UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE ADMINISTRACIÓN
DEPARTAMENTO DE CONTADURÍA

FUNDAMENTOS DE LA ESTADÍSTICA INFERENCIAL

ASESOR ACÁDEMICO:
Prof. Miguel Romero

AUTORES:
Mariangeles Gómez R.
C.I:13.836.535
Yaritza Ramos S.
C.I:13.499.178

Trabajo de Curso Especial de Grado presentado como requisito parcial para optar al título de Licenciada en Contaduría Pública

Cumaná, abril de 2008

UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE ADMINISTRACIÓN
DEPARTAMENTO DE CONTADURÍA

FUNDAMENTOS DE LA ESTADÍSTICA INFERENCIAL

Autores:
Mariangeles Gómez R. C.I: 13.836.535
Yaritza Ramos S. C.I: 13.499.178

ACTA DE APROBACIÓN DEL JURADO

Trabajo de grado aprobado en nombre de la Universidad de Oriente, por el siguiente jurado calificador, en la ciudad de Cumaná, a los 8 días de mes de abril de 2008

Prof. Miguel Romero
Jurado Asesor
C.I: 8.879.006

ÍNDICE GENERAL

AGRADECIMIENTOS i
DEDICATORIA iii
LISTA DE TABLAS v
LISTA DE FIGURAS vi
RESUMEN vii
INTRODUCCIÓN 1
PLANTEAMIENTO DE PROBLEMA 4
OBJETIVOS 6
Objetivo General 6
Objetivos Específicos 6
JUSTIFICACIÓN 7
MARCO METODOLÓGICO 9
Nivel de Investigación 9
Diseño de la Investigación 9
Fuentes de Información 10
CAPÍTULO I 11
ESTADÍSTICA INFERENCIAL, DISTRIBUCIONES MUESTRALES, INTERVALO DE CONFIANZA Y PRUEBAS DE HIPÓTESIS 11
1.1. Definición de Estadística Inferencial 11
1.2. Distribuciones Muestrales 12
1.2.1. Procedimientos de Muestreo 14
1.2.1.1. Errores y Sesgos 14
1.2.1.2. Métodos de Muestreo 15
1.3. Intervalo de Confianza 25
1.3.1. Intervalo de Confianza para la Media 26
1.3.2. Intervalo de Confianza para una Proporción 28
1.4. Pruebas de Hipótesis 29

1.4.1. Procedimientos para Probar una Hipótesis 30
1.4.2. Prueba de Hipótesis para la Media 36
1.4.3. Prueba de Hipótesis para una Proporción 37
CAPITULOII 39
ANÁLISIS DE REGRESIÓN Y CORRELACIÓN 39
2.1. Análisis de Regresión 39
2.1.1. Principio de Mínimos Cuadrados 44
2.1.2. Trazo de la Línea de Regresión 45
2.2. Análisis de Regresión Múltiple 50
2.3. Análisis de Correlación 51
2.3.1. Coeficiente de Correlación 52
2.3.2. Coeficiente de Determinación 56
2.3.3. Coeficiente de no Determinación 56
2.4. Análisis de Correlación Múltiple 59
CAPITULOIII 61
ANÁLISIS DE VARIANZA Y PRUEBAS NO PARAMETRICAS 61
3.1. Análisis de Varianza o Anova 61
3.1.1. Distribución F 65
3.1.2. Comparación de dos Varianzas Poblacionales 67
3.2. Pruebas no Paramétricas 68
3.2.1. Distribución de ji-cuadrado 69
3.2.1.1. Prueba de Bondad de Ajuste 69
3.2.1.2. Prueba de Independencia o Tablas de Contingencias 74
CONCLUSIONES 81
RECOMENDACIONES 83
BIBLIOGRAFÍA 84
ANEXOS 86

AGRADECIMIENTOS

Hoy día estoy convencida de que con empeño y optimismo podemos alcanzar lo que nos proponemos. Un millón de gracias les doy:

A Dios Todopoderoso, por otorgarme ese privilegio tan grande como es el “VIVIR” y darme la fortaleza para aprender a luchar por lo que se quiere.

A la Virgen del Valle y al Divino Niño Jesús, por ayudarme a mantener la fe en todo instante y, a creer en la esperanza y a que en el momento menos pensado las cosas se dan.

A mi madre, Ana Teresa Romero, por estar siempre en los momentos difíciles, y por darme el valor para no dejar de desistir en mi carrera.

A mi padre, Bautista Gómez, por tratar de motivarme y colaborar conmigo a lo largo de mi carrera.

A la Universidad de Oriente, por abrirme las puertas para pertenecer a esta gran casa de estudio.

A nuestro Asesor Académico, Profesor Miguel Romero, por ayudarnos a culminar de una manera satisfactoria nuestro trabajo.

A mis compañeros de curso, en especial a Yaritza Ramos, por mostrar ser un grupo unido durante toda la trayectoria del mismo.

Mariangeles Gómez R.

AGRADECIMIENTOS

En primer lugar le doy gracias a Dios todo poderoso por su fortaleza y por haberme permitido terminar mis estudios Universitarios. Por ser la luz que guía nuestros pasos, por darme la vida y permitirme realizar una de mis metas más anheladas, de obtener mi titulo.

A mis padres Orlando Ramos y Gloria Serrada de Ramos quienes con mucho amor, constancia y dedicación se hicieron presentes. Gracias a ustedes por tener tanto afecto y consejos. El triunfo que hoy obtengo es de ustedes, a ustedes le debo lo que soy hoy en la vida.

A mis hermanos Jacqueline, Yngrid, Jannifher, Orlando y la consentida Glorianny, por haber compartido todos los momentos buenos y los momentos difíciles, por su confianza y por estar ahí cuando los necesitos.

A mi novio José Manuel Pens por la colaboración de este trabajo.

A mi profesor Miguel Romero, por ser nuestro asesor en el desarrollo de la investigación, por orientarnos para la elaboración de este trabajo, apoyarnos en todo cuando necesitábamos y brindarnos su amistad.

A mis compañeros del curso de especial de grado, por todos los momentos compartidos en este curso. En especial a mi compañera de tesis Mariangeles, porque juntas emprendimos el camino con esfuerzos y esmero, para elaborar este trabajo.

A nuestra casa de estudio la Universidad de Oriente, Núcleo de sucre por brindarnos la oportunidad de realizarnos como profesionales.

Yaritza Ramos S.

DEDICATORIA

A mi abuela Ana del Valle Romero (mi querida MIMA); como me hubiese gustado que estuvieras hoy día compartiendo esta ¡alegría que me embarga!, pero aunque ya no estás presente, se que desde el cielo me has tendido tu mano para salir adelante.

A mis padres, Ana Teresa Romero y Bautista Gómez, por guiarme a lo largo de mi vida, dándome los principios y valores fundamentales de toda educación.

Mariangeles Gómez R.

DEDICATORIA

A mis padres Orlando Ramos y Gloria Serrada de Ramos, quienes siempre han estado a mi lado brindándome su confianza, comprensión, apoyo y amor, guiándome en todos los aspecto de mi vida.

A mis hermanos Jacqueline, Yngrid, Jannifher, Orlando y la consentida Glorianny , porque junto a ellos aprendí, que alcanzar una meta, hay que hacer muchos esfuerzos y sacrificio. Hermanos este triunfo es también de ustedes, ya que siempre han estado a mi lado, llenándome de fuerzas para luchar contra las adversidades y salir siempre triunfante.

A mis sobrinos Samuel y Michelle, espero que mi ejemplo los motive a lograr sus metas y ayude a obtener lo que quieren en la vida para que sean orgullo de sus padres.

A mi novio José Manuel Pens porque con su amor y comprensión, estuvo a mi lado dándome palabras de aliento para que luchara por mí meta. Mi amor te dedico mi triunfo, gracias por compartir conmigo los momentos más bello de mi vida.

A todos mis compañeros del curso especial de grado, en especial a Mariangeles, quien realizo junto a mi, este trabajo de investigación.

Yaritza Ramos S.

LISTA DE TABLAS

Tabla 1-1 Muestreo Estratificado 22
Tabla 1-2 Tipos de Muestreo 24
Tabla 2-1 Análisis de Regresión Lineal 48
Tabla 3-1 Frecuencias Esperadas 72
Tabla 3-2 Tabla de contingencia 2x2. Frecuencias Observadas 77
Tabla 3-3 Tabla de contingencia 2x2. Frecuencias Observadas y Esperadas 78
Tabla 3-4 Valor calculado de ji-cuadrado 79

LISTA DE FIGURAS

Figura 1-1 Clasificación de los Métodos de Muestreo... 16
Figura 1-2 Pasos para Efectuar una Prueba de Hipótesis.. 31
Figura 1-3 Zonas críticas en la curva normal.. 32
Figura 1-4 Posibilidades que se Tienen al Tomar una Hipótesis............................. 33
Figura 1-5 Prueba Unilateral Izquierda.. 34
Figura 1-6 Prueba Unilateral Derecha.. 35
Figura 2-1 Relación Positiva entre las Variables 42
Figura 2-2 Relación Negativa entre las Variables.. 42
Figura 2-3 No Existe Relación entre las Variables 43
Figura 2-4 Diferentes formas que toma el Trazo de la Línea de Regresión 45
Figura 2-5 Diagrama de Dispersión con respecto a las ventas y publicidad............. 48
Figura 2-6 Intensidad y Dirección del Coeficiente de Correlación.......................... 52
Figura 2-7 Correlación Positiva 53
Figura 2-8 Correlación Negativa.. 54
Figura 2-9 Correlación Negativa.. 54
Figura 2-10 Correlación Positiva Fuerte 54
Figura 2-11 Sin Correlación... 55
Figura 2-12 Valores del Coeficiente de Correlación Múltiple................................. 60
Figura 3-1 Grados de Libertad en la Distribución F 66
Figura 3-2 Representación Gráfica de ji-cuadrado como Prueba de Bondad de Ajuste
.... 73
Figura 3-4 Distribución ji-cuadrado para Diferentes Grados de Libertad 76
Figura 3-5 Representación Gráfica de ji-cuadrado como Prueba de Independencia 80

UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE ADMINISTRACIÓN
DEPARTAMENTO DE CONTADURÍA

FUNDAMENTOS DE LA ESTADÍSTICA INFERENCIAL

Autores:
Mariangeles Gómez R.
Yaritza Ramos S.

Asesor:
Prof. Miguel Romero

Fecha: 08-04-2008.

RESUMEN

Inferir significa deducir algo de otra cosa. Nuestra investigación se refiere, a los Fundamentos de la Estadística Inferencial, que se encarga de hacer deducciones de una población por medio de una muestra tomada a partir de ésta; sirviendo así para las organizaciones, porque le permite a la Gerencia tomar decisiones válidas, respecto a las predicciones futuras. Para realizar este análisis estadístico se requiere utilizar la distribución muestral, porque a partir de la muestra seleccionada de una población, puede construirse variables aleatorias alternativas, de cuyo análisis se desprenden interesantes propiedades estadísticas (distribución muestral de la media y de la proporción). Los problemas que se tratan en Inferencia Estadística, se basan en dos clases: la estimación o intervalo de confianza y las pruebas de hipótesis. En donde el intervalo de confianza viene dado por un rango de valores, dentro del cual se espera encontrar el valor del parámetro estudiado; y las pruebas de hipótesis, que son supuestos que se plantea el investigador antes de iniciar una investigación, partiendo de una muestra aleatoria significativa, para extraer conclusiones que permitan aceptar o rechazar una hipótesis previamente emitida, sobre el valor de un parámetro desconocido, el cual aborda una serie de pasos. El análisis de regresión y correlación, permite relacionar dos o más variables (variable independiente y variable dependiente). El análisis de varianza sirve para comparar si los valores de un conjunto de datos numéricos, son significativamente distintos a los valores de otro o más conjunto de datos. Como en la práctica todas las poblaciones no pueden tomarse como normales, por situación en donde no es posible formular una hipótesis segura sobre el valor de un parámetro, surgen las pruebas no paramétricas (ji cuadrado), éstas no dependen de un sólo tipo de distribución.

Palabras claves: Inferir, Estimación, Hipótesis, Regresión, Correlación.

INTRODUCCIÓN

La palabra estadística proviene del latín "statiscus" y significa "del estado".

La Estadística, desde su origen y a lo largo de la historia ha mostrado un respetable prestigio en las estrategias de hacer uso de la información recopilada, con la finalidad de analizarla por medio de los datos que se recogen de un evento cualquiera.

La Estadística es una ciencia de la rama de las Matemáticas, que estudia la recopilación, organización, presentación, análisis e interpretación de datos de manera cuantitativa y cualitativa, la cual nos sirve de gran utilidad para una toma de decisiones más efectiva por medio de resultados precisos y predicciones hacia el futuro. Dicha ciencia es aplicable a una amplia variedad de disciplinas que van desde la física hasta las ciencias sociales, así como la psicología, la medicina, entre otras; se puede decir que la Estadística se puede aplicar a casi todo el quehacer humano, que genere información cuantitativa y cualitativa.

La Estadística se divide en dos ramas, que van desde el cálculo más sencillo hasta el análisis más complejo: la Estadística Descriptiva y la Estadística Inferencial.

La Estadística Descriptiva, se basa en hechos ya ocurridos, y no es más que la aplicación de aquellos métodos que por medio de la recolección de datos nos van a permitir describir, resumir y presentar la información a través de tablas, gráficas o cualquier valor numérico, llamadas también medidas descriptivas.

Inferir significa deducir algo de otra cosa. Existen algunos eventos, que para ser analizados requieren de elementos estadísticos, que van más allá de la Estadística

Descriptiva, es decir, que no basta con recolectar, describir y resumir los datos, para presentarlos en gráficas, tablas o valor numérico, razón por la cual surge la Estadística Inferencial.

La Estadística Inferencial, es un método inductivo, que trata de estimar las características de universo estadístico o población total, a través del estudio de una parte del universo, a esta parte se le denomina muestra.

El investigador, pudiera estar interesado en conocer información que traspase los hechos ocurridos, a través del estudio de una porción de la población objeto de estudio. Es por eso que la Estadística Inferencial permite entre otras cosas:

- Comparar las actuaciones de dos o más grupos y comprobar la significación de cualquier diferencia entre ellos.
- Probar la significación de las relaciones entre variables.
- Predecir el comportamiento futuro de una o más variables.

La Estadística Inferencial, es de gran importancia para las organizaciones, ya que le permite a la Gerencia tomar decisiones más válidas, acerca de lo que puede acontecer en la empresa por medio de las predicciones futuras analizadas.

Este trabajo de investigación, está estructurado en tres capítulos, los cuales van a permitir comprender mejor los fundamentos de la Estadística Inferencial:

En el capítulo I, se da una definición de la Estadística Inferencial, para luego hablar de las distribuciones muestrales, en donde se analiza el por qué es necesario muestrear y se explica los procedimientos y métodos de muestreo. Además se hace referencia del intervalo de confianza y el uso de las pruebas de hipótesis.

El capítulo II, trata sobre el análisis de regresión y correlación más que todo lineal. En el análisis de regresión lineal, se habla del principio de mínimos cuadrados y el trazo de la línea recta. En el análisis de correlación lineal, se muestra el coeficiente de correlación, el coeficiente de determinación, etc.

El capítulo III, se refiere al análisis de varianza y las pruebas no paramétricas, en especial la distribución de ji-cuadrado.

PLANTEAMIENTO DE PROBLEMA

La Estadística es una ciencia perteneciente a la rama de las Matemáticas, que se encarga de reunir, organizar y analizar información cuantitativa y cualitativa, y deducir a través del análisis de los datos significados precisos o previsiones hacia el futuro; además que nos sirve de ayuda para la toma de decisiones, ya que nos proporciona información, así como la relación de datos económicos y administrativos, o cualquier otra variable.

La Estadística como todas las ciencias, no se originó de improviso, sino mediante un proceso largo de desarrollo y evolución, desde hechos que van de una simple recolección de datos hasta la diversidad rigurosa e interpretación de los mismos (datos) que se dan hoy en día.

Dependiendo del tipo de información, la Estadística se divide en dos grandes ramas: la Estadística Descriptiva y la Estadística Inferencial. La Estadística Descriptiva se encarga más que todo de la presentación de datos en forma de gráficas, tablas y valores numéricos; y la Estadística Inferencial que en este caso es nuestro problema de estudio, va más allá, se deriva de muestras, de observaciones hechas solo acerca de una parte de un conjunto numeroso de elementos, por lo cual su análisis necesita de generalidades más profundas.

La Estadística Inferencial, es la parte de la Estadística, que nos permite comprender las técnicas con las que, con base únicamente en una muestra sometida a observación, se toman decisiones sobre una población o proceso estadístico. Dado que estas decisiones se toman en condiciones de incertidumbre, suponen el uso de conceptos de probabilidad. Mientras que a las características medidas de una muestra se les llama estadísticas muestrales y a las características medidas de una población o universo, se conoce como parámetro de la población.

Dado nuestro objeto de estudio surgieron algunas interrogantes:

¿En qué se fundamenta la Estadística Inferencial?

¿A qué nos referimos cuando hablamos de Estadística Inferencial?

¿Cuáles son las técnicas fundamentales utilizadas en la Estadística Inferencial?

¿En qué consisten las pruebas de hipótesis?

¿Qué es el análisis de regresión y correlación simple y múltiple?

¿Cómo identificar una variable dependiente y una variable independiente en el análisis de regresión?

¿Cuáles son las pruebas no paramétricas?

OBJETIVOS

Objetivo General

Estudiar los Fundamentos de la Estadística Inferencial.

Objetivos Específicos

- Definir los aspectos conceptuales de la Estadística Inferencial.
- Definir las técnicas elementales de la Estadística Inferencial.
- Estudiar las distribuciones muestrales.
- Explicar las pruebas de hipótesis.
- Identificar la variable dependiente y la variable independiente en el análisis de regresión.
- Estudiar el análisis de regresión y correlación simple.
- Describir las pruebas no paramétricas.
- Analizar la distribución ji-cuadrado.

JUSTIFICACIÓN

La Estadística es una ciencia que está diseñada para aplicar algunos métodos y técnicas que nos van a permitir tomar decisiones más efectivas, a través de la recolección, organización, análisis e interpretación de datos.

La Estadística se divide en dos categorías o ramas: Estadística Descriptiva y Estadística Inferencial.

La primera (Estadística Descriptiva), se refiere a aquellos métodos mediante los cuales se organizan, resumen y presentan los datos en forma cuantitativa, a través de tablas, gráficas o valores numéricos, permitiendo así que la información sea interpretada cómoda y rápidamente, y de esta manera utilizarlas eficazmente para el fin que se desee.

Sin embargo, muchas veces se requiere ir más allá de los hechos ya ocurridos, es por eso que surge la Estadística Inferencial, la cual se dedica a la generación de los modelos, inferencias y predicciones asociados a los fenómenos en cuestión, teniendo en cuenta lo aleatorio e incertidumbre en las observaciones. Esta Estadística, trabaja con muestras, ya que son subconjuntos formados por algunos individuos de la población; y a partir del estudio de la muestra se pretende inferir aspectos relevantes de toda la población.

La Estadística Inferencial juega un papel importante en las organizaciones y en el mundo empresarial, ya que actualmente ésta se ha convertido en un método muy efectivo para estudiar con mucha precisión los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos; además, sirve como herramienta para relacionar y analizar dichos datos.

Por lo anteriormente mencionado, se puede decir que la Estadística Inferencial es de gran ayuda para la Gerencia, ya que por medio de la muestra que se toma de una población, se pueden hacer predicciones a futuro en una organización, analizando así la Gerencia los posibles cambios que pueda sufrir una empresa en cualquier departamento y por ende llegar a conclusiones válidas; permitiendo así lograr una adecuada planeación y control apoyados en los estudios de pronósticos, presupuestos, etc.

MARCO METODOLÓGICO

La metodología incluye el tipo o tipos de investigación, las técnicas y procedimientos que serán utilizados para llevar a cabo la indagación, con el fin de lograr el objetivo de la misma.

Nivel de Investigación

Se refiere al grado de profundidad con que se aborda un objeto o fenómeno.

El nivel de investigación en este caso fue exploratorio y descriptivo.

Exploratorio, porque se necesitó tener una visión general del tema a estudiar, además de ser éste un tema poco estudiado.

Descriptivo, porque consistió en la caracterización de un hecho con el fin de establecer su estructura o comportamiento.

Diseño de la Investigación

Se refiere a la estrategia que adopta el investigador, para responder al problema planteado.

Este estudio se realizó en base a una investigación documental.

> Según FIDIAS G. ARIAS (1999: p.47), señala: “la investigación documental es aquella que se basa en la obtención y análisis de datos provenientes de materiales impresos u otros tipos de documentos”.

Fuentes de Información

Tienen que ver con el suministro de datos o información, que se utilizan para el estudio.

Para este estudio se requirió de una revisión bibliográfica basada en documentos escritos o fuentes secundarias (trabajos de investigación, textos, enciclopedias, etc.), lo cual permitió darle soporte a la investigación.

CAPÍTULO I
ESTADÍSTICA INFERENCIAL, DISTRIBUCIONES MUESTRALES, INTERVALO DE CONFIANZA Y PRUEBAS DE HIPÓTESIS

De acuerdo con el diccionario de la Real Academia Española, inferir significa "Sacar una consecuencia o deducir algo de otra cosa"

Existen algunos eventos, que para ser analizados, necesitan elementos estadísticos que van más allá de la Estadística Descriptiva, es decir, que no basta con recolectar, describir y resumir los datos, para presentarlos en gráficas, tablas o valores numéricos. Es por eso que surge la Estadística Inferencial, la cual se encarga de hacer deducciones de una población por medio de una muestra tomada a partir de ésta.

1.1. Definición de Estadística Inferencial

> Para Mason, Lind, y Marchal (2001:p19), Se refiere al "conjunto de métodos utilizado para saber algo acerca de una población, basándose en una muestra".

> Según Berenson y Levine (1996:p3), "La Estadística Inferencial puede definirse como aquellos métodos que hacen posible la estimación de una característica de una población o la toma de decisión de una población, basándose sólo en los resultados de la muestra"

La Estadística Inferencial nos permite comprender las técnicas, con las que con base únicamente en una muestra sometida a observación, se toman decisiones sobre una población o proceso estadístico.

Ésta se dedica a la generación de los modelos, inferencias y predicciones asociados a los fenómenos en cuestión, teniendo en cuenta lo aleatorio e incertidumbre en las observaciones. Esta Estadística trabaja con muestra, que no son más que subconjuntos formados de algunos individuos de la población; y a partir del estudio de la muestra se pretende inferir aspectos relevantes de toda la población.

Además, sirve de gran ayuda para las organizaciones, ya que le permite a la Gerencia tomar decisiones válidas, acerca de lo que puede acontecer en una empresa, a través de las predicciones futuras analizadas.

1.2. Distribuciones Muestrales

Quizás a veces nos preguntamos: ¿Por qué se requiere muestrear? Pues el muestrear surge de la necesidad que se tiene cuando se evalúa la calidad de un producto, cuando se quiere conocer la opinión de los consumidores, si es eficaz o no el producto, etc. En muchos casos no es factible estudiar a la población entera, bien sea porque el costo sea muy alto, o se necesita de mucho tiempo para contactar a toda la población, entre otras, son algunas de las razones por lo que es indispensable muestrear.

Para hacer dicho estudio, se necesitan tomar muestras, que no son más que una parte de la población. Entendiéndose por población, el conjunto de elementos que son seleccionados para llevar a cabo una investigación.

Una muestra, es un método para inferir algo acerca de una población, es decir, no es más que una parte tomada de la población a estudiar. Por medio del muestreo, surge el uso de los métodos estadísticos inferenciales.

Una vez seleccionada la muestra, se pueden construir variables aleatorias alternativas, desprendiéndose de éstas, propiedades estadísticas de gran interés. Las dos formas más comunes de estas variables corresponden a las distribuciones muestrales de la media y las distribuciones muestrales de la proporción.

Cabe destacar, que cuando se usan valores muestrales o estadísticos para estimar parámetros o valores poblacionales, se puede correr el riesgo de que ocurran dos tipos de errores: el error muestral y el error no muestral (sesgo muestral).

> Para Webster, Allen (1996:p296), las distribuciones muestrales se refieren a la “lista de todos los valores posibles de un estadístico y la probabilidad asociada a cada valor.”

La distribución muestral, se refiere a la distribución de los valores que tomará el estimador al escoger diferentes muestras de la población. Ésta distribución se basa fundamentalmente en dos medidas: la media, que indica el valor promedio del estimador; y la desviación típica o error típico de estimación, que se refiere a la desviación promedio que se puede esperar entre el estimador y el valor del parámetro.

Existen dos tipos de distribuciones muestrales, de las cuales se definirán brevemente: la distribución muestral de la media y la distribución muestral de la proporción.

- Distribución Muestral de la Media: puede definirse como aquella distribución que consta de todas las medias muestrales posibles de un tamaño de muestra dado. En donde cada muestra de tamaño “n”, extraída de una población proporciona una media, la cual es considerada como una variable aleatoria para estudiar su distribución.

- Distribución Muestral de la Proporción: es una distribución que se da, cuando muchas veces se plantea estimar una proporción o porcentaje, en donde la variable aleatoria toma únicamente dos valores diferentes, que no son más, que el éxito o el fracaso, en otras palabras, sigue una distribución binomial.

1.2.1. Procedimientos de Muestreo

Antes de hablar en sí de los métodos de muestreo, es necesario mencionar los riesgos asociados al procedimiento de muestreo. Estos son: el error muestral y el sesgo muestral.

1.2.1.1. Errores y Sesgos

a) Error Muestral

> Para Mason, Lind y Marchal (2001:p.285), el error muestral "Es la diferencia entre un valor estadístico de muestra y su parámetro correspondiente".

> Linconl, Chao (1993:p.137), opina que "es la diferencia entre el valor de una estadística obtenido mediante los datos muestrales y el valor correspondiente al parámetro de la población debido a variaciones fortuitas en la selección de las unidades".

b) Sesgo Muestral

> Para Mason, Lind y Marchal (2001:p263), el sesgo muestral "es cuando los resultados de la muestra no probabilística pueden no ser representativos de la población".

Cuando se usan valores muestrales o estadísticos para estimar parámetros o valores poblacionales, se puede correr el riesgo de que ocurran dos tipos de errores; en donde las muestras no son representativas a la población, debido a que los datos no son cien por ciento reales. Entendiéndose por muestras no representativas, aquellas que dan lugar a una estimación errónea del parámetro y a un error muestral.

Estos tipos de errores son: el error muestral, que se conoce como azar de la extracción, el cual se refiere a la variación natural que se da con las muestras tomadas de la misma población y por ende se encarga de determinar que esas muestras no contengan elementos que no sean característicos de la población; y el error no muestral, que es a lo que se refiere el sesgo muestral, aunque para algunos autores este es un tipo de error muestral; no es más que una tendencia inherente a un método de muestreo, que da estimaciones de un parámetro, las cuales pueden ser menores o mayores en promedio que el parámetro real (sesgo positivo y sesgo negativo), debido a factores que dependen de la recolección, análisis, interpretación y revisión de los datos; permitiendo así, llegar a conclusiones que son sistemáticamente diferentes de la verdad o incorrectas acerca de los objetivos de una investigación.

1.2.1.2. Métodos de Muestreo

A continuación, se presenta un esquema, el cual muestra los diferentes métodos de muestreo:

Figura 1-1   Clasificación de los Métodos de Muestreo

Muestreo

- a. No probabilístico
  - a.1. Muestreo por juicio
  - a.2. Muestreo casual o fortuito
  - a.3. Muestreo de cuota
  - a.4. Muestreo de poblaciones móviles
- b. Probabilístico
  - b.1. Muestreo aleatorio simple
    - b.1.1 m.a.s sin reposición
    - b.1.2 m.a.s con reposición
  - b.2. Muestreo aleatorio sistemático
  - b.3. Muestreo estratificado
  - b.4. Muestreo por conglomerado

Fuente: Las autoras.

Ahora, se explicarán los diferentes tipos de muestreo, expuestos en la figura anterior:

Existen dos tipos de muestreo: el muestreo no probabilístico y el muestreo probabilístico.

a. - Muestreo no Probabilístico

Es conocido también como muestreo no aleatorio o de juicio, en él interviene solamente la opinión personal del investigador, para identificar así los elementos de la población que deben incluirse en la muestra, evitando el análisis estadístico requerido para realizar muestras de probabilidad.

a.1. - Muestreo por Juicio

El investigador toma la muestra seleccionando los elementos que a él le parecen representativos o típicos de la población, por lo que depende del criterio del investigador.

a.2. - Muestreo Casual o Fortuito

Se usa en los casos en donde no es posible seleccionar los elementos, y deben sacarse conclusiones con los elementos que estén disponibles.

a.3. - Muestreo de Cuota

Se utilizan en estudios de opinión del mercado. Los enumeradores, reciben instrucciones de obtener cuotas específicas, a partir de las cuales se constituye una muestra relativamente proporcional a la población.

a.4. - Muestreo de Poblaciones Móviles

Este tipo de muestreo utiliza métodos de captura, marca y recaptura. Se utiliza mucho en el estudio de migración de poblaciones de animales y otras características.

b. - Muestreo Probabilístico o Aleatorio

Es un tipo de muestreo en donde se elige una muestra, de modo que cada integrante de una población tenga una probabilidad conocida de ser incluida en la muestra. Este muestreo se basa en datos previamente obtenidos, de los cuales se sacan conclusiones, y se clasifica en cuatro tipos:

b.1. - Muestreo aleatorio simple

b.2. - Muestreo aleatorio sistemático

b.3. - Muestreo estratificado

b.4. - Muestreo por conglomerado

b.1. - Muestreo Aleatorio Simple

Es un tipo de muestreo en donde intervienen las leyes de la probabilidad, y por tal motivo, cada elemento de la población tiene la misma posibilidad de ser elegido para formar parte de la muestra.

En este muestreo existen dos modalidades:

b.1.1. - Muestreo Aleatorio sin Reposición

b.1.2. - Muestreo Aleatorio con Reposición b.1.1. - Muestreo Aleatorio Simple sin Reposición

Consiste en que una vez seleccionado un elemento de la población para formar parte de la muestra y, hecho el estudio correspondiente, dicho elemento no puede volver a formar parte de la población de origen.

b.1.2. - Muestreo Aleatorio Simple con Reposición

Aquí, una vez que es seleccionado un elemento de la población para formar parte de la muestra y, ya anotadas sus características, dicho elemento regresa a la población de origen.

b.2. - Muestreo Sistemático

En este tipo de muestreo, los elementos que van a formar parte de la muestra, son seleccionados de manera ordenada, tomando en cuenta el tamaño de la población y el tamaño de la muestra.

Se debe determinar un intervalo regular o una razón. Se elige un punto de partida, representando éste el primer elemento de la muestra, el cual es al azar y debe de estar comprendido o incluido en el valor de "k". A este punto de partida, se le suma el valor de "k" o intervalo regular, obteniendo de esta manera el segundo elemento de la muestra, al cual se le suma nuevamente el valor de "k" para encontrar el tercero y así sucesivamente hasta completar el tamaño de la muestra.

Ejemplo:

Si queremos tener una idea de la edad promedio de los estudiantes que cursan el último semestre en la Universidad de Oriente (Núcleo Sucre). Entonces, primeramente se toma un punto de partida y luego de forma sistemática se van eligiendo las posibles muestras.

Supongamos que son 1.000 estudiantes y vamos a usar una muestra de 20 estudiantes.

Como debemos determinar un intervalo regular o una razón antes de elegir el punto de partida, usamos la siguiente fórmula:

¡Error! Marcador no definido.

$$k = \frac{N}{n}$$

Donde:

k = Razón o intervalo regular

N = Tamaño de la población

n = Tamaño de la muestra

Sustituyendo la fórmula:

N = 1.000

n = 20

$$k = \frac{1.000}{20} = 50$$

Ahora para determinar los elementos muestrales de acuerdo al tamaño de la muestra(n), primeramente elegimos un punto de partida al azar con el tamaño de la población (N), luego de una forma sistemática vamos a ir sumando el valor de “k” hasta completar el tamaño de la muestra, que en este caso es 20.

Punto de partida = 2

Elementos Muestrales (20) { 2, 52, 102, 152, 202, 252, 302, 352, 402, 452, 502, 552, 602, 652, 702, 752, 802, 852, 902, 952

2 + 50 = 52
52 + 50 = 102
.
.
.
902 + 50 = 952

b.3. - Muestreo Estratificado

El muestreo estratificado, se encarga de separar a la población en diferentes estratos o grupos, para después elegir una muestra aleatoria de cada uno, lo cual dará como resultado una muestra global.

Ejemplo:

Supongamos que nos interesa obtener una muestra de la comunidad universitaria. En donde la población esta formada por 40.000 personas, para ello se toma una muestra de 600 personas, las cuales están clasificadas en: estudiantes, obreros, empleados y profesores.

Tabla 1-1 Muestreo Estratificado

Comunidad Universitaria (estratos)	**Nh**	**%**
Estudiantes (A)	30.000	75
Obreros (B)	3.000	7,5
Empleados (C)	2.000	5
Profesores (D)	5.000	12,5
Total	**40.000**	**100%**

nh = $\frac{Nh}{N}$ * n**¡Error! Marcador no definido.**

Donde:

nh = Tamaño de la muestra de cada estrato

Nh = Número de elementos de cada estratos

N = Tamaño de la población

n = Tamaño de la muestra

$$nh\ (A) = \frac{30.000}{40.000} * 600 = 450$$

$$nh\ (B) = \frac{3.000}{40.000} * 600 = 45$$

$$nh\ (C) = \frac{2.000}{40.000} * 600 = 30$$

$$nh\ (D) = \frac{5.000}{40.000} * 600 = 75$$

$$\overline{\underline{\underline{600}}}$$

b.4. - Muestreo por Conglomerado, de Racimo o de Agregados

El muestreo por conglomerado es un tipo de muestreo aleatorio, en el cual la población se divide en grupos o conglomerados de elementos, para luego seleccionar una muestra aleatoria de esos grupos.

Este tipo de muestreo su usa en áreas geográficas grandes, permitiendo así reducir el costo de muestrear una población dispersa.

Ejemplo:

Para una investigación de mercado, queremos determinar por muestreo el número promedio de ventiladores que hay en una casa. Como la ciudad es grande geográficamente, se puede usar un mapa de ésta, en donde se divida el territorio en manzanas, para luego escoger un cierto número de las manzanas y a través de cada casa poder encuestar a sus habitantes.

Una vez expuesto cada uno de los tipos de muestreo, ahora se muestra un cuadro, que presenta las características, ventajas y desventajas de los diferentes tipos de muestreo probabilístico:

Tabla 1-2 Tipos de Muestreo

TIPOS DE MUESTREO	CARACTERÍSTICAS	VENTAJAS	DESVENTAJAS
Aleatorio Simple	Se selecciona una muestra de tamaño n de una población de N unidades, cada elemento tiene una probabilidad de inclusión igual y conocida n/N	Sencillo y de fácil comprensión. Cálculo rápido de medias y varianzas. Se basa en la teoría estadística, y por tanto existen paquetes informáticos para analizar los datos.	Requiere que se posea de antemano un listado completo de toda la población. Cuando se trabaja con muestras pequeñas es posible que no represente a la población adecuadamente.
Sistemático	Conseguir un listado de los N elementos de la población. Determinar el tamaño muestral n. Definir un intervalo k=N/n. Elegir un número aleatorio, r, entre 1 y k (r=arranque aleatorio). Seleccionar los elementos de la lista.	Fácil de aplicar. No siempre es necesario tener un listado de toda la población. Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos.	Si la constante de muestreo está asociada con el fenómeno de interés, entonces, las estimaciones obtenidas a partir de la muestra pueden contener sesgo de selección.
Estratificado	En ciertas ocasiones resultará conveniente estratificar la muestra según ciertas variables de interés. Para ello debemos conocer la composición objetivo a muestrear. Una vez calculado el tamaño muestral adecuado, éste se reparte de manera proporcional entre los distintos estratos definidos en la población usando una simple regla de tres.	Tiende a asegurar que la muestra represente adecuadamente a la población en función de unas variables seleccionadas. Por medio de éste, se obtienen estimaciones más precisas. Su objetivo es obtener una muestra lo más semejante posible a la población en lo que a las variables estratificadotas se refiere.	Se ha de conocer la distribución en la población de las variables utilizadas para la estratificación.

Continuación de Tabla N° 1-2

Conglomerados	Se realizan varias fases de muestreo sucesivas. La necesidad de listados de las unidades de una etapa se limita a aquellas unidades de muestreo seleccionadas en la etapa anterior.	Es muy eficiente cuando la población es muy grande y dispersa. No es preciso tener un listado de toda la población, sólo de las unidades primarias de muestreo.	El error estándar es mayor que en el muestreo aleatorio simple o estratificado. El cálculo del error estándar es complejo.

Fuente:http://www.hsa.es/id/investigacion/uai/uai_docs/muestreo/muestreo.htm

1.3. Intervalo de Confianza

A partir de la normalización de estudios estadísticos por medio de las distribuciones muestrales, se pueden determinar parámetros de una población a través de sus valores estadísticos. Esto, es lo que se conoce como estimación, la cual se clasifica en: estimación puntual y estimación por intervalo.

La estimación puntual, es aquella donde se estima un parámetro poblacional por medio de un sólo estadístico o valor del estimador.

La estimación por intervalo o intervalo de confianza, es aquella que viene dada por un rango de valores dentro del cual se espera encontrar el valor del parámetro que se estime. En otras palabras, el intervalo de confianza, se refiere a todos los valores obtenidos a partir de los datos de una muestra, en el que hay una determinada probabilidad de que se encuentre el parámetro estudiado con cierta certeza. Dicho intervalo viene dado por un límite inferior (L_i) y un límite superior (L_s).

Pero antes, de empezar a hablar, tanto del intervalo de confianza de la media como el de la proporción, es necesario definir que es un nivel de confianza, así como el coeficiente de confianza, y el teorema central del límite.

Nivel de confianza: es la máxima probabilidad de encontrar el valor del parámetro que se estima dentro del intervalo establecido, es decir, es el coeficiente de confianza expresado en porcentaje.

Coeficiente de confianza: es la probabilidad que existe, de que el intervalo de confianza contenga el parámetro poblacional.

Teorema central del límite: permite el uso de probabilidad normal para crear los intervalos de confianza de la media poblacional y realizar pruebas de hipótesis.

1.3.1. Intervalo de Confianza para la Media

Es un intervalo cuya finalidad es hallar dos valores (límite inferior y límite superior), los cuales pueden permitir calcular la media poblacional verdadera. Si se conoce la desviación típica poblacional (σ) se usa la siguiente fórmula:

$$\overline{X} \pm z\frac{\sigma}{\sqrt{n}} \quad \begin{cases} L_i = \overline{X} - z\dfrac{\sigma}{\sqrt{n}} \\[2ex] L_s = \overline{X} + z\dfrac{\sigma}{\sqrt{n}} \end{cases}$$

Donde:

$\overline{X}_i$ = Media muestral

z = Valor encontrado en la tabla de área bajo la curva normal, según el nivel de confianza

σ = Desviación estándar poblacional

n = Tamaño de la muestra

En caso de desconocerse la desviación estándar de la población, la fórmula es la siguiente:

$$\overline{X} \pm z \frac{s}{\sqrt{n}}$$

Donde:

$\overline{X}_i$ = Media muestral

z = Valor encontrado en la tabla de área bajo la curva normal, según el nivel de confianza

s = Desviación estándar muestral

n = Tamaño de la muestra

El teorema central del límite, se aplica cuando la población es infinita o el muestreo es con reemplazamiento, utilizando el factor de corrección. Las fórmulas son las siguientes:

Si se conoce la desviación típica poblacional:

$$\overline{X} \pm z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{n-1}}$$

Si se desconoce la desviación típica poblacional:

$$\overline{X} \pm z \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{n-1}}$$

Todas las fórmulas anteriormente mencionadas, se usan cuando las muestras son grandes (mayor que 30 elementos) o si la fracción muestral (n/N) es mayor a 0,05 ó 5%.

Si se trata de muestras pequeñas, se utiliza la distribución de Student, utilizando las fórmulas anteriores pero "z" (área bajo la curva normal) se sustituye por "t".

1.3.2. Intervalo de Confianza para una Proporción

Una proporción, es una razón o porcentaje que indica la porción de la muestra o la población que posee una característica determinada. El procedimiento es semejante al de la media. Empleando la siguiente fórmula:

$$\overline{p} \pm z \sqrt{\frac{p(1-p)}{n}}$$

Donde:

$\overline{p} =$ Proporción muestral del evento estudiado

z = Valor encontrado en la tabla de área bajo la curva normal, asociado al nivel de confianza

n = Tamaño de la muestra

p = Proporción de la población

1.4. Pruebas de Hipótesis

Una hipótesis es una afirmación o suposición respecto al valor de un parámetro poblacional. Como por ejemplo: el ingreso mensual promedio de los trabajadores es Bs. 615.000,°°. Otro podría ser, que el 90% de las formas fiscales son llenadas correctamente.

Todas estas hipótesis tienen algo en común, las poblaciones de interés son tan grandes que no es factible estudiar todos sus elementos. Una solución para estudiar la población entera, es tomar una muestra de la población de interés; y de esta manera se puede probar una aseveración para determinar si la evidencia soporta o no la afirmación.

> Para Mason, Lind y Marchal (2001:p.353), una prueba de hipótesis “Es un procedimiento estadístico que se basa en evidencias muestrales y en la teoría probabilística, y se emplea para determinar si la declaración planteada acerca del parámetro poblacional es racionable”

La prueba de hipótesis, viene a ser un supuesto que se plantea el investigador antes de iniciar una determinada investigación. Supuesto que al final de la investigación puede ser cierto y aprobarse, así como también puede ser falso y rechazarse.

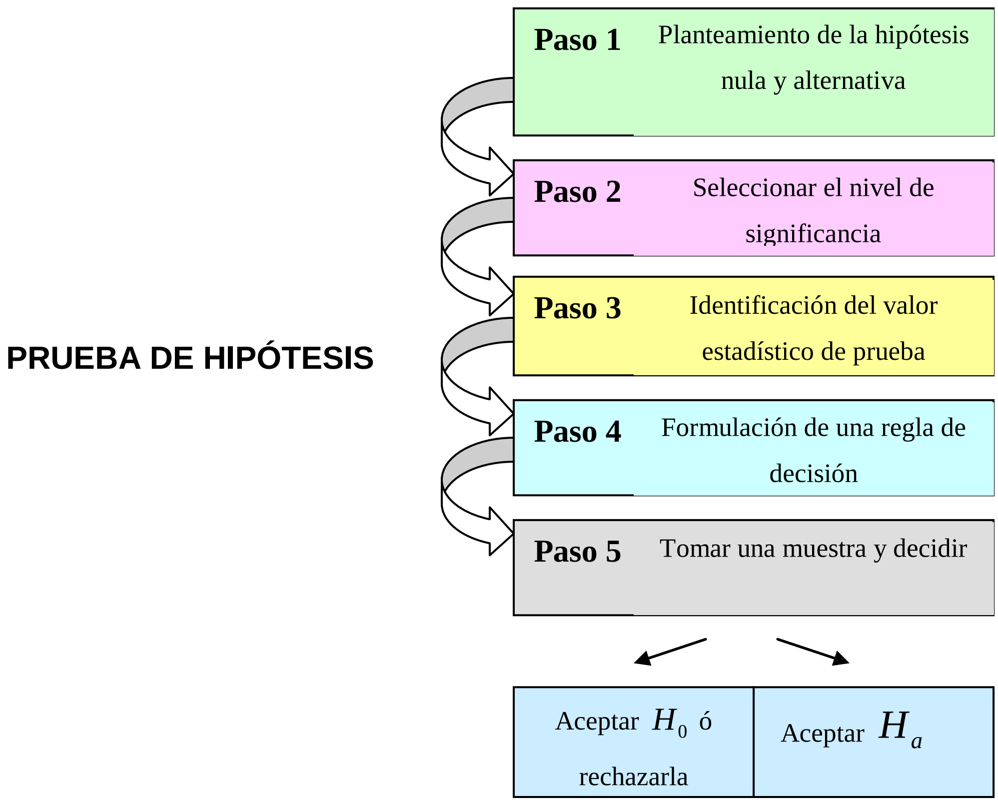
La finalidad de la prueba de hipótesis, no es cuestionar el valor calculado del estadístico muestral, sino que se encarga de hacer juicio con respeto a la diferencia entre el estadístico muestral y un valor planteado del parámetro.

1.4.1. Procedimientos para Probar una Hipótesis

Para probar una hipótesis, se sigue un procedimiento sistemático que consta de una serie de pasos, los cuales le van a permitir al investigador tomar una decisión.

A continuación se presenta una figura donde se muestran los pasos para realizar una prueba de hipótesis:

Figura 1-2 Pasos para Efectuar una Prueba de Hipótesis



Fuente: Las autoras

Explicando en profundidad cada paso, se tiene lo siguiente:

Paso 1: Plantear la hipótesis nula (H_0) y la hipótesis alternativa (H_a)

Toda investigación estadística requiere de la existencia de hipótesis o afirmaciones acerca de la población que se estudia.

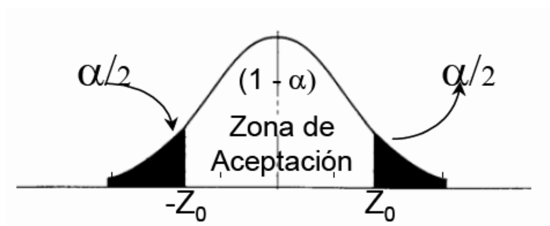
El primer paso, es establecer la hipótesis a ser probada. Ésta es llamada hipótesis nula (H_0), es una afirmación que no se rechaza tan solo que los datos de la muestra proporcionen evidencia convincente de que es falsa; indica que no existe diferencia significativa entre los resultados obtenidos esperados en la investigación determinada. Si se acepta la hipótesis nula, se dice que la evidencia no es suficiente para rechazarla pero tampoco se puede afirmar que es verdadera.

La hipótesis alternativa (H_a), es una afirmación que se acepta si se rechaza la hipótesis nula. Indica que existe diferencia significativa entre los resultados obtenidos y los resultados esperados en una determinada investigación. Se acepta si la evidencia proporcionada por la muestra es suficiente para afirmar que la hipótesis nula es falsa.

Paso 2: Seleccionar el nivel de significancia

Se determina el criterio de contraste, especificando el nivel de significancia, el tipo de distribución y los valores críticos, como se muestra en la siguiente figura:

Figura 1-3 Zonas críticas en la curva normal

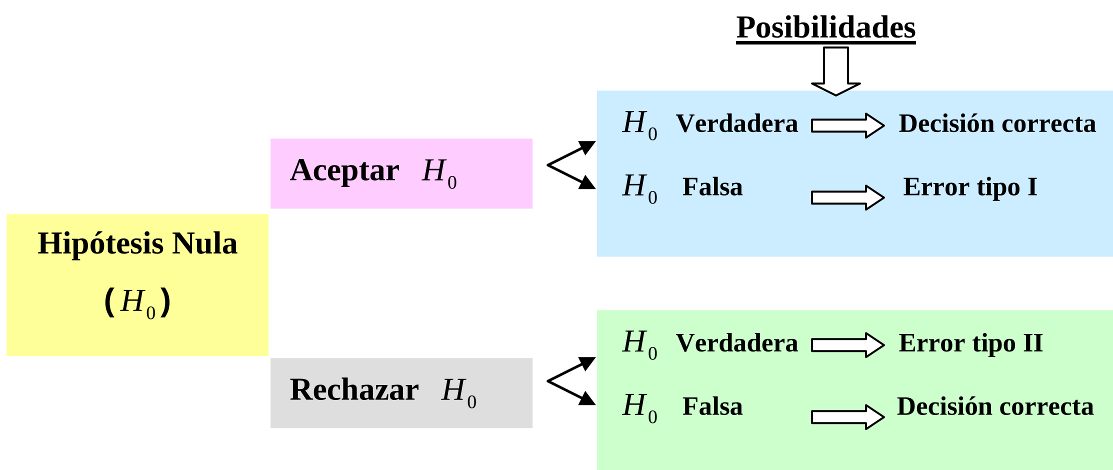


El nivel de significancia o de confianza, es la probabilidad de rechazar una hipótesis nula verdadera. Si la hipótesis planteada es verdadera, entonces, el nivel de confianza indicará la probabilidad de no aceptarla por estar en el área de aceptación. El nivel de confianza (1-α), indica la probabilidad de aceptar la hipótesis planteada cundo es verdadera en la población.

El tipo de distribución va a depender de la naturaleza de la hipótesis y del tamaño de la muestra. Si la hipótesis es relativa a las medias poblacionales y las muestras son grandes (n > 30) se usa la distribución normal. En caso contrario, de tratarse de una muestra pequeña ($n \leq 30$) se utiliza la distribución "t" de Student.

Toda decisión tomada por medio de una prueba de hipótesis, puede conllevar a un error. Existen cuatro posibilidades cuando se va a tomar una decisión respecto a una hipótesis, las cuales se observan en la figura siguiente:

Figura 1-4   Posibilidades que se Tienen al Tomar una Hipótesis



Fuente: Las autoras

El error tipo I, que se denota con la letra griega α, se da en el caso de que la hipótesis nula sea rechazada en vez de ser aceptada cuando es verdadera. Mientras que el error tipo II, denotado con la letra griega β, existe cuando se acepta la hipótesis nula en vez de ser rechazada, por ser ésta falsa.

Los valores críticos, separan a la región de no rechazo de la de rechazo. Son aquellos valores de la variable de la distribución que limitan el área crítica, que no es más que la parte de la curva que corresponde al nivel de significancia.

La prueba de hipótesis está conformada por dos regiones. Una región de rechazo, conocida como región crítica; y una región de no rechazo, conocida como región de aceptación.

Si el estadístico de prueba cae dentro de la región de aceptación, no se puede rechazar la hipótesis nula. Sin embargo, la región de rechazo es considerada como el conjunto de valores de la prueba de hipótesis que no tienen posibilidad de presentarse si la hipótesis nula es verdadera.

Por otra parte, es indispensable saber que existen dos tipos de prueba: la prueba de una cola y la prueba de dos colas.

La prueba unilateral, de una cola o de un extremo; es aquella en donde la hipótesis planteada indica una sola dirección, formulada con mayor ó igual que (≥) o menor ó igual que (≤).

Las siguientes figuras muestran, los diferentes tipos de prueba unilateral:

Figura 1-5 Prueba Unilateral Izquierda

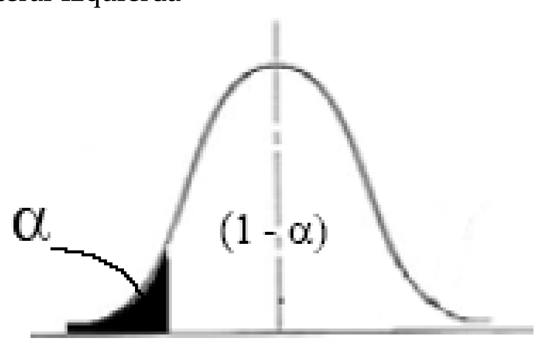
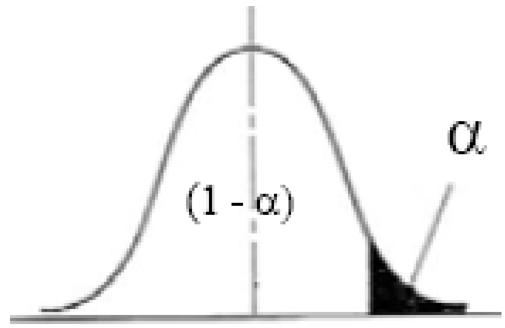


Figura 1-6   Prueba Unilateral Derecha



La prueba bilateral, de dos colas o de dos extremos; es aquella en donde la hipótesis planteada no indica dirección. H_0, se formula con la igualdad (=) y, H_a con diferencia (≠). La figura 1-3, muestra la prueba de dos colas o bilateral.

Paso 3. Calcular el valor estadístico de prueba

El estadístico de prueba, se refiere a un valor determinado por medio de la información de la muestra, el cual se debe comparar con el criterio de contraste, permitiendo así rechazar o aceptar la hipótesis.

Este estadístico, va a variar con la cantidad de muestras que se tomen. Si las muestras a utilizar son mayores a treinta, se utiliza el estadístico “z”, de ser el caso contrario, es utilizado el estadístico “t”.

Paso 4. Formular la regla de decisión

Es cuando se establecen las condiciones en las que se rechaza o se acepta la hipótesis nula. En donde, la región de rechazo, define la ubicación de todos los valores que son tan grandes o demasiado pequeños y, por lo tanto es muy remota la probabilidad que se de una hipótesis nula verdadera.

Paso 5. Tomar decisión

El último paso es tomar una decisión, para ver si se rechaza o se acepta la hipótesis nula. Si el estadístico de prueba queda dentro de la zona crítica, la hipótesis nula se tendrá que rechazar y se aceptará la hipótesis alternativa. Si dicho valor se encuentra fuera de la zona crítica, entonces, la hipótesis nula no deberá rechazarse.

1.4.2. Prueba de Hipótesis para la Media

Si se trata de una muestra grande, se usa el valor estadístico de prueba "z", es decir, se aplica la distribución normal, utilizando las siguientes fórmulas:

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Donde:

$\overline{X}$ = Media muestral

μ = Media poblacional

$\sigma_{\overline{X}}$ = Error estándar de la media

σ = Desviación típica poblacional

n = Tamaño de la muestra

Si se refiere a una muestra pequeña, se desconoce la desviación típica poblacional. En este caso se utiliza el valor estadístico "t" y para lo cual se debe conocer los grados de liberta (gl), además del nivel de significación.

$$t = \frac{\overline{X}_i - \mu}{S_{\overline{X}}}$$

$$S_{\overline{X}} = \frac{s}{\sqrt{n}}$$

gl = n -1

Donde:

$\overline{X}_i$ = Media aritmética muestral

μ = Media poblacional

$S_{\overline{X}}$ = Error estándar de media

n = Tamaño de la muestra

gl = Grados de libertad

1.4.3. Prueba de Hipótesis para una Proporción

Se utiliza para conocer el porcentaje de elementos de una población en una investigación. Sólo se utiliza para muestras grandes, por lo tanto se requiere de la distribución normal "z". Sus fórmulas son las siguientes:

$$z = \frac{\overline{p} - P}{\sigma_{\overline{p}}}$$

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\overline{p} = \frac{X}{n}$$

Donde:

$\overline{p}$ = Proporción muestral de evento

n = Tamaño de la muestra

P = Proporción poblacional

$\sigma_{\overline{p}}$ = Error estándar de la proporción poblacional

p = Proporción poblacional

X = Número de éxitos de la muestra

CAPÍTULO II
ANÁLISIS DE REGRESIÓN Y CORRELACIÓN

El análisis de regresión y correlación, son muy usados en la Investigación Científica, una herramienta muy útil cuando se trata de relacionar dos o más variables, relacionadas entre sí. La Correlación implica el grado de dependencia de una variable respecto a otra y la Regresión es otra técnica que ayuda en la investigación.

El análisis de regresión y correlación simple muestra la relación entre dos variables, la variable independiente y la variable dependiente. Al usar sólo una variable independiente estamos ignorando la relación que pudiera tener la variable dependiente con otras variables independientes. Al estudio de la influencia de dos o más variables independientes (x) sobre la variable dependiente (y) se le llama análisis de regresión y correlación múltiple.

Las técnicas de regresión, permiten hacer predicciones sobre los valores de cierta variable “y” *(dependiente)*, a partir de los de otra “x” *(independiente)*, entre las que intuimos que existe una relación**.**

2.1. Análisis de Regresión

La regresión, es un procedimiento por el medio del cual se trata de determinar si existe relación de dependencia o no entre dos o más variables.

En un Análisis de Regresión simple existe una variable dependiente (y) que puede ser el número de especies, la abundancia o la presencia-ausencia de una sóla especie y una variable explicativa o independiente (x).

El propósito es obtener una función sencilla de la variable explicativa (x), que sea capaz de describir lo más ajustadamente posible la variación de la variable dependiente (y). Como los valores observados de la variable dependiente difieren generalmente de los que predice la función, ésta posee un error. La función más eficaz es aquella que describe la variable dependiente con el menor error posible o, dicho en otras palabras, con la menor diferencia entre los valores observados y predichos. La diferencia entre los valores observados y predichos (el error de la función) se denomina variación residual o residuos. Para estimar los parámetros de la función se utiliza el ajuste por mínimos cuadrados. Es decir, se trata de encontrar la función en la cual la suma de los cuadrados de las diferencias entre los valores observados y esperados sea menor. Sin embargo, con este tipo de estrategia es necesario que los residuos o errores estén distribuidos normalmente y que varíen de modo similar a lo largo de todo el rango de valores de la variable dependiente. Estas suposiciones pueden comprobarse examinando la distribución de los residuos y su relación con la variable dependiente.

Cuando la variable dependiente es cuantitativa (por ejemplo, el número de especies) y la relación entre ambas variables sigue una línea recta, la función es del tipo $y_i{*} = a + bx_i$, en donde:

"*a*" es el intercepto o valor del punto de corte de la línea de regresión con el eje de la variable dependiente (una medida del número de especies existente cuando la variable ambiental tiene su mínimo valor).

"*b*" es la pendiente o coeficiente de regresión (la tasa de incremento del número de especies con cada unidad de la variable ambiental considerada).

Si la relación no es lineal pueden transformarse los valores de una o ambas variables para intentar linearizarla. Si no es posible convertir la relación en lineal, puede comprobarse el grado de ajuste de una función polinomial más compleja. La función polinomial más sencilla es la cuadrática ($y = a + bx + bx^2$) que describe una parábola, pero puede usarse una función cúbica u otra de un orden aun mayor capaz de conseguir un ajuste casi perfecto a los datos. Cuando la variable dependiente se expresa en datos cualitativos (presencia-ausencia de una especie) es aconsejable utilizar las regresiones logísticas ($Y= [\exp (a + bx)] / [1 + \exp (a + bx)]$).

De lo anteriormente dicho, se deduce, que el análisis de regresión simple, estudia el comportamiento de una variable dependiente (y) en función de una variable independiente (x) de manera tal que se pueda precisar la relación entre dichas variables, con el propósito de hacer pronósticos o predicciones.

La relación entre dos variables se puede determinar mediante un gráfico o diagrama de dispersión, un modelo de línea recta a través del método de los mínimos cuadrados y mediante un contraste de hipótesis.

Diagrama de Dispersión

Cuando los valores de la variable dependiente e independiente son llevados a un eje de coordenadas cartesianas, se puede apreciar un conjunto de puntos que muestran a simple vista la relación entre las variables, el cual recibe el nombre de diagrama de dispersión; es decir, no es más que graficar los pares de puntos de la variable, lo cual le permite al investigador, aproximar que tipo de relación hay entre las variables.

En otras palabras, el diagrama de dispersión, es un trazo en un sistema de coordenadas rectangulares o cartesianas de manera tal que se grafican los puntos apareados de las variables estudiadas.

Este gráfico permite precisar si existe relación o no entre las variables, y si la misma es positiva o negativa (ascendente o descendente).

Las figuras presentadas a continuación, muestran la relación existente entre las variables para el coeficiente de regresión:

Figura 2-1 Relación Positiva entre las Variables

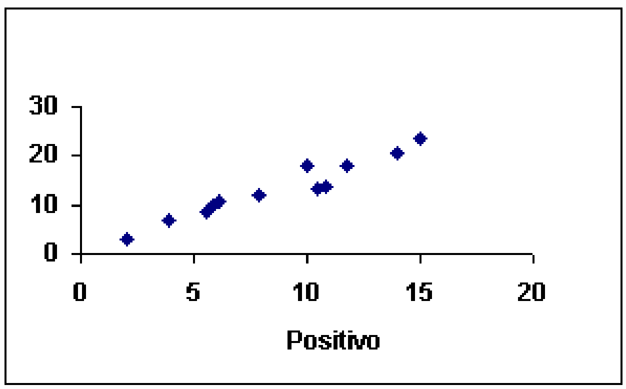


Figura 2-2 Relación Negativa entre las Variables

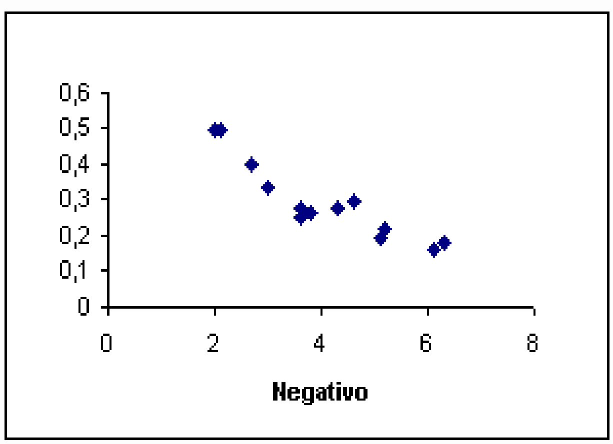
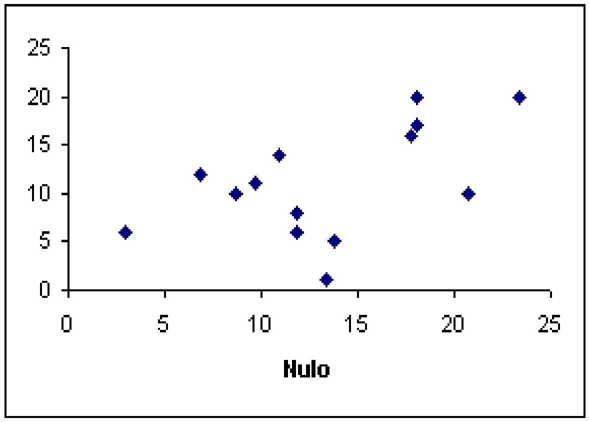


Figura 2-3 No Existe Relación entre las Variables



Ecuación de Regresión: es una ecuación que define la relación lineal entre dos variables.

$$y_i* = a + bx_i$$

Donde:

y_i* = es el valor pronosticado de la variable "y" para un valor seleccionado de "x".

a = es la ordenada de la intersección con el eje "y", o sea el valor estimado de "y" cuando x = 0. Es decir, corresponde al valor estimado de "y", donde la recta de regresión cruza el eje "y", cuando "x" es igual a cero.

b= es la pendiente de la recta, o sea, el cambio promedio en "y*"por unidad de cambio (incremento o decremento) en la variable independiente "x".

x = es cualquier valor seleccionado de la variable independiente.

Las fórmulas para b y a son:

Pendiente de la línea de regresión:

$$b = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum x_i^{\,2} - (\sum x_i)^2}$$

Intersección con el eje “x”:

$$a = \frac{\sum y_i - b\sum x_i}{n}$$

Donde:

x_i = es un valor de la variable independiente.

y_i = es un valor de la variable dependiente.

n = es el número de elementos en la muestra.

2.1.1. Principio de Mínimos Cuadrados

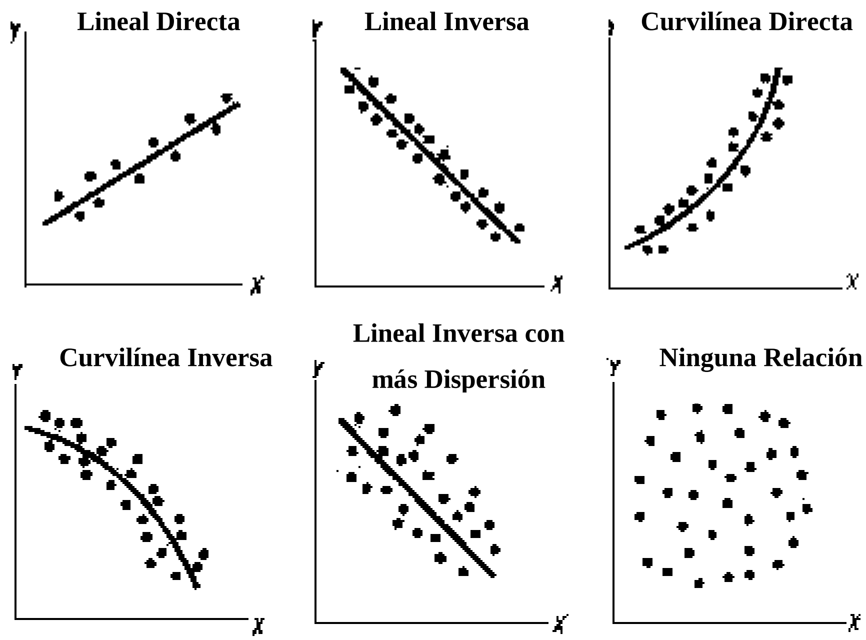
Es la técnica utilizada para lograr la ecuación de regresión, minimizando la suma de los cuadrados de las distancias verticales entre los valores verdaderos de "y" y los valores pronosticados "y".

Esta técnica nos permite seleccionar la línea recta que mejor se ajusta a los datos de manera tal que la suma de los cuadrados de la diferencia de cada valor observado y cada valor esperado (tendencia) sea la misma.

2.1.2. Trazo de la Línea de Regresión

El trazo de la línea recta, es una línea ajustada a un grupo de puntos para estimar la relación entre dos variables, como se puede observar en las figuras que se presentan a continuación:

Figura 2-4 Diferentes formas que toma el Trazo de la Línea de Regresión



Error Estándar de Estimación

Es la medida de confiabilidad de la ecuación de estimación, que indica la variabilidad de los puntos observados alrededor de la línea de regresión, esto, es hasta que punto los valores observados difieren de sus valores predichos sobre la línea de regresión.

En otras palabras, el error estándar de estimación, es la medida de la dispersión de los valores observados, con respecto a la línea de regresión.

Para comprender mejor la aplicación del error estándar de estimación en el análisis de regresión, deben enunciarse primero las consideraciones básicas con respectos a la regresión lineal y la correlación:

1.- Para cada valor de “x” existe un grupo de valores “y”, y estos valores “y” se distribuyen en forma normal.

2.- Las medias de estas distribuciones normales de valores “y”, se encuentran todas en la línea de regresión.

3.- Las desviaciones estándares de dichas distribuciones normales son iguales.

4.- Los valores de “y” son estadísticamente independientes. Esto significa que al seleccionar una muestra, los valores “y” seleccionados para un valor “x” específico no dependen de los valores “y” para cualquier otro valor “x”.

Fórmula para el cálculo del error estándar de estimación

- Método Directo

Si la muestra es pequeña:

$$S_{xy} = \sqrt{\frac{\sum (y_i - y_i*)^2}{n-2}}$$

Si la muestra es grande:

$$S_{xy} = \sqrt{\frac{\sum (y_i - y_i*)^2}{n}}$$

Donde:

S_{xy} = Error estándar

y_i = Valor de la variable dependiente

y_i* = Valor pronosticado de la variable dependiente

n = Número de la muestra

- Método Abreviado

Si la muestra es pequeña:

$$S_{xy} = \sqrt{\frac{\sum y_i^2 - a\sum y_i - b\sum y_i x_i}{n-2}}$$

Si la muestra es grande:

$$S_{xy} = \sqrt{\frac{\sum y_i^2 - a\sum y_i - b\sum y_i x_i}{n}}$$

Donde:

y_i = Valor de la variable dependiente

a = Valor estimado de “y” cuando “x” es igual a cero

b = Pendiente de la recta

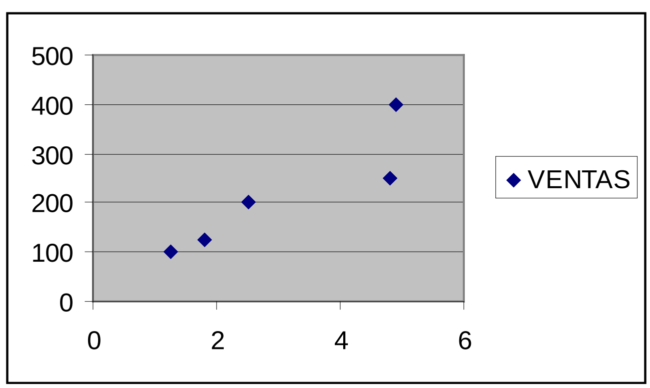
Para ilustrar el procedimiento del análisis de regresión lineal, se presenta a continuación un ejemplo:

Tabla 2-1 Análisis de Regresión Lineal

Ventas (miles Bs.) y_i	Publicidad (miles Bs.) x_i	$y_i x_i$	x_i^2	y_i^2	y_i*	$(y_i - y_i*)^2$
100	1,25	125	1,5625	10.000	101,4322692	2,0449
125	1,80	225	3,24	15.625	136,1335203	123,8769
200	2,5	500	6,25	40.000	180,2987489	388,09
250	4,8	1.200	23,04	62.500	325,4130716	5.686,6681
400	4,9	1.960	24,01	160.000	331,72239	4.656,6976
1.075	15,25	4.010	58,1025	288.125	1.075	10.857,3775

Primeramente realizamos el diagrama de dispersión:

Figura 2-5 Diagrama de Dispersión con respecto a las ventas y publicidad



$$y_i* = a + bx_i$$

n = 5

$$b = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum x_i^{\ 2} - (\sum x_i)^2}$$

$$b = \frac{5(4.010) - 1.075(15{,}25)}{5(58{,}1025) - (15{,}25)^2} = \frac{20.050 - 16.393{,}75}{290{,}5125 - 232{,}5625} = \frac{3.656{,}25}{57{,}95} = 63{,}09318378$$

$$a = \frac{\sum y_i - b\sum x_i}{n}$$

$$a = \frac{1.075 - 63{,}09318378(15{,}25)}{5} = \frac{1.075 - 962{,}1710526}{5} = \frac{112{,}8289474}{5} = 22{,}56578947$$

Sustituyendo la fórmula, se tiene:

$$y_i{*} = 22{,}56579947 + 63{,}09318378x_i$$

Este es un modelo matemático que permite predecir o estimar el valor de y_i (ventas), a través del valor de x_i (publicidad).

Supongamos que se quiere tener el monto de las ventas para cuando el monto invertido en publicidad sea igual a Bs.3.000.000,°°

$$y_i^{*}{}_{(3)} = 22{,}56578947 + 63{,}09318378(3) = 211{,}85$$

Se estima que las ventas serán de 211,85 miles de Bs. Para cuando la inversión de publicidad es de Bs.3.000.000,°°

Cálculo del error estándar de estimación:

Supongamos que la muestra es grande y el método aplicado es el directo.

$$S_{xy} = \sqrt{\frac{\sum (y_i - y_i*)^2}{n}}$$

Sustituyendo la fórmula:

$S_{xy} = \sqrt{\frac{10.857,38}{5}}$ = 46,60 miles de Bs.

Se estima, que el promedio de variaciones entre los valores observados y los estimados es de 46,60 miles de Bs., tanto por encima como por debajo de éstos.

2.2. Análisis de Regresión Múltiple

Se puede definir como el proceso, a través de cual, se utilizan varias variables para predecir otra. Para este análisis, la ecuación tiene varias variables independientes:

$$y* = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

Donde:

"$x_1, x_2, ...x_k$", son las variables independientes.

"a", es el punto donde la línea de regresión cruza el eje de las "y".

"$b_1, b_2, ...b_k$", son los coeficientes de regresión.

2.3. Análisis de Correlación

La correlación, es un conjunto de técnicas estadísticas utilizadas para medir la fuerza que existe entre la relación de dos variables, como por ejemplo: ¿Existe alguna relación entre los gastos de publicidad de una empresa y sus ventas?, ¿Hay relación entre la edad de los adultos y la estatura?, etc.

> Para Levin y Rubin (1996:p.680), "el análisis de correlación es una herramienta estadística que podemos usar para describir el grado hasta el cual una variable está linealmente relacionada con otra".

El análisis de correlación lineal o simple, permite medir y precisar la intensidad con que una variable se relaciona, es decir, su propósito primordial es encontrar que tan fuerte es la relación entre dos variables.

Usualmente, el análisis de correlación es, usado junto con el análisis de regresión para medir los cambios que explica la línea de regresión con respecto a la variable dependiente "y".

Pero, la correlación, también se puede utilizar para medir el grado de asociación entre dos variables; y para realizar dicho análisis, es necesario contar con varias medidas estadísticas como lo son: el diagrama de dispersión, el coeficiente de correlación, el coeficiente de determinación y el coeficiente de no determinación.

Diagrama de Dispersión

El diagrama de dispersión, no es más que una gráfica, en donde se describe en un plano cartesiano con una serie de puntos, la relación entre dos variables de interés. En donde, la variable dependiente se grafica sobre el eje vertical "y", y la variable independiente sobre el eje horizontal "x".

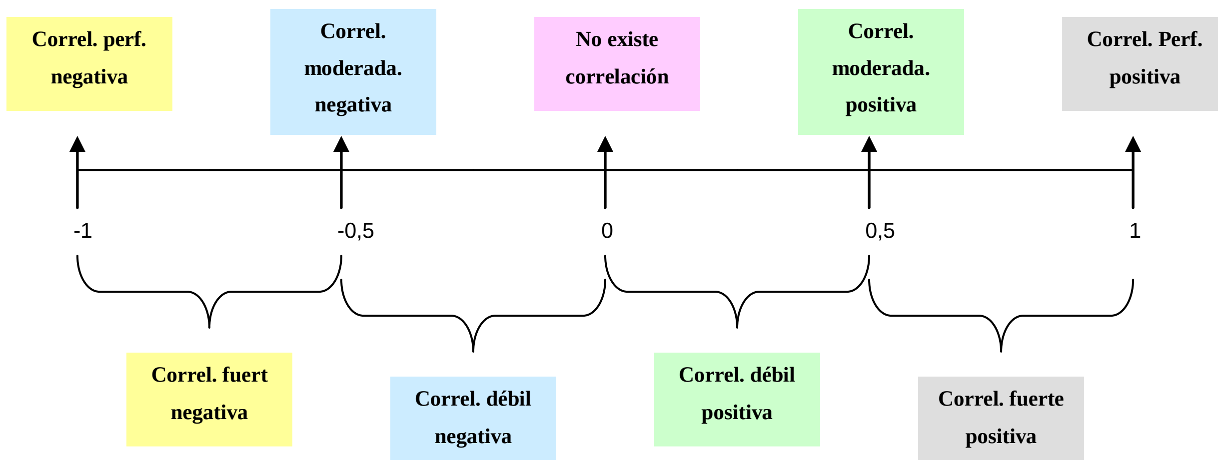
2.3.1. Coeficiente de Correlación

El coeficiente de correlación, llamado también coeficiente de correlación momento de Pearson, fue creado por Kart Pearson, aproximadamente en el año 1.900, el cual, es denotado con la letra “r”.

Dicho coeficiente, mide la intensidad de la relación entre dos variables y, tiene un campo de variabilidad o puede asumir valores entre -1 y +1, indicando éstos los puntos críticos de “r”.

La siguiente figura que se presenta a continuación, muestra la fuerza y dirección del coeficiente de correlación

Figura 2-6 Intensidad y Dirección del Coeficiente de Correlación



Fuente: Las autoras

Si r = -1, se dice que la correlación es perfecta negativa, es decir, entre las dos variables existe una relación matemática inversamente proporcional. En otras palabras, mientras una variable crece, la otra disminuye exactamente en la misma proporción. Véase figura 2-8

Ejemplo:

El salario real y la inflación.

Si r = 0, las variables no guardan relación y por lo tanto no existe correlación. Véase figura 2-11

Si r = 1, la correlación es perfecta positiva, implicando ésta una relación directamente proporcional, lo que quiere decir, que mientras una variable crece, la otra aumenta en la misma proporción. Véase figura 2-7

Ejemplo:

La inflación y el índice de precio.

A continuación, se muestran una serie de gráficas, en donde se presentan las diferentes formas del coeficiente de correlación:

Figura 2-7 Correlación Positiva

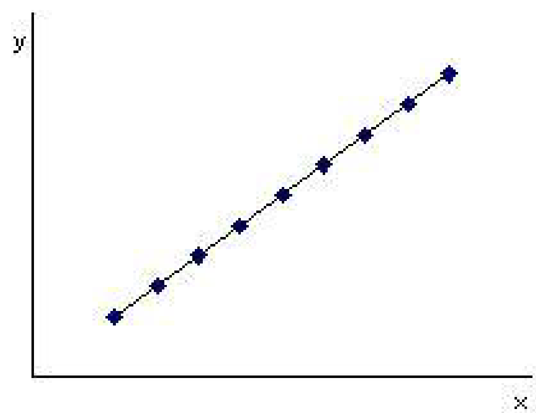


Figura 2-8   Correlación Negativa

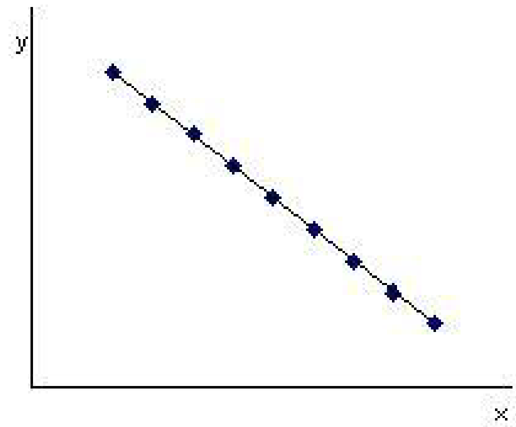


Figura 2-9   Correlación Negativa

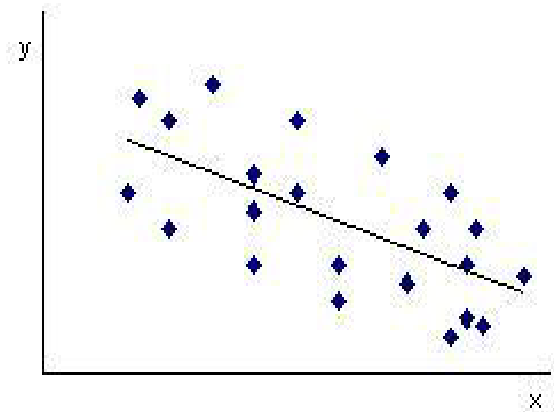


Figura 2-10   Correlación Positiva Fuerte

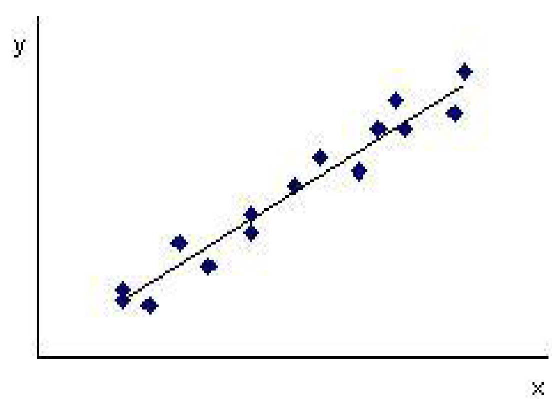
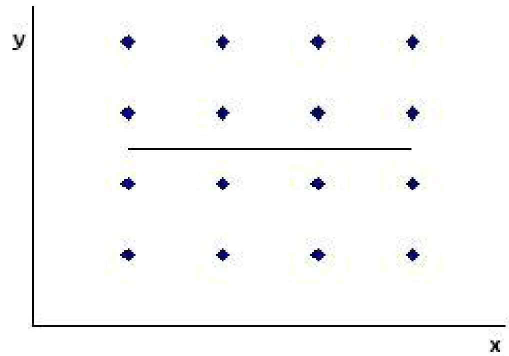


Figura 2-11 Sin Correlación



Fórmula del coeficiente de correlación:

$$r = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{\sqrt{\left[n\sum x_i^{\ 2} - \left(\sum x_i\right)^2\right]\left[n\sum y_i^{\ 2} - \left(\sum y_i\right)^2\right]}}$$

Donde:

n = Número de pares de observaciones

$\sum x$ = Suma de la variable independiente (x)

$\sum y$ = Suma de la variable dependiente (y)

$\sum x^2$ = Suma de los cuadrados de x

$\left(\sum x\right)^2$ = Cuadrado de la suma de x

$\sum y^2$ = Suma de los cuadrados de y

$\left(\sum y\right)^2$ = Cuadrado de la suma de y

2.3.2. Coeficiente de Determinación

Permite establecer si el modelo estudiado es confiable o no. Por medio del resultado de "r", es posible calcular el coeficiente de determinación, el cual mide la proporción en que la variable dependiente explica los cambios ocurridos por la variable independiente.

Su fórmula es la siguiente:

¡Error! Marcador no definido. $CD = r^2 * 100$

2.3.3. Coeficiente de no Determinación

El coeficiente de no determinación o indeterminación, es la proporción de la variación total en "y" que no esta explicada por la variación en "x". Este coeficiente se calcula con la siguiente fórmula:

$$CI = (1 - r^2.)*100$$

Los coeficientes de determinación y de no determinación pueden solamente ser positivos y pueden asumir valores entre 0 y 1 inclusive.

Siguiendo el mismo ejemplo del análisis de regresión lineal, ahora se ilustrará el análisis de correlación lineal.

Cálculo del coeficiente de correlación lineal de Pearson:

$$r = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{\sqrt{\left[n\sum x_i^{\ 2} - \left(\sum x_i\right)^2\right]\left[n\sum y_i^{\ 2} - \left(\sum y_i\right)^2\right]}}$$

Sustituyendo la fórmula con los valores calculados en la tabla, se tiene:

$$r = \frac{5*4.010 - 1.075*15,25}{\sqrt{\left[5*58,1025 - (15,25)^2\right]*\left[5*288.125 - (1.075)^2\right]}}$$

$$r = \frac{20.050 - 16.393,75}{\sqrt{\left[290,5125 - 232,5625\right]*\left[1.440.625 - 1.155.625\right]}}$$

$$r = \frac{3.656,25}{\sqrt{57,95*285.000}} = \frac{3.656,25}{4.063,957431} = 0,90$$

El valor del coeficiente obtenido, es positivo y se considera alto, ya que tiende a +1. En conclusión, la variable ventas, está fuertemente relacionada con la publicidad que se paga por el producto, mientras más publicidad haya respecto al producto, las ventas serán mayores.

Una vez calculado y analizado el coeficiente de correlación, se calcula el coeficiente de determinación:

$$CD = r^2 * 100$$

$$CD = (0{,}90)^2 * 100$$

$$CD = 81\%$$

Para hacer el análisis respectivo, calculemos ahora el coeficiente de no determinación o de indeterminación:

$$CI = (1 - r^2)*100$$

$$CI = (1 - 0{,}81)*100 = 0{,}19*100 = 19\%$$

Otra manera de calcular el coeficiente de indeterminación es:

$$CI = 100\% - CD$$

$$CI = 100\% - 81\%$$

$$CI = 19\%$$

Lo que significa, que el 81% de las variaciones de las ventas se debe a la publicidad, mientras que el 19% restante se deben a otras variables que han sido al azar o que no están dentro del modelo.

Ventas 100% → 81% publicidad

Ventas 100% → 19% otras variables o al azar

2.4. Análisis de Correlación Múltiple

Al igual que el análisis de correlación lineal o simple, éste utiliza las mismas medidas estadísticas pero en este caso múltiples, para describir la relación entre las variables (el diagrama de dispersión, el coeficiente de correlación, el coeficiente de determinación y el coeficiente de no determinación).

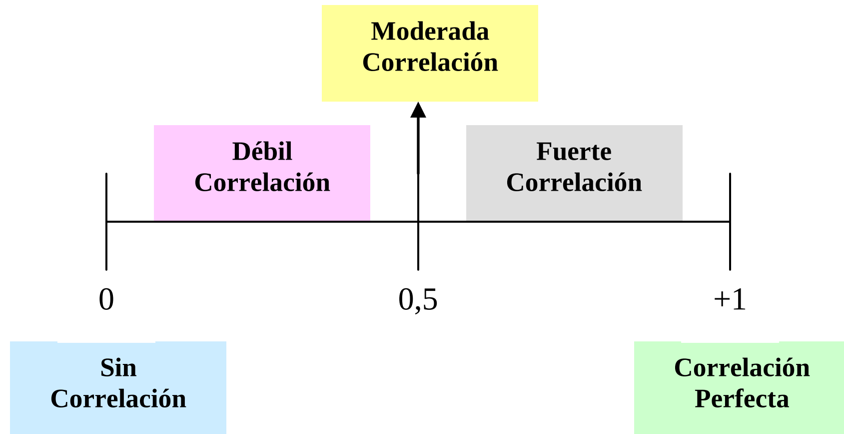
Coeficiente de Correlación Múltiple

Se puede definir como una medida de la fuerza de la asociación entre la variable dependiente y dos o más variables independientes. Este coeficiente está representado con la letra "R", y sólo puede tener valores comprendidos entre 0 y +1.

Si el coeficiente está cercano a +1, quiere decir que hay una fuerte correlación entre la variable dependiente y las variables independientes. En caso de estar cercano a 0, indica que existe una débil correlación.

La siguiente figura, es para visualizar los diferentes valores que toma el coeficiente de correlación múltiple:

Figura 2-12 Valores del Coeficiente de Correlación Múltiple



Fuente: Las autoras

Coeficiente de Determinación Múltiple

Se simboliza con "R^2" y, representa la proporción de la variación total en la variable dependiente "y", que es explicada por las variables independientes

Coeficiente de no Determinación Múltiple

Este coeficiente, se encarga de medir la proporción existente entre la variación en la variable dependiente, la cual no es explicada por las variables independientes.

CAPÍTULO III

ANÁLISIS DE VARIANZA Y PRUEBAS NO PARAMÉTRICAS

El análisis de varianza, es utilizado para comparar si los valores de un grupo de datos numéricos, son significativamente distintos a los valores de otro o más conjunto de datos. El procedimiento para comparar estos valores, está basado en la varianza global observada en los grupos de datos numéricos a comparar.

3.1. Análisis de Varianza o Anova

> Para Levin y Rubin (1996:p.631), el análisis de varianza, es una "técnica estadística utilizada para probar la equidad de tres o más medias de muestra y, de este modo, hacer inferencias sobre si las muestras provienen de poblaciones que tienen la misma media".

El análisis de varianza, es una herramienta que se aplica para probar simultáneamente si las medias de varias poblaciones son iguales. Dicho análisis tiene una serie de pasos a seguir:

1. Se debe determinar una estimación de la varianza de la población, a partir de la varianza entre las medias muestrales.

2. Se determina una segunda estimación de la varianza de la población desde la varianza dentro de de las muestras.

3. Se comparan las dos estimaciones, a través del cociente y; si un valor es aproximadamente igual, se acepta la hipótesis nula.

Ahora, se hablará un poco de los pasos antes mencionados:

Para obtener la primera estimación de la varianza de la población o varianza entre columnas, se utilizan las siguientes fórmulas:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Donde:

s^2 = Varianza de la muestra

n = Tamaño de la muestra

$\bar{x}$ = Media de muestra

Luego se busca la varianza entre las medias de muestras

$$s_{\bar{x}}^{\ 2} = \frac{\sum \left(x - \bar{\bar{x}} \right)^2}{k - 1}$$

Donde:

$s_{\bar{x}}^{\ 2}$ = Varianza entre las medias de muestras

k = Número de muestras

$\bar{\bar{x}}$ = Gran media

Para calcular el error estándar de la media, que no es más que la desviación estándar de todas las medias de muestras posibles a partir de un tamaño de muestra dado. Se usa la siguiente fórmula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Donde:

$\sigma_{\bar{x}}$ = Error estándar de la media

σ = Desviación estándar

$\sqrt{n}$ = Raíz cuadrada del tamaño de la muestra

Se puede simplificar la fórmula, multiplicándose de forma cruzada los término, y después elevar ambos lados al cuadrado, con la finalidad de cambiar la desviación estándar de la población (σ), en la varianza de la población (σ^2).

$$\sigma^2 = \sigma_{\bar{x}}{}^2 * n$$

Donde:

σ^2 = Varianza de la población

$\sigma_{\bar{x}}{}^2$ = Error estándar elevado al cuadrado (igual a la varianza entre las medias de muestras)

Entonces, se tiene que la fórmula general de la primera estimación de la varianza de la población es:

$$\hat{\sigma}^2 = \frac{\sum n_j \left(\bar{x}_j - \bar{\bar{x}} \right)^2}{k-1}$$

Donde:

$\hat{\sigma}^2$ = Primera estimación de la varianza de la población, basada en la varianza entre columnas

n_j = Tamaño de la j-ésima muestra

$\bar{\bar{x}}$ = Gran media

k = Número de muestras

Para buscar la segunda estimación de la varianza de la población o varianza dentro de columnas, se emplea la fórmula siguiente:

$$\hat{\sigma}^2 = \sum \left(\frac{n_j - 1}{n_t - k} \right) * s_j^{\ 2}$$

Donde:

$\hat{\sigma}^2$ = Segunda estimación de la varianza de la población, basada en la varianza dentro de columnas

n_j = Tamaño de le j-ésima muestra

$s_j^{\ 2}$ = Varianza de muestra de la j-ësima muestra

k = Número de muestras

$n_t = \sum n_j$ = Tamaño de muestra total

El último paso es comparar las dos estimaciones. Para esto, se calcula el cociente “F”.

> Según Levin y Rubin (1996:p.631), "es un cociente utilizado en el análisis de varianza, entre otras pruebas, para comparar la magnitud de dos estimaciones de la varianza de la población para determinar si las estimaciones son aproximadamente iguales".

Este cociente se calcula de la siguiente manera:

$$F = \text{Varianza entre columnas / varianza dentro de columnas}$$

Si la hipótesis nula es verdadera, el denominador y el numerador deben ser aproximadamente iguales. Mientras más cercano a "1" esté el cociente "F", se está más inclinado en aceptar la hipótesis nula. Si el cociente "F" se hace más grande, la hipótesis nula se inclina más al rechazo y se acepta la hipótesis alternativa.

Cuando las poblaciones no son las mismas, la varianza entre columnas tiende a ser mayor que la varianza dentro de columnas, y por ende, el valor del cociente "F" tenderá a ser grande, lo que implica que la hipótesis nula debe ser rechazada.

3.1.1. Distribución F

La distribución F o de Fisher, es utilizada como valor estadístico de prueba para problemas de Análisis de Varianza o ANOVA. Su nombre se debe a Sir Ronald Fisher, quien fue uno de los fundadores de la Ciencia Estadística moderna.

Esta distribución, es usada en los casos en donde se quiere probar si dos muestras se derivan de poblaciones con varianzas iguales, así como cuando se quieren comparar simultáneamente varias medias poblacionales.

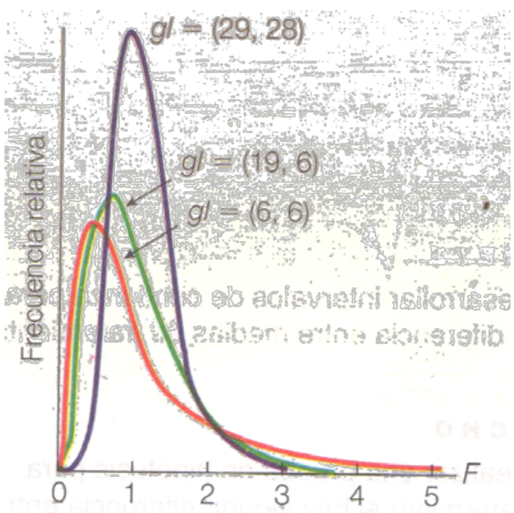
> Según Levin y Rubin (1996:p.631), definen la distribución de F como una "familia de distribuciones diferenciadas por dos parámetros (grados de libertad del numerador, grados de libertad del

denominador), utilizada principalmente para probar hipótesis con respecto a varianzas”.

Características de la distribución F:

- Existe una “familia” de distribuciones F. Cada distribución F, posee una pareja de grados de libertad, tanto para el numerador del cociente “F” como para el denominador. En donde, la forma de las curvas cambia a medida que varían los grados de libertad, como se muestra en la figura siguiente:

Figura 3-1 Grados de Libertad en la Distribución F



Para hallar los grados de libertad del numerador (varianza entre columnas), se realiza de la siguiente manera:

$$\textit{núm de gl en el numerador del cociente } "F" = (\textit{núm. de muestra} - 1)$$

En cambio, los grados de libertad del denominador, se obtienen de la siguiente forma:

Núm. De gl en el denominador del cociente "F" = $\sum\left(n_j - 1\right) = n_t - k$

Donde:

n_j = Tamaño de la j-ésima muestra

k = Número de muestras

$n_t = \sum n_j$ = Tamaño de muestra total

- El valor de F no debe ser negativo y se trata de una distribución continua.
- La distribución F tiene sesgo positivo.
- A medida que aumenta el valor de F, la curva se aproxima al eje "x", sin nunca tocarlo.

3.1.2. Comparación de dos Varianzas Poblacionales

En este caso, la distribución F, es utilizada para probar la hipótesis de que la varianza de una población normal, es igual a la variación de otra población normal, por lo cual, es importante comparar dos poblaciones para ver si una varía más que la otra en algunas medidas específicas.

Si una población posee más variación que otra, primeramente se tiene que realizar el planteamiento de hipótesis. En donde, la hipótesis nula, es que las dos poblaciones tienen la misma varianza; y la hipótesis alternativa, es que una población tiene mayor varianza que la otra, es decir, aquí las variaciones difieren.

Esta prueba de hipótesis se denota:

$$H_0 : \sigma_1^{\;2} = \sigma_2^{\;2}$$

$$H_0 : \sigma_1^{\;2} \neq \sigma_2^{\;2}$$

Para realizar las pruebas de hipótesis, es necesario obtener una muestra aleatoria de "n_1" (observaciones a partir de una población), una muestra de "n_2" (observaciones de una segunda población), y se calculan las variaciones muestrales o valor estadístico de prueba. Si la hipótesis nula es verdadera, entonces, el valor estadístico de prueba sigue la distribución F con (n_1 -1) y (n_2 -1) grados de libertad.

3.2. Pruebas no Paramétricas

Debido a que en la práctica todas las poblaciones no pueden tomarse como normales, ya que se generan situaciones en donde no es posible formular una hipótesis segura sobre el valor de un parámetro o la forma de la distribución poblacional, para las que las pruebas paramétricas de "z" y "t" no son adecuadas; surgen las pruebas no paramétricas, las cuales no dependen de un solo tipo de distribución o de unos valores específicos de los parámetros.

> Según Webster, Allen (1996:p.836), las pruebas no paramétricas, "son procedimientos estadísticos que se pueden utilizar para contrastar hipótesis cuando no es posible fijar ningún supuesto sobre parámetros o distribuciones poblacionales".

Existen diferentes tipos de pruebas no paramétricas, que se pueden utilizar para una necesidad determinada. Entre las cuales, se encuentra primordialmente ji-cuadrado, dividiéndose ésta en: prueba de bondad de ajuste y prueba de independencia o tablas de contingencias.

Cabe destacar, que existen otras pruebas no paramétricas como lo son: la prueba de los signos, la prueba de rachas, la prueba de Mann-Whitney, la correlación de rangos de Spearman y la prueba de Kruskal; las cuales no se van a abordar en el capítulo, porque se va hacer énfasis en la prueba de ji-cuadrado.

3.2.1. Distribución de ji-cuadrado

La distribución de ji-cuadrado, puede definirse como una familia de distribuciones de probabilidad que se diferencian por sus grados de libertad. Se encarga de probar un cierto número de hipótesis diferentes, referente de varianzas, porciones y bondad de ajuste de distribuciones.

Esta distribución abarca la prueba de bondad de ajuste, la cual comprende una serie de pruebas: distribución uniforme, estructura específica, distribución de Poisson, distribución binomial y distribución normal; y la prueba de independencia.

3.2.1.1. Prueba de Bondad de Ajuste

> Para Levin y Rubin (1996:p.632), la prueba de bondad de ajuste, es una “prueba estadística para determinar si existe una diferencia significativa entre una distribución de frecuencias observadas y una distribución de probabilidad teórica hipotetizada para describir la distribución observada”.

Esta prueba, mide el grado en que los datos muestrales observados cumplen una distribución hipotética determinada, y si el grado de cumplimiento es razonable, entonces se puede deducir que la distribución hipotética existe.

La prueba de ji-cuadrado como prueba de bondad de ajuste, es utilizada para decidir si una distribución de probabilidad en especial (distribución binomial, distribución de Poisson o la distribución normal), es la distribución apropiada.

Una vez elegida la prueba, se realiza el planteamiento de hipótesis:

H_0 : La muestra procede de la población especificada

H_a : La muestra no procede de la población especificada

Pasos para el cálculo de la prueba de bondad de ajuste:

1. Establecer la hipótesis nula y la hipótesis alternativa.
2. Calcular las frecuencias esperadas.
3. Determinar el valor crítico de ji-cuadrado.
4. Hallar el valor calculado de ji-cuadrado.
5. Comparar el valor crítico con el calculado.
6. Conclusiones.

Fórmulas para calcular la prueba de bondad de ajuste:

$$x^2 = \sum_{f_e}^{k} \frac{(f_o - f_e)^2}{f_e}$$

Donde:

f_0 = Frecuencia observada en la muestra

f_e = Frecuencia esperada

k = Número de categorías o clases

$$gl = k\text{-}1\text{-}c$$

Donde:

gl = Grados de libertad
k = Número de categorías
c = Número de parámetros a estimar

Ejemplo:

Supongamos, que una empresa de servicios varios, quiere contratar a un grupo de personas para que laboren en las diferentes actividades que ésta realiza. Para ello, utiliza la siguiente estructura, la cual va acorde con la edad:

30% entre 18-25 años
50% entre 25-35 años
20% de 35 años en adelante

La empresa toma una muestra de 70 personas, distribuidas así:

20 (entre 18-25 años)
40 (entre 25-35 años)
10 (de 35 años en adelante)

El nivel de significancia a utilizar es de 10%, si se mantienen las reglas establecidas.

Primeramente, se tiene que establecer el planteamiento de hipótesis:

H_0 = las reglas establecidas por la empresa se mantienen igual

H_a = las reglas establecidas por la empresa son diferentes

Luego, se determinan las frecuencias esperadas:

Tabla 3-1 Frecuencias Esperadas

Edades	**Frecuencias observadas (f_0)**	**Frecuencias esperadas (f_e)**
Entre 18-25 años	20	21
Entre 25-35 años	40	35
35 años en adelante	10	14
Totales	**70**	**70**

- 70 * 0,30 = 21
- 70 * 0,50 = 35
- 70 * 0,20 = 14

Ahora, se calcula el valor crítico de ji-cuadrado para un nivel de significancia del 10%:

gl = k-1-c = 3-1-0 = 2

k = 3 porque hay divisiones de las edades para seleccionar el grupos de personas.

c = 0 porque no hubo ninguna estimación de parámetro poblacional.

Buscando con los grados de libertad el valor de ji-cuadrado en la tabla, para un nivel de significancia del 10%, se tiene:

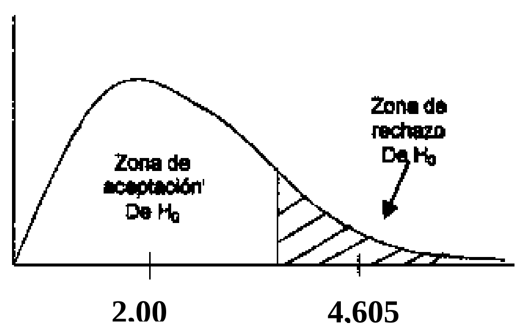
$$x^2 = 4{,}605$$

Seguidamente, se determina el valor calculado de ji-cuadrado, para luego comparar y sacar conclusiones:

$$x^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = \frac{(20-21)^2}{21} + \frac{(40-35)^2}{21} + \frac{(10-14)^2}{21}$$

$$x^2 = \frac{1}{21} + \frac{25}{21} + \frac{16}{21} = \frac{42}{21} = 2$$

Figura 3-2 Representación Gráfica de ji-cuadrado como Prueba de Bondad de Ajuste



Se puede observar en la gráfica, que el valor de ji-cuadrado, está dentro de la zona de rechazo de H_0. Por lo tanto, se puede concluir diciendo que se acepta la hipótesis nula (H_0) y se rechaza la hipótesis alternativa (H_a). Y se puede afirmar con un nivel de confianza del 90%, que las reglas establecidas por la empresa se mantienen igual.

3.2.1.2. Prueba de Independencia o Tablas de Contingencias

> Levin y Rubin (1996:p.632), definen la prueba de independencia como una "prueba estadística de porciones de frecuencias que se utiliza para determinar si la pertenencia de una variable a categorías es diferente como función de la pertenencia a la categoría de una segunda variable".

La prueba de independencia, permite determinar a través de unas tablas de contingencias, si dos variables categóricas se relacionan entre sí. Dichas tablas, poseen una serie de renglones y columnas; cada renglón corresponde a un nivel de una variable y cada columna a un nivel de otra variable. Las entradas del cuerpo de las tablas, vienen a ser las frecuencias, con que cada combinación de variable se presenta.

Esta prueba, es utilizada frecuentemente para el análisis de aspectos importantes de los datos investigados, los cuales consisten en medidas muestrales sobre dos variables categóricas. Una vez examinado los datos de la muestra, se procede al planteamiento de hipótesis, en donde la hipótesis nula y la alternativa permanecen constantes para todos los casos:

H_0: Las variables en las filas y las columnas son independientes

H_a: Las variables en las filas y columnas son dependientes

Para probar la hipótesis nula, se debe comparar las frecuencias que fueron observadas con las frecuencias que se esperarían en caso de que la hipótesis nula fuera verdadera. Si los grupos de frecuencias observadas y esperadas son casi iguales, de manera intuitiva se puede razonar que la hipótesis nula se aceptará. De lo contrario, si existe una diferencia grande entre las frecuencias observadas y las esperadas, intuitivamente la hipótesis nula se puede rechazar y por ende, se llega a la conclusión de que hay diferencias significativas.

La prueba de tablas de contingencias o de independencia, se calcula de la siguiente manera:

$$x^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Donde:

x^2 = Ji-cuadrado

f_0 = Una frecuencia observada

f_e = Una frecuencia esperada

Para hallar la frecuencia esperada, se utiliza la siguiente fórmula:

$$f_e = \frac{(total\ fila) * (total\ \ columna)}{n}$$

n = Tamaño de la muestra

Además, se debe determinar el número de grados de libertad en la tabla de contingencias, de la siguiente manera:

$$gl = (r-1)*(c-1)$$

Donde:

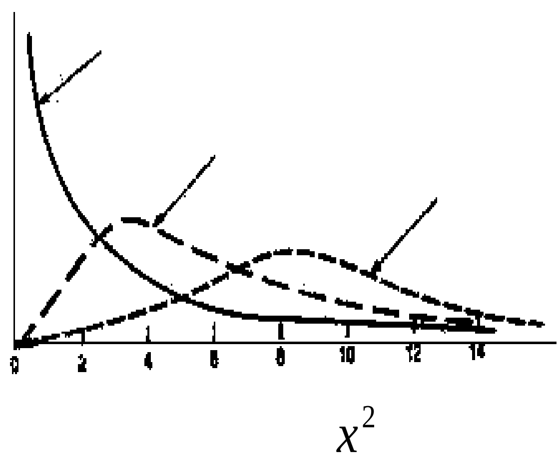
gl = grados de libertad

r = número de filas o de renglones

c = número de columnas

En la siguiente figura, se muestra una distribución ji-cuadrada, para cada grado de libertad (1, 5, 10):

Figura 3-4 Distribución ji-cuadrado para Diferentes Grados de Libertad



Se puede observar en la figura 3-4, que para un número muy pequeño de grados de libertad, la distribución ji-cuadrado está seriamente sesgada hacia la derecha; y mientras va aumentando el número de grados de libertad, rápidamente la curva se va haciendo cada vez más simétrica, hasta que el número de grados de libertad se torna bastante grande, permitiendo que la distribución se pueda aproximar con la normal.

De todo lo anteriormente dicho, se puede resumir que para el cálculo de la prueba de independencia o tablas de contingencias se utilizan los siguientes pasos:

1. Plantear de la hipótesis nula y la hipótesis alternativa.
2. Calcular las frecuencias esperadas, correspondientes a cada frecuencia observada.
3. Calcular el crítico de ji-cuadrado.
4. Determinar el valor calculado de ji-cuadrado.
5. Comparar el valor crítico con el valor esperado.
6. Conclusión.

Ejemplo:

Supongamos, que una muestra tomada con relación al consumo de un determinado producto, quedó estructurada de la siguiente forma:

Tabla 3-2 Tabla de contingencia 2x2. Frecuencias Observadas

Consumo del producto	**Sexo**		**totales**
	Mujeres	**Hombres**	
Sí	40	20	60
No	15	25	40
Totales	**55**	**45**	**100**

Primeramente, se realiza el planteamiento de hipótesis:

H_0 = El consumo del producto es independiente del sexo

H_a = El consumo del producto es dependiente del sexo

Luego, se procede a calcular las frecuencias esperadas, correspondiente a cada frecuencia esperada:

Tabla 3-3 Tabla de contingencia 2x2. Frecuencias Observadas y Esperadas

Consumo del producto	**Sexo**				**Totales**
	Mujeres		**Hombres**		
	f_0	f_e	f_0	f_e	
Sí	40	33	20	27	60
No	15	22	25	18	40
Totales	**55**		**45**		**100**

$$f_e = \frac{(total\ fila)*(total\ \ columna)}{n}$$

Mujeres

$$f_e = \frac{(60)*(55)}{100} = \frac{3.300}{100} = 33$$

$$f_e = \frac{(40)*(55)}{100} = \frac{2.200}{100} = 22$$

Hombres

$$f_e = \frac{(60)*(45)}{100} = \frac{2.700}{100} = 27$$

$$f_e = \frac{(40)*(45)}{100} = \frac{1.800}{100} = 18$$

Ahora, se calcula el valor crítico de ji-cuadrado, determinando los grados de libertad:

gl = (r-1)*(c-1)

Donde:

r = Número de filas = 2
c = Número de columnas = 2

gl = (2-1)*(2-1) = 1

Entonces, se busca en la tabla el valor que le corresponde al nivel de significación de 5%:

$x^2 = 3{,}841$

Se procede a determinar el valor calculado de ji-cuadrado:

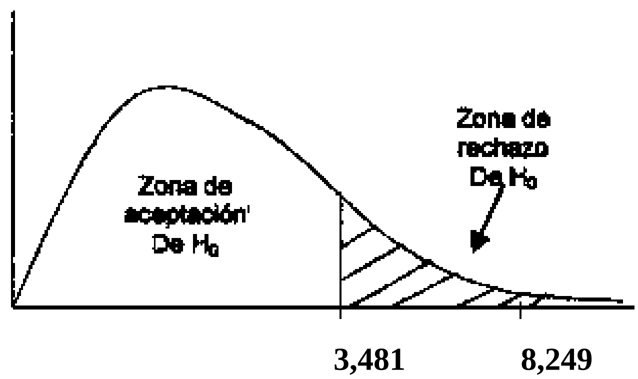
$$x^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Tabla 3-4 Valor Calculado de Ji-cuadrado

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
40	33	7	49	1,485
20	27	-7	49	1,815
15	22	-7	49	2,227
25	18	7	49	2,722
			x^2 **=**	**8,249**

Por último, se compara el valor crítico de ji-cuadrado con el valor calculado de ji-cuadrado, para así dar las conclusiones:

Figura 3-5 Representación Gráfica de Ji-cuadrado como Prueba de Independencia



Como el valor calculado está dentro de la zona de rechazo de la hipótesis nula, se concluye, que se puede afirmar que la demanda del producto depende del sexo del consumidor.

CONCLUSIONES

La Estadística Inferencial, es utilizada en aquellos casos en donde el investigador requiere ir más allá de una simple recolección de datos, lo cual le permite realizar un análisis profundo, por medio de una parte tomada de la población en estudio, conocida ésta como muestra.

La Estadística Inferencial, le permite a la Gerencia tomar decisiones ciertas, acerca de los acontecimientos futuros. Y de esta manera, obtener una adecuada planeación y control para el mejor funcionamiento de la empresa, en los respectivos departamentos que ésta conforma.

La Distribución Muestral, es de gran importancia, porque a través de la muestra tomada de la población, se puede por ejemplo, conocer la opinión que tienen los consumidores respecto a un producto determinado, permitiendo así al investigador evaluar la calidad del producto.

La Inferencia Estadística, genera una serie de problemas, basados fundamentalmente, en las estimaciones y las pruebas de hipótesis. En donde, las estimaciones, son utilizadas para determinar parámetros de una población por medio de sus valores estadísticos; y las pruebas de hipótesis, no son más que supuestos que se hace el investigador antes de empezar la investigación, para finalmente poder sacar conclusiones, aceptando o rechazando las hipótesis planteadas.

El Análisis de Regresión y Correlación, son muy usados, siendo éstos una herramienta de gran utilidad para la investigación, ya que se pueden relacionar dos o más variables, conocidas éstas como variable dependiente y variable independiente.

Existen dos pruebas estadísticas de gran importancia, las cuales están elaboradas para datos cualitativos y categóricos. Estas pruebas, abarcan a la conocida Distribución de Ji-cuadrado, entre las que se encuentran, la Prueba de Bondad de Ajuste y la Prueba de Independencia o Tablas de Contingencias.

RECOMENDACIONES

Como ya sabemos, la Estadística es utilizada en toda organización, bien sea desde una simple recolección de datos, hasta un análisis más profundo que le permita al investigador llegar a conclusiones que predigan el futuro de la empresa.

Como la Distribución Muestral se considera la parte más importante de la Estadística Inferencial, se recomienda a las organizaciones para realizar una determinada investigación hacer uso de ésta, ya que tomando una muestra de toda la población, le permite reducir en gran parte los costos que le generarían en caso de que estudiaran a toda la población. Además, por medio de la muestra, se puede reducir el tiempo, en aquellos casos en donde se tardaría mucho contactar a toda la población y por ende, el trabajo se haría más largo, lo que implicaría también más gastos para la empresa.

Una hipótesis, es una afirmación respecto al valor de un parámetro poblacional. Si el investigador quiere efectuar una Prueba de Hipótesis, se la recomienda que siga los diferentes pasos, los cuales lo van a conllevar a aceptar o rechazar las hipótesis, y de esta forma, a tomar una decisión cierta respecto a lo que acontece en la empresa.

Ya que el mundo empresarial, está íntimamente enlazado con todo lo referente a los datos numéricos, se le recomienda al investigador, examinar las relaciones entre las diferentes variables, para ver si éstas se relacionan o no.

BIBLIOGRAFÍA

Textos

Arias, F. (2004). El Proyecto de Investigación. Cuarta Edición. Episteme. Caracas.

Berenson, Mark y Levine, David. (1.982). Estadística para Administración y Economía. Primera Edición. Interamericana. México.

Berenson, Mark y Levine, David. (1.996). Estadística Básica en Administración. Segunda Edición. Prentice Hall Hispanoaméricana. México.

Chao, Lincoln. (1.993). Estadística para las Ciencias Administrativas. Segunda Edición. Mac. Graw-Hill. Colombia.

Levin, Richard y Rubin, David (1.996). Estadística para Administradores. Sexta Edición. Prentice Hall. México.

Mason, Lind y Marchal. (2.001). Estadística para Administración y Economía. Décima Edición. Alfaomega. Colombia.

Webster, Allen. (1.996). Estadística Aplicada a la Empresa y a la Economía. Segunda Edición. España.

Citas de Internet

Molina, Gonzalo. “Estadística”. http://html.rincondelvago.com/nociones-basicas-de-estadística.html (18-11-07).

Velasco,R.“EstadísticaInferencial”.<http://www.universidadabierta.edu.mx/SerEst/Apuntes/VelascoRoberto_EstadisticaInferencial.html> (15-11-07).

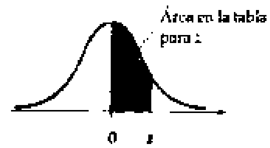
Wikipedia. “Probabilidad”. http://es.Wikipedia.org/wiki/Probabilidad (15-11-07).

ANEXOS

Tabla de números aleatorios

75421	11182	31304	08036	86922	77941	88944	30226	60766	90951
06692	19591	14171	04356	08744	46546	99184	97684	43285	86345
66065	12379	70386	09035	90126	74677	39885	84335	09442	21772
01098	06343	88773	94702	07203	60936	84445	12423	64860	99694
93526	56837	42025	45578	95193	97695	53146	51370	79923	83145
85129	31088	36253	40011	62078	72245	58783	47555	55681	45450
74312	81501	94303	30800	60660	69979	57625	00050	69795	15120
67348	11345	13361	40571	75687	78418	42407	87830	98069	98605
29241	77892	67728	60876	53046	75840	18933	18108	73509	76958
04366	94984	85131	22993	17240	63185	54786	31607	80705	61581
54205	61584	99698	74013	88263	96563	18003	77390	05762	40975
52801	44366	19748	74219	20982	91400	50685	56541	68392	96624
02573	59494	26362	40769	39340	19677	16923	04761	65952	03630
15096	32426	64984	99029	58073	28814	44849	39871	00825	29964
26032	33340	54573	55786	75383	14546	37499	43894	86358	19706
41349	18921	50836	65061	79821	38319	33909	74651	97319	17221
31246	35797	89051	36319	38137	11101	02808	36771	63163	00816
55704	87671	81987	18984	94617	89097	91628	49172	07106	06218
09107	53117	75464	25300	98186	29702	73632	77044	08238	08097
53779	05917	89367	88743	33981	64647	45685	11160	81086	29458
03252	99475	70537	29636	46984	49231	73571	64092	26162	26361
92966	81458	79782	39399	39278	20247	45367	76937	64563	71930
68109	88529	70116	11782	24198	68334	83184	26202	49315	38471
53118	70359	68973	95173	29213	29969	00445	24846	50957	80443
60924	44136	71034	80642	62977	93957	21006	66422	96753	69814
11151	59784	77446	64703	22038	40387	87749	82349	88010	20160
32731	14203	36222	13436	16935	26412	09878	27931	54679	35275
84037	48341	98595	26036	87821	16245	71204	44232	08527	49083
75807	89169	30622	23911	73689	50718	33796	30145	97763	75437
93509	68893	82351	84938	26829	04823	71697	48189	43465	99159
91528	38008	83069	29029	36617	09019	95788	52955	75018	83253
10603	93078	11673	36373	71987	89710	15378	89022	87934	86336
99185	30214	58351	16606	08569	19665	22531	88753	22759	90501
97288	87653	40124	51615	27365	26827	70265	23368	78952	05515
93564	66965	91850	25093	53517	39997	17521	57074	76743	11610
06959	27612	66188	10351	17367	84340	00247	49881	01987	33756
13172	61241	93558	59519	15082	75692	43138	22677	55844	70034
03690	87173	38088	03032	69496	42886	23096	43416	78732	12420
38005	70085	74744	32644	86440	12489	39538	64712	92792	51310
28758	45596	59049	79799	68763	49827	57854	76334	99237	11368
84260	58136	31250	88953	04929	06903	21175	42463	15327	15205
77800	77252	68397	37938	53941	89771	92875	37004	87044	18210
99905	24764	22807	54083	90303	43362	71223	96233	88058	03268
53803	68932	38510	87838	68543	73671	97403	80077	63351	85781
88379	47885	33501	10666	74222	81999	16699	51745	84672	11640
30033	45809	69655	31679	56931	40579	93867	22886	00794	67305
73888	69685	91050	60898	06171	01165	04192	03700	27979	76516
50935	51067	76172	52543	38383	43396	67725	68868	15571	78654
04689	09839	31801	18560	21328	87664	08203	82426	23946	82792
65860	04568	88383	49927	52267	63736	01964	86914	14949	55467

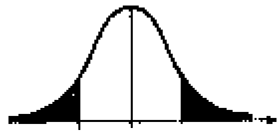
Áreas de la distribución normal estándar



Las áreas de la tabla son las probabilidades de que la variable normal estándar esté entre 0 y z.

	Dos decimales para z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4625	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998									
4.0	0.49997									
4.5	0.499997									
5.0	0.4999997									
6.0	0.499999999									

Distribución *t*-student



Área de dos colas

gl	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291
	.10	.05	.025	.01	.005	.0005

Área de una cola

La distribución F para $\alpha = 0.05$ y $\alpha = 0.10$ (**negritas**) para muchos grados de libertad posibles

Grados de libertad del denominador (k_2)	Numerador de grados de libertad (k_1): 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
	4,052	**5,000**	**5,403**	**5,625**	**5,764**	**5,859**	**5,928**	**5,981**	**6,022**	**6,056**	**6,083**	**6,106**	**6,143**	**6,170**	**6,209**	**6,235**	**6,261**	**6,287**	**6,303**	**6,324**	**6,334**	**6,350**	**6,360**	**6,366**
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.43	19.45	19.45	19.46	19.47	19.48	19.48	19.49	19.49	19.49	19.50
	98.50	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.36**	**99.37**	**99.39**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.47**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.59	8.58	8.56	8.55	8.54	8.53	8.53
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.35**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.28**	**26.24**	**26.18**	**26.15**	**26.13**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.87	5.84	5.80	5.77	5.75	5.72	5.70	5.68	5.66	5.65	5.64	5.63
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.55**	**14.45**	**14.37**	**14.25**	**14.15**	**14.02**	**13.93**	**13.84**	**13.75**	**13.69**	**13.61**	**13.58**	**13.52**	**13.49**	**13.46**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.41	4.39	4.37	4.37
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.46**	**10.29**	**10.16**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.08**	**9.04**	**9.02**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.73	3.71	3.69	3.68	3.67
	13.75	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.40**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.93**	**6.90**	**6.88**
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.53	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.27	3.25	3.24	3.23
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**6.99**	**6.84**	**6.72**	**6.62**	**6.54**	**6.47**	**6.36**	**6.28**	**6.16**	**6.07**	**5.99**	**5.91**	**5.86**	**5.79**	**5.75**	**5.70**	**5.67**	**5.65**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.24	3.20	3.15	3.12	3.08	3.04	3.02	2.99	2.97	2.95	2.94	2.93
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.18**	**6.03**	**5.91**	**5.81**	**5.73**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.12**	**5.07**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.03	2.99	2.94	2.90	2.86	2.83	2.80	2.77	2.76	2.73	2.72	2.71
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.61**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.01**	**4.92**	**4.81**	**4.73**	**4.65**	**4.57**	**4.52**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.86	2.83	2.77	2.74	2.70	2.66	2.64	2.60	2.59	2.56	2.55	2.54
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.20**	**5.06**	**4.94**	**4.85**	**4.77**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.51	2.47	2.46	2.43	2.42	2.40
	9.65	**7.21**	**6.22**	**5.67**	**5.32**	**5.07**	**4.89**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.81**	**3.74**	**3.71**	**3.66**	**3.62**	**3.60**
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.64	2.60	2.54	2.51	2.47	2.43	2.40	2.37	2.35	2.32	2.31	2.30
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.64**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.97**	**3.86**	**3.78**	**3.70**	**3.62**	**3.57**	**3.50**	**3.47**	**3.41**	**3.38**	**3.36**
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.22	2.21
	9.07	**6.70**	**5.74**	**5.21**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.86**	**3.78**	**3.66**	**3.59**	**3.51**	**3.43**	**3.38**	**3.31**	**3.27**	**3.22**	**3.19**	**3.17**

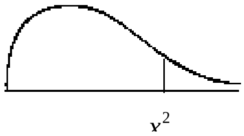
Grados de libertad del denominador (k_2)	Numerador de grados de libertad (k_1) 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
	8.86	**6.51**	**5.56**	**5.04**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.35**	**3.27**	**3.22**	**3.15**	**3.11**	**3.06**	**3.02**	**3.00**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.42	2.38	2.33	2.29	2.25	2.20	2.18	2.14	2.12	2.10	2.08	2.07
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.49**	**3.37**	**3.29**	**3.21**	**3.13**	**3.08**	**3.01**	**2.98**	**2.92**	**2.89**	**2.87**
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.37	2.33	2.28	2.24	2.19	2.15	2.12	2.09	2.07	2.04	2.02	2.01
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.62**	**3.55**	**3.45**	**3.37**	**3.26**	**3.18**	**3.10**	**3.02**	**2.97**	**2.90**	**2.86**	**2.81**	**2.78**	**2.75**
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.10	2.08	2.04	2.02	1.99	1.97	1.96
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.46**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.87**	**2.80**	**2.76**	**2.71**	**2.68**	**2.65**
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.06	2.04	2.00	1.98	1.95	1.93	1.92
	8.29	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.84**	**3.71**	**3.60**	**3.51**	**3.43**	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.92**	**2.84**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.26	2.21	2.16	2.11	2.07	2.03	2.00	1.96	1.94	1.91	1.89	1.88
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.71**	**2.64**	**2.60**	**2.55**	**2.51**	**2.49**
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.22	2.18	2.12	2.08	2.04	1.99	1.97	1.93	1.91	1.88	1.86	1.84
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.70**	**3.56**	**3.46**	**3.37**	**3.29**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.78**	**2.69**	**2.64**	**2.57**	**2.54**	**2.48**	**2.44**	**2.42**
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.16	2.10	2.05	2.01	1.96	1.94	1.90	1.88	1.84	1.83	1.81
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.64**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.64**	**2.58**	**2.51**	**2.48**	**2.42**	**2.38**	**2.36**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.17	2.13	2.07	2.03	1.98	1.94	1.91	1.87	1.85	1.82	1.80	1.78
	7.95	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.36**	**2.33**	**2.31**
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.15	2.11	2.05	2.01	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.54**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.13	2.09	2.03	1.98	1.94	1.89	1.86	1.82	1.80	1.77	1.75	1.73
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.26**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.37**	**2.33**	**2.27**	**2.24**	**2.21**
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.11	2.07	2.01	1.96	1.92	1.87	1.84	1.80	1.78	1.75	1.73	1.71
	7.77	**5.57**	**4.68**	**4.18**	**3.85**	**3.63**	**3.46**	**3.32**	**3.22**	**3.13**	**3.05**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.33**	**2.29**	**2.23**	**2.19**	**2.17**
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.09	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.73	1.71	1.69
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.18**	**3.09**	**3.02**	**2.96**	**2.86**	**2.78**	**2.66**	**2.58**	**2.50**	**2.42**	**2.36**	**2.29**	**2.25**	**2.19**	**2.16**	**2.13**

Grados de libertad del denominador (k_2)	Numerador de grados de libertad (k_1)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.08	2.04	1.97	1.93	1.88	1.84	1.81	1.76	1.74	1.71	1.69	1.67
	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.82	2.75	2.63	2.55	2.47	2.38	2.33	2.26	2.22	2.16	2.12	2.10
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.82	1.79	1.75	1.73	1.69	1.67	1.65
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.79	2.72	2.60	2.52	2.44	2.35	2.30	2.23	2.19	2.13	2.09	2.06
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.01	1.94	1.90	1.85	1.81	1.77	1.73	1.71	1.67	1.65	1.64
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.77	2.69	2.57	2.49	2.41	2.33	2.27	2.20	2.16	2.10	2.06	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.70	1.66	1.64	1.62
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.74	2.66	2.55	2.47	2.39	2.30	2.25	2.17	2.13	2.07	2.03	2.01
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.01	1.97	1.91	1.86	1.82	1.77	1.74	1.69	1.67	1.63	1.61	1.59
	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.86	2.80	2.70	2.62	2.50	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05	1.99	1.95	1.89	1.84	1.80	1.75	1.71	1.67	1.65	1.61	1.59	1.57
	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.82	2.76	2.66	2.58	2.46	2.38	2.30	2.21	2.16	2.08	2.04	1.98	1.94	1.91
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	1.98	1.93	1.87	1.82	1.78	1.73	1.69	1.65	1.62	1.59	1.56	1.55
	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72	2.62	2.54	2.43	2.35	2.26	2.18	2.12	2.04	2.00	1.94	1.90	1.87
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.68	1.63	1.61	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.69	2.59	2.51	2.40	2.32	2.23	2.14	2.09	2.01	1.97	1.90	1.86	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.56	2.48	2.37	2.29	2.20	2.11	2.06	1.98	1.94	1.87	1.83	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99	1.94	1.89	1.83	1.78	1.73	1.68	1.65	1.60	1.57	1.53	1.51	1.49
	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64	2.54	2.46	2.34	2.26	2.18	2.08	2.03	1.95	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.77	1.72	1.67	1.63	1.59	1.56	1.52	1.49	1.48
	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.07	2.01	1.93	1.89	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.76	1.71	1.65	1.62	1.57	1.53	1.51	1.48	1.46
	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.99	1.91	1.86	1.80	1.76	1.73
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.75	1.70	1.64	1.61	1.56	1.54	1.49	1.47	1.45
	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.12	2.02	1.97	1.89	1.84	1.78	1.73	1.70

Grados de libertad del denominador (k_2)	Numerador de grados de libertad (k_1) 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.89	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.19**	**3.02**	**2.89**	**2.78**	**2.70**	**2.63**	**2.56**	**2.46**	**2.38**	**2.27**	**2.18**	**2.10**	**2.01**	**1.95**	**1.87**	**1.82**	**1.76**	**1.71**	**1.68**
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.53	1.50	1.46	1.43	1.41
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.42**	**2.34**	**2.23**	**2.15**	**2.06**	**1.97**	**1.91**	**1.83**	**1.78**	**1.71**	**1.67**	**1.64**
60	4.00	3.15	2.76	2.51	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.82	1.75	1.70	1.65	1.59	1.56	1.51	1.48	1.44	1.41	1.39
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.39**	**2.31**	**2.20**	**2.12**	**2.03**	**1.94**	**1.88**	**1.79**	**1.75**	**1.68**	**1.63**	**1.60**
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.85	1.80	1.73	1.69	1.63	1.58	1.54	1.49	1.46	1.42	1.39	1.37
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.80**	**2.69**	**2.61**	**2.53**	**2.47**	**2.37**	**2.29**	**2.17**	**2.09**	**2.00**	**1.91**	**1.85**	**1.77**	**1.72**	**1.65**	**1.60**	**1.57**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.57	1.53	1.48	1.45	1.40	1.37	1.35
	7.01	**4.92**	**4.07**	**3.60**	**3.29**	**3.07**	**2.91**	**2.78**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.27**	**2.15**	**2.07**	**1.98**	**1.89**	**1.83**	**1.74**	**1.70**	**1.62**	**1.57**	**1.54**
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.43	1.38	1.35	1.33
	6.96	**4.88**	**4.04**	**3.56**	**3.26**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.42**	**2.31**	**2.23**	**2.12**	**2.03**	**1.94**	**1.85**	**1.79**	**1.70**	**1.65**	**1.58**	**1.53**	**1.50**
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.79	1.75	1.68	1.63	1.57	1.52	1.48	1.42	1.39	1.34	1.31	1.28
	6.90	**4.82**	**3.98**	**3.51**	**3.21**	**2.99**	**2.82**	**2.69**	**2.59**	**2.50**	**2.43**	**2.37**	**2.27**	**2.19**	**2.07**	**1.98**	**1.89**	**1.80**	**1.74**	**1.65**	**1.60**	**1.52**	**1.47**	**1.43**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83	1.77	1.73	1.66	1.60	1.55	1.49	1.45	1.40	1.36	1.31	1.27	1.25
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.66**	**2.55**	**2.47**	**2.39**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.76**	**1.69**	**1.60**	**1.55**	**1.47**	**1.41**	**1.37**
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.48	1.44	1.38	1.34	1.29	1.25	1.22
	6.81	**4.75**	**3.91**	**3.45**	**3.14**	**2.92**	**2.76**	**2.63**	**2.53**	**2.44**	**2.37**	**2.31**	**2.20**	**2.12**	**2.00**	**1.92**	**1.83**	**1.73**	**1.66**	**1.57**	**1.52**	**1.43**	**1.38**	**1.33**
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.74	1.69	1.62	1.57	1.52	1.46	1.41	1.35	1.32	1.26	1.22	1.19
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.89**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.27**	**2.17**	**2.09**	**1.97**	**1.89**	**1.79**	**1.69**	**1.63**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**
400	3.86	3.02	2.61	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.17	1.13
	6.70	**4.66**	**3.83**	**3.37**	**3.06**	**2.85**	**2.68**	**2.56**	**2.45**	**2.37**	**2.29**	**2.23**	**2.13**	**2.05**	**1.92**	**1.84**	**1.75**	**1.64**	**1.58**	**1.48**	**1.42**	**1.32**	**1.25**	**1.19**
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
	6.66	**4.63**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.27**	**2.20**	**2.10**	**2.02**	**1.90**	**1.81**	**1.72**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.12**
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.39	1.35	1.28	1.24	1.17	1.11	1.00
	6.63	**4.61**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.25**	**2.18**	**2.08**	**2.00**	**1.88**	**1.79**	**1.70**	**1.59**	**1.52**	**1.42**	**1.36**	**1.25**	**1.15**	**1.00**

Valores críticos de ji cuadrada

Esta tabla contiene los valores de χ^2 que corresponden a un área específica de una cola y a números de grados de libertad específicos gl



Valores posibles de X^2

GRADOS DE LIBERTAD gl	ÁREA DE LA COLA DERECHA 0.10	0.05	0.02	0.01
1	2.706	3.841	5.412	6.635
2	4 605	5.991	7.824	9.210
3	6.251	7 815	9.837	11.345
4	7.779	9.488	11.668	13.277
5	9.236	11 070	13.388	15.086
6	10.645	12.592	15 033	16.812
7	12.017	14 067	16.622	18.475
8	13.362	15 507	18.168	20.090
9	14.684	16.919	19 679	21 666
10	15.987	18.307	21.161	23.209
11	17.275	19.675	22.618	24.725
12	18.549	21 026	24.054	26.217
13	19 812	22 362	25.472	27.688
14	21 064	23.685	26.873	29.141
15	22.307	24.996	28.259	30.578
16	23.542	26.296	29.633	32.000
17	24.769	27 587	30 995	33.409
18	25.989	28 869	32.346	34.805
19	27 204	30 144	33.687	36 191
20	28.412	31.410	35.020	37.566
21	29.615	32 671	36.343	38.932
22	30.813	33 924	37 659	40.289
23	32.007	35 172	38.968	41.638
24	33 196	36.415	40 270	42.980
25	34.382	37.652	41.566	44.314
26	35 563	38 885	42 856	45.642
27	36.741	40 113	44.140	46.963
28	37.916	41 337	45.419	48.278
29	39.087	42.557	46.693	49.588
30	40.256	43 773	47.962	50.892

Hoja de Metadatos

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 1/5

Título	Fundamentos de la Estadística Inferencial
Subtítulo	

Autor(es)

Apellidos y Nombres	Código CVLAC / e-mail	
Gómez R., Mariangeles	**CVLAC**	13836535
	e-mail	Angeles23acuario@hotmail.com
	e-mail	
Ramos S., Yaritza	**CVLAC**	13499178
	e-mail	Yaritza78@hotmail.com
	e-mail	
	CVLAC	
	e-mail	
	e-mail	
	CVLAC	
	e-mail	
	e-mail	

Palabras o frases claves:

Estadística
Estadística Inferencial
Distribuciones muestrales
Hipótesis Hipótesis
Regresión
Correlación

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 2/5

Líneas y sublíneas de investigación:

Área	Subárea
Ciencias Económicas	Contaduría Pública

Resumen (abstract):

Inferir significa deducir algo de otra cosa. Nuestra investigación se refiere, a los fundamentos de la Estadística Inferencial, que se encarga de hacer deducciones de una población por medio de una muestra tomada a partir de ésta; sirviendo así para las organizaciones, porque le permite a le Gerencia tomar decisiones válidas, respecto a las predicciones futuras. Para analizar este análisis estadístico se requiere utilizar la distribución muestral porque a partir de la muestra seleccionada de una población, puede construirse variables aleatoria alternativa, de cuyo análisis se desprenden interesantes propiedades estadísticas (distribución muestral de la media y de la proporción). Los problemas que se tratan en Inferencia Estadística, se basan en dos clases: la estimación o intervalo de confianza y las pruebas de hipótesis. En donde el intervalo de confianza viene dado por un rango de valores, dentro del cual se espera encontrar el valor del parámetro estudiado; y las pruebas de hipótesis, que son supuestos que se plantea el investigador antes de iniciar una investigación, partiendo de una muestra aleatoria significativa, para extraer conclusiones que permitan aceptar o rechazar una hipótesis previamente emitida, sobre el valor de un parámetro desconocido, el cual aborda una serie de pasos. El análisis de regresión y correlación, permite relacionar dos o más variables (variable independiente y variable dependiente). El análisis de varianza, sirve para comparar si los valores de un conjunto de datos numéricos, son significativamente distintos a los valores de otros o más conjunto de datos. Como en la práctica todas las poblaciones no pueden tomarse como normales, por situaciones en donde no es posible formular una hipótesis segura sobre el valor de un parámetro, surgen las pruebas no paramétricas (ji-cuadrado), éstas no dependen de un solo tipo de distribución.

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 3/5

Contribuidores:

Apellidos y Nombres	ROL / Código CVLAC / e-mail	
Romero, Miguel	**ROL**	**CA** [ ] **AS** [ ] **TU** [x] **JU** [ ]
	CVLAC	8.879.006
	e-mail	
	e-mail	
Gòmez R., Mariangeles	**ROL**	**CA** [x] **AS** [ ] **TU** [ ] **JU** [ ]
	CVLAC	13.836.535.
	e-mail	Angeles23acuario@hotmail.com
	e-mail	
Ramos S., Yaritza	**ROL**	**CA** [x] **AS** [ ] **TU** [ ] **JU** [ ]
	CVLAC	13.499.178
	e-mail	Yaritza78@hotmail.com
	e-mail	
	ROL	**CA** [ ] **AS** [ ] **TU** [ ] **JU** [ ]
	CVLAC	
	e-mail	
	e-mail	

Fecha de discusión y aprobación:

Año	Mes	Día
2008	04	08

Lenguaje: spa

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 4/5

Archivo(s):

Nombre de archivo	Tipo MIME
Tesis-Gomez yRamos	Application/.doc

Alcance:

Espacial : Universal (Opcional)

Temporal: 6 meses (Opcional)

Título o Grado asociado con el trabajo:

Licenciada en Contaduría Pública

Nivel Asociado con el Trabajo: Licenciatura

Área de Estudio:

Ciencias Económicas

Institución(es) que garantiza(n) el Título o grado:

Universidad de Oriente

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 5/5

Derechos:

Nosotras, las autoras de esta tesis, garantizamos en forma permanente a la Universidad de Oriente el derecho de archivar y difundir a través de la biblioteca, el contenido de esta tesis; en caso de utilizar el medio de Internet, solamente mostrar el resumen y la bibliografía de ésta para que le permitan al lector consultar si así lo desea.

Las autoras nos reservamos los derechos de propiedad intelectual; así como cualquier otro derecho que pudiera derivarse.

Mariangeles Gómez R.
C.I:13.836.535
AUTOR 1

Yaritza Ramos S.
C.I:13.449.178
AUTOR 2

Miguel Romero
C.I:8.879.006
TUTOR

POR LA SUBCOMISIÓN DE TESIS: