pasajeros en caso de emergencia. Se estima que la probabilidad de que una burbuja esté disponible cuando se requiera es de 0.9. Si la disponibilidad de una burbuja es independiente de la otra calcule la probabilidad de que:

- Ambos vehículos estén disponibles.
- Ninguno esté disponible
- Un servicio requerido sea satisfecho

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- d). Uno de los dos esté disponible.
- e). Siquiera uno esté disponible.

9. La posibilidad de que una persona de determinada edad con el hábito de beber fallezca dentro de 10 años o más, es de $\frac{1}{4}$, y la posibilidad de que muera si tiene el hábito de fumar en ese mismo lapso de tiempo, es de $\frac{1}{3}$. Hallar la posibilidad de que:

- a). Ambos mueran dentro de 10 años
- b). Al menos uno muera dentro de 12 años
- c). Muera el que fuma y no el que bebe.
- d). No muera ninguno dentro de los próximos 10 años.
- e). Si se sabe que murió el que fuma, determine la probabilidad de que también muera el que bebe.

10 Un laboratorio está considerando comercializar una nueva droga contra la impotencia en los hombres. De acuerdo con una investigación realizada se encontró que: la probabilidad de que el producto tenga éxito es 0.80 si un laboratorio competidor no introduce un producto similar en el mercado, en tanto que la probabilidad de éxito es de 0.35 si la firma competidora comercializa el producto, además, el laboratorio estima que hay una probabilidad de 0.40 de que la firma competidora comercialice el producto.

a) Dado que el laboratorio tuvo éxito. Cual es la probabilidad de que el laboratorio de la competencia haya comercializado su producto?, b) Cual es la probabilidad de que no se tenga éxito?.

11. Un estudiante sabe que hay 20 preguntas que pueden formularle en un

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

examen. Sin embargo, sólo estudia a conciencia cinco de ellas, y las selecciona al azar, Suponga que la probabilidad de que gane si le preguntan una de las que estudio, es de 0.90, mientras que la probabilidad de que apruebe si figura una de las otras es solamente de 0.15. El examen contiene sólo una pregunta, y es una de las 20.

a). ¿Cuál es la probabilidad de que no aprueba el examen?

b) Suponga que él dice que aprobó el examen. ¿Cuál es la probabilidad de que la pregunta por él seleccionada halla sido la formulada en la prueba?

12. En un hospital ingresa un promedio de 50% de enfermos con problemas respiratorios, 30% con dificultades circulatorias y 20% con enfermedades referentes al corazón. La probabilidad de curación completa de afecciones respiratorias es de 0.7, para las afecciones circulatorias y del corazón, estas probabilidades son respectivamente 0.8 y 0.9. a) Un enfermo internado en el hospital fue dado de alta sano. Hallar la probabilidad de que éste enfermo sufría afección respiratoria, b) cuál es la probabilidad de que no sufra enfermedad del corazón, c) si se sabe que sufre de enfermedad circulatoria, cual es la probabilidad de que no se cure completamente?

13. En Medellín (Antioquia), el 50% de las personas son conservadoras, el 10% son liberales y el 40% son independientes. Los registros muestran que en las elecciones anteriores, votaron el 75 % de los Conservadores, el 80% de los Liberales y el 50% de los independientes. Si se selecciona una persona al azar.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- a) Cuál es la probabilidad de que vote?
- b) Si se sabe que no votó en las elecciones pasadas. Cuál es la probabilidad de que sea liberal?.
- c) Si es liberal, determine la probabilidad de que no haya votado.

14. El municipio, ha clasificado tres categorías de SISBEN el año pasado (A, B, C), de los carnés del Sisben, el 70% ha sido clasificado como A, el 20% como B y el 10% como C. De los clasificados como A, el 50% fue emitido por la Nación, el 40% por el Departamento y el 10% por el Municipio. De los clasificados como B, el 60% fue de la nación, el 20% por el departamento y el 20% por el municipio. De los clasificados como C, el 90% fue emitido por la Nación, el 5% por el Departamento y el 5% por el Municipio. a) Que proporción de carnés de Sisben han sido emitidos por el Departamento?, b) Si el departamento va a emitir un nuevo carné. Cuál es la probabilidad de que recibirá clasificación de A?, c) Que proporción de carnés de Sisben no han sido emitidos por el municipio?

15. Para poder realizar una operación de páncreas, se requiere de mínimo 2 especialistas en el tema. La probabilidad de que un especialista éste disponible es de 0.2. La clínica cuenta con 3 de ellos. Si la disponibilidad de un especialista es independiente de la otro calcule la probabilidad de que:

- a). Se pueda realizar la operación.
- b). Sólo uno esté disponible.
- c). La operación no se pueda llevar a cabo.
- d). Siquiera dos estén disponibles.

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

16. En tres regiones del país, se tienen diferentes posibilidades de que ataque un virus contagioso, estas son: Caribe, 30%, Cafetera 10% y Pacifico 60%. Además, se sabe que el porcentaje de personas que se han inmunizado contra el virus, por región es del 20%, 40% y 5% respectivamente. a) Calcule la probabilidad de que si se selecciona una persona al azar ésta no sea del Caribe, b) Probabilidad de que se haya inmunizado, c) si se sabe que es del eje cafetero, cuál es la probabilidad que no se haya inmunizado contra el virus?, d) si se sabe que se inmunizó, cuál es la probabilidad de que sea del Pacifico?.

17. El Congreso de la república aprobó una ley en la cual los empleados públicos se deben someter a una prueba para detectar su adicción a sustancias psicoactivas. Se estima que el 1.5% de los empleados públicos tienen esta dificultad. Se sugiere realizar una prueba en la que da positiva el 98% de las veces en los empleados que tienen el problema. De otro lado, la prueba da negativa en los empleados que no son adictos el 99% de las veces. Si se selecciona un empleado público, se aplica la prueba y da positiva, ¿cuál es la probabilidad de que sea adicta?. ¿De que no lo sea?.

18. Una prueba diagnóstica para el cáncer de mama tiene una probabilidad de 0.03, para detectar a sanos como enfermos, mientras que la probabilidad de dar un resultado negativo sabiendo que está enferma es de 0.08. En esa población, una mujer tiene una probabilidad de 0,10 de padecer la enfermedad. Encuentre la probabilidad de que si la prueba dio positiva, esté realmente enferma.

19. Una prueba se diseñó para diagnosticar tempranamente un tumor

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

maligno en el Colon, la prueba da falsos negativos y falsos positivos en un 5% de las veces. En la región del Oriente Antioqueño, el 2% tienen éste problema de Colon. ¿Cuál es la probabilidad de que una persona de la región que dio positiva en la prueba realmente la tenga?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

CAPITULO 6

DISTRIBUCIONES DE PROBABILIDAD

Con Excel

Es necesario, realizar algunos comentarios sobre las distribuciones de probabilidad, ya que son el soporte para entender de dónde salen los modelos probabilísticos, que son los que realmente se utilizan en el ámbito de la investigación.

Las distribuciones son discretas o continuas, las cuales se usan dependiendo del nivel de medición de la variable. Las discretas, tienen su soporte en la sumatoria de casos, mientras que las continuas, lo hacen de las integrales que permiten calcular áreas y por ende probabilidades.

Procurando ser lo más práctico posible, se presentan los conceptos de las distribuciones, teniendo como base las variables discretas, ya que las continuas, tienen sus propias ayudas para minimizar los cálculos.

6.1 VARIABLES ALEATORIAS (v.a).

Como su nombre lo indica, son valores que cambian y que de alguna manera, están influenciadas por el azar. Dado que el énfasis está en las variables discretas, se define la variable continua acá y no se mencionará más. Variable Aleatoria Continua es una variable que toma un número infinito de valores en un intervalo dado. Como ejemplos se pueden mencionar: Salario de empleados, peso de adolescentes, la estatura de las madres, etc. Cumplen las mismas características de las definidas en el capítulo 1.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.1.1 Variables Aleatorias Discretas: Son valores numéricos que representan los resultados de un espacio muestral, son resultados numerables y excluyentes. Como ejemplos se pueden mencionar: **Número de** accidentes de tránsito, **número de** reclamaciones de, **número de** artículos defectuosos, **número de** llamadas telefónicas, etc. Observe que todas las variables empiezan como **número de**.

6.1.2 Función de Probabilidad Discreta: Si a cada valor de la v.a se le puede asignar una probabilidad, entonces, se dice que es una **función de probabilidad**, y si además, la suma de estas probabilidades es uno, se dice que es una **función de densidad de probabilidad (f.d.p)**.

También, se puede decir que una **f.d.p** (minúsculas) es una tabla o un gráfico que permite apreciar la forma de la v.a con respecto a sus probabilidades. Se nombra así: $f(X) = P(X=x_i)$, donde x_i es el valor de la variable cuya probabilidad se quiere calcular.

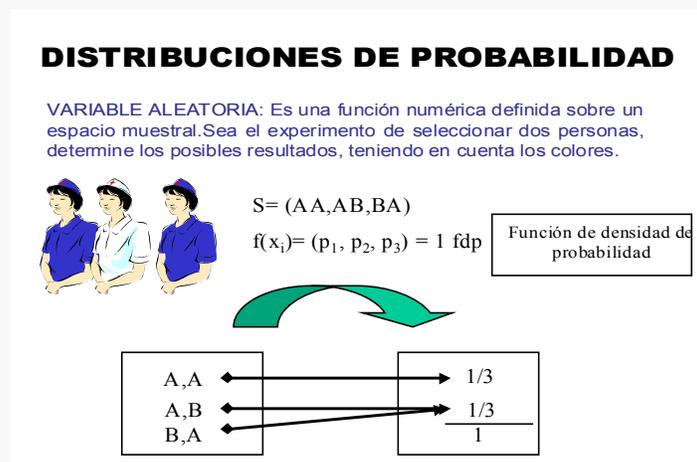
6.1.3 Función de Probabilidad Acumulada Discreta (FDA): Basados en la parte de construcción de tablas de frecuencia que se analizaron en el capítulo 2, se puede calcular las frecuencias acumuladas, es decir, sumando parcialmente la f.d.p.

$$F(x) = P(X \leq x_i).$$

Esta función representa la probabilidad de que la variable aleatoria sea menor o igual que un determinado valor:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Gráficamente, la función aumenta de "a saltos", ya que entre dos valores consecutivos de una variable discreta, no puede tomar valores intermedios. La diapositiva ilustrativa de este tema es la siguiente:



Ejemplo 1: De un grupo de investigadores conformado por 6 sociólogos y 4 psicólogos, se seleccionan 3 profesionales con el fin de entrar a negociar el presupuesto de un estudio, a) determine si el experimento es una f.d.p. teniendo en cuenta que nos interesa el número de sociólogos en la muestra, b) Realice el gráfico de la distribución, c) calcule la FDA.

Como ya se mencionó se construye una tabla con los posibles resultados de la variable en estudio (sociólogos).

X: Número de sociólogos en la muestra.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Número de Sociólogos (X)	$P(X=x_i) = f(x_i)$	$P(X \leq x_i) = F(x_i)$
0	$(4/10) \cdot (3/9) \cdot (2/8) = 0.033$	0.033
1	$(6/10) \cdot (4/9) \cdot (3/8) = 0.30$	0.333
2	$(6/10) \cdot (5/9) \cdot (4/8) = 0.50$	0.833
3	$(6/10) \cdot (5/9) \cdot (4/8) = 0.166$	1.00
	1.00	

f.d.p suma = 1.0

Parece evidente que si hay más de 3 psicólogos, entonces, cabe la posibilidad de que todos los seleccionados sean de dicha categoría, por lo tanto, puede suceder que ninguno sea sociólogo, pero también esta la opción de que todos sean de esta clase.

Con respecto a las probabilidades, se utiliza un muestreo sin reemplazo, ya que luego de seleccionar un profesional, este no puede ser seleccionado nuevamente, por lo tanto, el denominador va disminuyendo. En algunos casos, se multiplican las probabilidades por un término, que se origina de las posibles combinaciones que se pueden dar, valga decir, para $X=1$, la condición es que uno sea sociólogo, pero ese puede ser cualquiera de los tres seleccionados, es decir, pudo ser el 1º ó el 2º ó el último.

La interpretación de las probabilidades es idéntica a la mencionada anteriormente. Para $X=2$, la probabilidad de que dos sociólogos queden en la muestra es de **0.50**, es decir, es igualmente probable que quede 2 sociólogos a que esto no suceda.

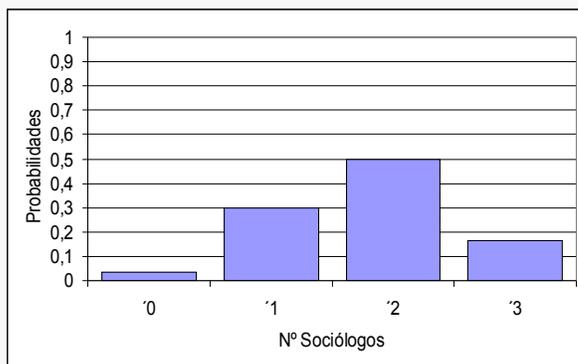
Dado que la suma de las probabilidades es igual a 1, se puede afirmar que

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

las parejas de X_i y P_i forman una f.d.p. Esto es importante en la medida que nos sirve para calcular otros indicadores.

Al sumar de manera parcial la f.d.p, se construye la FDA, la cual entrega las probabilidades desde un valor de la v.a hacia abajo, por ejemplo: 0.333, que se lee: la probabilidad de que en una muestra de 3 personas queden 1 o ninguno es de 0.333, por lo tanto, la posibilidad de que se seleccionen 2 o más es de 0.667. Dicho de otra manera es más probable que ocurra esto último.

El gráfico no es más que un histograma separado, entregando la misma información, es decir, permite observar que es más probable que suceda. Para el caso se tiene:



Se aprecia que es más probable que queden 2 sociólogos en la muestra (0.50), mientras que lo menos probable es que no quede ninguno de ellos (0.033).

6.1.4 Valor Esperado o Esperanza Matemática Discreta: Es el promedio que se espera suceda dado unas condiciones de incertidumbre. Para su

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

calculo se requiere multiplicar los valores de la v.a por sus respectivas probabilidades.

Entre otras aplicaciones, se tiene la de encontrar la mejor alternativa de un grupo de posibles opciones. Su fórmula matemática está dada por:

$$E(X) = \sum x_i f(x_i)$$

Que al desarrollarse se puede escribir así:

$$E(X) = \mu_x = x_1 f(x_1) + x_2 f(x_2) + x_3 f(x_3) + \dots + x_n f(x_n)$$

Ejemplo 1: Se selecciona una muestra al azar de dos participantes de un comité editorial con el fin de elaborar un anteproyecto, el grupo está compuesto por dos mujeres y cuatro hombres. Encuentre el número esperado de hombres en la muestra.

X: Número de hombres en la muestra.

Número de Hombres (X)	P(X=x _i)= f(x _i)	X*f(X)
0	(2/6)*(1/5)=0.0667	0
1	(4/6)*(2/5)*2=0.5333	0.5333
2	(4/6)*(3/5)=0.4000	0.8000
	1.00	1.33

Valor esperado	f.d.p suma = 1.0
----------------	------------------

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

El análisis para 1 hombre se puede expresar así: La probabilidad de que la primera persona seleccionada sea un hombre es $4/6$, pero entonces, la segunda tiene que ser mujer para que se cumpla la condición ($X=1$), por lo tanto la probabilidad para el segundo caso es $2/5$. Pero, puede ser que el hombre se seleccione en la segunda oportunidad, por lo tanto, se tienen dos opciones de cumplir el requisito, de ahí, que se multiplique por 2.

Para calcular el valor esperado, se tiene que verificar primero si la función de probabilidad es una f.d.p. Tal como sucede en éste caso.

Se espera que al seleccionar 2 personas de 6, una sea hombre, esto esta influenciado por el tamaño de muestra pequeño, si se incrementa el tamaño de muestra, será cada vez más probable seleccionar más hombres que mujeres, si se mantiene la misma proporción de personas en la población

Ejemplo 2: Un estudio de factibilidad requiere identificar cual de tres alternativas es la más probable, teniendo en cuenta 3 posibles escenarios. La siguiente tabla de rendimientos muestra las utilidades estimadas de cada opción, según el escenario económico.

<i>Situación</i>	<i>Alternativa</i>			<i>Probabilidad</i>
	<i>A</i>	<i>B</i>	<i>C</i>	
Recesión	500	-2000	-7000	0.3
Sin Cambios	1000	2000	-1000	0.5
Expansión	2000	5000	20000	0.2

Determine la mejor elección para el inversionista.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

A	Prob.	A*Prob.	B	B*Prob.	C	C*Prob.
500	0.3	150	-2000	-600	-7000	-2100
1000	0.5	500	2000	1000	-1000	-500
2000	0.2	400	5000	1000	20000	4000
		1050		1400		1400

Teniendo en cuenta la fórmula para calcular el valor esperado, se calculó para las diferentes alternativas, encontrándose que para las opciones B y C, el valor esperado es igual, por lo tanto, se tendrán que tener en cuenta otros aspectos, como son: La política del inversionista, el cual puede preferir correr más riesgos, en tal caso, preferiría la opción C, que pierde con dos opciones pero gana más con una opción.

PROPIEDADES DEL VALOR ESPERADO.

Cumple las mismas propiedades definidas para el promedio aritmético definido anteriormente, es decir:

$E(c) = c$ la esperanza de una constante es la misma constante

$E(X \pm c) = E(X) \pm c$ La esperanza de una variable más o menos una constante, es igual a la esperanza de la variable más o menos la constante.

$E(cX) = c E(X)$ La esperanza de una constante por una variable es igual a la constante por la esperanza de la variable. Propiedad de homogeneidad.

$E(X + Y) = E(X) + E(Y)$ la esperanza de una suma es la suma de las esperanzas. Propiedad aditiva.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.1.5 DESVIACIÓN ESTÁNDAR.

Cumple la misma función de la desviación típica calculada en la parte descriptiva, es decir, mide la variabilidad de los valores con respecto al valor esperado. No obstante, se requiere calcular primero la varianza, cuya fórmula es:

$$\text{Var}(X) = s^2_x = E(x^2) - (E(X))^2 ; \text{ donde } E(X^2) = \sum X^2 f(X)$$

Con el siguiente ejemplo se repasa los últimos temas tratados, mostrando así, que la clave está en construir la función de probabilidad, es decir, la tabla de probabilidades.

Ejemplo: Tres pacientes reciben inyecciones de desensibilización contra la picadura de insectos. Se calcula que este suero tiene una eficacia del 95%. Sea X el número de pacientes desensibilizados. Encuentre el número esperado de X y su desviación estándar.

X: Número de pacientes desensibilizados.

Número de Pacientes(X)	$P(X=x_i) = f(x_i)$	$X \cdot f(X)$	$X^2 \cdot f(X)$
0	$0.05 \cdot 0.05 \cdot 0.05 = 0.0001$	0	0
1	$(0.95 \cdot 0.05 \cdot 0.05) \cdot 3 = 0.0048$	0.0048	0.0048
2	$(0.95 \cdot 0.95 \cdot 0.05) \cdot 3 = 0.1354$	0.2708	0.5416
3	$0.95 \cdot 0.95 \cdot 0.95 = 0.8574$	2.5721	7.7166
	1.00	2.85	8.2630

f.d.p

Valor esperado

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$\text{Var}(X) = s_x^2 = S_X^2 f(X) - (E(X))^2 = 8.2630 - (2.85)^2 = 0.1405$$

$$\text{Desviación} = \sqrt{0.1405} = 0.3748.$$

Se espera que prácticamente todos los pacientes queden desensibilizados, no obstante, con la desviación de 0.3748, se puede pensar que el número de desensibilizados esté entre 2 y 3 personas.

Ejercicios

1. Se seleccionan cuatro profesionales del área de la salud con el fin de asistir a una capacitación en investigación cualitativa, la administración puede seleccionar de seis médicos y tres enfermeras profesionales. Si X es una v.a que denota el número total de enfermeras.

- Encuentre la f.d.p
- ¿Cuál es la probabilidad de que dos sean enfermeras?.
- ¿Cuál es la probabilidad de que mínimo dos sean enfermeras?.
- Encuentre el número esperado de enfermeras en la delegación.
- Calcule la desviación (s_x) con respecto al valor esperado, Interprete.

2. Se desarrolla un medicamento para aliviar el estrés. El laboratorio afirma que es eficaz el 90% de los casos. Se prueba en 3 pacientes. Sea X el número de pacientes que obtienen alivio. a) Diga si la función es f.d.p, si la afirmación del fabricante es correcta. b) Encontrar P ($X < 2$). Interprete en términos del problema. c) Encuentre el número esperado de pacientes que sienten alivio. d) Calcule la desviación estándar e interprete.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

3. Un brote está extendiéndose en niños de edad preescolar. El 20% de ellos está infectado. Un pediatra atiende a 4 niños de esta edad durante la primera hora en su día de trabajo. Sea X el número de los que tienen dicho brote. Suponga independencia para calcular la f.d.p. Utilícela para calcular la probabilidad de que ninguno de los cuatro niños tenga el brote, y la de que sólo uno lo tenga. Calcule el número esperado de niños con brote.
4. En un laboratorio se hicieron 10 pruebas en pacientes para detectar el VIH, en una de ellas el resultado dio positivo. Por un error involuntario los resultados se confundieron. Se seleccionan al azar tres resultados, encuentre la función de probabilidad si X , siendo ésta la v.a del número de pruebas positivas, representela gráficamente y encuentre el número esperado.
5. Se seleccionan cuatro profesionales del área de la salud con el fin de asistir a una capacitación en registros médicos, la administración puede seleccionar de seis médicos y tres enfermeras profesionales. Si X es una v.a que denota el número total de médicos.
- Encuentre la f.d.p
 - ¿Cuál es la probabilidad de que dos sean médicos?.
 - ¿Cuál es la probabilidad de que mínimo dos sean médicos?.
 - Encuentre el número esperado de médicos en la delegación.
 - Calcule la desviación con respecto al valor esperado. S_x
6. Por invertir en unas acciones una persona puede obtener ganancias de \$ 2.000.000 con probabilidad de 0.4 o una pérdida de \$500.000 con el resto

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

de probabilidad. Cuál es la ganancia esperada de la persona?

6.2 MODELOS PROBABILISTICOS.

Son modelos matemáticos apropiados para situaciones reales en condiciones específicas, son importantes por que nos ayudan a predecir la conducta de futuras repeticiones de un experimento aleatorio. Los modelos pueden ser discretos o continuos.

Los modelos o distribuciones discretas más comunes son: La Uniforme, Binomial, Poisson y la Hipergeométrica Dado que el enfoque del texto es presentar los modelos más usados en investigación, y más específicamente en áreas sociales y humanísticas, acá se abordarán los temas de la Binomial, la cual es base para definir los tamaños muestrales y la Poisson, de gran utilidad en teoría de colas o fenómenos de espera. Además, la Hipergeométrica en muchos casos (n grande) se aproxima con el modelo Binomial.

En cuanto a las continuas, se utilizan fundamentalmente las siguientes: Z de la Normal, T de Student, F de Snedecor y la Ji cuadrado (X^2), las cuales serán objeto de estudio.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.2.1 MODELO BINOMIAL

Es un modelo discreto, es decir, la variable toma valores conocidos y finitos. Es utilizado en gráficos de control para análisis de número o porcentaje defectuoso, adicionalmente se utiliza para calcular probabilidades de aceptación o rechazo de lotes en muestreo de aceptación, de ahí su importancia.

Sintaxis en Excel:

`DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)`

Núm_éxito: X

Ensayos: n

Prob_éxito: p

Acumulado. Si el argumento es VERDADERO, devuelve la función acumulada; si es FALSO, devuelve la densidad de probabilidad.

Se utiliza además, para calcular probabilidades cuando los posibles resultados sólo pueden ser 2 sucesos excluyentes.

Un modelo Binomial cumple las siguientes propiedades:

- 1) El experimento tiene un número fijo de pruebas.
- 2) De cada prueba solo puede presentarse uno de dos resultados. Un éxito o un fracaso.
- 3) La probabilidad de éxito y la probabilidad de fracaso son constante para cada prueba.
- 4) Cada una de las pruebas es independiente de todas las demás, es decir, lo que ocurre en una prueba cualquiera, no afecta los resultados de las otras pruebas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Si una variable aleatoria definida previamente cumple las condiciones anteriores, se dice que ella se distribuye binomialmente con parámetros n , p y se escribe: $X \sim b(x, n, p)$.

La función de probabilidad de densidad y de distribución acumulada de una variable aleatoria $X \sim b(x, n, p)$ esta dada por:

$$fdp = f(X) = P(X = x) = C_x^n p^x (1 - p)^{n-x}, x = 0, 1, 2, \dots$$

$$FDA = F(X) = P(X \leq x) = \sum_{K=0}^x C_x^n p^x (1 - p)^{n-x}, x = 0, 1, 2, \dots, K$$

Donde

$$C_x^n = \frac{n!}{x!(n-x)!}$$

$n! = 2 * 3 * 4 * \dots * (n - 1) * n$, por ejemplo $5! = 2 * 3 * 4 * 5$

$0! = 1$

p = probabilidad de éxito y es igual a: x/n .

$q = 1 - p$. Es decir, es el complemento de la probabilidad de éxito.

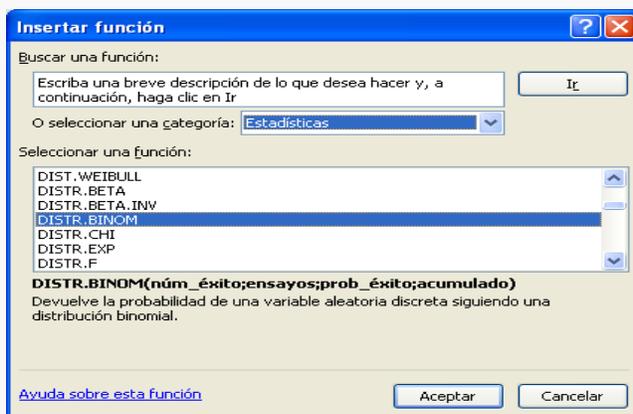
Los parámetros relevantes de la Binomial son: $m_x = E[x] = np$ y su **varianza** $V[x] = s_x^2 = np(1 - p)$.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

Si bien prácticamente todos los libros poseen los valores tabulados para valores de n pequeños y de p comunes, ahora es posible calcular estas probabilidades con la ayuda del Excel.



Sólo se requiere dar Click en el botón **fx** para luego seleccionar funciones estadísticas, tal como se aprecia en el cuadro de la izquierda.

Ejemplo 1: Según el último estudio sobre favorabilidad del actual gobernante, éste tiene a su favor el 52% de todos los habitantes mayores de edad. Si de esta población se selecciona al azar 5 personas, a) ¿cuál es la probabilidad de que exactamente 3 de personas estén a favor del mandatario?, b) máximo 3, c) entre 2 y 4 inclusive estén a favor, d) cuál es el valor esperado y su desviación.

Solución manual.

Sea X = número de personas a favor del mandatario.

Lo primero a determinar es si cumple los requerimientos del modelo Binomial.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- La prueba se realiza 5 veces.
- Lo que diga una persona es independiente de la opinión de otra.
- Sólo se puede presentar un solo resultado, a favor o no.

Al cumplir los requerimientos, entonces, $X \sim b(x, 5, 0.52)$

Donde $p = \text{éxito} = 0.52$

$q = \text{fracaso} = 0.48$

Reemplacemos los valores en la fórmula.

$$a) P(x=3) = 10 \cdot (0.52)^3 \cdot (0.48)^2 = 0.3239$$

Existe una probabilidad moderada de que tres personas estén a favor del gobernante.

$$b) P(X \leq x) = P(x \leq 3) = F(3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$P(X \leq 3) = 0.7865$$

Es muy probable que tres o menos estén a favor del gobernante.

$$c) P(2 \leq X \leq 4) = F(4) - F(1) = P(X \leq 4) - P(X \leq 1) = 0.9619 - 0.1634 = 0.7985$$

$$\text{ó } P(X=2) + P(X=3) + P(X=4) = 0.2990 + 0.3239 + 0.1754 = 0.7983$$

Estadística Descriptiva y Distribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Existe una alta probabilidad de que entre 2 y 4 personas inclusive, estén a favor del gobernante.

$$d) E(X) = np = 5 \cdot 0.52 = 2.6 \text{ personas.}$$

$$V[x] = s_x^2 = npq = 5 \cdot 0.52 \cdot 0.48 = 1.248. \quad s_x = 1.1171$$

De 5 personas se espera que aproximadamente 3 estén a favor del gobernante, con una desviación de 1.11, es decir, realmente, las personas a favor estarán entre 2 y 4.

Solución Excel.

El programa permite calcular tanto la probabilidad en un punto como la acumulada. A continuación se muestra la sintaxis para resolver el problema anterior.

=DISTR.BINOM(3;5;0.52;falso). Indica que se quiere determinar la probabilidad de que 3 personas estén a favor de 5 encuestados, donde la probabilidad de éxito es de 0.52 y el **falso**, quiere decir que no es acumulada. $P(x = 3) = 0.3239$.

=DISTR.BINOM(3;5;0.52;verdadero). Indica que se quiere determinar la probabilidad de que máximo 3 personas estén a favor de 5 encuestados, donde la probabilidad de éxito es de 0.52 y el **verdadero**, quiere decir que la probabilidad es acumulada. $P(x \leq 3) = 0.7865$.

$$c) P(2 \leq X \leq 4) = F(4) - F(1) = (\text{=DISTR.BINOM}(4;5;0.52;1) - \text{=DISTR.BINOM}(1;5;0.52;1)).$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$0.9619 - 0.1634 = 0.7985$$

Para resolver las probabilidades con Excel, se da la siguiente información:

Argumentos de función

DISTR.BINOM

Núm_éxito 3 = 3

Ensayos 5 = 5

Prob_éxito ,52 = 0,52

Acumulado 1 = VERDADERO

= 0,786500813

Devuelve la probabilidad de una variable aleatoria discreta siguiendo una distribución binomial.

Acumulado es un valor lógico: para usar la función de distribución acumulativa = VERDADERO; para usar la función de probabilidad bruta = FALSO.

Resultado de la fórmula = 0,786500813

[Ayuda sobre esta función](#)

Para solicitar éste pantallazo, se da clic en la opción de funciones (fx), ubicada en la barra de formato. Igual que se puede hacer para las demás funciones estadísticas analizadas hasta ahora y las que vienen.

Ejemplo 2: Se estima que en una cirugía al hígado el 10% de los pacientes muere. Si la operación se realiza a 3 personas, cual es la probabilidad de que:

- Exactamente 2 mueran.
- No más de 2 se mueran.
- Al menos 1 pero no más de 2 se mueran.

El objetivo fundamental de éste ejemplo, es mostrar como funciona la tabla diseñada para calcular las probabilidades Binomiales, en la dirección de Internet http://personal5.iddeo.es/ztt/prap1_binomial000.htm se encuentra la tabla usando la f.d.p. es decir, muestra las probabilidades puntuales. El pantallazo se presenta luego del ejemplo.

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$P(X=2) = 0.0270$$

$$P(X \leq 2) = 0.7290 + 0.1354 + 0.0071 = 0.8715$$

$$P(X=1) + P(X=2) = 0.2430 + 0.0270 = 0.27$$

Página nueva 1 - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Historial Correo Imprimir Modificar Discutir

Dirección http://personal5.iddeo.es/ztt/prg/p1_binomial000.htm Ir a Vínculos >>

n k p 0,01 0,05 0,10 0,15 0,20 0,25 0,30 0,33 0,35 0,40 0,45 0,4

TABLA DE DISTRIBUCIÓN BINOMIAL

Probabilidad de obtener k-éxitos

$$p(X = k) = \binom{n}{k} \cdot p^k q^{n-k}$$

n	k	p	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,33	0,35	0,40	0,45	0,4
2	0		0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4449	0,4225	0,3600	0,3025	0,2600
	1		0,0198	0,0950	0,1800	0,2550	0,3200	0,3750	0,4200	0,4442	0,4550	0,4800	0,4950	0,4950
	2		0,0001	0,0025	0,0100	0,0225	0,0400	0,0625	0,0900	0,1109	0,1225	0,1600	0,2025	0,2400
3	0		0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2967	0,2746	0,2160	0,1664	0,1320
	1		0,0294	0,1354	0,2430	0,3251	0,3840	0,4219	0,4410	0,4444	0,4436	0,4320	0,4084	0,3820
	2		0,0003	0,0071	0,0270	0,0574	0,0960	0,1406	0,1890	0,2219	0,2389	0,2880	0,3341	0,3670
	3		0,0000	0,0001	0,0010	0,0034	0,0080	0,0156	0,0270	0,0369	0,0429	0,0640	0,0911	0,1170

Internet

Inicio Explorando - Texto Pagina nueva 1 - Mic... 05:30 p.m.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.2.2. Modelo Poisson.

Es una distribución discreta, tiene su principal aplicación en teoría de colas o fenómenos de espera y en control estadístico de calidad, así, como para calcular probabilidades en eventos escasos y en intervalos de tiempo y espacio. Ejemplo, cuál es la probabilidad de que lleguen 5 llamadas en un minuto a un conmutador?, cuál de que se encuentren 4 huecos en un kilometro?

Sintaxis en Excel:

POISSON(x;media;acumulado).

X es el número de sucesos. Media es el valor numérico esperado, usualmente se conoce como **lambda λ** .

Acumulado. Si el argumento es VERDADERO, se devuelve la probabilidad de que un suceso aleatorio ocurra un número de veces comprendido entre 0 y x inclusive (F.D.A); si el argumento es FALSO, se devuelve la probabilidad de que un suceso ocurra exactamente x veces (f.d.p)

Se deben cumplir los siguientes supuestos, el número de llegadas de los sucesos en el intervalo objeto de estudio, ocurren de manera aleatoria y por ende son independientes entre sí, el modelo está definido completamente por su promedio (λ).

Por último, se debe resaltar, que el promedio, λ , es proporcional al intervalo

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

que representa, es decir, si $\lambda = 2$ muertos/semana, entonces, $\lambda = 8$ muertos/mes.

Si una variable aleatoria definida previamente cumple las condiciones anteriores, se dice que ella se distribuye Poisson con parámetro λ y se denota así:

$$X \sim P(x, \lambda)$$

La función de probabilidad de una variable aleatoria $X \sim P(x, \lambda)$ esta dada por:

$$f(x) = P(x = x_i) = \frac{e^{-\lambda} \lambda^x}{x!}$$

para $x = 0, 1, 2, \dots$

Donde:

$e = 2.72782$ base de los logaritmos neperianos.

Los parámetros relevantes de la Poisson son: $\lambda_x = E[x]$ y su varianza $V[x] = s_x^2 = \lambda_x$

La función de distribución acumulada esta dada por:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$P(X \leq x) = F(x) = FDA = F(X) = \sum_{i=0}^x P(X = x_i) = \sum_{i=0}^x \frac{e^{-\lambda} \lambda^x}{x!}$$

Por último, los parámetros relevantes de la Poisson son: $\mu_x = E[x] = \lambda$ y su varianza $V[x] = s_x^2 = \lambda$, luego la desviación es $= \sqrt{\lambda}$.

No sobra recordar que λ es el promedio, luego se calcula como la sumatoria de los valores dividido por el número de ellos.

Ejemplo 1: El número promedio de individuos que llegan a solicitar información sobre un nuevo producto es de 5 por hora. Determine la probabilidad de que, en una hora determinada, lleguen: a) exactamente 2 usuarios, b) a lo sumo 4, c) más de 4, d) 2 en media hora.

Solución manual.

Sea X = número de personas que solicitan información sobre el producto.

Lo primero a determinar es si cumple los requerimientos del modelo Poisson.

- Las personas llegan de manera aleatoria y sus llegadas son independientes.
- Los sucesos ocurren en un intervalo determinado.

Al cumplir los requerimientos, entonces, $X \sim P(x, \lambda)$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Donde $\lambda = 5$ personas/hora

Reemplacemos los valores en la fórmula.

$$a) P(X = 2) = \frac{e^{-5} 5^2}{2!} = 0.08422$$

Poco probable que pase.

$$b) P(X \leq 4) = F(4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 0.4404$$

c) $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.4404 = 0.5596$. Es más probable que lleguen 5 o más que lo hagan 4 o menos.

d) Dado que se cambio la unidad de medida, el promedio se altera.

$\lambda = 2.5$ personas/30 minutos.

$$P(X = 2) = \frac{e^{-2.5} 2.5^2}{2!} = 0.2651$$

Existe aproximadamente $\frac{1}{4}$ de posibilidad de que lleguen dos usuarios en media hora a solicitar información sobre el producto.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Solución Excel.

El programa permite calcular tanto la probabilidad en un punto como la acumulada. A continuación se muestra la sintaxis para resolver el problema anterior.

=POISSON(2;5;0). Indica que se quiere determinar la probabilidad de que 2 personas soliciten información sobre el producto, donde el promedio es de 5 y 0, quiere decir que no es acumulada. El resultado es 0.08422.

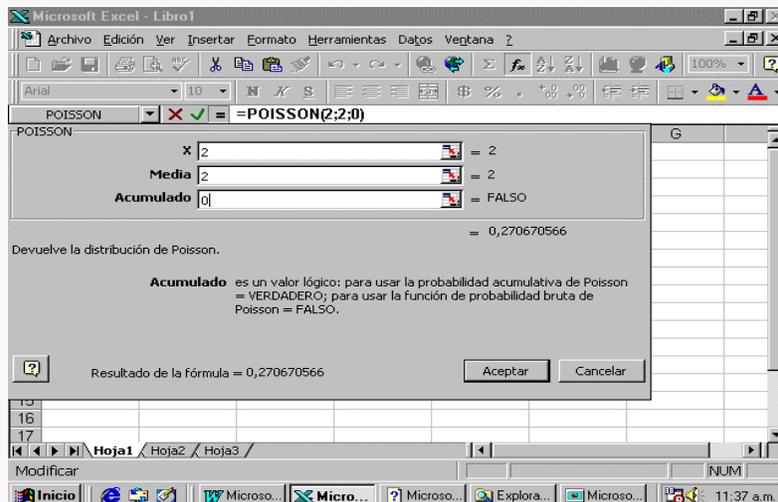
= POISSON(4;5;1). Indica que se quiere determinar la probabilidad de que máximo 4 personas pidan información en un intervalo de una hora, el promedio es 5 y 1, quiere decir que la probabilidad es acumulada. El resultado es 0.4404

1 - =POISSON(4;5;1). Indica la probabilidad de que más de 4 personas pregunten por el producto en un intervalo de una hora, el promedio es de 5. El resultado es $1 - 0.4404 = 0.5596$.

=POISSON(2;2.5;0). Calcula la probabilidad de que 2 personas soliciten información en un intervalo de **media hora**, el promedio cambia a **2.5** personas/30 minutos. El resultado es 0.2651.

Si bien se puede digitar la sintaxis de la orden, también se puede llenar el siguiente cuadro que se activa al solicitar las funciones del programa.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.



Ejemplo 2: El número promedio de errores que se detectan en las páginas de los periódicos locales es de 3 por página, determine la probabilidad de que en la página 8 del periódico matutino encuentre:

a. Exactamente 3 errores, b) a lo sumo 4, c) más de 4 en la página 10, d) 2 en dos páginas.

Solución manual.

Sea X = número de errores por página

Requerimientos del modelo Poisson.

- Los errores ocurren de manera aleatoria.
- Los sucesos ocurren en un intervalo de espacio – página -

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$X \sim P(x, \lambda)$. Donde $\lambda = 3$ errores /página.

Reemplacemos los valores en la fórmula.

a) $P(X = 3) = \frac{e^{-3} 3^3}{3!} = 0.2240$ Existe una moderada posibilidad de que se encuentren 3 errores en una página.

b) $P(X \leq 4) = F(4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 0.8152$

c) $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.8152 = 0.1848$. Es poco probable que se presenten más de 5 errores.

d) Dado que se cambio la unidad de medida, el promedio se altera.

$\lambda = 6$ errores /2 páginas.

$$P(X = 2) = \frac{e^{-6} 6^2}{2!} = 0.1338$$

Existe una moderada posibilidad de encontrar exactamente 2 errores en 2 páginas.

Solución Excel.

```
=POISSON(3;3;0) = 0.2240
```

```
= POISSON(4;3;1) = 0.8152
```

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$1 - =\text{POISSON}(4;3;1) = 1 - 0.8152 = 0.1848$$

$$=\text{POISSON}(2;6;0) = 0.1338$$

Solución Tablas.

Valores de λ

X	1	1.5	2	2.5	3	3.5	4	4.5	5
0	0.36789	0.22314	0.13534	0.08209	0.04979	0.03020	0.01832	0.01111	0.00674
1	0.36789	0.33471	0.27069	0.20523	0.14937	0.10570	0.07327	0.05000	0.03369
2	0.18395	0.25103	0.27069	0.25653	0.22406	0.18498	0.14654	0.11249	0.08424
3	0.06132	0.12552	0.18046	0.21378	0.22406	0.21581	0.19539	0.16874	0.14040
4	0.01533	0.04707	0.09023	0.13361	0.16805	0.18883	0.19539	0.18983	0.17549
5	0.00307	0.01412	0.03609	0.06681	0.10083	0.13218	0.15631	0.17085	0.17549
6	0.00051	0.00353	0.01203	0.02784	0.05041	0.07711	0.10421	0.12814	0.14624
7	0.00007	0.00076	0.00344	0.00994	0.02161	0.03855	0.05955	0.08237	0.10446
8	0.00001	0.00014	0.00086	0.00311	0.00810	0.01687	0.02977	0.04634	0.06529
9	0.00000	0.00002	0.00019	0.00086	0.00270	0.00656	0.01323	0.02317	0.03627
10	0.00000	0.00000	0.00004	0.00022	0.00081	0.00230	0.00529	0.01043	0.01814
11	0.00000	0.00000	0.00001	0.00005	0.00022	0.00073	0.00192	0.00426	0.00824
12	0.00000	0.00000	0.00000	0.00001	0.00006	0.00021	0.00064	0.00160	0.00343
13	0.00000	0.00000	0.00000	0.00000	0.00001	0.00006	0.00020	0.00055	0.00132
14	0.00000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00006	0.00018	0.00047
15	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00002	0.00005	0.00016
16	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00002	0.00005

La tabla anterior, se construyó utilizando la hoja electrónica Excel, y se optó por la tabla no acumulada, sin embargo, la construcción de la tabla acumulada se hace de manera similar.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

a) $P(x=3) = 0.2240$. El número se ubica en la Intersección de $\lambda = 3$ y $X=3$

b) $P(X \leq 4) = F(4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$$= 0.04979 + 0.14937 + 0.22406 + 0.22406 + 0.16805 = 0.8152$$

c) $P(X > 4) = 1 - P(X \leq 4) = 1 - 0.8152 = 0.1848$.

El manejo de tabla es similar para cada valor de X , valga decir, si cambia el promedio, la probabilidad se encuentra en el cruce entre l y el valor de X .

Aproximación del modelo Binomial al modelo Poisson: Como ya observó, con ayuda de software con aplicaciones estadísticas es innecesario realizar aproximaciones, ya que el objetivo de éstas, es disminuir la parte operativa del proceso, situación que en la actualidad no es necesario realizarse.

Una persona que según sus hábitos tiene una probabilidad de infectarse de 0.01 por día. En muestras de 100 sujetos cada una. Cuál es la probabilidad de que al finalizar el primer día no haya ningún infectado?

El modelo es Binomial, sin embargo, no se tienen valores tabulados para una muestra de 100, además, por fórmula, se presume que es tedioso realizar el cálculo, por esa razón, se ha usado la aproximación al modelo Poisson. No obstante, con la ayuda del Excel, se procede así, sin necesidad de ajustes.

$$= \text{DISTR.BINOM}(0;100;0.01;0) = 0.3660$$

ESTADISTICA COMO APOYO A LA INVESTIGACIÓN.

6.2.3 MODELO NORMAL.

Sin duda, es la más usada de las distribuciones estadísticas, dado que tiene múltiples aplicaciones en todas las áreas del conocimiento. La sintaxis en Excel es:

`DISTR.NORM(x;media;desv_estándar;acum)`. *X* es el valor bajo el cual se encontrará el área (probabilidad) deseada, *Media* es la media aritmética de la distribución, *Desv_estándar* es la desviación estándar de la distribución y *Acum* es un valor lógico que determina la forma de la función. Si el argumento *acum* es VERDADERO, la función `DISTR.NORM` devuelve la función de distribución acumulada.

Si el argumento *media* = 0 y *desv_estándar* = 1, la función `DISTR.NORM` devuelve la distribución normal estándar, `DISTR.NORM.ESTAND`.

Es el modelo continuo más utilizado en inferencia estadística, dado que muchos fenómenos socio demográfico y de otra índole tienen un comportamiento acampanado y por ende cumplen la teoría de la distribución normal, sin embargo, es importante aclarar que antes de proceder a aplicar los métodos sugeridos por la teoría estadística, es imprescindible identificar primero si en realidad los datos si se comportan como tal, es decir, se debe saber a ciencia cierta si los datos son aproximadamente acampanados, para ello se disponen de procedimientos descriptivos ya descritos, como son: el histograma, comparar las medidas de posición relevantes (media, mediana y moda), calcular los coeficientes de asimetría y curtosis. Además, de otras pruebas más avanzadas como la de Smirnov Kolmogorov y la de Shapiro Wilks. Por lo ya expresado, es importante observar el OA sobre éste importante tema: www.leondariobello.com/OA/normal

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Si X se distribuye normal con parámetros μ y s^2 , entonces se denota así,

$$X \sim N(\mu, s^2)$$

CARACTERÍSTICAS:

- Es simétrica, además, de ser asintótica con respecto al eje X , es decir, no toca el eje.
- $Mo \equiv Me \equiv Media \equiv \mu$
- Es unimodal
- Si tomamos intervalos centrados en μ , se cumple:
- $\mu \pm \sigma$ se tiene aproximadamente el 68% de las observaciones.
- $\mu \pm 2\sigma$ se tiene aproximadamente el 95% de las observaciones.
- $\mu \pm 3\sigma$ se tiene aproximadamente el 99.73% de las observaciones.
- Para calcular cualquier otro porcentaje, se utilizan las tablas de la distribución normal y/o las funciones del programa Excel.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

TEOREMA DE TIPIFICACION O ESTANDARIZACION.

Si $X \sim N(\mu, s^2)$, entonces la variable aleatoria $Z = \frac{x - \mu}{\sigma} \sim N(0, 1)$

Esto se hace con el fin de uniformar las tablas y poder calcular todas las probabilidades requeridas, ya que si no se hace este cambio de escala, se tendría que calcular la fórmula de la distribución de densidad que implica calcular integrales y por tanto, se volvería inmanejable para un gran número de investigadores que no tienen formación para ello. En la actualidad, se tiene la ayuda de software para su cálculo.

En el **Excel**[\[1\]](#) la función devuelve la probabilidad de una variable aleatoria continua siguiendo una distribución acumulativa normal para la media y desviación estándar especificadas. Esta función tiene un gran número de aplicaciones en estadística, incluyendo las pruebas de hipótesis.

Ejemplo 1: El tiempo para realizar una intervención quirúrgica tiene un comportamiento aproximadamente normal, con media = 140 minutos y desviación de 50 minutos. Calcule la probabilidad de que una intervención quirúrgica se demore: a) A lo sumo 100 minutos, b) Mínimo 200 minutos, c) Entre 110 y 190 y d) Mayor que 40 y menor de 240 y e) ¿Cuál es el tiempo requerido para que el 50% de las intervenciones terminen?.

Solución manual:

Sea X = tiempo que se demora una intervención quirúrgica

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$X \sim N(\mu, \sigma^2); X \sim N(140, (50)^2)$$

Si no se realiza el cambio de variable $Z = \frac{x - \mu}{\sigma}$ se tendría que calcular

la integral, situación poco práctica.

Tipifiquemos:

$$Z \sim N(0, 1)$$

$$a) P(x < 100) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{100 - 140}{50}\right) = 0.2119 \text{ Valor de la tabla normal.}$$

Es relativamente probable que la cirugía se demore menos de 100 minutos.

$$b) P(x \geq 200) = 1 - P(x < 200).$$

$$1 - P(x \leq 200) = 1 - P\left(Z \leq \frac{200 - 140}{50}\right) = 1 - 0.8849 = 0.1151$$

Es poco probable que una intervención se demore más de 200 minutos, ya que la probabilidad es de 0.1151, aproximándose al cero.

$$c) P(110 < x < 190) = ?$$

$$P(x \leq 190) - P(x \leq 110) = P\left(Z \leq \frac{190 - 140}{50}\right) - P\left(Z \leq \frac{110 - 140}{50}\right) = P(Z \leq 1) - P(Z \leq -0.6) = 0.8413 - 0.2743$$

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$= 0.567$$

$$d) P(40 < x < 240) = P(-2 < Z < 2)$$

$$= P(Z < 2) - P(Z \leq -2) = 0.9772 - 0.0228 = 0.9544$$

e) Para el caso en que se tiene la probabilidad y el interés es determinar el valor de x , el proceso es inverso, es decir, se busca el valor de Z y se despeja el de x . Para el caso se tiene:

$$P(x \leq x_0) = 0.50$$

$$P(Z < (x_0 - \mu)/s) = P(Z < (x_0 - 140)/50) = 0.50$$

$(x_0 - 140)/50 = 0.00$ probabilidad de la tabla normal. Al despejar x_0 se tiene:

$$x_0 = \mu + p(Z) * s$$

$$x_0 = 140 + 0.00 * 50 = 140.$$

Se requieren de 140 minutos para terminar el 50% de las intervenciones quirúrgicas, con un poco más de experiencia, dicho valor se deduce al observar que el 50% de las observaciones se encuentran por debajo de la media, de ahí, la importancia de la figura para intuir, no sólo el lugar del área a encontrar, sino, la probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Solución Excel:

Sea X = tiempo que se demora una intervención quirúrgica

$$X \sim N(\mu, s^2); \quad X \sim N(140, (50)^2)$$

a) $P(x \leq 100)$.

DISTR.NORM

x	100	=	100
Media	140	=	140
Desv_estándar	50	=	50
Acum	VERDADERO	=	VERDADERO

= 0,211855334

Devuelve la distribución acumulativa normal para la media y desviación estándar especificadas.

x es el valor cuya distribución desea obtener.

Resultado de la fórmula = 0,211855334

Aceptar Cancelar

=DISTR.NORM(100;140;50;VERDADERO)

b) $P(x \geq 200) = 1 - P(x < 200)$.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

DISTR.NORM

X	200	=	200
Media	140	=	140
Desv. estándar	50	=	50
Acum	verdadero	=	VERDADERO

= 0,884930268

Devuelve la distribución acumulativa normal para la media y desviación estándar especificadas.

Acum es un valor lógico: para usar la función distribución acumulativa = VERDADERO; para usar la función de probabilidad bruta = FALSO.

Resultado de la fórmula = 0,884930268

Aceptar Cancelar

$$=DISTR.NORM(200;140;50;VERDADERO) = 0.884930$$

$$P(x \geq 200) = 1 - P(x < 200) = 1 - 0.884930 = 0.1151$$

$$c) P(110 < x < 190) = P(z < 190) - P(z < 110)$$

$$=DISTR.NORM(190;140;50;VERDADERO)$$
$$=DISTR.NORM(110;140;50;VERDADERO)$$

$$0,84134474 - 0,27425306 = 0.567$$

$$d) P(40 < x < 240) = P(z < 240) - P(z < 40)$$

$$=DISTR.NORM(240;140;50;VERDADERO)$$
$$=DISTR.NORM(40;140;50;VERDADERO)$$

$$0,97724994 - 0,02275006 = 0.9544$$

e) $P(x \geq x_0) = 0.50$. En éste caso, se requiere entrar la probabilidad para

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

que el procedimiento entregue el valor de x . Note que la sintaxis cambia, adicionando la palabra INV.

$$=DISTR.NORM.INV(0,5;140;50) = 140$$

Parece evidente que dado la popularidad del programa Excel, es posible que sea ésta la herramienta de más fácil uso para los investigadores cuando requieren calcular las diversas probabilidades, incluso por encima de las tablas que proveen la totalidad de textos de estadística.

Ejemplo 2: Los pesos de una población de adultos mayores indigentes tienen un comportamiento que se asemeja a una campana normal con una media de 132 libras y una varianza de 225. Calcular la probabilidad de que una persona seleccionada al azar de entre esa población, pese: a) Más de 155, b) 100 libras o menos, c) Entre 105 y 143 libras y d) Cuál es el peso de tal forma que sólo el 5% de los adultos mayores sobrepasan dicho valor..

Solución manual:

Sea X = peso de mayores indigentes en libras.

$$X \sim N(m, s^2); \quad X \sim N(132, 225)$$

Como en todo problema práctico, lo primero es tipificar:

$$Z = (x_0 - \mu)/s \sim N(0, 1).$$

$$\mu = 132 ; s = 15.$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$a) P(x > 155) = 1 - P(x < 155) = 1 - P(z < (155 - 132)/15) = 1 - P(Z < 1.53)$$

$$1 - 0.9370 = 0.063.$$

$$b) P(x < 100) = P(x < 100) = P(Z < (100 - 132)/15) = P(Z < -2.13) = 0.0166$$

c) $P(105 < x < 143) = P(x < 143) - P(x < 105)$. Luego de tipificar, se tiene:

$$P(Z < 0.73) - P(Z < -1.8) = 0.7673 - 0.0359 = 0.7324$$

d) $P(X > x_0) = 0.05$ Se pide el valor de tal manera que el 5% de los indigentes mayores pesen más. Esto quiere decir que el 95% pesarán menos, luego la expresión en términos de menor que, (para usar las tablas acumuladas) es:

$P(x < x_0) = 0.95$. Se busca la probabilidad (0.95) dentro de la tabla normal, encontrando el valor de 1.65. El cual se ajusta ya que esta a la misma distancia el valor de $Z = 1.64$.

$$P(Z < (x_0 - \mu)/\sigma) = P(Z < (x_0 - 132)/15) = 0.95$$

$$(x_0 - 132)/15 = 1.645$$

$$x_0 = \mu + p(Z) * \sigma$$

$$x_0 = 132 + 1.645 * 15 = 156.67$$

Solución Excel:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Sea X = peso de mayores indigentes en libras.

$$X \sim N(\mu, s^2); X \sim N(132, 225)$$

$$a) P(x > 155) = 1 - P(x < 155) = 1 - 0.9374 = 0.626$$

$$= \text{DISTR.NORM}(155; 132; 15; \text{VERDADERO})$$

$$b) P(x < 100) = P(x < 100) = 0.01644864$$

$$= \text{DISTR.NORM}(100; 132; 15; \text{VERDADERO}) = 0.01644864$$

$$c) P(105 < x < 143) = P(z < 143) - P(z < 105)$$

$$= \text{DISTR.NORM}(143; 132; 15; \text{VERDADERO})$$

$$= \text{DISTR.NORM}(105; 132; 15; \text{VERDADERO})$$

$$0.7683225 - 0.03593027 = 0.7323923$$

d) $P(x > x_0) = 0.05$. En éste caso, se requiere entrar la probabilidad para que el procedimiento entregue el valor de x .

$$P(x < x_0) = 0.95.$$

$$= \text{DISTR.NORM.INV}(0.95; 132; 15) = 156.67$$

En todos los casos se obtienen resultados similares y/o más precisos que los encontrados usando los valores de tablas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Solución Tablas:

Al ser una variable continua, se pueden calcular infinitos valores de probabilidad, el objetivo de éste texto es ilustrar el manejo para aquellos casos en los que no pueda contar con el apoyo de sistemas electrónicos, por lo tanto, la tabla normal se construyó con Excel y se calcularon probabilidades para valores de Z múltiplos de 0.05, no obstante, para ilustrar el manejo con el ejemplo, se procedió a colocar los valores de Z para el ejemplo planteado, planteando de paso, que al tener la tabla en Excel, puede calcular cualquier valor de Z . No sobra comentar, que la tabla normal es acumulada y además, para valores menores de -3.1 la probabilidad va tendiendo a cero, mientras que para valores mayores de 3.1 la probabilidad tiende a 1.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

TABLA ACUMULADA NORMAL

Z	Prob.	Z	Prob.	Z	Prob.	Z	Prob.	Z	Prob.	Z	Prob.
0.00	0.5000	-1.05	0.1469	-2.10	0.0179	0.00	0.5000	1.05	0.8531	2.10	0.9821
-0.05	0.4801	-1.10	0.1357	-2.13	0.0166	0.01	0.5040	1.10	0.8643	2.15	0.9842
-0.10	0.4602	-1.15	0.1251	-2.20	0.0139	0.10	0.5398	1.15	0.8749	2.20	0.9861
-0.15	0.4404	-1.20	0.1151	-2.25	0.0122	0.15	0.5596	1.20	0.8849	2.25	0.9878
-0.20	0.4207	-1.25	0.1056	-2.30	0.0107	0.20	0.5793	1.25	0.8944	2.30	0.9893
-0.25	0.4013	-1.30	0.0968	-2.35	0.0094	0.25	0.5987	1.30	0.9032	2.35	0.9906
-0.30	0.3821	-1.35	0.0885	-2.40	0.0082	0.30	0.6179	1.35	0.9115	2.40	0.9918
-0.35	0.3632	-1.40	0.0808	-2.45	0.0071	0.35	0.6368	1.40	0.9192	2.45	0.9929
-0.40	0.3446	-1.45	0.0735	-2.50	0.0062	0.40	0.6554	1.45	0.9265	2.50	0.9938
-0.45	0.3264	-1.50	0.0668	-2.55	0.0054	0.45	0.6736	1.50	0.9332	2.55	0.9946
-0.50	0.3085	-1.55	0.0606	-2.60	0.0047	0.50	0.6915	1.53	0.9370	2.60	0.9953
-0.55	0.2912	-1.60	0.0548	-2.65	0.0040	0.55	0.7088	1.60	0.9452	2.65	0.9960
-0.60	0.2743	-1.65	0.0495	-2.70	0.0035	0.60	0.7257	1.65	0.9505	2.70	0.9965
-0.65	0.2578	-1.70	0.0446	-2.75	0.0030	0.65	0.7422	1.70	0.9554	2.75	0.9970
-0.70	0.2420	-1.75	0.0401	-2.80	0.0026	0.70	0.7580	1.75	0.9599	2.80	0.9974
-0.75	0.2266	-1.80	0.0359	-2.85	0.0022	0.73	0.7673	1.80	0.9641	2.85	0.9978
-0.80	0.2119	-1.85	0.0322	-2.90	0.0019	0.80	0.7881	1.85	0.9678	2.90	0.9981
-0.85	0.1977	-1.90	0.0287	-2.95	0.0016	0.85	0.8023	1.90	0.9713	2.95	0.9984
-0.90	0.1841	-1.95	0.0256	-3.00	0.0013	0.90	0.8159	1.95	0.9744	3.00	0.9987
-0.95	0.1711	-2.00	0.0228	-3.05	0.0011	0.95	0.8289	2.00	0.9772	3.05	0.9989
-1.00	0.1587	-2.05	0.0202	-3.10	0.0010	1.00	0.8413	2.05	0.9798	3.10	0.9990

a) $P(x > 155) = 1 - P(x < 155) = 1 - P(z < (155 - 132) / 15) = 1 - P(Z < 1.53)$

$1 - 0.9370 = 0.063.$

b) $P(x < 100) = P(x < 100) = P(Z < (100 - 132) / 15) = P(Z < -2.13) = 0.0166$

Al ser un valor para < la probabilidad se calcula de manera directa, es decir,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

se ubica el valor de $Z=-2.13$ correspondiendo la probabilidad de 0.0166.

c) $P(105 < x < 143) = P(x < 143) - P(x \leq 105)$. Luego de tipificar, se tiene:

$$P(Z < 0.73) - P(Z < -1.8) = 0.7673 - 0.0359 = 0.7324$$

d) $P(X > x_0) = 0.05$ Se pide el valor de tal manera que el 5% de los indigentes mayores pesen más. Esto quiere decir que el 95% pesarán menos, luego la expresión en términos de menor que, (para usar las tablas acumuladas) es:

$P(x < x_0) = 0.95$. Se busca la probabilidad (0.95) dentro de la tabla normal, encontrando el valor de 1.65.

$$P(Z < (x_0 - \mu) / \sigma) = P(Z < (x_0 - 132) / 15) = 0.95$$

$$(x_0 - 132) / 15 = 1.65. \quad x_0 = 156.75 \text{ libras.}$$

Ejercicios

1. Usando Excel calcule las siguientes probabilidades, teniendo en cuenta que el promedio es \$755.000 y la desviación estándar es \$45.000. Asuma comportamiento acampanado en los datos.

a) $P(X > 755.000)$.

b) $P(X < 900.000)$

c) $P(680.350 < X < 800.000)$

d) $P(755.000 < X < 800.000)$

e) $P(X > x_0) = 0.75$

f) $P(X < x_0) = 0.25$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

g) $P(x_0 < X < x_1) = 0.95$

Compruebe los resultados anteriores de manera manual.

2. Los siguientes datos representan las tasas de interés cobradas por un prestamista durante varios meses.

5.3 5.8 5.3 5.5 5.5 5.4 6.0 6.1 5.5 5.4 5.6
5.2 5.4 5.6 5.4 5.3 6.1 4.9 5.5 5.3 5.4 6.0.

Compruebe si los datos son aproximadamente normales (cualquier método) y encuentre las siguientes probabilidades.

a) $P(X > 5.6)$. b) $P(X < 5.6)$. c) $P(5.0 < X < 6.0)$

d) $P(X < x_0) = 0.15$ e) $P(X > x_0) = 0.90$ f) $P(x_0 < X < x_1) = 0.90$

3. Suponga que el 25% de cierta población tiene sangre tipo B. Si se selecciona una muestra de 15 personas calcular la probabilidad de:

- Tres o más personas tengan ese tipo de sangre.
- Menos de dos tengan ese tipo de sangre.
- Entre dos y cuatro inclusive tengan ese tipo de sangre.
- Siquiera cuatro tengan ese tipo de sangre.
- Cual es el número esperado de personas con ese tipo de sangre.

4. Si hay una probabilidad de 0.2 de que un niño cumpla los tiempos mínimos para pertenecer al equipo de natación. Cuál es la probabilidad de

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

que de cuatro niños seleccionados al azar:

- a. No se encuentre ninguno apto.
- b. Exactamente dos estén aptos.
- c. Máximo dos estén aptos.
- d. Mínimo dos estén aptos.
- e. Menos de cuatro pero más de uno.

5. Por experiencia pasada se sabe que el 80% de las personas que se preparan adecuadamente para una actividad física, clasifican para representar al departamento en algún evento. Para un grupo de 20 personas, encuentre la probabilidad de que.

- a. A lo sumo 15 nos representen.
- b. Puedan representarnos siquiera la mitad.
- c. Si el grupo fuera de 120 personas, cuantos se espera que nos representen.

6. El número promedio de llamadas que llegan al PBX de la Universidad es de 4.5 por minuto. Determine la probabilidad de:

- a. En un minuto determinado ocurran 3 ó 4 llamadas.
- b. Mínimo 3 pero máximo 5.
- c. A lo sumo 5.
- d. Recalcule las probabilidades anteriores si el intervalo de tiempo fuera 30 segundos. Analice los cambios y utilice la herramienta más práctica para calcularlas.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

7. Un centro médico tiene un promedio de 12 usuarios todas las mañanas entre las 10:00 y las 11:00. Si las llegadas siguen una distribución Poisson, calcule la probabilidad de que:

- Lleguen máximo 12 usuarios en esa hora.
- Lleguen entre 8 y 12 usuarios inclusive.
- Lleguen más de 10 pero menos de 15.
- Menos de 8 o más de 12.

8. En la facultad de Psicología, las calificaciones de sus 500 estudiantes tienen una distribución normal con media 3.2 y varianza de 0.1444.

- Cuál es la probabilidad de que un estudiante seleccionado al azar pueda habilitar?
- Cuántos estudiantes se espera tenga una calificación de como mínimo 3.2?
- Qué porcentaje de estudiantes no habilitan?
- Si a los estudiantes de más de 4.0 los clasifican como muy buenos, cuantos de ellos están en dicha categoría?
- Qué índice de calificación alcanzará el 88% del cuerpo estudiantil?

9. Se sabe que en una determinada cirugía el 10% de los pacientes muere. Si la operación se realiza a 10 personas, cuál es la probabilidad de que:

- No más de 5 se mueran.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- b. Exactamente 5 mueran.
- c. Al menos 2 pero no más de 6 se mueran.

10. Una máquina de cerveza está regulada para que descargue un promedio de 207 ml. Por vaso. Si la cantidad del líquido está distribuida normalmente con desviación estándar igual a 15 ml?.

- a. Que fracción de los frascos contendrá mas de 231 ml.?
- b. Cuál es la probabilidad de que un vaso contenga entre 198 y 210 mililitros?
- c. Bajo que valor se obtendrá el 20% de los vasos con menor cantidad.

11. El número de días entre la facturación y el pago de las cuentas corrientes de crédito en el Hospital Metropol tiene una distribución aproximadamente normal con una media de 18 días y una varianza de 16. Que proporción de las facturas será pagada.

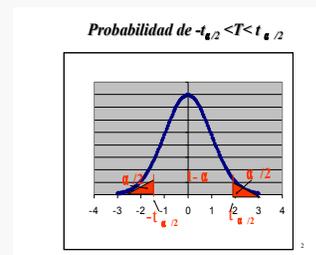
- a. Entre 12 y 18 días?
- b. Dentro de cuantos días estará pagado el 99.5% de las facturas?
- c. Entre cuales dos valores simétricamente distribuidos alrededor de la media recaerán el 98% de las facturas?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.2.4 Distribución T de Student.

Tiene características similares a la distribución normal, su diferencia principal radica en las áreas de los extremos las cuales son más amplias, como consecuencia de que usualmente se trabaja con muestras pequeñas. La sintaxis en Excel es: `DISTR.T(x;grados_de_libertad;colas)`

X es el valor numérico al que se ha de evaluar la distribución. Grados_de_libertad es un entero que indica el número de grados de libertad. Colas especifica el número de colas de la distribución que se ha de devolver. Toma los valores de 1 o 2.



El nombre de la distribución se debe a su autor W.S. Gosset, quien escribió bajo el seudónimo de Student ante la imposibilidad de presentar sus trabajos so pena de perder su empleo, esto sucedió a principio del siglo XX.

Esta distribución es recomendada cuando se requiere estimar la media poblacional y no se conoce la desviación estándar y por lo tanto, hay que estimarla, eso si, siempre y cuando la distribución original sea aproximadamente normal.

Otro término utilizado en esta distribución continua, es el de grados de

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

libertad (g.l), el cual de manera intuitiva se expone así:

$Y = x_1 \pm x_2 \pm x_3 \pm x_4$, para satisfacer la ecuación, tres variables se pueden cambiar a libertad, pero un de ellos no, por eso, cuando se tiene una sola muestra, se hable de $n-1$ g.l. A medida que se aumenten los g.l. la distribución t , se aproxima a la distribución Z de la normal. Otra lectura que se puede dar es que los g.l es una medida del número de observaciones independientes en la muestra, que se usan para estimar la desviación estándar. Cuando el tamaño de muestra no sea muy pequeño y la simetría no sea alta, se puede usar para estimar la media poblacional cuando no se conoce la desviación.

En general, las características relevantes del modelo se sintetizan de la siguiente manera:

- En general cumple con las características de la distribución normal, es decir, es simétrica, unimodal y asintótica.
- El valor crítico de T es mayor que el valor crítico de z .
- Presenta mayor variabilidad en las colas.
- Se utiliza cuando no se conoce la desviación estándar poblacional, independiente del tamaño de muestra, se tipifica de la siguiente manera cuando se trata de promedios.

$$T = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

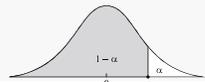
Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

TABLA DE LA DISTRIBUCION t-Student

La tabla da áreas $1 - \alpha$, para valores menores o iguales a t y n.g.l, se construyó con Excel.



N	1 - α							
	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Es importante resaltar que al ser una distribución simétrica al tener información sobre un valor positivo, se obtiene el dato para el mismo valor con signo negativo.

Un hecho de relevancia significativa, es que se utiliza para calcular probabilidades con respecto al promedio, en estos casos, el divisor al estandarizar los valores se divide sobre S/\sqrt{n} , término que se conoce como el error estándar de la media y mide la variabilidad de la media entre muestra y muestra. A mayor tamaño de muestra, menor es el error estándar de la media.

Por último, se puede afirmar, la distribución t es útil para realizar inferencias acerca de la media poblacional cuando no se conoce s y la población es normal, independiente del n , no obstante, aún cuando la distribución sea un tanto sesgada, la t sigue siendo apropiada, esto se conoce como una distribución robusta, es decir, a cambios moderados de los supuestos, el modelo sigue siendo válido. Como en el caso de la distribución normal, ésta distribución también usa valores tabulados, tal como se aprecian en la tabla precedente, teniendo en cuenta, que a medida que los g.l aumenten los valores tienden a ser igual a los encontrados en la tabla Z.

Ejemplo 1: Los valores de las matriculas de estudiantes en una universidad privada tienen un comportamiento aproximadamente normal, donde el promedio es de 2.100.000. Se seleccionan 8 liquidaciones, siendo los valores los siguientes: 1.950.000, 2.100.000, 2.250.000, 1.890.000, 2.250.000, 1.950.000, 2.050.000, 2.350.000. Determine la probabilidad de que:

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

- El promedio sea menor de 2.000.000.
- El promedio se encuentre entre 2.000.000 y 2.200.000
- El promedio sea mayor o igual a 2.500.000

Solución manual:

Sea X = Liquidación matriculas.

$\mu = 2.100.000$; $\sigma = ?$

$\bar{x} = 2.098.750$ $s = 168.644.8085$ $n = 8$

$$T = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

a) $P(\bar{x} < 2.000.000) = P(\bar{x} < 2.000.000)$

$$P(t < (2.000.000 - 2.100.000) / (168644.8085 / 2.8284)) = P(t < -1.677)$$

La probabilidad se encuentra entre 0.9 y 0.95, según la tabla T que se encuentra más adelante, no obstante, al t ser negativo, la probabilidad está entre 0.1 y 0.05, es decir, los valores complementarios.

Para buscar en la tabla, se tiene en cuenta la fila con 7 g.l y se ubica el 1.677, el cual se encuentra entre los valores mencionados. De ahí que sea importante utilizar el Excel. que nos permite calcular la probabilidad exacta.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

$$b) P(2.000.000 < \bar{x} < 2.200.000) = P(\bar{x} < 2.200.000) - P(\bar{x} < 2.000.000).$$

Luego de tipificar, se tiene:

$$P(t < 3.35) - P(t < -1.677) = 0.995 - 0.075 = 0.92$$

Existe una alta probabilidad de que el promedio de las matriculas se encuentre entre 2.000.000 y 2.200.000.

$$c) P(\bar{x} > 2.500.000) = P(t > 6.70) = 1 - P(t < 6.70) = 1 - 1 = 0$$

Dado que el valor de 6.70 es mucho mayor que el ubicado en la tabla de 3.49 y corresponde a 0.995, es claro, entonces, que para valores mayores de 3.49, la probabilidad será de 1.

Por lo tanto, la probabilidad de que el promedio de matricula sea superiora a 2.500.000 es cero.

Solución Excel:

$$a) P(\bar{x} < 2.000.000) = P(\bar{x} < 2.000.000)$$

$$P(t < (2.000.000 - 2.100.000) / (168644.8085 / 2.8284)) = P(t < -1.677)$$

$$= \text{DISTR.T}(1.677; 7; 1) = 0.0687$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Como se dijo utilizando la tabla, la probabilidad está entre 0.1 y 0.05. La probabilidad exacta es de 0.0687. Es decir, la probabilidad de que el promedio de matrícula que pagan los estudiantes sea menor de 2.000.000 es baja.

$$b) P(2.000.000 < \bar{x} < 2.200.000) = P(\bar{x} < 2.200.000) - P(\bar{x} < 2.000.000).$$

Luego de tipificar, se tiene:

$$P(t < 3.35) - P(t < -1.677) = 0.995 - 0.075 = 0.92$$

$$(1 - =\text{DISTR.T}(3.35;7;1) - =\text{DISTR.T}(1.677;7;1) = ?$$

$$(1 - 0.006125) - 0.06872 = 0.9251$$

Los resultados son similares a los ya presentados. Por la forma de calcular el Excel las probabilidades, se resta a uno la probabilidad de 3.35, es decir, el programa calcula la cola de la derecha.

$$c) P(\bar{x} > 2.500.000) = P(t > 6.70) = =\text{DISTR.T}(6.7;7;1) = 0.00013$$

Se observa fácil, que el Excel permite calcular las probabilidades de manera más exactas que las usadas comúnmente (tablas). Esto es importante tenerlo en cuenta, ya que cuando se tienen poblaciones muy grandes, esas pequeñas diferencias se convierten en significativas.

Ejemplo 2: Los puntajes de un grupo de estudiantes se comportan normal,

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

con promedio de 50, sin embargo, no se conoce la desviación. Se tomó una m.a de 9 estudiantes encontrando una varianza de 36 y un promedio de 52. Cuál es la probabilidad de que el promedio:

- Sea mayor de 54?
- Sea menor que 54?
- Esté comprendido entre 48 y 52 puntos?

Solución manual:

Sea X = Puntaje estudiantes.

$\mu = 50$ puntos; $\sigma = ?$

$\bar{x} = 52$ $s^2 = 36$ $s = 6$ $n = 9$

a) $P(\bar{x} > 54) = 1 - P(t < (54 - 50) / (6 / \sqrt{9})) = 1 - P(t < 2) = 1 - 0.9625 = 0.0375$

a)

	1 - α							
n	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355

Como se observa en la tabla, el 2.0 se encuentra entre 1.86 y 2.306, valores que corresponden a las áreas de 0.95 y 0.975. Realizando una estimación

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

burda, se promedian los dos valores correspondientes a las áreas. Encontrando que la probabilidad de que el promedio del puntaje de los estudiantes sea mayor de 54 es muy baja, 0.0375.

c) $P(\bar{x} < 54) = P(t < (54-50)/(6/3)) = P(t < 2) = 0.9625$. Por el contrario de lo anterior, es muy probable que el promedio del puntaje de los estudiantes sea menor de 54, dicha probabilidad equivale al 0.9625.

d) $P(48 < \bar{x} > 52) = P(\bar{x} < 52) - P(\bar{x} < 48) = P(t < (52-50)/(6/3)) - P(t < (48-50)/(6/3)) =$

$$P(t < 1) - P(t < -1) = 0.825 - (1 - 0.825) = 0.65$$

La probabilidad es de 0.65. Se aprecia que al ser simétrica la distribución t, se calcula la utilizando el inverso.

Solución Excel:

a) $P(\bar{x} > 54) = P(t > 2) = =\text{DISTR.T}(2;8;1) = 0.04025$

b) $P(\bar{x} < 54) = P(t < 2) = (1 - =\text{DISTR.T}(2;8;1)) = 0.95975$

c) $P(48 < \bar{x} > 52) = P(\bar{x} < 52) - P(\bar{x} < 48) = P(t < 1) - P(t < -1) =$

$$(1 - =\text{DISTR.T}(1;8;1)) - =\text{DISTR.T}(1;8;1))$$

$$= (1 - 0.173297) - 0.173297 = 0.6534$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Los valores de t, están diseñados para valores mayores, por eso, se le resta la unidad cuando se quiere calcular un valor hacia la izquierda, situación diferente a la tabla.

6.2.5 Distribución Ji cuadrado - χ^2 -

Tiene un sólo parámetro denominado grados de libertad. La función de densidad es asimétrica positiva, como las anteriores es continua y su función de densidad se hace más simétrica cuando aumenta el número de grados de libertad.

DISTR.CHI(x;grados_de_libertad)

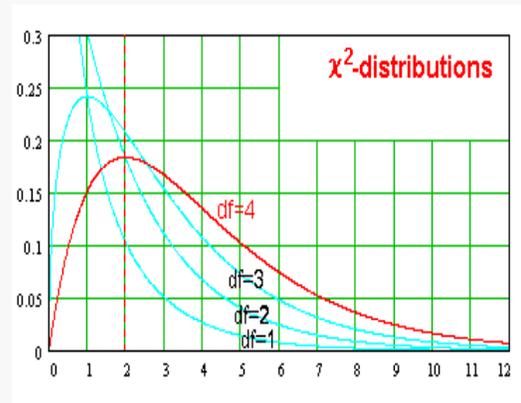
X es el valor al que desea evaluar la distribución.

Grados_de_libertad es el número de grados de libertad, su número depende del tipo de prueba que se este utilizando.

DISTR.CHI se calcula como
DISTR.CHI = $P(X > x)$,

Ejemplo.

DISTR.CHI(18,307;10) es igual a
0,050001



istribuciones de Probabilidad.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Esta distribución tiene una gran aplicación en investigación, ya que se utiliza para probar independencia entre variables categóricas, además, se prueba si unos datos provienen de una determinada distribución. También es clave para las pruebas que involucran la varianza de una población o la razón de varianzas de dos poblaciones

La distribución ji-cuadrado muestra como es el comportamiento de s^2 , si se seleccionan todas las muestras posibles de una población normal y a cada muestra se le calcula su varianza.

Cuando se requiere estimar la varianza o la desviación estándar poblacional, así como calcular probabilidades en función de la desviación se calcula el estadístico χ^2 , de la siguiente manera:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \approx \chi^2_{(n-1)}$$

Los grados de libertad para el caso de una muestra, es igual al numero de ella menos uno.

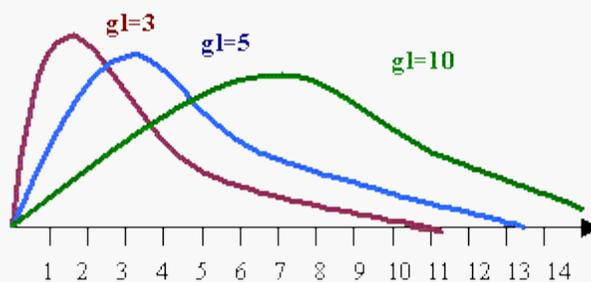
Donde n es el tamaño de la muestra, s^2 la varianza muestral y σ^2_x la varianza de la población de donde se extrajo la muestra.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Propiedades de la distribución Ji-cuadrado

1. Los valores de χ^2 son mayores o iguales que 0.
2. La forma de una distribución χ^2 depende del $gl=n-1$. En consecuencia, hay un número infinito de distribuciones χ^2 .
3. Las distribuciones χ^2 son asimétricas positivas.
4. Cuando $n>2$, la media de una distribución χ^2 es $n-1$ y la varianza es $2(n-1)$.
5. El valor modal de una distribución χ^2 se da en el valor $(n-3)$.

De manera gráfica se presentan tres distribuciones χ^2 , resaltando que a medida que los gl aumentan, la forma de la curva se asemeja a la normal.



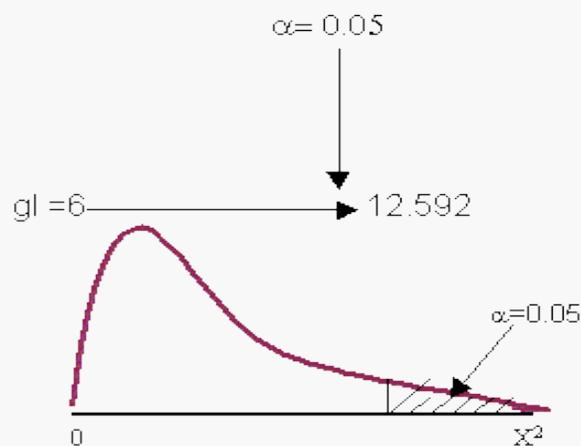
Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

La ji-cuadrada es de gran aplicación en la investigación científica, debido a que además, de lo ya expuesto, apoya múltiples pruebas de hipótesis, por lo tanto, en este apartado, se hará referencia a como calcular los valores de la χ^2 . Para ello se utilizará el manejo de la tabla χ^2 , y el Excel. Es importante anotar acá, que el llamado a está en función de la probabilidad que se quiera calcular, valga decir, si el área a calcular es del 5%, entonces, se entiende que es un área muy pequeña y viceversa. También, hay que tener conocimiento sobre la tabla a usar, ya que algunas tienen valores tabulados para mayores y otras para menores, en el texto se usarán las que presentan los valores hacia la derecha, siendo coherentes con lo entregado por el programa Excel.

Ejemplo manejo de tabla: Encontrar $\chi^2_{(0.05;6)}$: en la tabla se localiza 6 gl en el lado izquierdo y 0.05 en la parte superior, tal como se muestra en la gráfica y en la tabla que se presenta posteriormente.



Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

En Excel la función es la siguiente:

$$=PRUEBA.CHI.INV(0.05;6) = 12.5915$$

Si se quiere calcular $\chi^2_{(0.025,10)}$, en la tabla se muestra la intersección de los dos puntos requeridos, 0.025 y 10.

$$\text{En Excel, se tiene: } =PRUEBA.CHI.INV(.025;10) = 20.4832$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

TABLA DE LA DISTRIBUCION Chi-cuadrado

La tabla da áreas α , para valores mayores o iguales a χ^2 y n g.l.

g.l	α									
	0,995	0,975	0,900	0,500	0,100	0,050	0,025	0,010	0,005	0,001
1	0,000	0,000	0,016	0,455	2,706	3,841	5,024	6,635	7,879	10,828
2	0,010	0,051	0,211	1,386	4,605	5,991	7,378	9,210	10,597	13,816
3	0,072	0,216	0,584	2,366	6,251	7,815	9,348	11,345	12,838	16,266
4	0,207	0,484	1,064	3,357	7,779	9,488	11,143	13,277	14,860	18,467
5	0,412	0,831	1,610	4,351	9,236	11,070	12,832	15,086	16,750	20,515
6	0,676	1,237	2,204	5,348	10,645	12,592	14,449	16,812	18,548	22,458
7	0,989	1,690	2,833	6,346	12,017	14,067	16,013	18,475	20,278	24,322
8	1,344	2,180	3,490	7,344	13,362	15,507	17,535	20,090	21,955	26,124
9	1,735	2,700	4,168	8,343	14,684	16,919	19,023	21,666	23,589	27,877
10	2,156	3,247	4,865	9,342	15,987	18,307	20,483	23,209	25,188	29,588
11	2,603	3,816	5,578	10,341	17,275	19,675	21,920	24,725	26,757	31,264
12	3,074	4,404	6,304	11,340	18,549	21,026	23,337	26,217	28,300	32,910
13	3,565	5,009	7,042	12,340	19,812	22,362	24,736	27,688	29,819	34,528
14	4,075	5,629	7,790	13,339	21,064	23,685	26,119	29,141	31,319	36,123
15	4,601	6,262	8,547	14,339	22,307	24,996	27,488	30,578	32,801	37,697
16	5,142	6,908	9,312	15,338	23,542	26,296	28,845	32,000	34,267	39,252
17	5,697	7,564	10,085	16,338	24,769	27,587	30,191	33,409	35,718	40,790
18	6,265	8,231	10,865	17,338	25,989	28,869	31,526	34,805	37,156	42,312
19	6,844	8,907	11,651	18,338	27,204	30,144	32,852	36,191	38,582	43,820
20	7,434	9,591	12,443	19,337	28,412	31,410	34,170	37,566	39,997	45,315

Ejemplo: La prueba HG de laboratorio, requiere de mucha precisión, por lo tanto, es importante determinar la probabilidad de que la varianza se pase de un valor predeterminado. El tiempo de la prueba tiene un comportamiento normal con una $\sigma_x = 1$ minuto. Si se elige al azar una

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

muestra de 17 pruebas, calcule la probabilidad de que la S^2_x del tiempo sea mayor de 2.

Solución manual:

Tenga presente que los g.l son $n-1$, es decir, $17-1=16$.

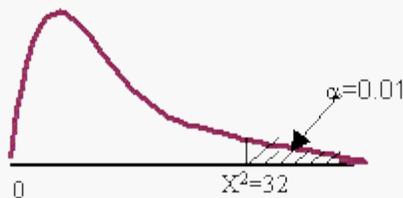
Partiendo del estadístico,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \approx \chi^2_{(n-1)}$$

Entonces:

$$P(\sigma^2_x > (16 \cdot 2) / 1) = P(\sigma^2_x > 32) = 0.01$$

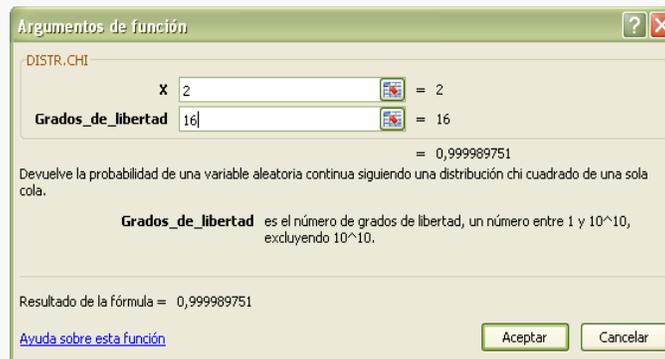
En la gráfica se aprecia como la probabilidad de que la varianza sea mayor de 32 es muy pequeña.



La búsqueda se hace de manera inversa a la anterior, es decir, se tiene el valor dentro de la tabla, 32, y los g.l 16, donde se intercepten, se busca el valor del α , encontrando el número 0.01. En caso de no estar el valor exacto, se procede como se trabajó en la distribución t.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Solución Excel:



El resultado es similar al encontrado en las tablas: 0.0099.

Dado que ésta distribución tiene múltiples aplicaciones en la investigación, más adelante, se presentarán nuevos ejemplos.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

6.2.6 Distribución F.

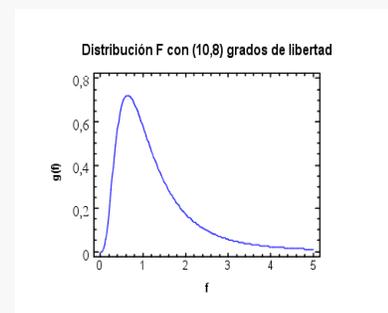
Se utiliza para determinar si existe diferencia en el grado de variabilidad de dos poblaciones con distribuciones normales.
Sintaxis en Excel.

$DISTR.F(x; \text{grados_de_libertad1}; \text{grados_de_libertad2})$

X es el valor al que desea evaluar.
 $\text{Grados_de_libertad1}$ es el número de grados de libertad del numerador.

$\text{Grados_de_libertad2}$ es el número de grados de libertad del denominador.

$DISTR.F$ se calcula como $DISTR.F = P(F < x)$, donde F es una variable aleatoria con una distribución F.



Tiene dos parámetros denominados grados de libertad, los cuales dependen de los tamaños de muestra, esto implica, que se utiliza cuando se trabaja con dos poblaciones que se distribuyen aproximadamente normal y donde interesa analizar el cociente de sus varianzas.

Sólo toma valores positivos. Es asimétrica. Normalmente se consideran valores anómalos los de la cola de la derecha, siempre y cuando se alejen del valor crítico.

Se aplica fundamentalmente para probar la igualdad de dos varianzas, por lo tanto, es una prueba necesaria para determinar diferencia de medias que

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

se verá en el capítulo 8. El estadístico esta definido así:

Una variable aleatoria se distribuye según el modelo de probabilidad F de Snedecor con (V_1, V_2) grados de libertad, donde V_1, V_2 son enteros positivos. La distribución F es el cociente de 2 Chi cuadrados. y se aplica fundamentalmente para probar la razón entre dos varianzas, el estadístico esta definido así:

$$F_{(n_1-1, n_2-1)} = \frac{S_1^2}{S_2^2}$$

$n_1 =$ tamaño muestra 1

$n_2 =$ tamaño muestra 2

$n_{1-1} =$ g.l grupo 1

$n_{2-1} =$ g.l grupo 2

$S_1^2 =$ Varianza grupo 1

$S_2^2 =$ Varianza grupo 2

$V_1 = n_{1-1}$

$V_2 = n_{2-1}$

De momento, se muestra la manera de calcular los valores de F, teniendo en cuenta los g.l y las diferentes áreas. Para ello, es importante tener en cuenta la siguiente relación:

$$F_{(1-\alpha, v_1, v_2)} = \frac{1}{F_{(\alpha, v_2, v_1)}}$$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

Lo anterior se usa debido a que usualmente sólo se tienen tablas diseñadas para valores de 0.05 y 0.01, por lo tanto y dado que la distribución F no es simétrica, se requiere calcular valores para áreas grandes, valga decir, 0.95 y 0.99.

Notará que se invierten los g.l en el denominador.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

TABLA DE LA DISTRIBUCION F

La tabla entrega áreas de a , para valores mayores o iguales a F y para 0.05 y 0.01.

alfa =	0.05	g.l numerador									
g.l denominador	1	2	3	4	5	6	7	8	9	10	
1	161.446	199.499	215.707	224.583	230.160	233.988	236.767	238.884	240.543	241.882	
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	
alfa =	0.01	g.l numerador									
g.l denominador	1	2	3	4	5	6	7	8	9	10	
1	4052.18	4999.34	5403.53	5624.25	5763.95	5858.95	5928.33	5980.95	6022.39	6055.92	
2	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397	
3	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228	
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	

Estadística Descriptiva y Distribuciones de Probabilidad.

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

9	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368

Ejemplo 1: Encuentre los siguientes valores de F.

a) $F_{(0.05;8,10)}=3.072$, b) $F_{(0.05,5,7)}=3.972$

c) $F_{(0.01;7,15)}= 4.142$, d) $F_{(0.01;8,8)} = 6.029$

Ejemplo Excel:

a) $F_{(0.05;8,10)} = =\text{DISTR.F.INV}(0.05,8,10)=3.071666$

b) $F_{(0.05,5,7)} = =\text{DISTR.F.INV}(0.05,5,7)=3.9715$

c) $F_{(0.01;7,15)} = =\text{DISTR.F.INV}(0.01,7,15)=4.142$

d) $F_{(0.01;8,8)} = =\text{DISTR.F.INV}(0.05,8,8)=6.0288$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

e) $F_{(0.95,6,10)} = \text{DISTR.F.INV}(0.95,6,10) = 0.2463$

f) $F_{(0.95;8,10)} = \text{DISTR.F.INV}(0.95,8,10) = 0.2987$

g) $F_{(0.99;7,9)} = \text{DISTR.F.INV}(0.99,7,9) = 0.1488$

Las ventajas de usar Excel son evidentes, no sólo en lo ágil, sino, en la precisión.

Ejercicios

1. Se selecciona una muestra al azar de dos participantes de un comité editorial con el fin de elaborar un anteproyecto, el grupo esta compuesto por dos mujeres y cuatro hombres. Encuentre el número esperado de hombres en la muestra. Halle el número esperado de mujeres en la muestra.
2. Se seleccionan tres profesores con el fin de asistir a una capacitación en registros médicos, la administración puede seleccionar de cuatro médicos y tres enfermeras profesionales. Si X es una v.a que denota el número total de médicos.
 - a. Encuentre la f.d.p
 - b. ¿Cuál es la probabilidad de que dos sean médicos?.
 - c. ¿Cuál es la probabilidad de que mínimo dos sean médicos?.
 - d. Encuentre el número esperado de médicos en la delegación.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

e. Calcule la desviación con respecto al valor esperado. s_x

3. Una muestra al azar de tres participantes en una conferencia escogida, sin reposición, de un grupo de cuatro mujeres y tres hombres. Encuentre el número esperado de hombres en la muestra. Si se utilizara muestras con reposición, halle el número esperado de mujeres en la muestra.

4. Un computador contiene 6 “CHIPS”, dos de los cuales son defectuosos. Se seleccionan tres “CHIPS” al azar, se sacan del equipo y se inspeccionan. Sea X el Número de defectuosos observados. Encuentre la distribución de X , representélo gráficamente y encuentre el número esperado de defectuosos.

5. Los siguientes datos representan las horas de espera para ser atendidos en una Unidad Intermedia de Salud.

5.3 5.8 5.3 5.5 5.5 5.4 6.0 6.1 5.5 5.4 5.6

5.2 5.4 5.6 5.4 5.3 6.1 4.9 5.5 5.3 5.4 6.0.

6. Compruebe si los datos son aproximadamente normales (cualquier método) y encuentre las siguientes probabilidades.

a) $P(X > 5.6)$. b) $P(X < 5.6)$. c) $P(5.0 < X < 6.0)$

d) $P(X < x_0) = 0.15$ e) $P(X > x_0) = 0.90$ f) $P(x_0 < X < x_1) = 0.90$

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

7. Los siguientes datos representan los días que transcurren para recibir el pago de glosas en una institución de salud.

20 25 30 35 25 30 26 30 35 34 30 25 35
30 25 35 29 35 30 24 32 30 29 33 42 38

Compruebe si los datos son aproximadamente normales (cualquier método) y encuentre las siguientes probabilidades.

a) $P(X > 35)$. b) $P(X < 25)$ c) $P(25 < X < 36)$

d) $P(X < x_0) = 0.25$ e) $P(X > x_0) = 0.80$ f) $P(x_0 < X < x_1) = 0.93$

8. Las ventas diarias en un restaurante tienen una distribución aproximadamente normal con media $m=500$ (miles) por día y desviación estándar igual a 120 (miles).

a) ¿Cuál es la probabilidad de que las ventas excedan los \$ 700 en un día dado?

b) ¿Cuál es la venta diaria de tal forma que sólo sea superada por el 5% de las ventas?

9. Las notas en Introducción a la Psicología de un grupo de estudiantes de la maestría se distribuyen según una distribución normal con media 3.8 y desviación típica 0.95. Hallar la probabilidad de que la media de una muestra tomada al azar de 25 estudiantes esté comprendida entre: a) 2.5 y 3. b) La probabilidad de que la media sea mayor de 4.0. c) ¿Cuál es la

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

probabilidad de que un estudiante pierda? Use tablas y Excel. Compare los resultados.

10. Responda si las siguientes afirmaciones son verdaderas o no y justifique su respuesta.

- La distribución normal se utiliza cuando no se conoce s .
- A medida que se incrementan los g.l la distribución t se aproxima a la normal.
- A medida que se incrementan los g.l la distribución χ^2 se aproxima a la normal.
- Para calcular probabilidades con respecto a una s se utiliza la χ^2
- Para calcular probabilidades con respecto a una $2s$ se utiliza la F
- Si el supuesto de normalidad falla moderadamente, la t puede seguir utilizando.

11. Los pesos de una población de mujeres jóvenes con la mayoría de edad tienen una distribución aproximadamente normal con una media de 130 libras y una desviación típica de 15. Calcular la probabilidad de que una joven elegida al azar de entre esa población, pese: a) Más de 155 pero menos de 100 libras, b) Si son 1650 mujeres, cuantas de ellas pesan siquiera 160 libras, c) ¿Cuál es el peso de tal manera que el 50% de las mujeres están por encima de ese valor? Manual y Excel.

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

12. Los puntajes obtenidos en un test para medir el estrés se distribuyen según una distribución normal con media 72 y desviación típica 9.5. Hallar la probabilidad de que la media de una muestra tomada al azar de 36 personas esté comprendida entre: a) 68 y 73. b) La probabilidad de que la media sea menor de 72. c) ¿Cuál es la probabilidad de que una persona sea considerada con problemas de estrés, si el estándar sugiere que el que tenga menos de 68 puntos tiene el problema?

13. El número de días entre la facturación y el pago de las cuentas corrientes de crédito en el Hospital Metropól tiene una distribución aproximadamente normal con una media de 18 días y una varianza de 16. Que proporción de las facturas será pagada.

a. Entre 12 y 18 días?

b. Dentro de cuántos días estará pagado el 99.5% de las facturas?

c. Entre cuáles dos valores simétricamente distribuidos alrededor de la media recaerán el 98% de las facturas?

15. En una población de 1200 adolescentes la cantidad media de dinero gastado en recreación por semana es de \$8000, se selecciona una muestra de 52 adolescentes de esta población y se encuentra una desviación típica de \$7300. ¿Cuál es la probabilidad de que un promedio muestral este comprendido entre \$6000 y \$10000?. Compare los resultados manual y con Excel.

16. Usando tanto tablas como el programa Excel, encontrar los siguientes

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

valores:

a) $F(0.01;7,9)$ b) $F(0.01,8,6)$ c) $F(0.05;9,10)$

d) $F(0.05;8,8)$ e) $F(0.95,6,8)$ f) $F(0.95;8,6)$

g) $F(0.99;8,7)$ h) $F(0.99,7,7)$ i) $F(0.02,8,6)$

17. Si S_1^2 y S_2^2 son las varianzas muestrales encontradas en dos muestras de tamaño 15 y 20 respectivamente, provenientes de distribuciones normales con varianzas poblacionales iguales. Calcule la probabilidad de que:

a) $P(S_1^2/S_2^2 < 4.89)$. b) $P(S_1^2/S_2^2 > 14.19)$. c) $P(4.89 < S_1^2/S_2^2 < 14.19)$

18. Se toma una muestra aleatoria de $n = 25$ observaciones de una población normal que tiene una varianza poblacional de 10. Encuentre la probabilidad de que la varianza de la muestra sea:

a) $P(S^2 < 16.4)$. b) $P(S^2 > 7.19)$ c) $(7.19 < S^2) < 16.4)$

19. Un laboratorio produce un medicamento cuya variación excesiva podría producir efectos secundarios graves en los usuarios. Se sabe que la desviación estándar del proceso es 1.2°F . Si la desviación de la muestra es mayor de 2.1°F , el lote no debe salir al mercado, si se ensayan 30 píldoras, ¿cuál es la probabilidad de que no se salga al mercado con dicho medicamento?

ESTADÍSTICA COMO APOYO A LA INVESTIGACIÓN.

20. Se toma una muestra aleatoria de $n = 50$ observaciones de una población normal que tiene una varianza poblacional de 10. Encuentre la probabilidad de que la varianza de la muestra sea:

a) $P(S^2 < 16.4)$. b) $P(S^2 > 7.19)$. c) $P(7.19 < S^2 < 16.4)$

[\[1\]](#) Tomado de las ayudas de Excel.