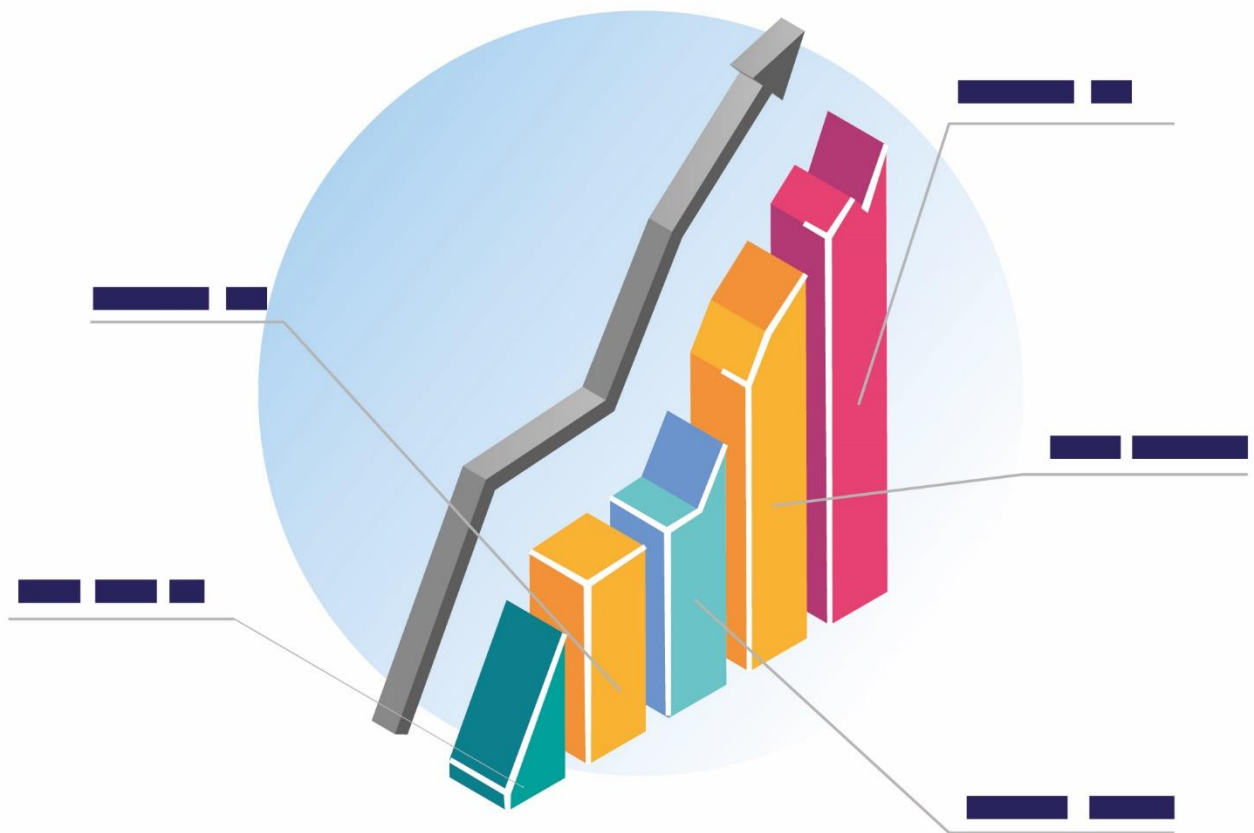


CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA INFERENCIAL

Notas de Clase



Alberto Javier Mesa Guerrero
Segundo Javier Caicedo-Zambrano



Editorial
Universidad de Nariño



Editorial

Universidad de Nariño

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA INFERENCIAL

NOTAS DE CLASE

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA INFERENCIAL

NOTAS DE CLASE

Alberto Javier Mesa Guerrero
Segundo Javier Caicedo Zambrano



Editorial
Universidad de **Nariño**

Mesa Guerrero, Javier Alberto

Cálculo de probabilidades y Estadística inferencial: Notas de clase / Javier Alberto Mesa Guerrero, Segundo Javier Caicedo Zambrano. -- 1ª ed. -- San Juan de Pasto: Editorial Universidad de Nariño, 2022

196 p. : il. col., gráficos, tablas

Incluye bibliografía p. 169 y datos de los autores p. 3 ISBN: 978-628-7509-39 Digital

1. Probabilidades 2. Estadística descriptiva 3. Teoría de conjuntos
4. Estadística--Problemas, ejercicios, etc. 5. Estadística matemática
6. Varianza I. Caicedo Zambrano, Segundo Javier

519.53 M578 – SCDD-Ed. 22



SECCIÓN DE BIBLIOTECA
"Alberto Quintero Guerrero"

Universidad de Nariño
FUNDADA EN 1904

Cálculo de Probabilidades y Estadística Inferencial. Notas de clase

© Alberto Javier Mesa Guerrero
soundmesa@yahoo.com

© Segundo Javier Caicedo-Zambrano
jacaza1@gmail.com; jacaza1@udenar.edu.co

© Editorial Universidad de Nariño

ISBN: 978-628-7509-39
Primera edición

Corrección de Estilo: Germán Chaves Jurado
Diseño y Diagramación: Diana Salas

Fecha de publicación: 2022
San Juan de Pasto - Nariño - Colombia

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin la autorización escrita de su Autor o de la Editorial Universidad de Nariño.

Tabla de contenido

| | |
|--|-----------|
| INTRODUCCIÓN..... | 10 |
| CAPÍTULO 1 INTRODUCCIÓN A LA TEORÍA DE CONJUNTOS | 13 |
| 1.1 OPERACIONES CON CONJUNTOS..... | 13 |
| 1.1.1 Unión | 13 |
| 1.1.2 Intersección | 13 |
| 1.1.3 Conjunto vacío | 13 |
| 1.1.4 Conjuntos ajenos, excluyentes o disyuntos..... | 14 |
| 1.1.5 Complemento | 14 |
| 1.1.6 Complemento respecto al conjunto universal | 14 |
| 1.1.7 Diferencia..... | 15 |
| 1.1.8 Diferencia simétrica..... | 15 |
| 1.1.9 Resumen de operaciones con conjuntos | 15 |
| 1.2 PRODUCTO CARTESIANO..... | 16 |
| CAPÍTULO 2 INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES..... | 18 |
| 2.1 TÉCNICAS DE CONTAR | 18 |
| 2.1.1. Principio fundamental..... | 18 |
| 2.1.2 Permutaciones..... | 19 |
| 2.1.3 Permutaciones con repetición..... | 20 |
| 2.1.4 Pruebas ordenadas..... | 21 |
| 2.1.5 Combinaciones | 22 |
| 2.2 ESPACIO MUESTRAL | 23 |
| 2.2.1 Ocurrencia de un suceso | 23 |
| 2.2.2 Álgebra de conjuntos..... | 24 |

| | |
|---|-----------|
| 2.3 DEFINICIÓN DE PROBABILIDAD | 24 |
| 2.3.1. Probabilidad clásica o a priori | 24 |
| 2.3.2 Probabilidad frecuencial o a posteriori | 26 |
| 2.3.3. Probabilidad subjetiva..... | 26 |
| 2.3.4. Definición axiomática de probabilidad | 26 |
| 2.4 LEYES DE PROBABILIDAD..... | 27 |
| 2.5 ESPACIOS FINITOS DE PROBABILIDAD..... | 28 |
| | |
| CAPÍTULO 3 PROBABILIDAD CONDICIONAL | 30 |
| 3.1 PROCESOS ESTOCÁSTICOS FINITOS | 31 |
| 3.2 PROBABILIDAD TOTAL..... | 31 |
| 3.3 PROBABILIDAD DE BAYES..... | 33 |
| | |
| CAPÍTULO 4 VARIABLES ALEATORIAS..... | 38 |
| 4.1 VARIABLES ALEATORIAS DISCRETAS | 38 |
| 4.1.1 Espacio de probabilidad | 39 |
| 4.1.2 Función acumulativa..... | 40 |
| 4.1.3 Media, varianza y desviación estándar | 41 |
| 4.2 VARIABLES ALEATORIAS CONTINUAS..... | 44 |
| 4.2.1 Espacio de probabilidad de una variable aleatoria continua | 44 |
| 4.2.2 Función acumulativa..... | 45 |
| 4.2.3 Media, Varianza y Desviación Estándar | 46 |
| | |
| CAPÍTULO 5 MODELOS DE PROBABILIDAD..... | 48 |
| 5.1 MODELOS DISCRETOS | 48 |
| 5.1.1 Modelo de Bernoulli | 48 |
| 5.1.2 Modelo Binomial | 49 |
| 5.1.3 Modelo de Poisson..... | 52 |
| <i>Cálculo de probabilidad de la distribución de Poisson con Microsoft Excel.....</i> | <i>53</i> |
| 5.1.4 Modelo Hipergeométrico..... | 54 |
| <i>Probabilidad distribución hipergeométrica con Microsoft Excel</i> | <i>55</i> |
| 5.1.5 Modelo Binomial negativo..... | 56 |
| <i>Probabilidad distribución binomial negativa con Microsoft Excel.....</i> | <i>58</i> |
| 5.1.6 Modelo Geométrico | 58 |

| | |
|--|------------|
| <i>Cálculo de probabilidad de la Distribución Geométrica (binomial negativa) con Microsoft Excel</i> | 60 |
| 5.1.7 Modelo Multinomial | 60 |
| 5.2 MODELOS CONTINUOS | 62 |
| 5.2.1 Distribución Normal | 62 |
| 5.2.3 Distribución t de Student..... | 68 |
| 5.2.4 Distribución F de Fisher | 71 |
| | |
| CAPÍTULO 6 DISTRIBUCIONES MUESTRALES | 75 |
| 6.1 RESUMEN DE DISTRIBUCIONES MUESTRALES | 75 |
| 6.1.1 Variables aleatorias con distribución normal..... | 75 |
| 6.1.2 Variables aleatorias con distribución t de Student..... | 76 |
| 6.1.3 Distribución Ji-Cuadrado..... | 76 |
| 6.2 DISTRIBUCIÓN MUESTRAL DE LA MEDIA | 77 |
| 6.3 DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN DE ÉXITOS | 77 |
| 6.4 DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE MEDIAS | 78 |
| 6.5 DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE PROPORCIONES | 78 |
| | |
| CAPÍTULO 7 APLICACIONES DE LA DISTRIBUCIÓN NORMAL | 81 |
| 7.1 CÁLCULO DE PROBABILIDADES | 81 |
| 7.1.1 Distribución Normal | 81 |
| 7.1.2 Distribución Muestral de Medias | 85 |
| 7.1.3 Distribución Muestral de Proporción de Éxitos | 86 |
| 7.1.4 Distribución Muestral de Diferencia de Medias..... | 87 |
| 7.2 ESTIMACIÓN DE PARÁMETROS | 92 |
| 7.2.1 Estimación por intervalos para la media (μ) | 92 |
| 7.2.3 Estimación para diferencia de medias ($\delta = \mu_1 - \mu_2$)..... | 98 |
| 7.2.4 Límites de confianza para diferencia de proporciones ($\Delta p = p_1 - p_2$) | 99 |
| 7.3 PRUEBAS DE HIPÓTESIS | 101 |
| 7.3.1 PROCEDIMIENTO | 101 |
| <i>Primer método</i> | 104 |
| <i>Segundo método</i> | 104 |
| <i>Tercer método</i> | 104 |
| 7.3.1.1 Prueba de hipótesis para la media (μ) | 107 |
| 7.3.1.2. Prueba de hipótesis para la proporción (p)..... | 109 |
| 7.3.1.3 Prueba de hipótesis para la diferencia de medias..... | 111 |

| | |
|---|------------|
| 7.3.1.4 Prueba de hipótesis para la diferencia de proporciones (Δp)..... | 112 |
| 7.3.1.5 Potencia de una prueba..... | 114 |
| Potencia de una prueba de hipótesis con Microsoft Excel..... | 121 |
| 7.3.2 Distribución t de Student..... | 123 |
| 7.3.2.1 Estimación por intervalos para la media de la población | 124 |
| 7.3.2.2 Límites de confianza para diferencia de medias poblacionales..... | 125 |
| 7.3.2.3 Prueba de hipótesis para la media de población con t de Student | 130 |
| 7.3.2.4 Prueba de hipótesis con t de Student para dos muestras independientes | 133 |
| 7.3.2.5 Prueba de hipótesis con t de Student para dos muestras dependientes | 134 |
| 7.3.3 Distribución Ji-Cuadrado (χ^2) | 136 |
| 7.3.3.1 Estimación por intervalos para la varianza poblacional con Ji-Cuadrado (χ^2)..... | 136 |
| 7.3.3.2 Prueba de hipótesis para la varianza (σ^2)..... | 138 |
| 7.3.3.3 Prueba de bondad de ajuste | 140 |
| 7.3.3.4 Prueba de independencia | 141 |
| 7.3.4 Distribución F (Fisher) | 145 |
| CAPÍTULO 8 PROBLEMAS PROPUESTOS..... | 151 |
| 8.1 CÁLCULO DE PROBABILIDADES..... | 151 |
| 8.2 MODELOS DE PROBABILIDAD | 159 |
| 8.3 ESTADÍSTICA INFERENCIAL..... | 164 |
| BIBLIOGRAFÍA..... | 174 |
| ANEXOS | 175 |
| ANEXO A: PRUEBA DE HIPÓTESIS PARA μ | 175 |
| ANEXO B: TABLA PARA DISTRIBUCIÓN NORMAL | 176 |
| ANEXO C. TABLA PARA DISTRIBUCIÓN t | 178 |
| ANEXO D. TABLA PARA DISTRIBUCIÓN JI-CUADRADO | 180 |
| ANEXO E. TABLA DISTRIBUCIÓN F | 182 |
| Acerca de los Autores..... | 185 |

INTRODUCCIÓN

Esta obra surge por el interés de los autores de poner a disposición de la comunidad académica unas notas de clase, *en formato de libro de texto*, con el fin de apoyar el desarrollo del curso de Cálculo de Probabilidades y Estadística, que los autores ofrecen en la Universidad de Nariño, de modo que se constituya en una fuente de consulta y de estudio de temáticas relacionadas con conceptos básicos del cálculo de probabilidades, de los modelos de distribución de probabilidad y de la estadística inferencial. Concretamente, el libro está orientado a estudiantes de la Universidad de Nariño de las áreas de Ciencias Básicas, Técnicas e Ingenierías; sin embargo, dada la forma clara y didáctica en la que se presentan los conceptos y ejemplos, lo pueden utilizar estudiantes de otras áreas del conocimiento, inclusive programas de educación no formal. Los temas que se tratan son exactamente los que desarrollan los autores en los cursos de Cálculo de Probabilidades y Estadística de la Universidad de Nariño; destacando que uno de los coautores tiene experiencia de más de 30 años en el desarrollo de estas temáticas.

El libro contiene exactamente las temáticas que se trabaja en los cursos en mención de la Universidad de Nariño, por lo cual, en dichos cursos, esta obra, escrita en formato de libro de texto, podría ser asumida como texto guía. Esta aclaración es muy importante, por cuanto, no se trata de un libro de texto en sí mismo, sino de unas notas de clase constitutivas en un curso de Cálculo de Probabilidades y Estadística de la Universidad de Nariño. En este contexto, respetuosamente, sugerimos que la evaluación se realice considerando que se trata del desarrollo de un curso, que en términos del Decreto 1279 de 2002, corresponde a unas notas de clase, por lo tanto, en sí mismo, no se debe evaluar como un libro de texto, que en otros contextos, puede requerir nuevas temáticas, otros enfoques, etc.

El libro está organizado en ocho (8) capítulos. El primero, "Introducción a la teoría de conjuntos", presenta las operaciones entre conjuntos y el concepto de producto cartesiano. El segundo capítulo, "Introducción al cálculo de probabilidades", presenta los fundamentos de las técnicas de contar, los conceptos de espacio muestral, definición de probabilidad, las leyes de probabilidad y finaliza con el estudio de

espacios finitos de probabilidad. El capítulo tercero, trata sobre probabilidad condicional, procesos estocásticos finitos, probabilidad total y probabilidad de Bayes. El Capítulo cuarto, "Variables aleatorias", desarrolla el tema de variables aleatorias discretas y continuas. El capítulo quinto, "Modelos de probabilidades", trata algunos modelos de distribución de probabilidad, tanto discretos como continuos. En el capítulo sexto, "Distribuciones muestrales", se estudia la distribución muestral de la media, de la proporción de éxitos, de la diferencia de medias y de la diferencia de proporciones. En el capítulo séptimo, se trabaja la aplicación de las distribuciones en el cálculo de probabilidades, la estimación de parámetros, la prueba de hipótesis y el análisis de varianza. Finalmente, el capítulo ocho, "Problemas propuestos", presenta diversos problemas sobre cálculo de probabilidades, modelos de probabilidad y aplicaciones de la estadística inferencial.

Los autores, Abril de 2021.



Capítulo 1
Introducción a La
Teoría De Conjuntos



CAPÍTULO 1

INTRODUCCIÓN A LA TEORÍA DE CONJUNTOS

Algunas ideas, propiedades y operaciones de conjuntos son necesarias para comprender y aplicar los conceptos de probabilidad y sus leyes.

Un conjunto es una lista o colección bien definida de objetos, llamados elementos, puntos o resultados.

La pertenencia de un elemento a a un conjunto A se indica así: $a \in A$.

Si B es un conjunto cuyos elementos pertenecen a un conjunto A , entonces B es un subconjunto de A y se indica como $B \subseteq A$.

Ejemplo:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

1.1 OPERACIONES CON CONJUNTOS

1.1.1 Unión

$$A \cup B = \{x/x \in A \vee x \in B\}.$$

\cup : representa la unión de conjuntos, se lee "o".

$A \cup B$ es el conjunto formado por elementos que pertenecen al conjunto A o al conjunto B o a los dos. Los elementos comunes que se repitan se escriben una sola vez.

1.1.2 Intersección

$$A \cap B = \{x/x \in A \wedge x \in B\}.$$

El símbolo " \cap " representa la intersección de conjuntos, se lee "y".

$A \cap B$ es el conjunto formado por elementos comunes del conjunto A y del conjunto B .

El símbolo " $\{$ " se lee "tales que".

1.1.3 Conjunto vacío

Es un conjunto que no tiene elementos; se representa con la letra \emptyset o también así $\{\}$. El vacío es subconjunto de todo conjunto. Si A es cualquier conjunto, entonces $\emptyset \subset A$.

1.1.4 Conjuntos ajenos, excluyentes o disyuntos

Si la intersección de dos conjuntos es igual al conjunto vacío, estos conjuntos se denominan Conjuntos Ajenos, Excluyentes o Disyuntos; es decir:

Si $A \cap B = \emptyset$, entonces A y B son disyuntos.

1.1.5 Complemento

Si $A \subset B$, entonces el complemento de A respecto a B se indica así: C_B^A

$$C_B^A = \{x / x \in B \wedge x \notin A\}$$

De la definición se deduce que: $C_B^A = B - A$

C_B^A está formado por todos los elementos que le "faltan" a A para ser "igual" al conjunto B .

1.1.6 Complemento respecto al conjunto universal

Si U es el conjunto de referencia (universal) en una determinada teoría, por ejemplo, en la teoría de conjuntos, entonces cualquier conjunto A, B, C, D en esta teoría será subconjunto de U .

En este caso, el conjunto C_U^A se indica así A^c .

$$\text{Luego, } C_U^A = A^c = U - A.$$

El símbolo " $-$ " representa la diferencia de conjuntos, se lee "menos".

$U - A$: es el conjunto formado por los elementos del conjunto universal que no pertenecen al subconjunto A .

Además: $x \in A^c \Leftrightarrow x \notin A$; esto es equivalente a: $x \in A \Leftrightarrow x \notin A^c$.

Esto es, si un elemento pertenece a un conjunto, dicho elemento no pertenece al complemento del conjunto y si un elemento no pertenece a un conjunto, dicho elemento pertenece al complemento del conjunto.

Otras notaciones para C_U^A son las siguientes: A' , \bar{A} .

1.1.7 Diferencia

$$A - B = \{x/x \in A \wedge x \notin B\}$$

$$B - A = \{x/x \in B \wedge x \notin A\}$$

1.1.8 Diferencia simétrica

$$A \Delta B = (A - B) \cup (B - A) = (A \cup B) - (A \cap B)$$

El símbolo " Δ " se lee "diferencia simétrica".

1.1.9 Resumen de operaciones con conjuntos

La Tabla 1 contiene un resumen de las propiedades y operaciones más usuales entre conjuntos.

Tabla 1. Resumen de propiedades y operaciones entre conjuntos

| Propiedad: | Unión | intersección |
|-----------------|--|--|
| Asociativa | $(A \cup B) \cup C = A \cup (B \cup C)$ | $(A \cap B) \cap C = A \cap (B \cap C)$ |
| Conmutativa | $A \cup B = B \cup A$ | $A \cap B = B \cap A$ |
| Idempotente | $A \cup A = A$ | $A \cap A = A$ |
| Distributiva | $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ | $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ |
| Neutralidad | $A \cup \emptyset = A$ | $A \cap U = A$ |
| | $A \cup U = U$ | $A \cap \emptyset = \emptyset$ |
| Complementación | $A \cup A^c = U$ | $A \cap A^c = \emptyset$ |
| De D'Morgan | $(A \cup B)^c = A^c \cap B^c$ | $(A \cap B)^c = A^c \cup B^c$ |

Además, se cumple que: $(A^c)^c = A$ y $A - B = A \cap B^c$

1.2 PRODUCTO CARTESIANO

Definición. Sean A y B dos conjuntos. Al conjunto formado por todos los pares ordenados donde la primera componente pertenece al conjunto A y la segunda al conjunto B , se denota por $A \times B$ y se denomina *producto cartesiano* de A y B .

Simbólicamente:

$$A \times B = \{(x, y) / x \in A \wedge y \in B\}$$

En consecuencia,

$$(x, y) \in A \times B \Leftrightarrow x \in A \wedge y \in B \quad \text{y} \quad (x, y) \notin A \times B \Leftrightarrow x \notin A \vee y \notin B$$

Ejemplo:

Sean los conjuntos $A = \{1, 2\}$ y $B = \{3, 4, 5\}$; entonces, el producto cartesiano de A y B es:

$$A \times B = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}$$

El Gráfico 1 contiene una representación de los conjuntos A y B en el Plano Cartesiano $A \times B$.

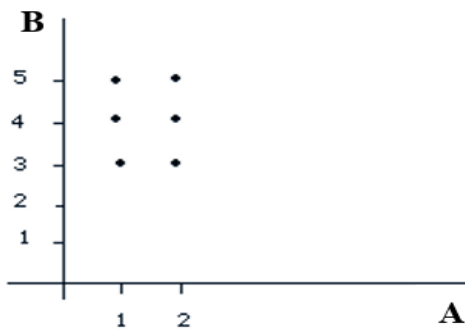


Gráfico 1. Representación gráfica del producto cartesiano de A y B



Capítulo 2

Introducción al Cálculo de Probabilidades

CAPÍTULO 2

INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES

La probabilidad estudia el azar a partir de experimentos o fenómenos aleatorios, es decir, a partir de situaciones donde los resultados no se pueden predecir, aun en el caso en que el fenómeno o experimento se repita bajo las mismas condiciones. En muchos casos, con anticipación, se conocen todos los posibles resultados del experimento.

2.1 TÉCNICAS DE CONTAR

Permiten determinar el número de elementos de un conjunto o el número de resultados posibles de un experimento, sin necesidad de acudir a la enumeración directa de las opciones. Un conjunto formado con posibles resultados de un experimento se denomina suceso o evento.

2.1.1. Principio fundamental

Sean los siguientes eventos: $E_1, E_2, E_3, \dots, E_m$.

Supongamos que E_1 puede ocurrir de n_1 maneras diferentes; E_2 puede ocurrir de n_2 maneras diferentes; E_3 puede ocurrir de n_3 maneras diferentes y así sucesivamente, entonces el número de maneras en que pueden ocurrir todos los eventos se determina así:

$$n_1 \times n_2 \times n_3 \times \dots \times n_m.$$

Ejemplo:

Determinar el número de placas para automóvil que se pueden fabricar, si cada placa consta de 3 letras y de 3 dígitos.

Solución:

El alfabeto inglés consta de 26 letras y se dispone de 10 dígitos, numerados del 0 al 9. Entonces, cada letra se puede escoger de 26 maneras y cada número de 10 formas, por lo cual, el número de placas que se pueden fabricar viene dado por la siguiente expresión:

$$26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17\,576\,000.$$

R/ Con las condiciones dadas, se puede formar 17 576 000 placas diferentes.

2.1.2 Permutaciones

Una ordenación de los n elementos de un conjunto, tomados en un orden determinado, se denomina *Permutación*.

El número de permutaciones de n elementos, tomados todos a la vez, se denota y define por:

$$P_n = n!$$

Donde $n! = n(n-1)(n-2) \times \dots \times 3 \times 2 \times 1$, se denomina *factorial de n* .

Ejemplo:

Determinar el número de maneras en que pueden ordenarse 8 libros diferentes en un estante.

Solución:

$${}_8P_8 = P(8,8) = 8! = 40\,320.$$

El número de permutaciones de tamaño r , con $r < n$, se denota por $P(n,r)$ o también por ${}_nP_r$ y se define como sigue:

$$P(n,r) = \frac{n!}{(n-r)!} = \frac{n(n-1)(n-2)\dots(n-(r-1))(n-r)!}{(n-r)!}.$$

$$P(n,r) = n(n-1)(n-2)\dots(n-(r-1)) = n(n-1)(n-2)\dots(n-r+1).$$

Ejemplos:

En una competencia de atletismo en los juegos olímpicos participan 30 deportistas. Determinar el número de formas distintas en que los atletas pueden ocupar el podio.

Solución:

La medalla de oro la puede ganar cualquiera de los 30 atletas, la de plata cualquiera de los 29 que quedan, y la de bronce cualquiera de los otros 28 participantes.

$$P(30, 3) = \frac{n!}{(n-r)!} = \frac{30!}{(30-3)!} = \frac{30 \times 29 \times 28 \times 27!}{27!} = 30 \times 29 \times 28 = 24\,360.$$

R/ El número de forma en que se pueden entregar las tres medallas, corresponde a la siguiente permutación: ${}_{30}P_3 = 24\,360$.

2.1.3 Permutaciones con repetición

El número de permutaciones de los n elementos de un conjunto que contiene n elementos, en el cual n_1 son iguales, n_2 son iguales, ..., n_r son iguales, se denota y define así:

$$P^*_{(n,n)} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

Ejemplo:

Suponga que, en un proceso de elección estudiantil, el tarjetón tiene tres opciones: candidato A, candidato B y voto en blanco, y que son 10 estudiantes los autorizados para votar. Determinar el número de maneras mediante las cuales se pueden obtener 5 votos por el candidato A, 3 por el B y 2 en blanco.

$$P^*_{(10;5,3,2)} = \frac{10!}{5! 3! 2!} = 2520.$$

R/ Existen 2520 posibilidades para obtener el resultado indicado.

2.1.4 Pruebas ordenadas

Cuando se escoge un elemento de un conjunto, uno tras otro, r veces, esta escogencia se denomina *Prueba Ordenada* de tamaño r . Se presentan dos casos dependiendo de la forma de la escogencia, bien sea con sustitución o sin sustitución de los elementos.

a) *Escogencia con sustitución*: se escoge un elemento y se regresa al conjunto, con lo cual, ese elemento tiene la posibilidad de ser escogido nuevamente. El número de pruebas ordenadas CON sustitución está dado por la siguiente fórmula:

$$n \times n \times n \times \cdots \times n = n^r.$$

Ejemplo:

Diez personas deben votar a favor o en contra de un proyecto. Determinar el número de resultados posibles.

Solución:

$$n^r = 2^{10} = 2 \times 2 \times 2 \times \cdots \times 2 = 1024.$$

b) *Escogencia sin sustitución*: se escoge un elemento y *no* se regresa al conjunto, con lo cual, ese elemento no tiene la posibilidad de ser escogido nuevamente. El número de pruebas ordenadas de tamaño r SIN sustitución coincide con las permutaciones y está dado por la siguiente expresión:

$${}_n P_r = P(n, r) = \frac{n!}{(n - r)!}.$$

Ejemplo:

Determinar la cantidad de números de 3 cifras que se pueden formar con 5 dígitos diferentes, si no se admite repetición de cifras.

Solución:

$$P(n, r) = \frac{n!}{(n - r)!} = \frac{5!}{(5 - 3)!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 5 \times 4 \times 3 = 60.$$

R/ Se pueden formar 60 números diferentes de tres cifras distintas cada uno.

2.1.5 Combinaciones

Una combinación de tamaño r es un subconjunto de r elementos de un conjunto de tamaño n . El número de combinaciones de tamaño r , tomados de r en r , se denota y define por:

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{(n-r)! r!}.$$

Se puede observar que el número de combinaciones de r elementos de un conjunto de r elementos diferentes equivale al número de permutaciones de r elementos, dividido entre $r!$

Vale destacar que en las combinaciones no importa el orden de los elementos en cada grupo.

También se denota $C(n, r) = \binom{n}{r}$, entonces,

$$C(n, r) = \binom{n}{r} = \frac{n!}{(n-r)! r!}.$$

Ejemplo:

Si en un curso hay 50 estudiantes, determinar el número de grupos que se pueden formar de cuatro (4) estudiantes cada uno.

Solución:

$$C(n, r) = C(50, 4) = \binom{50}{4}.$$

$$\binom{50}{4} = \frac{50!}{(50-4)! \times 4!} = \frac{50 \times 49 \times 48 \times 47 \times \mathbf{46!}}{\mathbf{46!} \times 4!} = \frac{50 \times 49 \times 48 \times 47}{4 \times 3 \times 2 \times 1} = 230\,300.$$

R/ Con 50 estudiantes se pueden formar 230 300 subgrupos de 4 estudiantes cada uno.

2.2 ESPACIO MUESTRAL

El espacio muestral S de un experimento aleatorio es el conjunto de todos los resultados posibles del experimento. El conjunto puede ser finito, infinito contable o continuo. Un elemento del conjunto S se denomina *Punto Muestral*. Un *subconjunto* del espacio muestral es un *suceso* o *evento*.

Si un conjunto tiene n elementos, entonces el número de subconjuntos es igual a 2^n .

El conjunto vacío \emptyset constituye un suceso imposible; por su parte, el conjunto S constituye un suceso seguro o cierto.

2.2.1 Ocurrencia de un suceso

Se dice que *ocurre un suceso* A si el punto muestral pertenece al conjunto A . Por ejemplo, suponga que A es un suceso de un espacio muestral S ; además que, al realizar el experimento correspondiente, resulta el punto muestral x , entonces:

Si $x \in A$, se dice que ocurre el *evento* A .

Los sucesos A y B son *excluyentes* sí y solo si $A \cap B = \emptyset$.

Utilizando operaciones de conjuntos, se pueden generar nuevos sucesos.

Ejemplos:

Sean los sucesos A y B de un espacio muestral S .

Ocurre al menos uno de los sucesos A, B si $x \in (A \cup B)$.

Ocurren ambos sucesos si $x \in (A \cap B)$.

Ocurre solamente A si $x \in (A - B)$.

Ocurre solamente uno de los sucesos si $x \in (A \cup B) - (A \cap B)$.

No ocurre A si $x \in A^c$.

No ocurre ninguno de los dos sucesos si $x \in (A \cup B)^c$.

No ocurren los dos sucesos simultáneamente si $x \in (A \cap B)^c$.

2.2.2 Álgebra de conjuntos

Una clase no vacía de conjuntos constituye un *álgebra* si es cerrada bajo la unión y la complementación de conjuntos.

& es álgebra si:

& es diferente de vacío.

$\forall A, B \in \&, (A \cup B) \in \&$.

$\forall A \in \&, A^c \in \&$.

Ejemplo:

Sea el conjunto $X = \{a, b, c\}$, entonces,

$\& = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

2.3 DEFINICIÓN DE PROBABILIDAD

2.3.1. Probabilidad clásica o a priori

Si un experimento tiene n resultados mutuamente excluyentes e igualmente probables y si r de ellos poseen una propiedad A , entonces la probabilidad de que A ocurra, que se lee probabilidad de A , se anota y se calcula así:

$$P(A) = \frac{r}{n} = \frac{\#A}{\#S}.$$

Donde $\#A$ y $\#S$ representan, respectivamente, la cardinalidad de A y la cardinalidad de S ; es decir, representan el número de elementos de A y de S , respectivamente.

Ejemplo:

Sea un curso de Estadística de 30 estudiantes, entre los cuales 10 son mujeres. Determinar el número de grupos que se pueden formar de 5 estudiantes cada uno; además, calcular la probabilidad de que un grupo esté conformado exactamente por 2 hombres y 3 mujeres.

Solución:

Número total de grupos de 5 estudiantes:

$$\#S = \text{número de grupos de 5 estudiantes} = C(30,5) = \binom{30}{5}.$$

$$\binom{30}{5} = \frac{30!}{(30-5)! \times 5!} = \frac{30 \times 29 \times 28 \times 27 \times 26 \times \mathbf{25!}}{\mathbf{25!} \times 5!} = \frac{30 \times 29 \times 28 \times 27 \times 26}{5 \times 4 \times 3 \times 2 \times 1} = 142\,506.$$

Número total de grupos de 5 estudiantes, a los cuales pertenecen 2 hombres y 3 mujeres:

Sea el suceso A : se seleccionan grupos de 5 personas, entre las cuales aparecen 2 hombres y 3 mujeres.

$$\#A = \text{número de elementos de } A = C(20,2) \times C(10,3).$$

$$\#A = \binom{20}{2} \times \binom{10}{3} = \frac{20 \times 19}{2} \times \frac{10 \times 9 \times 8}{6} = 120 \times 120 = 22\,800.$$

De modo que,

$$P(A) = \frac{r}{n} = \frac{22\,800}{142\,506} = 0,16 = 16\%.$$

R/ La probabilidad de escoger un grupo de 5 estudiantes, al que pertenezcan 2 hombres y 3 mujeres es del 16 %.

2.3.2 Probabilidad frecuencial o a posteriori

Consiste en la probabilidad de un suceso A con base en la experimentación, para lo cual se divide el número de veces que se presenta el suceso A (frecuencia) entre el número de veces que se realiza el experimento, esto es,

$$P(A) = \frac{f}{n} = \frac{\text{frecuencia}}{\text{total de casos}}.$$

Ejemplo:

Calcular la probabilidad de que un estudiante de la Universidad de Nariño pague menos de un salario mínimo por concepto de matrícula.

Solución:

Sea S una muestra de 300 estudiantes de la Universidad de Nariño, y el evento A , estudiantes de la Universidad de Nariño que pagan menos de un salario mínimo. Se encontró que 120 pagan menos de un salario mínimo.

$$P(A) = \frac{f}{n} = \frac{120}{300} = 0,4 = 40\%.$$

R/ La probabilidad de que un estudiante de la Universidad de Nariño pague menos de un salario mínimo es del 40 %.

2.3.3. Probabilidad subjetiva

Las probabilidades de los sucesos pueden ser arbitrarias y dependen del grado de conocimiento que tenga un sujeto acerca de una determinada situación.

Ejemplo:

La probabilidad de que un estudiante apruebe Cálculo Diferencial es 0,80.

2.3.4. Definición axiomática de probabilidad

Sea S un espacio muestral de un experimento E y sea una \mathcal{A} un álgebra de sucesos.

Una función P constituye una función de probabilidad de \mathcal{E} sobre el intervalo $[0,1]$ si se cumplen los siguientes axiomas:

$$\forall A, 0 \leq P(A) \leq 1.$$

$$P(S) = 1.$$

$$P(A \cup B) = P(A) + P(B) \text{ si } A, B \text{ son excluyentes.}$$

Si A_1, A_2, \dots, A_n son sucesos excluyentes, entonces,

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

2.4 LEYES DE PROBABILIDAD

Son propiedades que se cumplen para todo suceso A y B de un espacio muestral S .

Probabilidad de que ocurra un suceso imposible (un suceso que nunca ocurre)

$$P(\emptyset) = 0.$$

Probabilidad de que no ocurra A :

$$P(A^c) = 1 - P(A).$$

Probabilidad de que ocurra alguno de los sucesos A o B :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Probabilidad de que ocurra solamente A :

$$P(A - B) = P(A) - P(A \cap B); \text{ pero } A - B = A \cap B^c, \text{ entonces,}$$

$$P(A \cap B^c) = P(A - B).$$

De las propiedades anteriores, se derivan las siguientes relaciones:

Probabilidad de que ocurra sólo uno de los dos sucesos:

$$P(A \Delta B) = P((A \cup B) - (A \cap B)) = P(A \cup B) - P(A \cap B).$$

Probabilidad de que no ocurra ninguno de los sucesos (ni A , ni B):

$$P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - P(A \cup B).$$

Probabilidad de que no ocurran los dos sucesos a la vez:

$$P(A^c \cup B^c) = P[(A \cap B)^c] = 1 - P(A \cap B).$$

2.5 ESPACIOS FINITOS DE PROBABILIDAD

Sea S un espacio muestral finito, definido como sigue: $S = \{a_1, a_2, a_3, \dots, a_n\}$.

Se obtiene un espacio de probabilidad si a cada punto muestral a_i se le asigna un número real p_i denominado probabilidad de a_i , que cumpla las siguientes condiciones:

$$p_i \geq 0.$$

$$p_1 + p_2 + p_3 + \dots + p_n = 1.$$

La probabilidad de un suceso A se obtiene sumando las probabilidades de todos los puntos muestrales que conforman A .

Si todos los valores p_i son iguales, se obtiene un espacio finito equiprobable.

Ejemplo:

Al lanzar un dado se obtiene 6 resultados posibles y la probabilidad de cada punto es $\frac{1}{6}$.

Solución:

Si el suceso A consiste en obtener un número mayor que 3, entonces, $A = \{4, 5, 6\}$. Por tanto,

$$P(A) = P(\{4\}) + P(\{5\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$



Capítulo 3

Probabilidad Condicional



CAPÍTULO 3

PROBABILIDAD CONDICIONAL

Sea A un suceso arbitrario de un espacio muestral S , con $P(A) > 0$, y sea B otro suceso del mismo espacio muestral S . La probabilidad de que ocurra B una vez haya sucedido A , se llama *Probabilidad Condicional* de B dado A , se denota por $P(B/A)$ y se define así:

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

De aquí se deduce la *Ley del Producto* para sucesos dependientes, así:

$$P(A \cap B) = P(A)P(B/A).$$

Si la ocurrencia del suceso A no afecta a la ocurrencia de B , los sucesos A y B son *independientes*; en este caso, se tiene la siguiente *Ley del producto*:

$$P(A \cap B) = P(A)P(B).$$

Ejemplo:

Una urna contiene 4 balotas azules y 6 blancas. Si se extraen 2 balotas al azar, determinar la probabilidad de que la primera sea azul y la segunda blanca, con las siguientes condiciones:

Sin sustitución.

Con sustitución.

Solución:

Sean los siguientes sucesos:

A: se extrae una balota azul.

B: se extrae una balota blanca.

$$a) P(A \cap B) = P(A)P(B/A) \text{ entonces, } P(A \cap B) = \frac{4}{10} \times \frac{6}{9} = \frac{24}{90} = 0,266.$$

R/ La probabilidad de que la segunda balota sea blanca dado que la primera fue azul es del 26,6 %.

b) Si se extrae una balota y se regresa a la urna antes de extraer la segunda balota, entonces los sucesos A y B son independientes, por lo cual, $P(A \cap B) = P(A)P(B)$.

$$P(A \cap B) = \frac{4}{10} \times \frac{6}{10} = \frac{24}{100} = 0,24.$$

R/ Si se extrae una balota y se regresa a la urna antes de extraer la segunda balota, entonces La probabilidad de que la segunda balota sea blanca dado que la primera fue azul es del 24 %.

3.1 PROCESOS ESTOCÁSTICOS FINITOS

Una sucesión finita de experimentos, en los cuales cada uno tiene un número finito de resultados, con probabilidades conocidas, se llama *Proceso Estocástico Finito* y se representa por medio de un diagrama de árbol. Para encontrar la probabilidad en cualquiera de las ramas, se puede aplicar la *Ley del Producto*.

3.2 PROBABILIDAD TOTAL

Sean los sucesos A_1, A_2, \dots, A_n que forman una partición del espacio muestral S , entonces los eventos A_i son mutuamente excluyentes y su unión es S ; es decir, se cumple lo siguiente:

$$A_i \cap A_j = \emptyset \quad \forall i \neq j \quad \text{y} \quad \bigcup_{i=1}^n A_i = S.$$

Por lo tanto, para un evento cualquiera B de S , se cumple la siguiente relación:

$$B = (A_1 \cup A_2 \cup \dots \cup A_n) \cap B.$$

Aplicando propiedad distributiva de la intersección respecto a la unión de conjuntos, se tiene la siguiente igualdad de conjuntos:

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B).$$

Esta expresión se puede representar de la siguiente manera:

$$B = \bigcup_{i=1}^n (A_i \cap B).$$

Debido a que los sucesos A_i son independientes, entonces también lo son los sucesos $A_i \cap B, \forall i = 1, 2, \dots, n$.

Por lo tanto, $P(B)$ se puede expresar mediante la siguiente expresión:

$$P(B) = \sum_{i=1}^n P(A_i \cap B).$$

Al desarrollar la sumatoria, se obtiene la siguiente expresión:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B).$$

Aplicando la *Ley del Producto*, se obtiene la *Ley de Probabilidad Total*, que se expresa así:

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \dots + P(A_n)P(B/A_n).$$

La probabilidad $P(A_i)$ se denomina *Probabilidad Marginal*; $P(B/A_i)$ es una *Probabilidad Condicional*, donde la ocurrencia de B está condicionada por la ocurrencia de los eventos. A_i donde $i = 1, 2, \dots, n$.

La expresión $P(A_i \cap B)$ se denomina *Probabilidad Conjunta*; y $P(B)$ se denomina *Probabilidad Total*.

3.3 PROBABILIDAD DE BAYES

Dadas las condiciones de la *Probabilidad Total*, se trata de encontrar la probabilidad de que ocurra A_i dada la ocurrencia de B .

De la relación de *Probabilidad Condicional* aplicada a cada A_i dado B , se tiene la siguiente igualdad:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} \quad (*)$$

Dado que $P(A_i \cap B) = P(A_i)P(B/A_i)$, reemplazando en (*) se obtiene la siguiente expresión:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}.$$

Reemplazando $P(B)$, resulta la siguiente igualdad:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^n P(A_i \cap B)}, \forall i = 1, 2, \dots, n.$$

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \dots + P(A_n)P(B/A_n)}, \forall i = 1, 2, \dots, n.$$

Ejemplo:

Mediante estudios de tránsito en una vía, se sabe que el 25 % de los vehículos que transitan son de servicio público pesado; el 30 %, de servicios público liviano, y el 45 %, de servicio particular. También se sabe que la probabilidad de accidente en cada clase de vehículos es 0.2, 0.3 y 0.15, respectivamente. Se desea conocer:

La probabilidad de que ocurran accidentes en dicha vía.

Si hubo un accidente, calcular la probabilidad de que el vehículo accidentado sea de servicio particular.

Calcular la probabilidad de que no ocurra un accidente en la vía.

Solución:

Sean los siguientes eventos:

A_1 : Vehículos de servicio público pesado; entonces, $P(A_1) = 0,25$.

A_2 : Vehículos de servicio público liviano; entonces, $P(A_2) = 0,30$.

A_3 : Vehículos de servicio particular; entonces, $P(A_3) = 0,45$.

Sea el evento B : ocurre un accidente en la vía.

Con los datos del ejemplo y teniendo en cuenta que B ocurre por la ocurrencia de algunos de los eventos A_i con $i = 1, 2, 3$, se tiene las probabilidades que se indican en seguida. Vale destacar que, los problemas de probabilidad condicional se resuelven con facilidad utilizando una hoja de cálculo, tal como Excel de Microsoft Office.

| Tipo de vehículo | Eventos | Probabilidad marginal | Probabilidad condicional | Probabilidad conjunta |
|--------------------------|---------|-----------------------|--------------------------|----------------------------------|
| | | $P(A_i)$ | $P(B/A_i)$ | $P(A_i \cap B) = P(A_i)P(B/A_i)$ |
| Servicio público pesado | A_1 | 0,25 | 0,20 | 0,05 |
| Servicio público liviano | A_2 | 0,30 | 0,30 | 0,09 |
| Servicio particular | A_3 | 0,45 | 0,15 | 0,0675 |

$$a) P(B) = \sum_{i=1}^3 P(A_i \cap B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3).$$

$$P(B) = 0,25 \times 0,20 + 0,30 \times 0,30 + 0,45 \times 0,15 = 0,2075.$$

R/ La probabilidad de que ocurra un accidente en la vía es del 20,75 %.

$$b) P(A_3/B) = \frac{P(A_3)P(B/A_3)}{P(B)} = \frac{0,45 \times 0,15}{0,2075} = 0,32.$$

R/ La probabilidad de que, ocurrido un accidente, este corresponda a un vehículo de servicio particular es del 32 %.

c) Ahora se calcula la probabilidad de que no ocurra un accidente en la vía; esto es, $P(B^c)$ (ver Gráfico 2).

$$P(B^c) = P(A_1 \cap B^c) + P(A_2 \cap B^c) + P(A_3 \cap B^c), \text{ donde:}$$

$$P(A_i \cap B^c) = P(A_i - B) = P(A_i) - P(A_i \cap B), i = 1, 2, 3.$$

En el siguiente resumen, se presentan las probabilidades requeridas.

| Tipo de vehículo | Eventos | $P(A_i \cap B)$ | $P(A_i \cap B^c)$ | Probabilidad total $P(A_i)$ | Probabilidad total |
|--------------------------|---------|-----------------|-------------------|-----------------------------|--------------------|
| Servicio público pesado | A_1 | 0,05 | 0,20 | 0,25 | $P(A_1)$ |
| Servicio público liviano | A_2 | 0,09 | 0,21 | 0,30 | $P(A_2)$ |
| Servicio particular | A_3 | 0,0675 | 0,3825 | 0,45 | $P(A_3)$ |
| Probabilidad total | | 0,2075 | 0,7925 | 1,00 | |
| | | $P(B)$ | $P(B^c)$ | | |

$$P(B^c) = 1 - P(B) = 1 - 0,2075 = 0,7925.$$

R/ La probabilidad de que no se presenten accidentes en la vía es del 79,25 %.

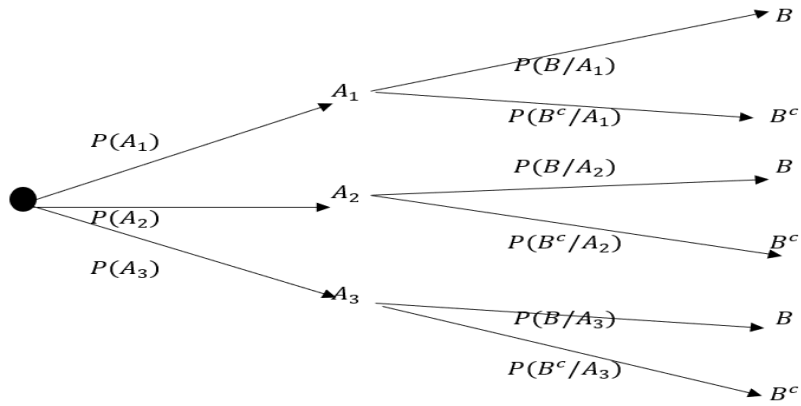


Gráfico 2. Ejemplo diagrama de árbol para cálculo de probabilidad total



Capítulo 4

Variables Aleatorias



CAPÍTULO 4

VARIABLES ALEATORIAS

Una variable aleatoria X de un espacio muestral S es una función que transforma los resultados de un experimento en números reales. Las variables aleatorias se clasifican en *Discretas* y en *Continuas*.

4.1 VARIABLES ALEATORIAS DISCRETAS

Una variable aleatoria X de un espacio muestral S es *Discreta* si su conjunto imagen es finito o infinito contable, así:

Conjunto finito: $X(S) = \{x_1, x_2, \dots, x_n\}$

Conjunto infinito contable: $X(S) = \{x_1, x_2, \dots\}$

Ejemplos:

Diez (10) personas se disponen a votar en favor o en contra de un proyecto.

Determinar: a) el espacio muestral correspondiente; b) el número de resultados favorables.

Solución:

a) $S = \{CCCCCCCCC, FCCCCCCCC, FFCCCCCCC, \dots, FFFFFFFF\}$

b) Sea X la variable aleatoria que indica el número de votos a favor, entonces,

$X = \{0, 1, 2, 3, \dots, 10\}$.

Cero resultados favorables: el evento {CCCCCCCCC} se presenta de $\binom{10}{0} = 1$ forma.

Un resultado favorable: el evento {FCCCCCCCC} se presenta de $\binom{10}{1} = 10$ formas.

Dos resultados favorables: el evento {FFCCCCCCCC} se presenta de $\binom{10}{2} = 45$ maneras.

⋮

Diez resultados favorables: el evento {FFFFFFFFFFFF} se presenta de $\binom{10}{10} = 1$ forma.

Por lo tanto, el número total de resultados favorables es:

$$\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \dots + \binom{10}{9} + \binom{10}{10} = 2^{10} = 1024.$$

Sea desea determinar el número de clientes que hacen cola en la ventanilla de un banco. Determinar el espacio muestral.

Solución:

Sea X la variable aleatoria que indica el número de clientes que hacen fila en la ventanilla de un banco.

$$X(S) = \{0, 1, 2, 3, \dots\}.$$

4.1.1 Espacio de probabilidad

Se obtiene un *Espacio de Probabilidad Discreto*, si a cada punto x_i se le asigna la probabilidad de que la variable aleatoria X tome el valor x_i , esto es:

$$f(x_i) = P(X = x_i).$$

La función f debe cumplir las siguientes propiedades:

a) $f(x_i) \geq 0$.

b) $\sum f(x_i) = 1$.

En este caso, la función f se denomina *Función de Probabilidad* y en casos sencillos se expresa en forma de tabla (ver Tabla 2).

Tabla 2. Ejemplo de función de probabilidad de variable aleatoria discreta

| | | | | | |
|----------|----------|----------|-----|----------|-------------------|
| x_i | x_1 | x_2 | ... | x_n | |
| $f(x_i)$ | $f(x_1)$ | $f(x_2)$ | ... | $f(x_n)$ | $\sum f(x_i) = 1$ |

4.1.2 Función acumulativa

La *Función Acumulativa* F de la variable aleatoria X se define de la siguiente manera:

$$F(a) = P(X \leq a) = \sum_{x \leq a} f(x).$$

Ejemplo:

Diez personas votarán a favor o en contra de un proyecto. Sea X la variable aleatoria que indica el número de votos a favor. Las probabilidades se calculan como se indica en seguida.

| X | Número de casos favorables | $f(X)$ | $F(X)$ |
|-------|----------------------------|----------|---------|
| 0 | 1 | 0,00098 | 0,00098 |
| 1 | 10 | 0,00977 | 0,01074 |
| 2 | 45 | 0,04395 | 0,05469 |
| 3 | 120 | 0,11719 | 0,17188 |
| 4 | 210 | 0,20508 | 0,37695 |
| 5 | 252 | 0,24609 | 0,62305 |
| 6 | 210 | 0,20508 | 0,82813 |
| 7 | 120 | 0,11719 | 0,94531 |
| 8 | 45 | 0,04395 | 0,98926 |
| 9 | 10 | 0,00977 | 0,99902 |
| 10 | 1 | 0,00098 | 1,00000 |
| Total | 1024 | 1 | |

La probabilidad de que haya exactamente 6 votos a favor está dada por la siguiente expresión:

$$P(X = 6) = f(6) = 0,20508.$$

La probabilidad de que máximo se obtengan 6 votos a favor, o equivalentemente, que haya menos de 7 votos a favor, se obtiene como sigue:

$$F(6) = P(X \leq 6) = \sum_{x \leq 6} f(x) = 0,82813$$

4.1.3 Media, varianza y desviación estándar

El valor esperado, promedio o media de una variable aleatoria discreta se denota por $E(X) = \mu$, y se define de la siguiente manera:

$$E(X) = \mu = \sum_{x_i \in S} x_i f(x_i).$$

En forma abreviada se escribe así:

$$E(X) = \mu = \sum x f(x).$$

La *Varianza* de una variable aleatoria, se denota por $V(X)$ o por σ^2 y se define como el promedio de los cuadrados de las desviaciones (diferencias) respecto a la media. Esto es,

$$Var(X) = \sigma^2 = \sum_{x_i \in S} (x_i - \mu)^2 f(x_i).$$

Esta expresión, se escribe en forma abreviada, así:

$$\sigma^2 = \sum (X - \mu)^2 f(x).$$

Al desarrollar el binomio y simplificar términos, se obtiene la siguiente expresión:

$$\sigma^2 = \sum (X^2 - 2X\mu + \mu^2) f(x) = \sum X^2 f(x) - 2\mu \sum X f(x) + \mu^2 \sum f(x);$$

$$\sigma^2 = \sum X^2 f(x) - \left[\sum X f(x) \right]^2 = E(X^2) - [E(X)]^2;$$

$$\sigma^2 = E(X^2) - [E(X)]^2.$$

La *Desviación Típica* o *Desviación Estándar* se denota por σ , y se define como la raíz cuadrada no negativa de la varianza; esto es, $\sigma = \sqrt{\text{Var}(X)} = \sqrt{\sigma^2}$.

Ejemplo:

Diez (10) personas votarán a favor o en contra de un proyecto. La variable aleatoria X representa los votos a favor. Determinar lo siguiente:

Número de resultados posibles.

Probabilidad de cada uno de los puntos.

Valor esperado $E(X)$, Varianza $\text{Var}(X)$ y Desviación Estándar $\sqrt{\text{Var}(X)}$.

Interpretar los resultados.

Solución:

| X | # | $f(X)$ | $Xf(X)$ | $(X - \mu)^2 f(X)$ |
|-----|------|---------|---------|--------------------|
| 0 | 1 | 0,00098 | 0,00000 | 0,02441 |
| 1 | 10 | 0,00977 | 0,00977 | 0,15625 |
| 2 | 45 | 0,04395 | 0,08789 | 0,39551 |
| 3 | 120 | 0,11719 | 0,35156 | 0,46875 |
| 4 | 210 | 0,20508 | 0,82031 | 0,20508 |
| 5 | 252 | 0,24609 | 1,23047 | 0,00000 |
| 6 | 210 | 0,20508 | 1,23047 | 0,20508 |
| 7 | 120 | 0,11719 | 0,82031 | 0,46875 |
| 8 | 45 | 0,04395 | 0,35156 | 0,39551 |
| 9 | 10 | 0,00977 | 0,08789 | 0,15625 |
| 10 | 1 | 0,00098 | 0,00977 | 0,02441 |
| | 1024 | 1,00000 | 5 | 2,5 |
| | | | $E(X)$ | $\text{Var}(X)$ |

R/

Número de resultados posibles: $n = 2^{10} = 1024$.

Valor esperado: $E(X) = \mu = 5$ votos a favor.

Varianza: $Var(X) = \sigma^2 = 2,5$.

Desviación estándar: $\sigma = \sqrt{Var(X)} = \sqrt{2,5} = 1,58$.

¿Cuál es la probabilidad de que el número de votos a favor se encuentre dentro del rango definido por la media \pm una desviación estándar?

Dado que la Desviación Estándar $\sigma = 1,58$, entonces el número de votos a favor puede estar entre: $\mu - \sigma = 5 - 1,58 = 3,42$ y $\mu + \sigma = 5 + 1,58 = 6,58$.

Como la variable aleatoria es discreta, entonces se puede afirmar que la votación puede estar entre 3 y 7 votos a favor. La probabilidad de que esto ocurra está dada por:

$$P(3 \leq X \leq 7) = f(3) + f(4) + f(5) + f(6) + f(7) = 89,06 \%$$

Esta probabilidad confirma que, efectivamente, los resultados de la votación en cuanto a votos a favor, oscila entre 3 y 7.

En el Gráfico 3, la función $f(X)$ representa la probabilidad de ocurrencia de cada uno de los puntos del espacio muestral de variable aleatoria discreta.

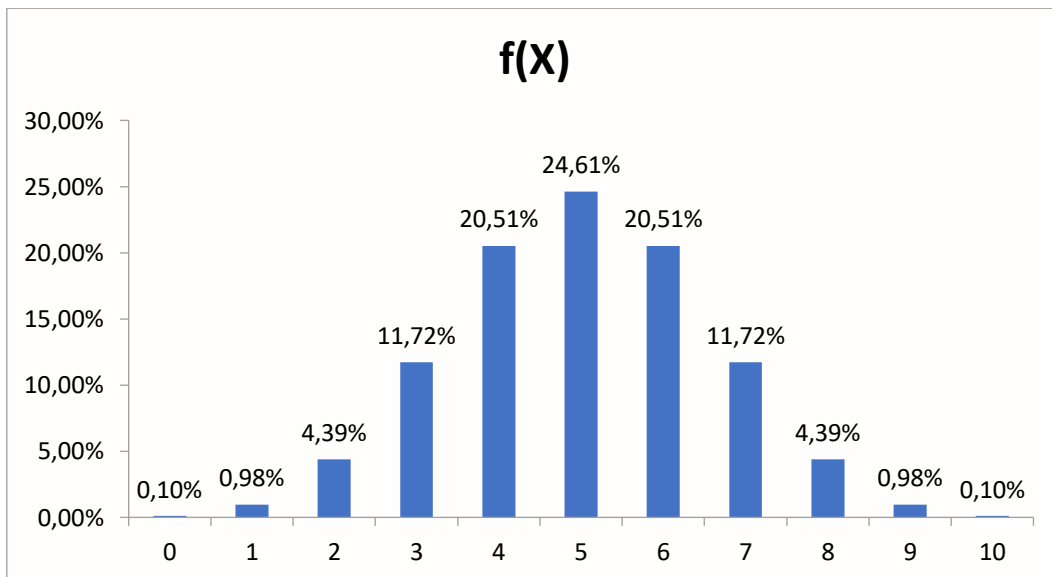


Gráfico 3. Ejemplo de representación de una función de probabilidad

4.2 VARIABLES ALEATORIAS CONTINUAS

Una variable aleatoria X de un espacio muestral S es *Continua* si su conjunto imagen $X(S)$ es continuo. En este caso:

$$X(S) = \{x \in \mathbb{R} / -\infty < X < +\infty\}.$$

4.2.1 Espacio de probabilidad de una variable aleatoria continua

Se obtiene un *Espacio de Probabilidad Continuo* si a cada punto x se le asigna un número real $f(x)$, de modo que se cumplan las siguientes condiciones:

$$a) f(x) \geq 0, \forall x \in X(S).$$

$$b) \int f(x)dx = 1.$$

En este caso, la función f se denomina *Función de Probabilidad* o *Función de Densidad*. Esta función no representa la probabilidad, sino las ordenadas de la curva en cualquier punto x . Para determinar una probabilidad se requiere un intervalo (a, b) .

Ejemplo:

Sea la variable aleatoria X definida como sigue:

$$f(x) = \begin{cases} kx^2 & \text{si } 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Determinar el valor de k , de modo que se cumplan las dos condiciones siguientes:

$$a) f(x) \geq 0.$$

$$b) \int f(x)dx = 1.$$

Solución:

$$\int_0^2 kx^2 dx = k \left(\frac{x^3}{3} \right) = k * \frac{8}{3} = 1 \rightarrow k = \frac{3}{8}.$$

Con este valor de k se cumple que: $\int_0^2 kx^2 dx = \int_0^2 \frac{3}{8} x^2 dx = \frac{3}{8} \int_0^2 x^2 dx = 1$.

Entonces, la función de densidad correspondiente es:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{si } 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Nótese que $\frac{3}{8}x^2 \geq 0 \forall x: 0 < x < 2$.

4.2.2 Función acumulativa

La Función Acumulativa F se define de la siguiente manera:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Las probabilidades se calculan como *Áreas Bajo la Curva $f(x)$ en un intervalo (a, b)* , así:

$$P(a < X < b) = \int_a^b f(x) dx = F(b) - F(a).$$

Ejemplo:

Sea la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{si } 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Determinar la probabilidad de que la variable aleatoria continua X cumpla la condición: $1 < X < 2$.

Solución:

$$P(1 < X < 2) = \int_1^2 f(x) dx = F(2) - F(1) = \frac{3}{8} \left(\frac{1}{3} 2^3 - \frac{1}{3} 1^3 \right) = 1 - \frac{1}{8} = \frac{7}{8} = 0,875.$$

$$P(1 < X < 2) = 87,5 \%$$

R/ La probabilidad pedida es del 87,5 %.

4.2.3 Media, Varianza y Desviación Estándar

El valor esperado, promedio o *Media* de una variable aleatoria continua se denota por $E(X)$ o por μ , y se calcula de la siguiente forma:

$$E(X) = \mu = \int xf(x)dx$$

La *Varianza* se define como el promedio de los cuadrados de las desviaciones (diferencias) con respecto a la *Media*; entonces:

$$Var(X) = \sigma^2 = \int (x - \mu)^2 f(x)dx$$

Al desarrollar el cuadrado y simplificar términos, se obtiene:

$$Var(X) = \int x^2 f(x)dx - \left[\int xf(x)dx \right]^2 = E(X^2) - [E(X)]^2.$$

La *Desviación Típica* o *Desviación Estándar* se denota por σ y se define como la raíz cuadrada no negativa de la Varianza; esto es, $\sigma = \sqrt{var(X)} = \sqrt{\sigma^2}$.

Ejemplo:

Determinar la Media, Varianza y Desviación Estándar para la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{si } 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Solución:

$$E(X) = \mu = \int xf(x)dx = \frac{3}{8} \int_0^2 x^3 dx = \frac{3}{8} \left[\frac{x^4}{4} \right] = \frac{3}{8} * \frac{16}{4} = 1,5.$$

$$Var(X) = \int x^2 f(x)dx - \left[\int xf(x)dx \right]^2 = E(X^2) - [E(X)]^2.$$

$$E(X^2) = \int x^2 f(x)dx = \frac{3}{8} \int_0^2 x^4 dx = \frac{3}{8} \left[\frac{x^5}{5} \right] = \frac{3}{8} * \frac{32}{5} = 2,4.$$

Entonces:

$$Var(X) = 2,4 - 1,5^2 = 0,15.$$

$$\sigma = \sqrt{var(X)} = \sqrt{\sigma^2} = \sqrt{0,15} = 0,3873.$$



Capítulo 5

Modelos de Probabilidad



CAPÍTULO 5

MODELOS DE PROBABILIDAD

Según el tipo de variable aleatoria, bien sea discreta o continua, existen sendos modelos de distribución de probabilidad, discretos y continuos.

5.1 MODELOS DISCRETOS

En esta sección, se analizan los siguientes modelos de distribución de variable aleatoria discreta:

Bernoulli

Binomial

Poisson

Hipergeométrico

Binomial negativo

Geométrico

Multinomial

5.1.1 Modelo de Bernoulli

Todo experimento que tenga sólo dos resultados posibles, que sean exhaustivos y excluyentes, constituye una prueba de Bernoulli; a uno de los resultados se le denomina Éxito y al otro Fracaso.

La variable aleatoria asigna el valor 1 al éxito y 0 al fracaso; las probabilidades correspondientes son p y q .

Dado que la suma de las probabilidades debe ser 1, entonces $q = 1 - p$.

Los estadígrafos (medidas estadísticas) para la variable aleatoria del modelo de Bernoulli, son:

Media: $\mu = p$.

Varianza: $\sigma^2 = pq$.

Desviación estándar: $\sigma = \sqrt{pq}$.

Ejemplo:

Sea el experimento: lanzar una vez una moneda corriente. Determinar los estadígrafos correspondientes.

Solución:

La probabilidad de que resulte cara es $P = \frac{1}{2} = 0,5$.

Entonces, media: $\mu = p = 0,5$. Como $p + q = 1$, entonces $q = 1 - p = 1 - \frac{1}{2} = 0,5$.

Varianza: $\sigma^2 = pq = 0,5 \times 0,5 = 0,25$.

Desviación estándar: $\sigma = \sqrt{pq} = \sqrt{0,25} = 0,5$.

5.1.2 Modelo Binomial

Este modelo se aplica cuando la variable aleatoria cumple las siguientes condiciones:

Se trata de una prueba de Bernoulli que se repite n veces.

Las pruebas son independientes (p constante).

La variable aleatoria X indica el número de éxitos en la muestra.

La función de probabilidad se define de la siguiente manera:

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{(n-x)} \text{ donde } x = 0, 1, 2, \dots, n.$$

Se puede demostrar que se cumple la siguiente propiedad: $\sum f(x) = 1$.

Para la variable aleatoria del Modelo Binomial, se tienen los siguientes estadígrafos:

$$E(X) = \mu = np .$$

$$Var(X) = \sigma^2 = npq .$$

$$\sigma = \sqrt{npq} .$$

La función acumulativa correspondiente es:

$$F(a) = \sum_{x \leq a} \frac{n!}{x!(n-x)!} p^x q^{n-x} .$$

Ejemplo:

Debido a las altas tasas de interés, una compañía informa que el 30 % de sus cuentas por cobrar están vencidas. Si un contador escoge aleatoriamente 5 de esas cuentas, determinar la probabilidad que se cumpla lo siguiente:

Ninguna cuenta esté vencida.

Exactamente 2 cuentas estén vencidas.

Solución:

Sea la variable aleatoria X , número de cuentas vencidas, entonces,

Probabilidad de éxito: $p = 0,3$.

Probabilidad de fracaso: $q = 0,7$.

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{(n-x)} = \frac{5!}{x!(5-x)!} 0,3^x 0,7^{(5-x)} .$$

Probabilidad de que ninguna cuenta resulte vencida:

$$f(0) = \frac{5!}{0!(5-0)!} (0,3)^0 (0,7)^{(5-0)} = 0,16807 .$$

Probabilidad de que exactamente 2 cuentas resulten vencidas:

$$f(2) = \frac{5!}{2!(5-2)!} (0,3)^2 (0,7)^{(5-2)} = 0,3087 .$$

Cálculo de probabilidad de la distribución Binomial con Microsoft Excel

En la barra de herramienta de Excel, aparece el símbolo: fx (insertar función), se despliega una ventana de diálogo, Buscar Estadísticas, seleccionar DISTR.BINOM.N,

Aceptar; digitar X : número de éxitos, ensayos o tamaño de muestra, probabilidad de éxito y en la opción acumulado, 0 o falso, para obtener la probabilidad individual del punto X es decir $f(x)$ y si teclaea 1 o verdadero, se obtiene la probabilidad acumulada hasta ese punto $F(x)$ (ver Imagen 1).

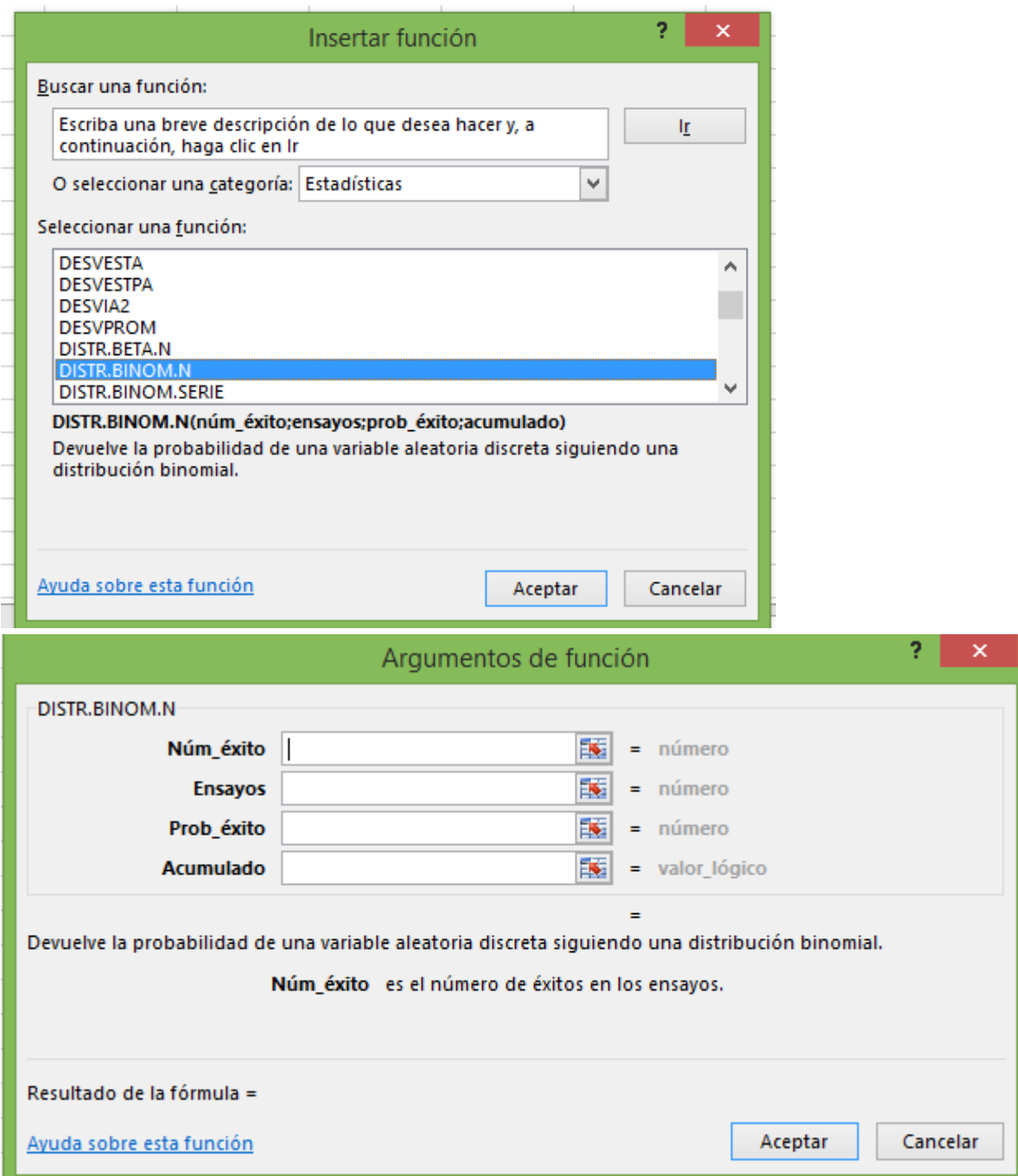


Imagen 1. Microsoft Excel: argumentos de función DISTR.BINOM.N

5.1.3 Modelo de Poisson

El Modelo de Poisson se aplica cuando se observa un fenómeno en una unidad de tiempo o de espacio, donde el valor esperado $E(x) = \mu$ es el promedio de éxitos por unidad de tiempo.

El Modelo de Poisson es una aproximación del binomial cuando $n \rightarrow \infty$ y $p \rightarrow 0$; esto es,

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{e^{-\mu} \mu^x}{x!}.$$

En consecuencia, la función de probabilidad de Poisson se define de la siguiente manera:

$$f(x) = \frac{e^{-\mu} \mu^x}{x!}, \text{ donde } X = 0, 1, 2, \dots \text{ corresponde al número de éxitos.}$$

Los parámetros del Modelo de Poisson son:

$$\text{Media: } \mu = \lambda = np.$$

$$\text{Varianza: } \text{Var}(X) = \sigma^2 = npq = np(1-p).$$

Ahora bien, teniendo en cuenta que $p \rightarrow 0$, entonces, $1-p \rightarrow 1$, en este caso se cumple que, $\sigma^2 = np(1-p) = np = \mu = \lambda$.

Ejemplo:

Se sabe que el promedio de clientes que llegan a la ventanilla de un banco es de 3 por minuto. Determinar la probabilidad de que en el próximo minuto lleguen 4 clientes.

Solución:

$$\text{La función de probabilidad correspondiente es: } f(x) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-3} 3^x}{x!}.$$

$$\text{Por tanto, la probabilidad pedida es: } f(4) = \frac{e^{-3} 3^4}{4!} = 0,1680.$$

R/ Si el promedio de clientes que llega a la ventanilla de un banco es de 3 por minuto, entonces, la probabilidad de que al siguiente minuto lleguen 4 clientes, es 0,1680 = 16,80 %.

Cálculo de probabilidad de la distribución de Poisson con Microsoft Excel

En la barra de herramienta de Excel, aparece el símbolo: f_x (insertar función), se despliega una ventana de diálogo, Buscar Estadísticas, seleccionar POISSON.DIST, Aceptar; digitar número de éxitos, promedio o media y en la opción acumulado, 0 o falso para obtener la probabilidad individual del punto X , esto es, $f(x)$; y si teclea 1 o verdadero, se obtiene la probabilidad acumulada hasta ese punto, $F(x)$ (ver Imagen 2).

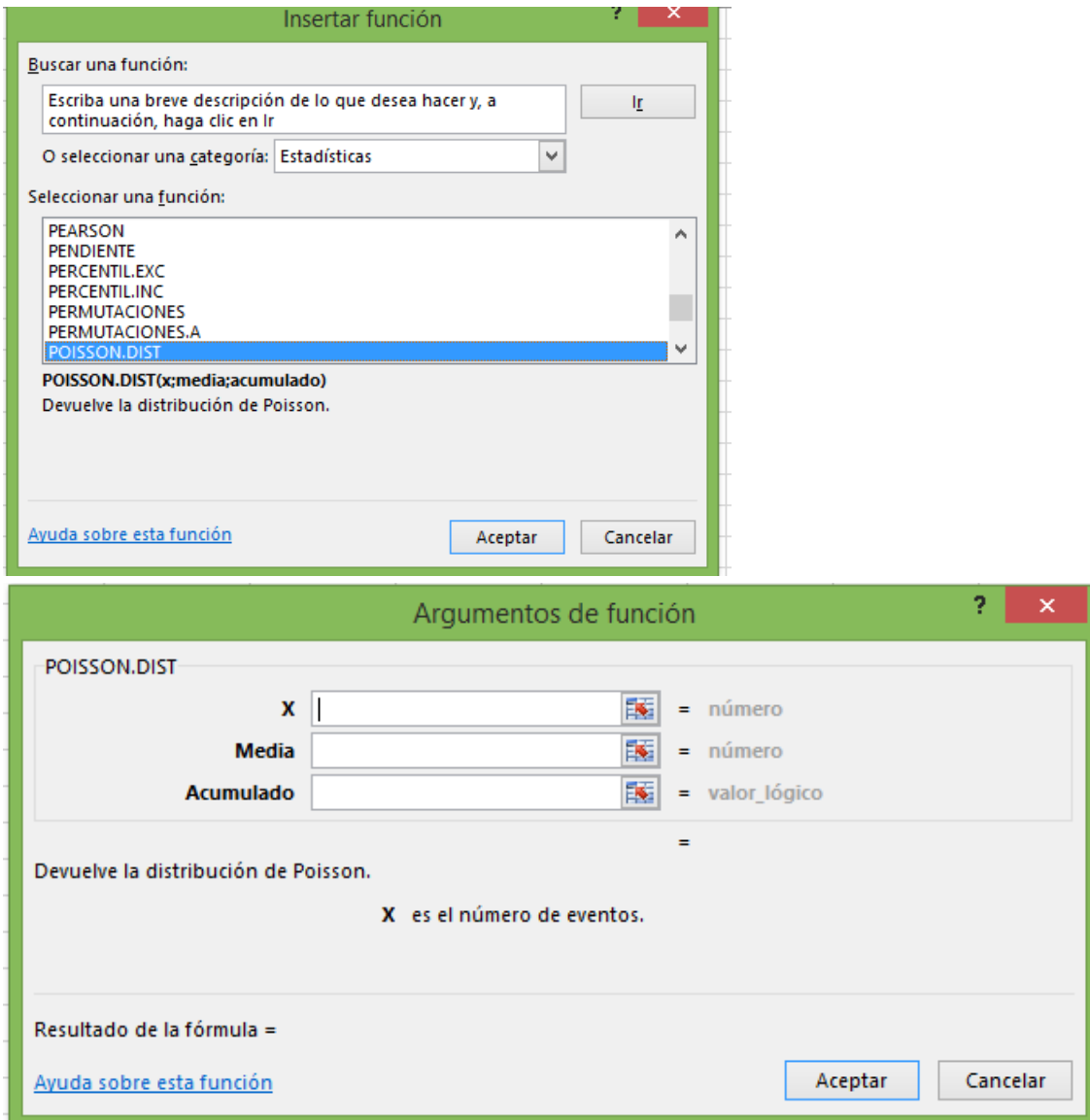


Imagen 2. Microsoft Excel: argumentos de la función POISSON.DIST.

5.1.4 Modelo Hipergeométrico

Para aplicar un Modelo Hipergeométrico, el problema debe cumplir las siguientes condiciones:

Consistir en una prueba de Bernoulli que se repita n veces.

Las pruebas deben ser dependientes (p es variable).

La variable aleatoria X indica el número de éxitos en la muestra.

Para analizar la dependencia o independencia de las pruebas, se debe tener en cuenta el tamaño de la población y el tipo de muestra seleccionada, de la siguiente forma:

Si la población es muy grande o infinita, no importa que la muestra se haya tomado con o sin sustitución; en ambos casos las pruebas son independientes.

Si la población es finita y pequeña, se presentan dos casos:

Si la muestra se toma con sustitución, entonces las pruebas son *independientes*.

Si la muestra se toma sin sustitución, entonces, las pruebas son *dependientes*.

La función de probabilidad hipergeométrica se expresa de la siguiente manera:

$$f(x) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}}.$$

Donde:

N_1 : representa el total de éxitos en la población.

N_2 : total de fracasos en la población.

$$N = N_1 + N_2$$

n : tamaño de la muestra.

x : número de éxitos en la muestra.

$n - x$: número de fracasos en la muestra.

Los estadígrafos de la variable aleatoria del Modelo Hipergeométrico son:

$$\text{Media: } E(X) = n \times \frac{N_1}{N}.$$

$$\text{Varianza: } \text{Var}(X) = \sigma^2 = n \left(\frac{N_1}{N} \right) \left(\frac{N_2}{N} \right) \left(\frac{N-n}{N-1} \right).$$

Ejemplo:

Un curso de Estadística Inferencial está conformado por 10 hombres y 8 mujeres; si se toma una muestra al azar de 5 estudiantes, determinar la probabilidad de que la muestra contenga 3 hombres y 2 mujeres.

Solución:

Sea la variable X que hace corresponder el número de hombres en la muestra.

La función de probabilidad correspondiente es:

$$f(x) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{x} \binom{8}{5-x}}{\binom{18}{5}}.$$

$$f(3) = \frac{\binom{10}{3} \binom{8}{5-3}}{\binom{18}{5}} = 0,3921.$$

R/ La probabilidad que la muestra contenga 3 hombres y 2 mujeres es: 0,3921.

Probabilidad distribución hipergeométrica con Microsoft Excel

En la barra de herramienta de Excel, aparece el símbolo: fx (insertar función), se despliega una ventana de diálogo, Buscar Estadísticas, seleccionar DISTR.HIPERGEOM.N, Aceptar; digitar número de éxitos, tamaño de muestra, número de éxitos en la población, población y, en la opción acumulado, 0 o falso para obtener la probabilidad individual del punto X , es decir $f(x)$ y si teclea 1 o verdadero, se obtiene la probabilidad acumulada hasta ese punto, $F(x)$ (ver Imagen 3).

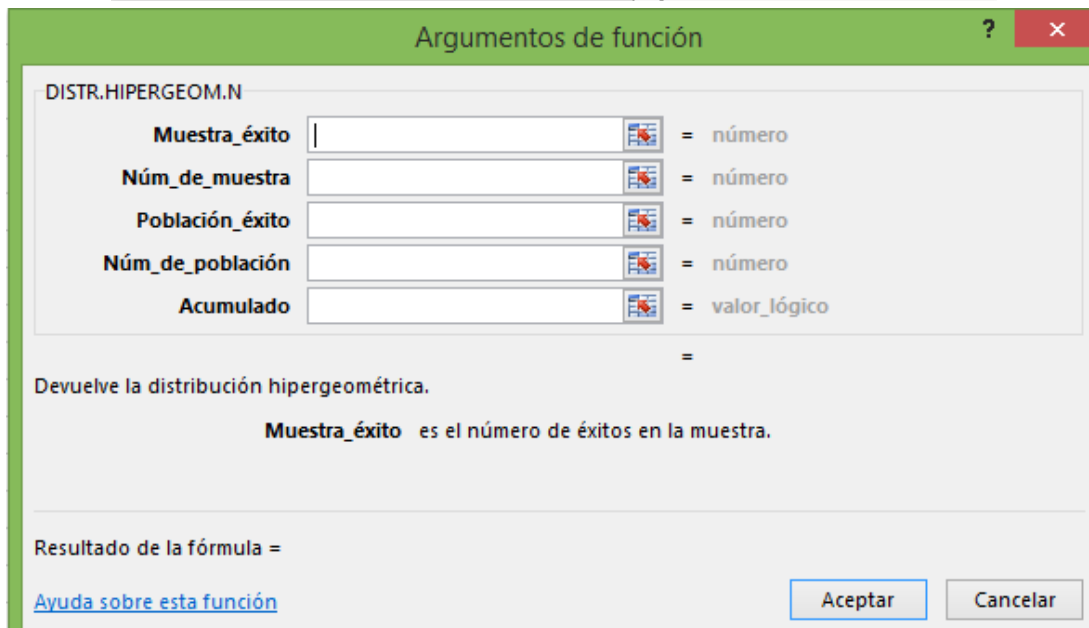
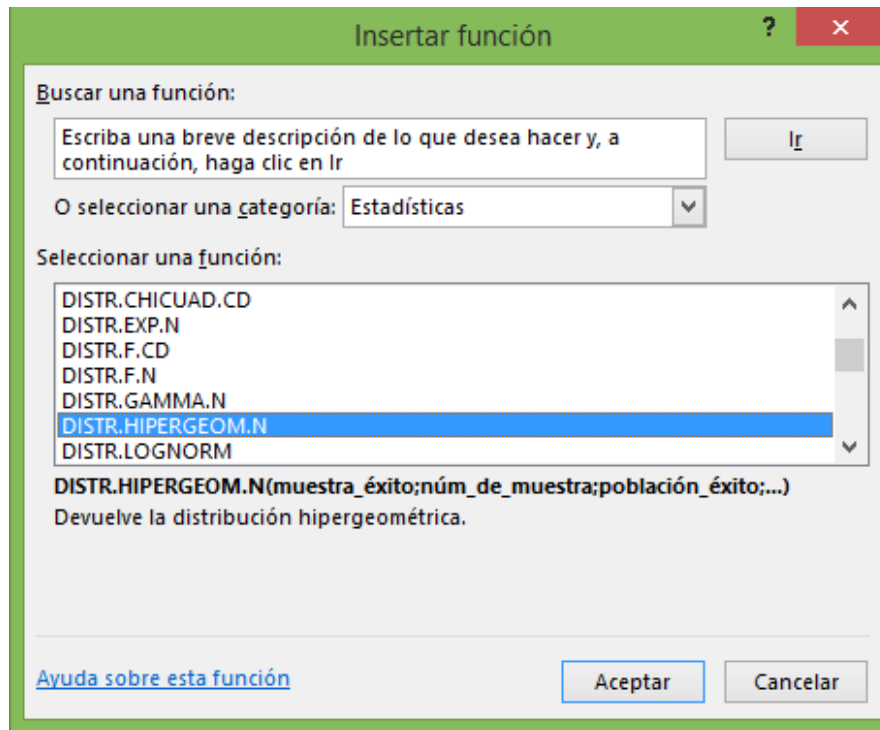


Imagen 3. Microsoft Excel: función DISTR.HIPERGEOM.N

5.1.5 Modelo Binomial negativo

El modelo Binomial negativo se aplica cuando se desea calcular la probabilidad de obtener el k -ésimo éxito, en la n -ésima prueba de Bernoulli.

La variable aleatoria X es el número de pruebas que se realizan y K el número de éxitos, p es la probabilidad de éxito, $q = 1 - p$ probabilidad de fracaso en cada prueba.

$$f(x, k, p) = \binom{x-1}{k-1} p^k q^{x-k}.$$

k = número de éxitos; $K = 1, 2, 3, \dots$

X = número de pruebas; $X = 1, 2, \dots$

p = probabilidad de éxito.

q = probabilidad de fracaso.

$$E(x) = \frac{k}{p}.$$

$$Var(x) = \frac{kq}{p^2}.$$

Ejemplo:

Si al programa de licenciatura en matemáticas del 2010, se matricularon 60 estudiantes, de los cuales el 15 % son mujeres, ¿cuál es la probabilidad de que el 7° matriculado sea la 2ª mujer.

k = número de mujeres matriculadas; $k = 1, 2, \dots$

X = número de alumnos matriculados; $X = 1, 2, \dots, 60$.

$p = 0,15$.

$q = 0,85$.

$$f(x, k, p) = \binom{x-1}{k-1} p^k q^{x-k}.$$

Que el 7° matriculado sea la 2ª mujer:

$$f(7, 2, 0,15) = \binom{6}{1} \times 0,15^2 \times 0,85^5 = 0,0599 = 5,99 \%.$$

R/ La probabilidad es del 5,99 %.

En la barra de herramienta de Excel, aparece el símbolo: f_x (insertar función); al pulsar clic, se despliega una ventana de diálogo; Buscar Estadísticas, seleccionar NEGBINOM.DIST, Aceptar. Luego, digitar número de fracasos, número de éxitos, probabilidad de éxito y en la opción acumulado, se digita 0 o falso, para obtener la probabilidad individual del punto X , es decir $f(x)$; y si digita 1 o verdadero, se obtiene la probabilidad acumulada hasta ese punto, $F(x)$ (ver Imagen 4).

Probabilidad distribución binomial negativa con Microsoft Excel

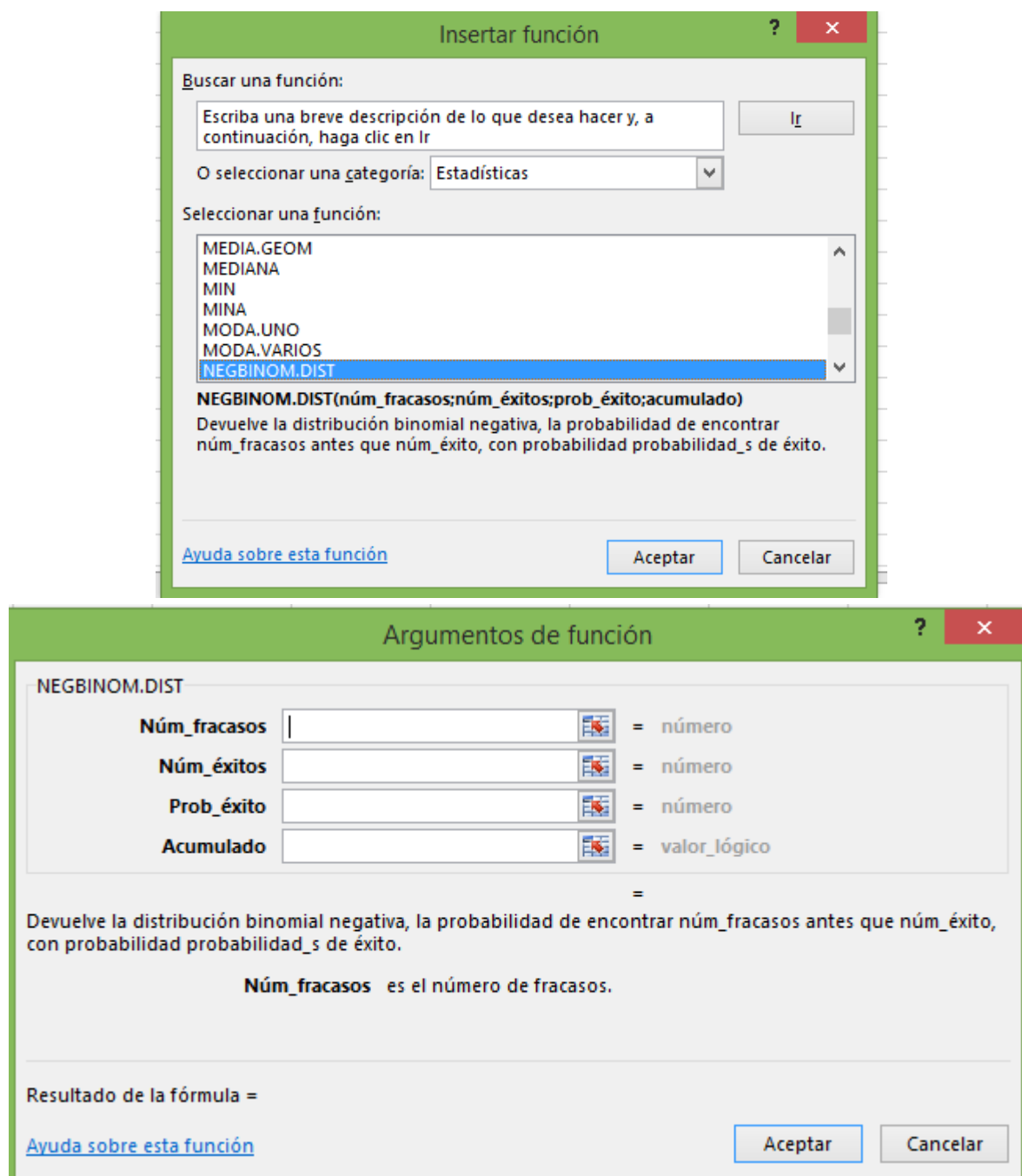


Imagen 4. Microsoft Excel: argumentos de la función NEGBINOM.DIST.

5.1.6 Modelo Geométrico

Un experimento se ajusta al Modelo Geométrico si se cumple las siguientes condiciones:

Existe una prueba de Bernoulli que se repite n veces.
Las pruebas son independientes (p constante).

La variable aleatoria X , indica el número de pruebas que se realizan hasta lograr el primer éxito.

El modelo geométrico es un caso particular del modelo binomial negativo cuando $K = 1$.

La función de probabilidad correspondiente es: $f(x) = pq^{x-1}$.

Para la variable aleatoria X del Modelo Geométrico, se tienen los siguientes parámetros:

$$\text{Media } E(X) = \mu = \frac{1}{p}.$$
$$\text{Varianza: } Var(X) = \sigma^2 = \frac{q}{p^2}.$$

Ejemplo:

Un vendedor realiza visita a 10 clientes potenciales; si la probabilidad de que hagan una compra es del 80 %, determinar la probabilidad que el objetivo se logre con el tercer cliente.

Solución:

Sea X el número de clientes visitados.

La función de probabilidad correspondiente es:

$$f(x) = pq^{x-1} = (0,8)(0,2)^{x-1}.$$

Por tanto, la probabilidad pedida es: $f(3) = (0,8)(0,2)^{3-1} = 0,032$.

R/ La probabilidad de que se logre una venta con el tercer cliente visitado es del 3,2 %.

Cálculo de probabilidad de la Distribución Geométrica (binomial negativa) con Microsoft Excel (ver Imagen 5).

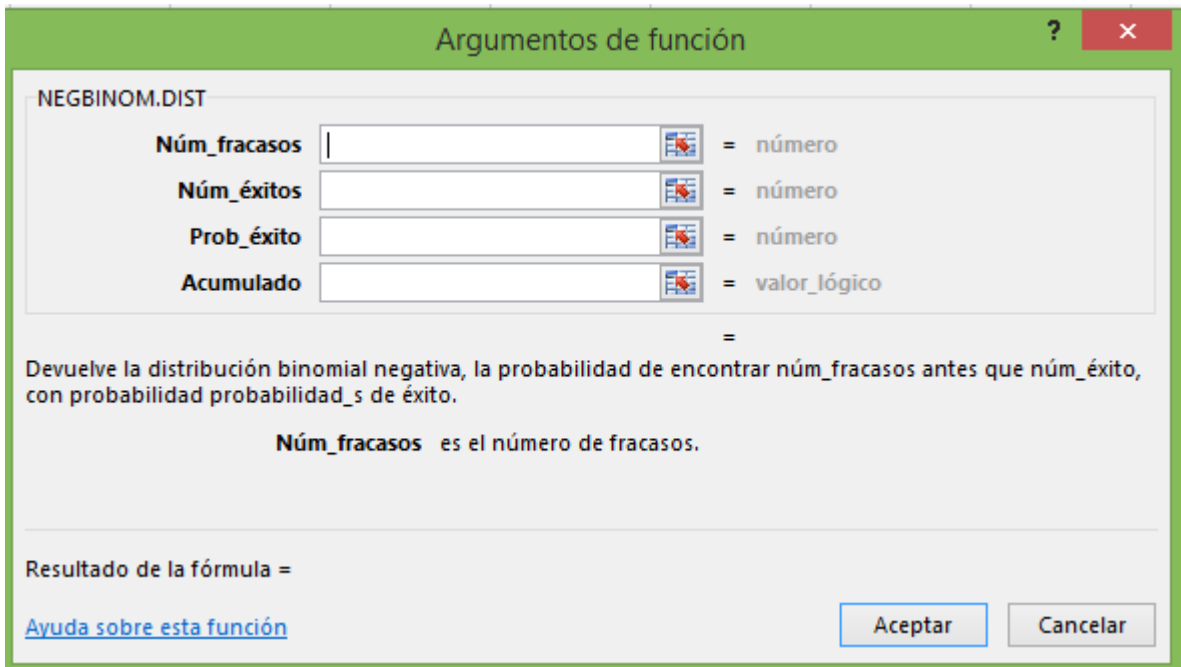


Imagen 5. Microsoft Excel: argumentos de función, NEGBINOM.DIST.

5.1.7 Modelo Multinomial

Si en una prueba hay más de 2 resultados posibles, no se puede hablar de éxito ni fracaso; en su lugar, se designa por x_i , $i = 1, 2, \dots, k$, al número de veces que se repite cada uno de ellos en la muestra de tamaño n .

Sean las variables $x_1, x_2, x_3, \dots, x_k$ como se indica enseguida:

x_1 : representa el número de veces que aparece el primer resultado.

x_2 : representa el número de veces que aparece el segundo resultado.

x_3 : representa el número de veces que aparece el tercer resultado.

⋮

x_k : representa el número de veces que aparece el k -ésimo resultado.

Además, con cada $x_i = 0, 1, 2, \dots, n$ se cumple que: $\sum_{i=1}^k x_i = n$

La función de probabilidad multinomial correspondiente es:

$$P(X_1 = a_1, X_2 = a_2, \dots, X_k = a_k) = \frac{n!}{a_1! a_2! \dots a_k!} P_1^{a_1} P_2^{a_2} \dots P_k^{a_k}.$$

La Media y la Varianza para cada variable x_i se calculan así:

$$\text{Media: } E(X_i) = np_i.$$

$$\text{Varianza: } \text{Var}(X_i) = np_i q_i$$

Ejemplo:

En una empresa de la ciudad de Pasto, los empleados se clasifican según su grado de preparación así: el 40% son profesionales, el 35 % son bachilleres y el 25 % restante tienen solamente educación primaria. Si se toma una muestra de 12 empleados, determinar la probabilidad de que en la muestra haya 5 profesionales, 4 bachilleres y 3 de primaria.

Solución:

Función de probabilidad multinomial:

$$P(X_1 = a_1, X_2 = a_2, \dots, X_k = a_k) = \frac{n!}{a_1! a_2! \dots a_k!} P_1^{a_1} P_2^{a_2} \dots P_k^{a_k}.$$

Con los datos del problema, se tiene:

$$P(X_1 = 5, X_2 = 4, X_3 = 3) = \frac{12!}{5! 4! 3!} (0,40)^5 (0,35)^4 (0,25)^3 = 0,0662.$$

R/ La probabilidad de obtener una muestra con 5 profesionales, 4 bachilleres y 3 de primaria es 0,0662.

5.2 MODELOS CONTINUOS

En esta sección se analizan los siguientes modelos de distribución de variable aleatoria continua:

Normal.

Ji-cuadrado.

t de Student.

F de Fisher.

5.2.1 Distribución Normal

Una variable aleatoria x de una distribución normal toma cualquier valor entre $-\infty$ y $+\infty$.

La función de probabilidad o función de densidad (planteada por Gauss), denominada Función Normal, es la siguiente:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ donde } -\infty < x < +\infty; \mu: \text{ media y } \sigma \text{ desviación estándar de } x.$$

La función $f(x)$ representa la ordenada de la variable en el punto x .

La probabilidad en una función continua se determina como el área bajo la curva de la función en un intervalo (a, b) , es decir

$$P(a < x < b) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = F(b) - F(a).$$

La gráfica de esta función es una curva en forma de campana; comúnmente se la conoce como Campana de Gauss.

El

Gráfico 4 indica probabilidades en diferentes intervalos. En los valores $X_1 = \mu - \sigma$ y $X_2 = \mu + \sigma$ se ubican los puntos de inflexión, y la probabilidad de que $P(\mu - \sigma < X < \mu + \sigma) = 0.6823 = 68,23 \%$.

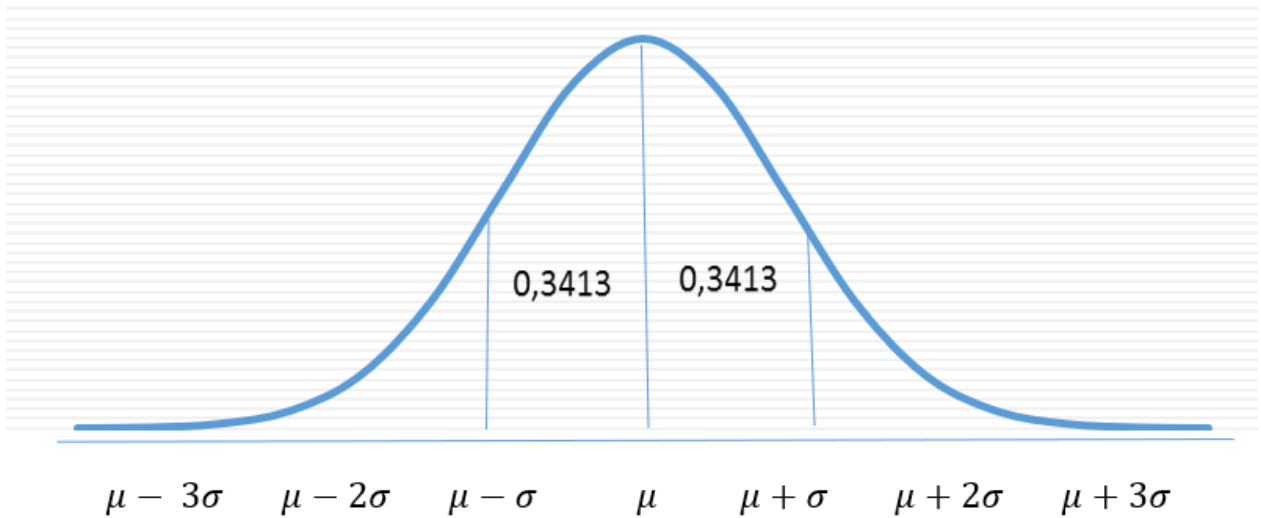


Gráfico 4. Representación de una distribución normal

Por ejemplo:

$$P(\mu < X < \mu + \sigma) = 0,3413 = 34,13 \%$$

Partiendo del hecho que la curva es simétrica alrededor de la media μ y que el área total es 1, se puede afirmar que:

$$P(x > \mu) = \int_{\mu}^{+\infty} f(x)dx = 0,5 .$$

$$P(x < \mu) = \int_{-\infty}^{\mu} f(x)dx = 0,5 .$$

Se observa que, si cambia la media μ , entonces la curva se desplaza en el eje x , a la derecha o a la izquierda, y si cambia la desviación estándar σ , la curva se concentra o se dispersa de la media, es decir, los datos se agrupan alrededor de la media o se alejan de ella. En las gráficas 5 y 6, se puede observar lo indicado.

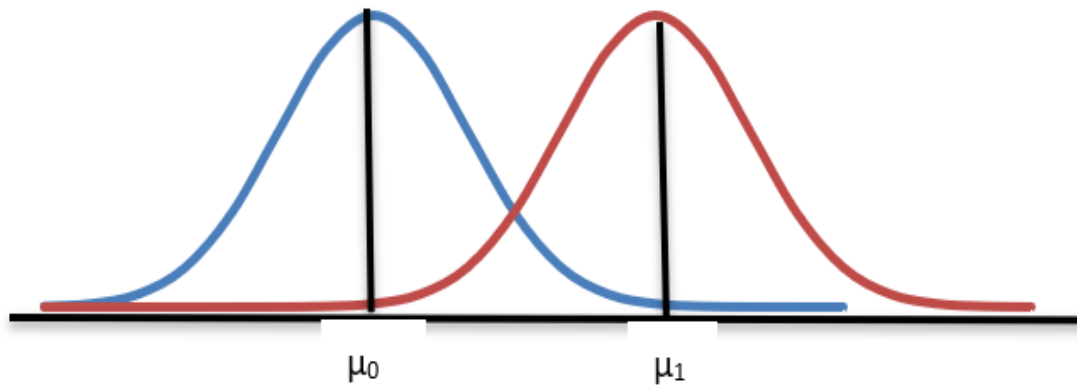


Gráfico 5. Representación de distribución normal comparando dos medias

Se puede observar en el Gráfico 5 que la curva de la izquierda tiene una media μ_0 y la curva de la derecha media μ_1 , donde $\mu_1 > \mu_0$.

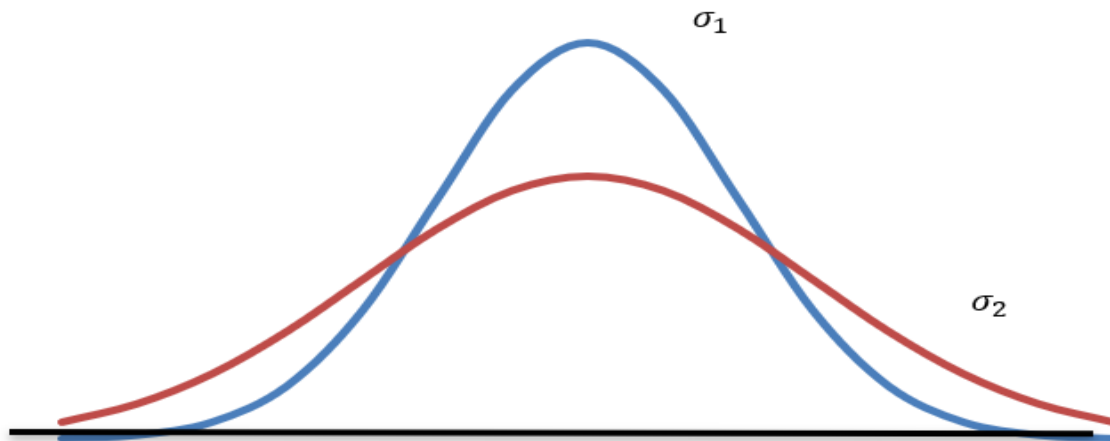


Gráfico 6. Representación de distribución normal comparando desviaciones estándar

Se puede observar en el Gráfico 6 que, la curva en color azul tiene desviación estándar σ_1 y la curva en color rojo, desviación estándar σ_2 , donde $\sigma_1 < \sigma_2$.

Dado que los valores de μ y σ determinan infinidad de curvas normales, entonces se utiliza una *Transformación Lineal*, con el fin de estandarizar la distribución en una sola curva, con media $\mu = 0$ y desviación estándar $\sigma = 1$.

La variable que transforma a x se llama *Variable Estandarizada* y se representa por z . La ecuación para la transformación es la siguiente:

$$z = \frac{x - \mu}{\sigma}$$

La función correspondiente para la variable z estandarizada es como sigue:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \text{donde } E(Z) = 0 \text{ y } Var(z) = \sigma^2 = 1.$$

Observaciones:

En la naturaleza existen muchos fenómenos que se distribuyen normalmente.

La distribución normal, bajo ciertas condiciones, es una excelente aproximación para distribuciones discretas como la binomial y Poisson.

Algunas distribuciones muestrales con tamaños de muestras grandes se pueden modelar con la función normal.

Cuando una variable aleatoria x es normal, para calcular probabilidades, es necesario conocer la media μ y la desviación estándar σ , con el fin de estandarizar la variable mediante la expresión: $z = \frac{x-\mu}{\sigma}$.

Una vez determinado Z , las probabilidades (que en este caso representan áreas bajo la curva normal) se determinan en una tabla de distribución normal.

Para convertir un Modelo Binomial en Modelo Normal, se procede así:

Se corrige la variable discreta x a continua, aumentando o restando 0,5. En el modelo binomial, $f(x)$ es la ordenada en el punto x ; en cambio, en la normal es el área en el intervalo $x - 0,5$ a $x + 0,5$.

Por ejemplo, $f(10)$ en binomial, en la normal equivale a $P(9,5 < x < 10,5)$.

$P(x \leq 10) = F(10)$ en la binomial, en la normal equivale a $P(x < 10,5)$.

$P(x < 10) = F(9)$ en la binomial, en la normal equivale a $P(x < 9,5)$.

Se determina la media $\mu = np$ y la desviación estándar $\sigma = \sqrt{npq}$.

Se calcula z con la siguiente fórmula:

$$z = \frac{(x \pm 0,5) - np}{\sqrt{npq}}.$$

Se determina las probabilidades en una tabla normal estandarizada.

Cuando las muestras son grandes, es decir, $n \geq 30$, las distribuciones muestrales para la media, la proporción, la diferencia de medias y diferencia de proporciones, también se distribuyen normalmente.

Ejemplo:

Suponga que el 40 % de los estudiantes universitarios no están de acuerdo con una política institucional. Determinar la probabilidad de que, en una encuesta realizada a 200 estudiantes, menos del 50 % no estén de acuerdo con la política institucional.

Solución:

$$p = 40 \% = 0,40; n = 200.$$

$$q = 1 - p = 1 - 0,40 = 0,60.$$

$$\text{Media: } \mu = np = 200 \times 0,40 = 80.$$

$$\text{Desviación estándar: } \sigma = \sqrt{npq} = \sqrt{200 \times 0,40 \times 0,60} = 6,928.$$

Variable $x < 100$.

$$\text{Variable corregida: } x = x - 0,5 = 100 - 99,5.$$

$$z = \frac{(x \pm 0,5) - np}{\sqrt{npq}} = \frac{(x \pm 0,5) - \mu}{\sigma} = \frac{99,5 - 80}{6,928} = 2,81.$$

$$P(x < 50 \%) = P(x < 99,5) = P(z < 2,81) = 0,9975 = 99,75 \%$$

R/ La probabilidad que menos del 50% de los estudiantes tenga la misma posición es del 99,75 %.

Solución con Microsoft Excel

Insertar función estadística DISTR.NORM.ESTAND.N, digitar el valor de z y, en acumulado, 1, con lo cual se obtiene la probabilidad de que z sea menor que el valor dado. Para determinar la probabilidad de que z sea mayor que el número dado, se lo hace por complemento.

$$P(z < 1,96) = 0,9750$$

$$P(z > 1,96) = 0,0250$$

5.2.2 Distribución Ji-Cuadrado

La distribución Ji-Cuadrado es un caso particular de la distribución Gamma, cuya función de densidad se define de la siguiente manera:

$$f(x, \alpha, \beta) = \begin{cases} \frac{x^{\alpha} e^{-\frac{x}{\beta}}}{\alpha! \beta^{\alpha+1}} & \text{si } 0 < x < \infty, \beta > 0, \alpha > 1. \\ 0 & \text{en otro caso.} \end{cases}$$

La media y varianza se determinan de la siguiente manera:

$$\begin{aligned} \text{Media: } \mu &= \beta(\alpha + 1). \\ \text{Varianza: } \sigma^2 &= \beta^2(\alpha + 1)^2. \end{aligned}$$

Se obtiene la Distribución Ji- Cuadrado (χ^2), cuando en la distribución Gamma se cumplen las siguientes condiciones:

$$\begin{aligned} \beta &= 2. \\ \alpha &= \frac{n}{2} - 1. \end{aligned}$$

En este caso, la función de densidad de Ji- Cuadrado está dada por la siguiente expresión:

$$f(x) = \begin{cases} \frac{\chi^{\frac{\nu}{2}-1} e^{-\frac{\chi}{2}}}{\Gamma\left(\frac{\nu}{2}\right) * 2^{\frac{\nu}{2}}} & \text{si } \chi^2 > 0 \\ 0 & \text{si } \chi^2 < 0 \end{cases}$$

Donde $\nu = n - 1$, grados de libertad.

Los grados de libertad corresponden al número de valores que libremente se puede asignar para

obtener un resultado. Por ejemplo, si la suma de 5 números debe dar 50, se puede asignar cuatro valores libremente y el quinto queda determinado por la diferencia.

La media y varianza respectivas son:

$$\begin{aligned} \text{Media: } \mu &= \nu. \\ \text{Varianza: } \sigma^2 &= 2\nu. \end{aligned}$$

El estadígrafo Ji-Cuadrado que se utiliza en las aplicaciones se determina con la siguiente expresión:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}; \text{ donde } S^2 \text{ es la varianza sesgada de la muestra.}$$

$$\chi^2 = \frac{n\hat{S}^2}{\sigma^2}; \text{ donde } \hat{S}^2 \text{ es la varianza insesgada de la muestra.}$$

Donde $n - 1 = \nu$, grados de libertad (ver Gráfico 7).

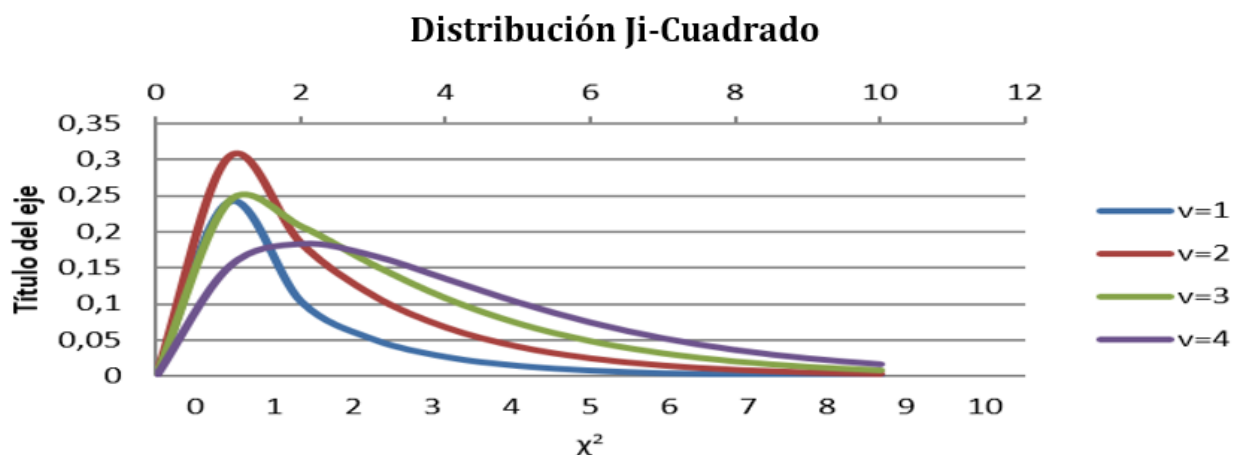


Gráfico 7. Distribución Ji-Cuadrado con diversos grados de libertad

Los valores ν corresponden a los grados de libertad, que dependen del tamaño de la muestra; a medida que aumenta ν , la distribución tiende a ser simétrica.

Probabilidad a la izquierda distribución Ji-Cuadrado con Microsoft Excel

Insertar función Estadística, DISTR.CHICUAD, digitar valor de la variable y grados de libertad. En acumulado digitar 0, para obtener la ordenada $f(x)$; y, 1, para obtener $F(x)$, que corresponde a la probabilidad acumulada de la variable (ver Imagen 6).

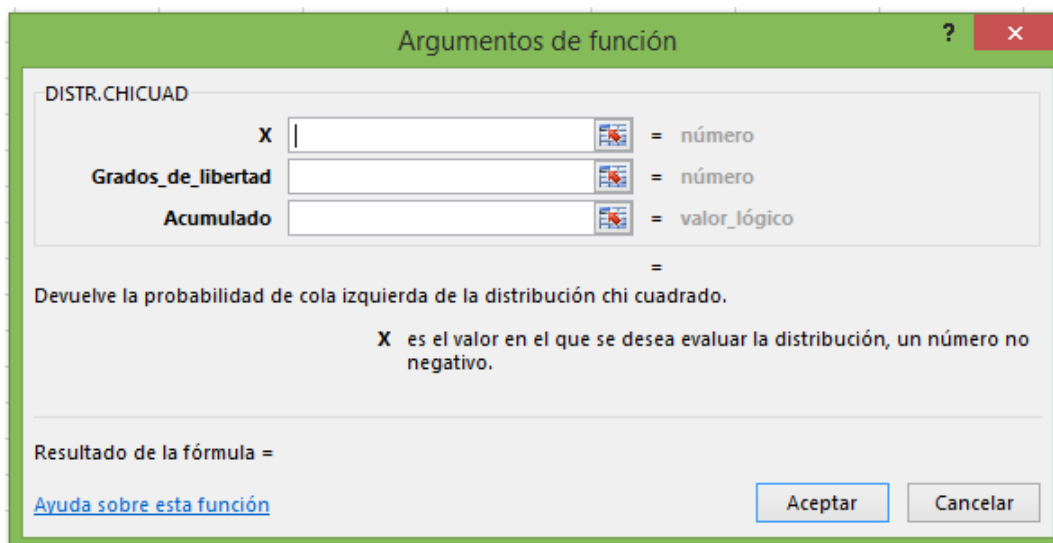


Imagen 6. Microsoft Excel: argumentos de la función DISTR.CHICUAD

5.2.3 Distribución t de Student

La distribución t de Student se expresa así:

$$T = \frac{Z}{\sqrt{\frac{V}{v}}} = Z \sqrt{\frac{v}{V}}.$$

Z es una variable aleatoria con distribución normal estándar, es decir, con media $\mu = 0$ y varianza $\sigma^2 = 1$.

V es una variable continua que tiene una distribución Ji-Cuadrado con ν grados de libertad.

Las variables z y V son independientes.

Las funciones de densidad y acumulativa correspondientes son las siguientes:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1+x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

$$F(x) = \frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right).$$

Valor esperado: $\mu = \begin{cases} 0 & \text{para } \nu > 1. \\ \text{Indefinida} & \text{para otros casos.} \end{cases}$

Varianza: $\sigma^2 = \begin{cases} \frac{\nu}{\nu-2} & \text{para } \nu > 2. \\ \text{Indefinida} & \text{para otros casos.} \end{cases}$

El estadígrafo t que se utiliza en las aplicaciones es similar a Z , tal como se indica enseguida:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}}, \text{ donde } S \text{ es desviación estándar sesgada.}$$

$$t = \frac{\bar{X} - \mu}{\frac{\hat{s}}{\sqrt{n}}}, \text{ donde } \hat{S} \text{ es desviación estándar insesgada.}$$

Donde $n - 1 = \nu$, representa los grados de libertad.

Distribución t

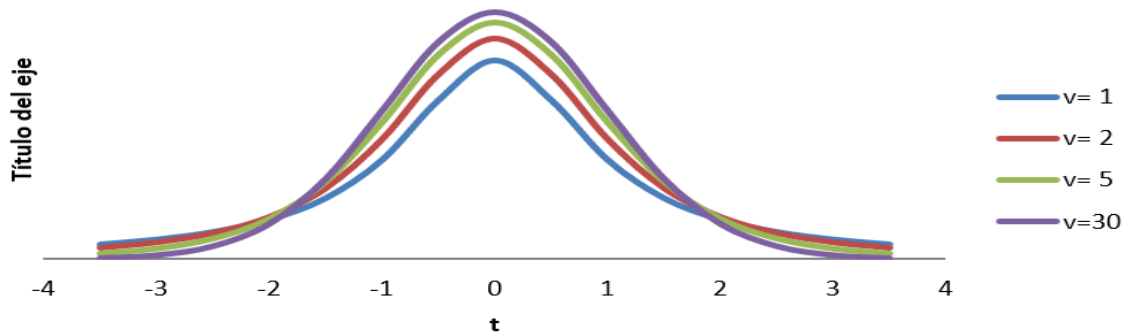


Gráfico 8. Distribución t de Student con diversos grados de libertad

En el Gráfico 8, ν representa los grados de libertad, los cuales dependen del tamaño de la muestra. La distribución t se aproxima a la normal cuando las muestras son grandes, es decir, cuando $n > 30$.

La distribución de probabilidad t de Student se aplica para estimar y hacer prueba de hipótesis de la media de una población cuando el tamaño de la muestra es pequeño ($n \leq 30$) y se desconoce la desviación estándar; por lo cual, para calcular el error estándar del promedio muestral, se utiliza siguiente la fórmula:

$$\frac{s}{\sqrt{n-1}} = \frac{\hat{s}}{\sqrt{n}}$$

Probabilidad a la izquierda distribución T de Student con Microsoft Excel

Insertar función, DISTR.T.N, digitar valor de X y grados de libertad; 0, en acumulado para obtener $f(x)$ y, 1, para $F(x)$ (ver Imagen 7).

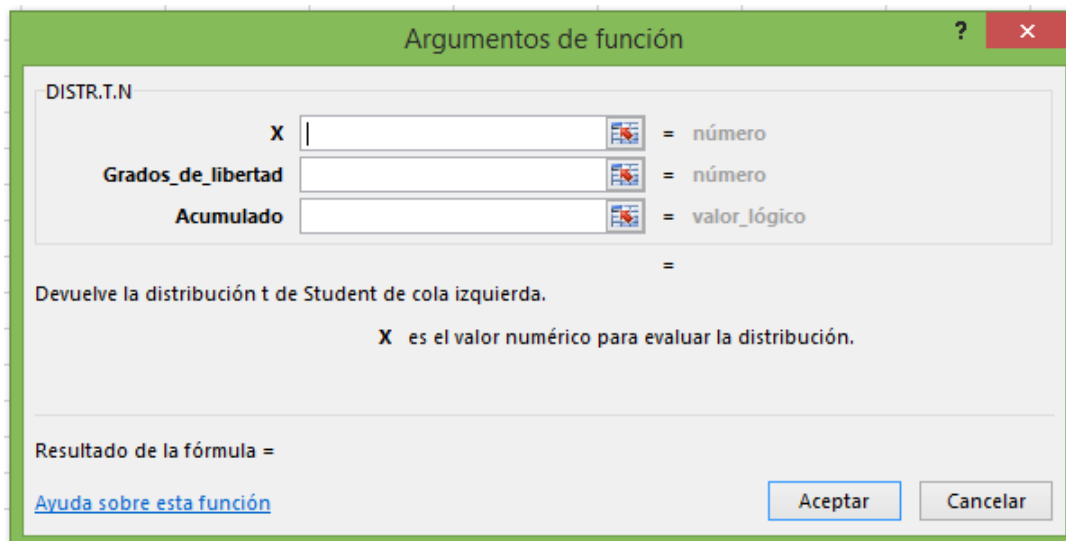
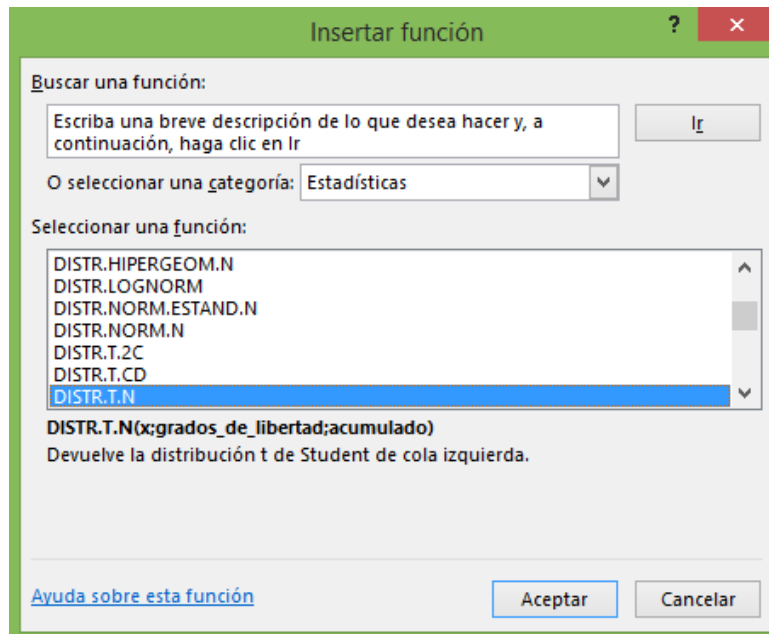


Imagen 7. Microsoft Excel: argumentos de la función DISTR.T.N

5.2.4 Distribución F de Fisher

La distribución F es una distribución de probabilidad de una variable aleatoria continua; se la conoce como F de Fisher.

Una variable aleatoria de distribución F se define como sigue:

$$F = \frac{U_1}{d_1} \cdot \frac{U_2}{d_2}$$

Donde, U_1 y U_2 tienen una distribución Ji-Cuadrado con d_1 y d_2 grados de libertad, respectivamente.

El estadígrafo F utilizado en las aplicaciones es: $F = \frac{S_1^2}{S_2^2}$.

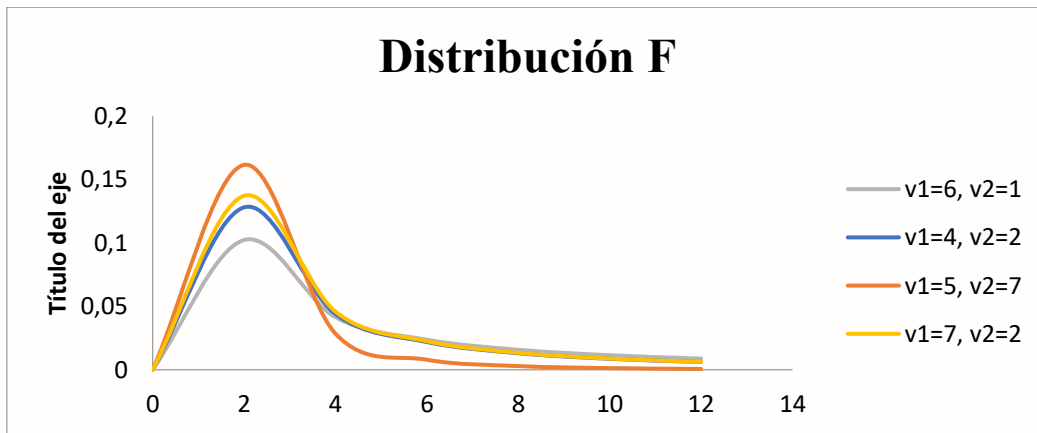


Gráfico 9. Distribución F con diversos grados de libertad

La distribución F se aproxima a una distribución normal cuando los tamaños de las muestras son mayores que 100; es decir: $n_1 > 100$ y $n_2 > 100$ (ver Gráfico 9).

Probabilidad a la izquierda Distribución F, con Microsoft Excel

Insertar función estadística, DISTR.F.N, digitar el valor de x y grados de libertad 1 y grados de libertad 2; y , en acumulado, 0 para $f(x)$, y 1, para $F(x)$ (ver Imagen 8).

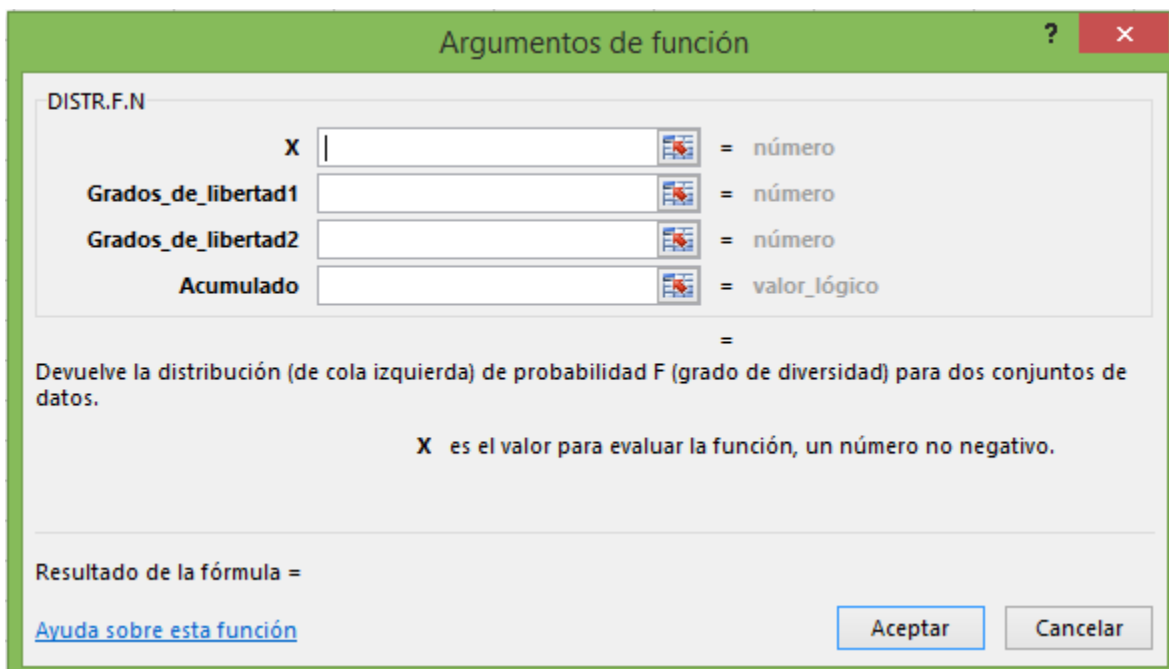
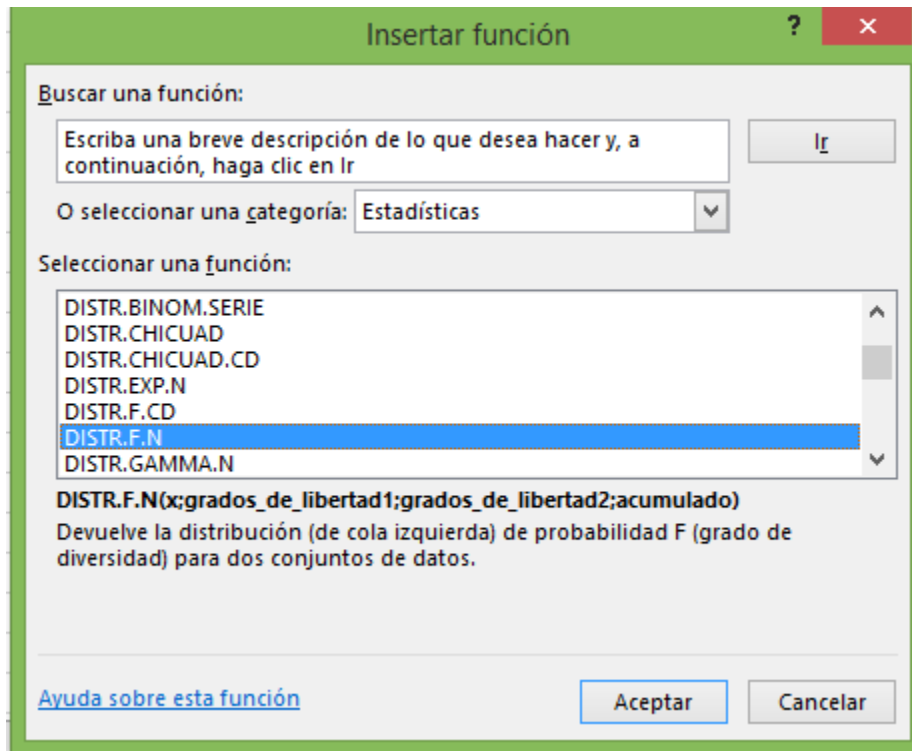


Imagen 8. Microsoft Excel: argumentos de la función DISTR.F.N



Capítulo 6

Distribuciones Muestrales

CAPÍTULO 6

DISTRIBUCIONES MUESTRALES

Si en una población se consideran todas las muestras posibles del mismo tamaño n y en cada muestra posible se calcula un estadístico, ya sea el Promedio, la Varianza, la Desviación Estándar, la Proporción de Éxitos, etc., entonces los resultados son diferentes en cada una de las muestras y conforman una nueva variable aleatoria, con su correspondiente distribución de probabilidad llamada Distribución Muestral de Promedios, Varianzas, etc., según el caso. Estas distribuciones se aproximan a una distribución normal, t de Student, Ji-Cuadrado o F de Fisher. En cada variable, es importante determinar el valor esperado y el error estándar.

6.1 RESUMEN DE DISTRIBUCIONES MUESTRALES

La Tabla 3 y la Tabla 4 contienen un resumen de variables aleatorias que se pueden modelar con una distribución normal y t de Student, respectivamente, con sus correspondientes valores esperados y desviación estándar. Por su parte, la Tabla 5 contiene la distribución Ji-Cuadrado, con los correspondientes límites para la estimación y pruebas de hipótesis de varianza.

6.1.1 Variables aleatorias con distribución normal

La Tabla 3 presenta un resumen de variables aleatorias con distribución normal.

Tabla 3. Resumen de variables aleatorias con distribución normal

| Variable | Valor esperado | Error estándar |
|---|---|---|
| $\bar{X} = \frac{\sum X}{n}$ | $E(\bar{X}) = \mu$ | $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ |
| $\bar{p} = \frac{X}{n}$ | $E(\bar{p}) = p$ | $\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$ |
| $\bar{D} = \bar{X}_1 - \bar{X}_2$ | $E(\bar{D}) = \delta = \mu_1 - \mu_2$ | $\sigma_{\bar{D}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
| $\Delta\bar{p} = \bar{p}_1 - \bar{p}_2$ | $E(\Delta\bar{p}) = \Delta p = p_1 - p_2$ | $\sigma_{\Delta\bar{p}} = \sqrt{\frac{\bar{p}_1 * \bar{q}_1}{n_1} + \frac{\bar{p}_2 * \bar{q}_2}{n_2}}$ |

6.1.2 Variables aleatorias con distribución t de Student

La Tabla 4 presenta un resumen de variables aleatorias con distribución t.

Tabla 4. Resumen de variables aleatorias con distribución t

| Variable | Valor esperado | Error estándar estimado |
|-----------------------------------|---------------------------------------|---|
| $\bar{X} = \frac{\sum X}{n}$ | $E(\bar{X}) = \mu$ | $S_{\bar{X}} = \frac{S}{\sqrt{n-1}}$ |
| $\bar{D} = \bar{X}_1 - \bar{X}_2$ | $E(\bar{D}) = \delta = \mu_1 - \mu_2$ | $S_{\bar{D}} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}$ |
| $\bar{D} = \frac{\sum D}{n}$ | $E(\bar{D}) = \delta = \mu_1 - \mu_2$ | $S_{\bar{D}} = \frac{\sqrt{\frac{\sum D^2 - (\sum D)^2}{n-1}}}{\sqrt{n}}$ |

6.1.3 Distribución Ji-Cuadrado

La Tabla 5 presenta un resumen de límites de ajuste de la función Ji-Cuadrado.

\hat{s}^2 varianza insesgada. $\hat{S}^2 = \sum \frac{(X-\bar{X})^2}{n-1}$.

$$s^2 \text{ varianza sesgada. } s^2 = \sum \frac{(X-\bar{X})^2}{n}.$$

Tabla 5. Resumen de límites de ajuste de la función Ji-Cuadrado

| Variable | Límite inferior | Límite superior |
|--|---|---|
| $\chi^2 = \frac{(n-1)\hat{s}^2}{\sigma^2}$ | $L_i = \frac{(n-1)\hat{s}^2}{\chi_2^2}$ | $L_s = \frac{(n-1)\hat{s}^2}{\chi_1^2}$ |
| $\chi^2 = \frac{ns^2}{\sigma^2}$ | $L_i = \frac{ns^2}{\chi_2^2}$ | $L_s = \frac{ns^2}{\chi_1^2}$ |

6.2 DISTRIBUCIÓN MUESTRAL DE LA MEDIA

La Media \bar{x} de una muestra de tamaño n , para $n \geq 30$, tomada de una población con distribución normal, con media μ y desviación estándar σ , se distribuye normalmente con media μ y desviación $\sigma_{\bar{x}}$ dada por la siguiente expresión:

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

La variable Z tipificada se obtiene mediante la siguiente relación:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Ver ejemplo de la sección "7.1.2 Distribución Muestral de Medias" (p. 85).

6.3 DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN DE ÉXITOS

La proporción de éxitos \bar{p} en una muestra de tamaño grande, $n \geq 100$, se distribuye normalmente con $E(\bar{p}) = p$ y error estándar $\sigma_{\bar{p}}$, dado por la siguiente expresión:

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}.$$

La variable z tipificada está dada por la siguiente relación: $Z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}}$.

Ver ejemplo de la sección “7.1.3 Distribución Muestral de Proporción de Éxitos” (p. 86).

6.4 DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE MEDIAS

Si se toma dos muestras de dos poblaciones normales o de la misma población normal, con promedios y varianzas conocidas, la distribución muestral de la diferencia de medias se distribuye normalmente.

| Descripción | Población No. 1 | Población No. 2 |
|-------------------|-----------------|-----------------|
| Media | μ_1 | μ_2 |
| Varianza | σ_1^2 | σ_2^2 |
| Tamaño de muestra | n_1 | n_2 |
| Promedio muestral | \bar{X}_1 | \bar{X}_2 |

Diferencia de medias muestrales: $\bar{D} = \bar{X}_1 - \bar{X}_2$.

Diferencia de medias poblacionales: $\delta = \mu_1 - \mu_2$.

Si las muestras fueran de la misma población, $\delta = 0$, ya que $\mu_1 = \mu_2 = \mu$.

Desviación estándar de la diferencia: $\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Variable Z tipificada: $z = \frac{\bar{D} - \delta}{\sigma_D}$.

Ver ejemplo de la sección “7.1.4 Distribución Muestral de Diferencia de Medias” (p. 87).

6.5 DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE PROPORCIONES

Si se toma dos muestras de dos poblaciones o de la misma población, se encuentra la proporción de éxitos en cada muestra, dividiendo el número de éxitos entre el tamaño de la muestra, así:

$$\bar{p}_1 = \frac{x_1}{n_1} ; \bar{p}_2 = \frac{x_2}{n_2} .$$

La distribución muestral de diferencia de proporciones, $\Delta\bar{p}$, se distribuye normalmente con media Δp , calculada así:

$$E(\Delta\bar{p}) = \Delta p = p_1 - p_2 .$$

$$\text{Desviación estándar: } \sigma_{\Delta\bar{p}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} .$$

$$\text{Variable Z tipificada: } z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sigma_{\Delta\bar{p}}} .$$

Ver ejemplo de la sección "7.1.5 Distribución Muestral de Diferencia de Proporciones" (p. 88).



Capítulo 7

Aplicaciones de la Distribución Normal



CAPÍTULO 7

APLICACIONES DE LA DISTRIBUCIÓN NORMAL

La Distribución Normal se puede utilizar para realizar cálculo de probabilidades, estimación por intervalos y para contraste de hipótesis.

7.1 CÁLCULO DE PROBABILIDADES

7.1.1 Distribución Normal

Para calcular una probabilidad, se encuentra Z con la fórmula que corresponda y se determina los valores en la tabla normal Z .

Ejemplos:

Sea la variable X que mide los puntajes del ICFES en el último año, donde la media $\mu = 300$ puntos y la desviación estándar $\sigma = 20$.

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 300}{20}.$$

Determinar el porcentaje de estudiantes que obtuvieron menos de 330 puntos.

$$Z = \frac{X - \mu}{\sigma} = \frac{330 - 300}{20} = 1,5.$$

$$P(X < 330) = P(Z < 1,5) = 0,9332 = 93,32\%.$$

Determinar el porcentaje de estudiantes que obtuvieron más de 330 puntos.

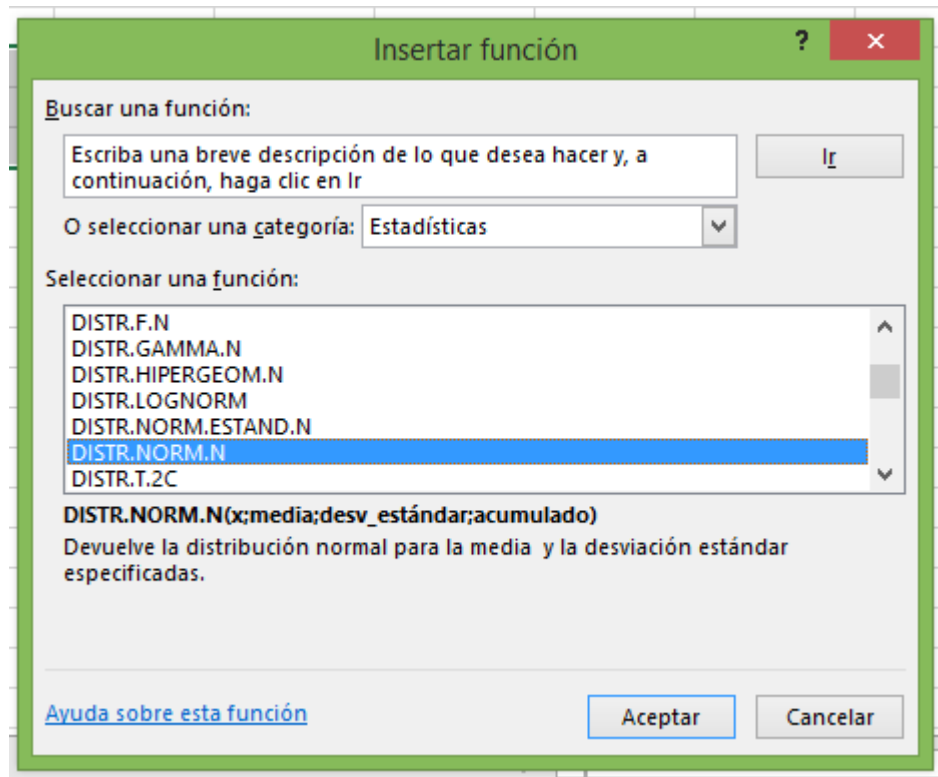
$$P(X > 330) = P(Z > 1,5) = 1 - P(Z < 1,5) = 1 - 0,9332 = 0,0668 = 6,68\%$$

Solución con Microsoft Excel

Insertar función DISTR.NORM.N, digitar el valor de la variable X , la media y la desviación estándar; en acumulado, 1 Aceptar (ver Imagen 9).

Igual en todas las distribuciones muestrales con Normal, en X digitar el valor de la variable, en media el valor esperado y en lugar de desviación estándar teclear el valor del error estándar.

Cálculo de la probabilidad de la cola izquierda distribución normal, utilizando Excel



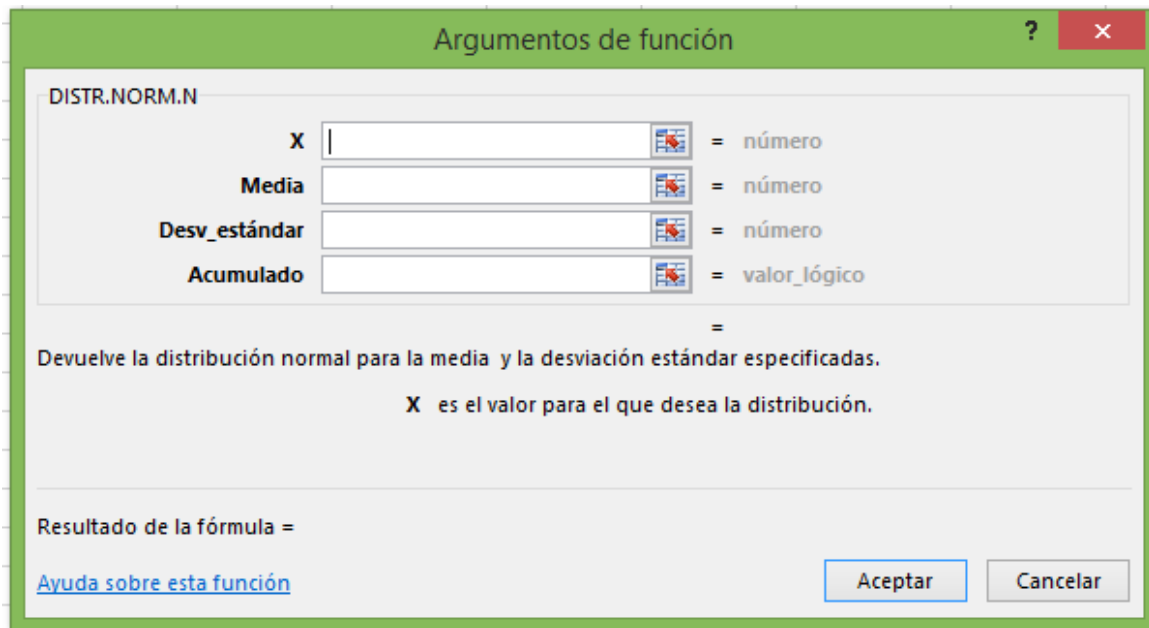
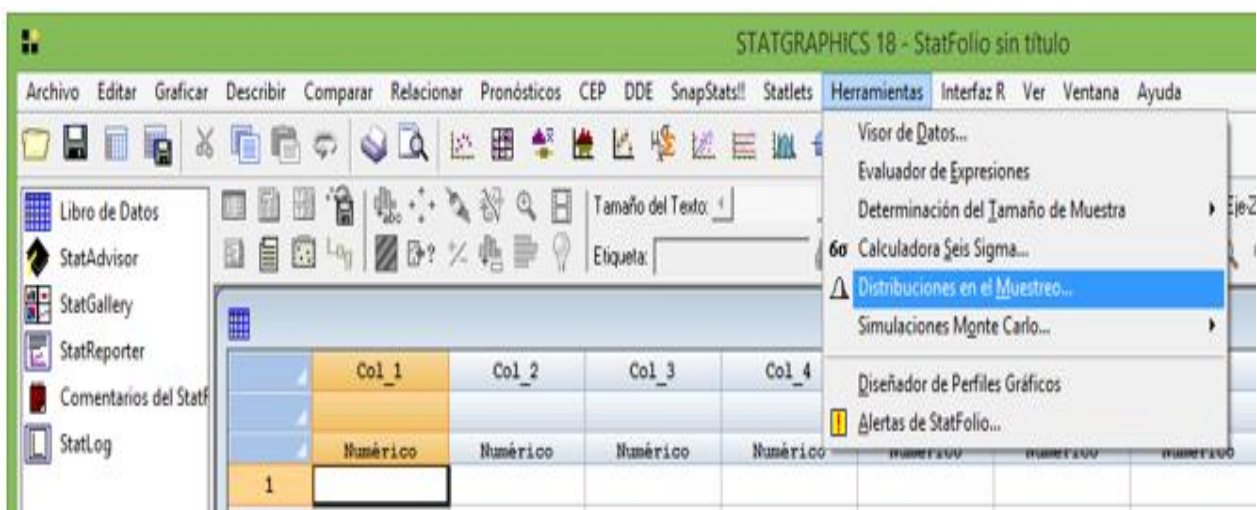


Imagen 9. Microsoft Excel: argumentos de la función DISTR.NORM.N

En el menú, escoger la opción herramientas, distribuciones muestrales, distribución normal.

Digitar la media y la desviación estándar. Aceptar. Digitar límite inferior, seleccionar área a la izquierda límite inferior y pulsar clic en Aceptar.

Cálculo de la probabilidad en la distribución normal con Statgraphics (ver Imagen 10).



Distribuciones en el muestreo ✕

Distribución

Normal

Media: Desviación típica:

t de Student

Grados de libertad:

Chi-Cuadrada

Grados de libertad:

F (Razón de Varianzas)

G.l. numerador: G.l. denominador:

Opciones de distribuciones en el muestreo ✕

Límite Inferior: Límite Superior: Área:

Calcular

Área límite inferior

Área límite superior

Área entre límites

Área fuera de límites

Valor crítico inferior para un área dada

Valor crítico superior para un área dada

Valor crítico conteniendo un área dada

Valor crítico excluyendo un área dada

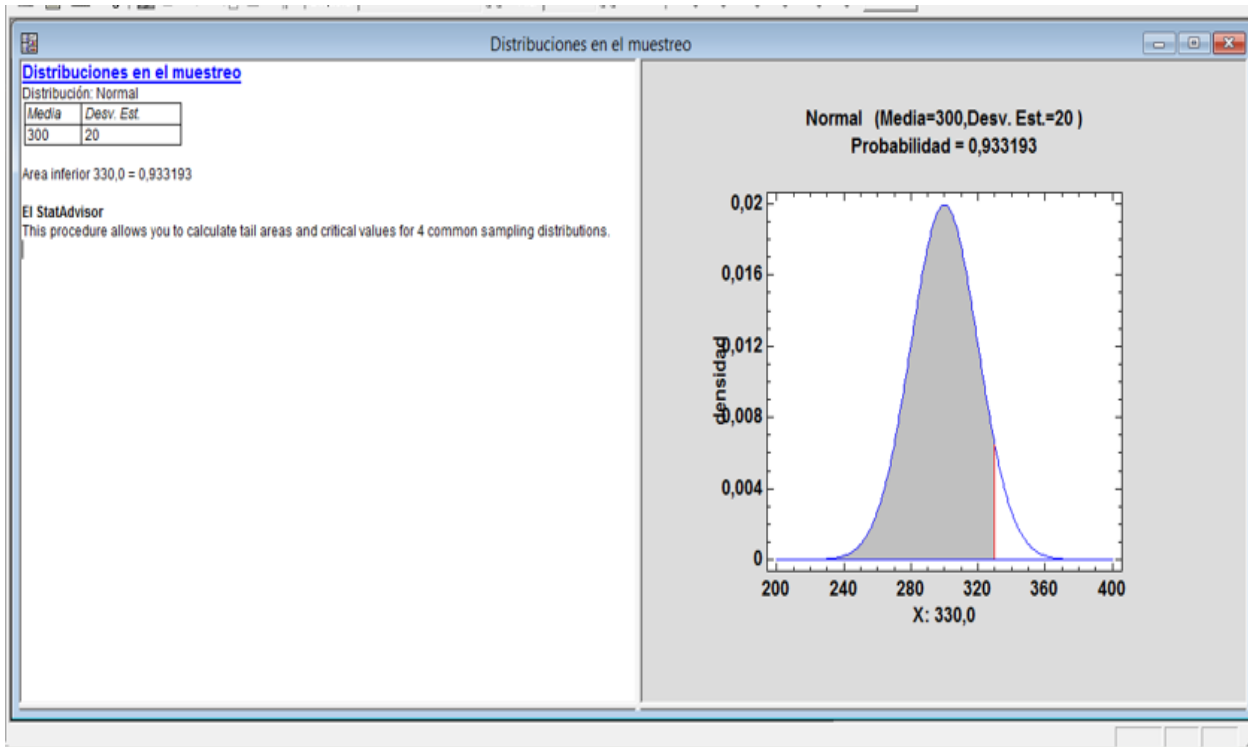


Imagen 10. Respuesta y gráfica mediante Statgraphics

7.1.2 Distribución Muestral de Medias

Corresponde a la distribución muestral de \bar{X} .

Ejemplo:

La diferencia con el problema anterior es que ahora se toma una muestra y se busca la probabilidad del promedio. Si se toma una muestra de 36 estudiantes, determinar la probabilidad que el *Promedio* sea inferior a 310 puntos.

Solución:

Promedio muestral: $\bar{X} = 310$.

Promedio poblacional: $\mu = 300$.

Desviación poblacional: $\sigma = 20$.

Tamaño de la muestra: $n = 36$.

Error estándar o error típico: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{36}} = \frac{20}{6} = 3,33$.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{310 - 300}{3,33} = 3.$$

$$P(\bar{X} < 310) = P(Z < 3) = 0,9987 = 99,87 \%$$

R/ El 99,87 % de las muestras de tamaño 36 tendrían una media menor a 310 puntos.

7.1.3 Distribución Muestral de Proporción de Éxitos

Corresponde a la distribución muestral de proporción de éxitos \bar{p} .

Ejemplo:

El 40 % de los pacientes hospitalizados son mujeres. Si se toma una muestra de 200 pacientes, determinar la probabilidad de que menos de 90 sean mujeres.

Solución:

Porcentaje de población: 0,40 .

$$\text{Porcentaje de mujeres: } \bar{p} = \frac{90}{200} = 0,45.$$

Desviación de la proporción:

$$\sigma_{\bar{p}} = \sqrt{\frac{0,4 \times 0,6}{200}} = 0,03.$$

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0,45 - 0,40}{0,03} = 1,67.$$

$$P(X < 90) = P(\bar{p} < 0,45) = P(Z < 1,67) = 0,9525 = 95,25 \%$$

R/ La probabilidad que haya menos de 90 mujeres en una muestra de 200 pacientes, o equivalentemente, la probabilidad que menos del 45 % sean mujeres, es de 0,9525. También significa que el 95,25 % de las muestras aleatorias del mismo tamaño tendrán menos del 45 % de mujeres.

La probabilidad de que un trabajador ascienda de cargo es del 60%. Si en una empresa hay 40 trabajadores, determinar la probabilidad de que la proporción de la muestra de personas que ascienden de cargo sea mayor del 65%.

Solución:

Probabilidad de ascenso: $p = 0,60$.

Número de trabajadores: $n = 40$.

$P(\bar{p} > 65\%) = ?$

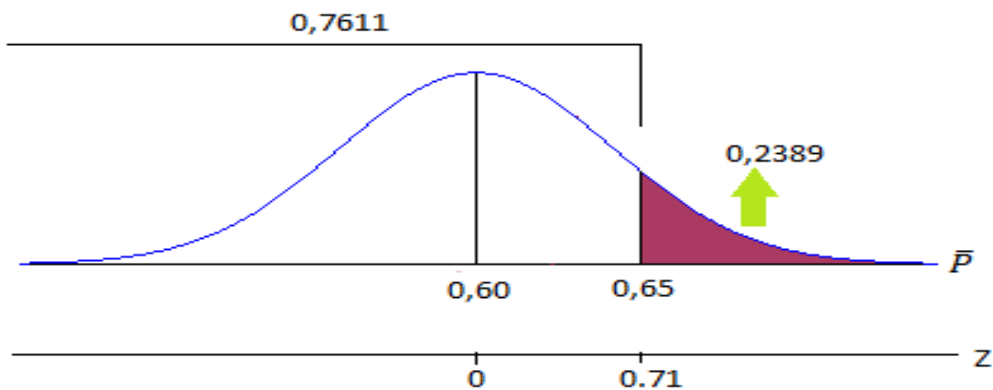


Gráfico 10. Proporción de que una muestra de personas sea mayor del 65 %

$P(\bar{p} > 65\%) = 1 - P(Z < 0,71)$

$P(\bar{p} > 65\%) = 1 - 0,7611$

$P(\bar{p} > 65\%) = 0,2389$.

$P(\bar{p} > 65\%) = 23,89 \%$ (ver Gráfico 10).

R/ El 23,89 % de todas las muestras del mismo tamaño, tienen un porcentaje de ascensos mayor que el 65 %.

7.1.4 Distribución Muestral de Diferencia de Medias

Corresponde a la distribución muestral de diferencia de medias \bar{D} .

Ejemplo:

Para el proyecto de acueducto de una vereda de Nariño, la gobernación destina fondos a la compra de tubos de transporte de fluidos a dos compañías diferentes, a saber, las compañías Alfa S. A y la Beta S. A. Los tubos distribuidos por la primera compañía tienen una vida media de 7,2 años, mientras que los de la segunda, de 6,7 años. Cuentan con una desviación estándar de 0,8 y 0,7 años, respectivamente.

Determine el error estándar y la probabilidad de que una muestra aleatoria de 40 tubos de la compañía Beta S. A. tenga una vida promedio que sea un año menor que la vida promedio de una muestra aleatoria de 34 tubos de la compañía Alfa S.A.

Solución:

$$\mu_1 = 7,2; \sigma_1 = 0,8; n_1 = 34.$$

$$\mu_2 = 6,7; \sigma_2 = 0,7; n_2 = 40.$$

$$\text{Variable: } \bar{D} = \bar{X}_1 - \bar{X}_2 = 1.$$

$$E(\bar{D}) = \mu_1 - \mu_2 = 7,2 - 6,7 = 0,5.$$

$$\text{Error estándar: } \sigma_{\bar{D}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0,8^2}{34} + \frac{0,7^2}{40}} = 0,17.$$

$$Z = \frac{\bar{D} - E(\bar{D})}{\sigma_{\bar{D}}} = \frac{1 - 0,5}{0,17} = 2,94.$$

$$P(\bar{D} < 1) = P(Z < 2,94) = 0,9984.$$

R/ La probabilidad que la diferencia de medias de vida entre tubos de la compañía Beta S. A. sea de un año inferior a los de la compañía Alfa es del 99,84 %.

7.1.5 Distribución Muestral de Diferencia de Proporciones

Corresponde a la distribución muestral de diferencia de proporciones $\Delta \bar{p}$.

Ejemplo:

Se cree que el 80% de los habitantes de la ciudad 1 y el 65 % de los habitantes de la ciudad 2 aceptan un producto X. Se realizan encuestas en estas dos ciudades, considerando 100 personas de la ciudad 1 y 100 personas de la ciudad 2, elegidos al azar. Determinar el error estándar y la probabilidad de que la diferencia de proporciones sea menor del 20 %.

Solución:

Datos ciudad 1:

$$p_1 = 0,8; q_1 = 1 - 0,8 = 0,2; n_1 = 100 .$$

Datos ciudad 2:

$$p_2 = 0,65; q_2 = 1 - 0,65 = 0,35; n_2 = 100 .$$

Valor esperado:

$$\Delta p = p_1 - p_2 = 0,8 - 0,65 = 0,15 .$$

$$\sigma_{\Delta \bar{p}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{0,8 * 0,2}{100} + \frac{0,65 * 0,35}{100}} = 0,062 .$$

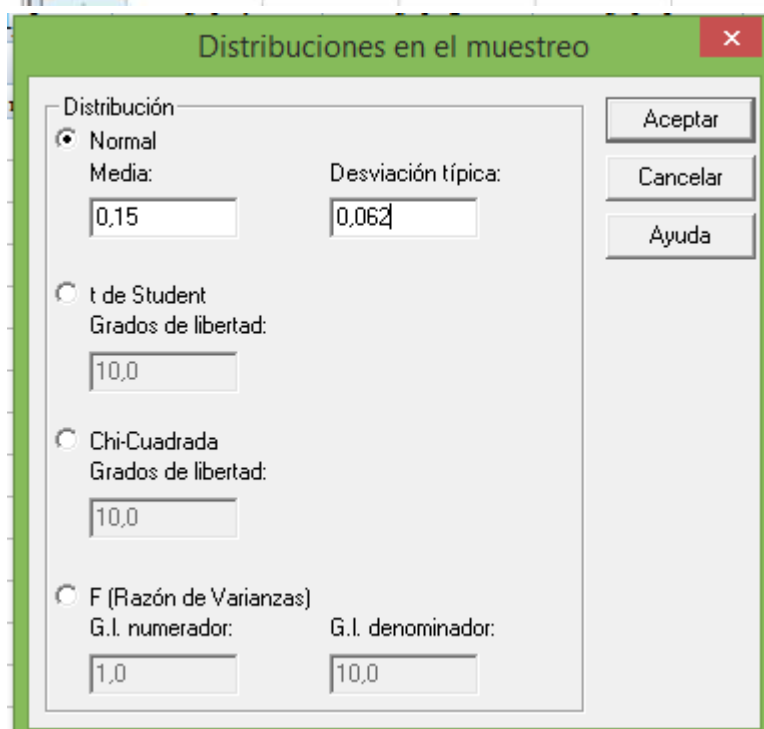
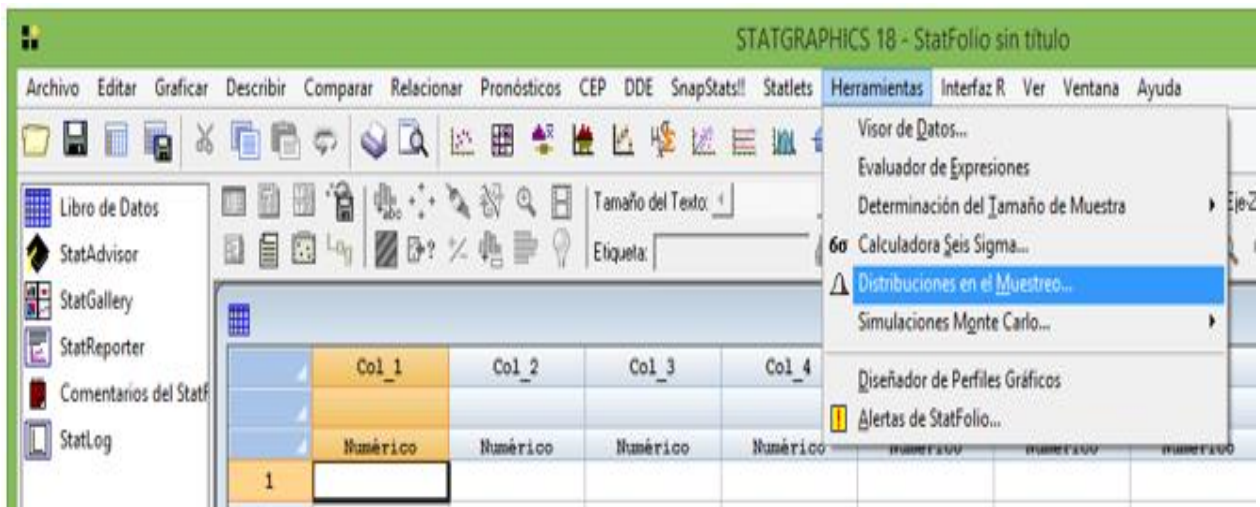
$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sigma_{\Delta \bar{p}}} = \frac{0,20 - 0,15}{0,062} = 0,81 .$$

$$P(\Delta \bar{p} < 20\%) = P(Z < 0,81) = 0,7910 = 79,10 \% .$$

R/ El 78,81 % de las muestras de 100 habitantes de la ciudad 1 y 100 habitantes de la ciudad 2, tendrán una diferencia de porcentaje menor al 20 %.

Cálculo de probabilidad de la distribución muestral diferencia de proporciones con Statgraphics

En el Menú, seleccionar herramientas, Distribuciones Muestrales, se puede encontrar probabilidades con Distribución Normal, t, Ji-Cuadrado y F de Fisher, digitando los valores solicitados en cada caso; probabilidad a la izquierda, a la derecha o entre dos valores y su representación gráfica para cada distribución (ver Imagen 11).



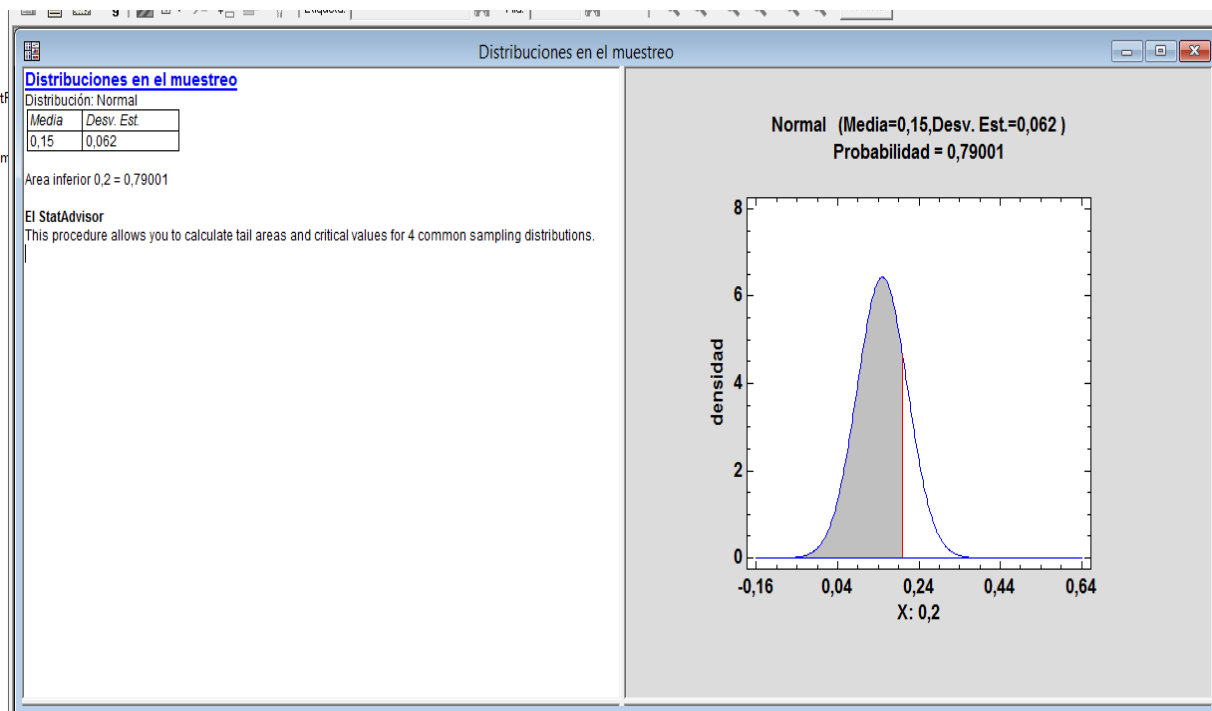
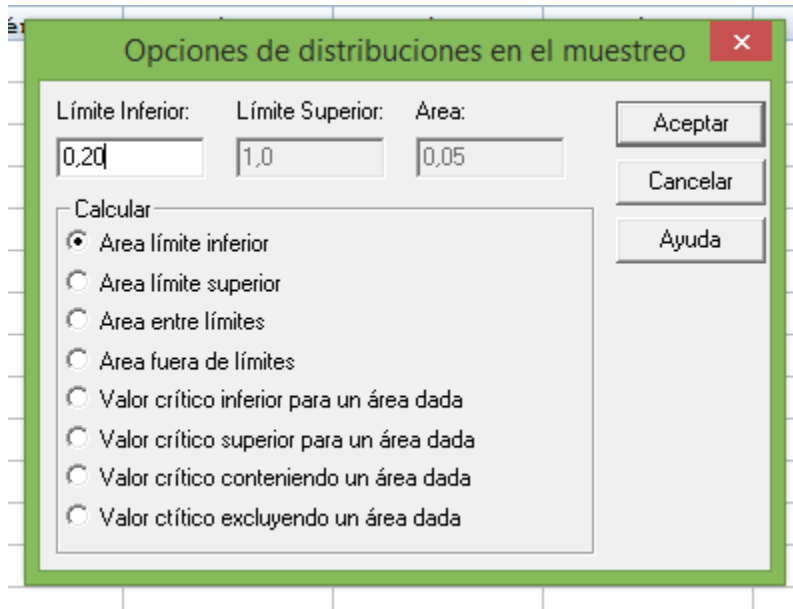


Imagen 11. Respuesta y gráfico, Distribuciones en el muestreo

7.2 ESTIMACIÓN DE PARÁMETROS

La estimación de un parámetro puede ser puntual, para lo cual basta calcular el estadígrafo en la muestra y se toma ese resultado como estimación del parámetro.

Mediante esta estrategia, no es posible detectar el margen de error de la estimación, ni la precisión, ni el grado de confiabilidad; en cambio, si se realiza una estimación por intervalos, buscando un límite inferior y otro superior, dentro de los cuales, se espera, esté incluido el parámetro, se puede medir el margen de error, la precisión y la confiabilidad del parámetro estimado.

Un estadígrafo se considera un *buen estimador del parámetro* si cumple las siguientes condiciones:

Insesgabilidad. Un estadígrafo $\hat{\theta}$ es insesgado si su valor esperado es igual al parámetro; es decir: $E(\hat{\theta}) = \theta$.

Consistencia o precisión. Un estadígrafo es consistente si cumple que: $P(|\hat{\theta} - \theta| \leq e) \rightarrow 1$ cuando $n \rightarrow \infty$.

Eficiencia. Si existen dos estadígrafos para estimar un parámetro, es más eficiente el que tenga menor variabilidad.

Suficiencia. Un estadígrafo cumple este criterio si para calcularlo se utiliza toda la información de la muestra.

Los siguientes estadígrafos: promedio aritmético, proporción de éxitos, diferencia de promedios y diferencia de proporciones, cumplen los criterios para ser un buen estimador del parámetro.

7.2.1 Estimación por intervalos para la media (μ)

Los límites de confianza para la media poblacional se obtienen de la siguiente forma:

Promedio de muestra $\pm Z_{\alpha/2} \times$ *error estándar de la media.*

Margen de error: $Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$.

Límites de confianza: $\bar{X} \pm Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$.

Si la población es muy grande ($N \geq 30$) o infinita, entonces,

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

La expresión $1 - \alpha$ se denomina *Grado de Confianza* y es la probabilidad de que la media μ se encuentre entre los dos límites determinados.

Enseguida, para cada grado de confianza, se muestra el valor $Z_{\frac{\alpha}{2}}$ correspondiente:

Si $1 - \alpha = 90\%$, entonces $Z_{\frac{\alpha}{2}} = 1,65$.

Si $1 - \alpha = 95\%$, entonces $Z_{\frac{\alpha}{2}} = 1,96$ (ver Gráfico 11).

Si $1 - \alpha = 98\%$, entonces $Z_{\frac{\alpha}{2}} = 2,33$.

Si $1 - \alpha = 99\%$, entonces $Z_{\frac{\alpha}{2}} = 2,58$.

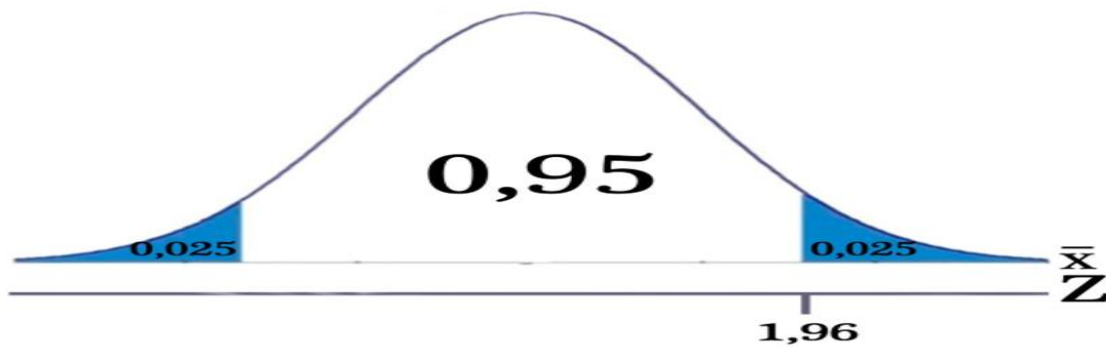


Gráfico 11. Estimación por intervalos para la media

El margen de error para poblaciones grandes o infinitas es:

$$E = Z_{\frac{\alpha}{2}} \times \left(\frac{\sigma}{\sqrt{n}}\right).$$

De aquí se puede obtener el tamaño n de la muestra para poblaciones infinitas:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2}{E^2}.$$

Para poblaciones finitas de tamaño $N < 30$, el margen de error se determina por la fórmula:

$$E = Z_{\frac{\alpha}{2}} \times \left(\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{(N-n)}{(N-1)}} \right).$$

El último radical se conoce como *Factor de Corrección para Poblaciones Finitas*. De esta expresión se obtiene, el valor de n que representa el tamaño de la muestra para poblaciones finitas:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2 N}{(N-1)E^2 + Z_{\frac{\alpha}{2}}^2 \sigma^2}.$$

Ejemplo:

Se desea conocer el grado de anemia en pacientes de una ciudad; para lograrlo, se toma una muestra de 100 pacientes diabéticos que acuden a la unidad renal del hospital local. Después de medir la hemoglobina en miligramos por decilitro, se encuentra un promedio de 8,64 ml/dl y una desviación estándar 2,32 ml/dl. Utilice un nivel de significancia del 98%.

Solución:

Promedio de la muestra: $\bar{X} = 8,64$.

Valor $Z_{\frac{\alpha}{2}}$ para un nivel de significancia del 98%: $Z_{\frac{\alpha}{2}} = 2,33$.

Desviación estándar: $\sigma = 2,32 \frac{ml}{dl}$.

Tamaño de la muestra: $n = 100$.

Límite inferior: $L_i = \bar{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 8,64 - 2,33 * \frac{2,32}{\sqrt{100}} = 8,64 - 0,53 = 8,01 \text{ ml/dl}$.

Límite superior: $L_s = \bar{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 8,64 + 2,33 * \frac{2,32}{\sqrt{100}} = 8,64 + 0,53 = 9,27 \text{ ml/dl}$.

Entonces, se tiene que:

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \rightarrow P(8,01 < \mu < 9,27) = 98\% .$$

Con una confiabilidad del 98%, se puede afirmar que el grado medio de anemia en dicha población, oscila entre 8,01 ml/dl y 9,27 ml/dl.

En esta investigación, el *margen de error* es el siguiente:

$$E = Z_{\frac{\alpha}{2}} * \left(\frac{\sigma}{\sqrt{n}}\right) = 2,33 * \frac{2,32}{\sqrt{100}} = 0,53 \text{ ml/dl}$$

R/ Con una confiabilidad del 98 %, se puede afirmar que el grado medio de anemia de la población estudiada oscila entre 8,01 ml/dl y 9,27 ml/dl.

Si el margen de error $E = 0,53$ se reduce a 0,30, entonces el nuevo tamaño de la muestra sería:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2}{E^2} = \frac{2,33^2 * 2,32^2}{0,30^2} = 325 .$$

7.2.2 Estimación por intervalos para la proporción (p)

Los límites de confianza para la proporción de la población se obtienen así:

Proporción muestral $\pm Z_{\frac{\alpha}{2}} \times$ *error estándar de la proporción* .

$$\text{Límites de confianza: } \bar{p} \pm Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} .$$

$$\text{Margen de error: } E = Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} .$$

$$\text{La probabilidad es: } P\left(\bar{p} - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} \leq p \leq \bar{p} + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}}\right) = 1 - \alpha .$$

$1 - \alpha$: se denomina *Grado de Confianza*; corresponde a la probabilidad de que el parámetro se encuentre entre los límites inferior y superior indicados.

El valor α representa el riesgo de la estimación y corresponde a la probabilidad de que el parámetro no se encuentre entre los límites inferior y superior planteados.

Valores de $1 - \alpha$ frecuentes:

Si $1 - \alpha = 90\%$, entonces $Z_{\frac{\alpha}{2}} = 1,65$.

Si $1 - \alpha = 95\%$, entonces $Z_{\frac{\alpha}{2}} = 1,96$.

Si $1 - \alpha = 98\%$, entonces $Z_{\frac{\alpha}{2}} = 2,33$.

Si $1 - \alpha = 99\%$, entonces $Z_{\frac{\alpha}{2}} = 2,58$.

Margen de error: $E = Z_{\frac{\alpha}{2}} * \sqrt{\frac{pq}{n}}$.

Despejando n , que corresponde al tamaño de la muestra, se obtiene:

Tamaño de la muestra para poblaciones infinitas o $N \geq 30$:

$n = \frac{Z_{\frac{\alpha}{2}}^2 pq}{E^2}$, en lugar de pq se puede utilizar el máximo producto $pq = 0,25$.

Para poblaciones finitas de tamaño $N < 30$: $E = Z_{\frac{\alpha}{2}} \times \sqrt{\frac{pq}{n}} \times \sqrt{\frac{(N - n)}{(N - 1)}}$.

El último radical se conoce como Factor de Corrección para Poblaciones Finitas - FCPF.

Despejando n , se obtiene el tamaño de la muestra para poblaciones finitas de tamaño $N < 30$:

$n = \frac{Z_{\frac{\alpha}{2}}^2 pqN}{(N - 1)E^2 + Z_{\frac{\alpha}{2}}^2 pq}$.

Ejemplo:

Una encuesta realizada a 500 amas de casa en la ciudad de Pasto dio como resultado que 350 de ellas consumen leche en polvo. Determinar un intervalo del 95% de confianza para la verdadera proporción de consumidores de leche en dicha ciudad.

Solución:

$$\text{Proporción muestral: } \bar{p} = \frac{350}{500} = 0,75 = 75\% .$$

Dado que el intervalo de confianza es del 95%, entonces $Z_{\frac{\alpha}{2}} = 1,96$.

Límites de confianza para p:

Dado que el parámetro p es desconocido, en el error estándar se utiliza su estimación, es decir, \bar{p} .

$$\text{Límite inferior: } L_i = \bar{p} - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} .$$

$$L_i = 0,75 - 1,96 \times \sqrt{\frac{0,75 \times 0,25}{500}} = 0,75 - 0,019 = 0,731 = 73,1\% .$$

$$\text{Límite superior: } L_s = \bar{p} + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} .$$

$$\text{Límite superior} = 0,75 + 1,96 \times \sqrt{\frac{0,75 \times 0,25}{500}} = 0,75 + 0,019 = 0,769 = 76,9\% .$$

$$\text{Margen de error: } E = Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} = 1,96 \times \sqrt{\frac{0,75 \times 0,25}{500}} = 0,019 = 1,9\% .$$

$$P\left(\bar{p} - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}} \leq p \leq \bar{p} + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p} \times \bar{q}}{n}}\right) = 1 - \alpha \rightarrow P(0,731 < p < 0,769) = 95\% .$$

R/ El verdadero porcentaje de consumo de leche oscila entre el 73,1% y el 76,9%, afirmación que se efectúa con un 95% de confiabilidad y con un margen de error del 1,9%.

Si se acepta un margen de error del 3%, el nuevo tamaño de muestra sería:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 pq}{E^2} = \frac{1,96^2 \times 0,5 \times 0,5}{0,03^2} = 1067 .$$

7.2.3 Estimación para diferencia de medias ($\delta = \mu_1 - \mu_2$)

Los límites de confianza para la distribución de la diferencia de medias son:

Variable: $\bar{D} = \bar{X}_1 - \bar{X}_2$.

Error estándar de la diferencia: $\sigma_{\bar{D}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Margen de error: $E = Z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Límites de confianza: $\bar{D} \pm Z_{\frac{\alpha}{2}} \times \sigma_{\bar{D}}$.

La probabilidad es: $P\left(\bar{D} - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \delta \leq \bar{D} + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$.

Ejemplo:

Suponga que dos grupos de estudiantes resuelven un examen idéntico. El primer grupo está formado por 30 estudiantes de la institución A, con promedio 85 y una varianza de 240; el segundo grupo está conformado por 40 estudiantes de la institución B, con promedio 80 y varianza 340. Determinar los límites del 95 % de confianza para la verdadera diferencia de promedios.

Solución:

$$\bar{D} = \bar{X}_1 - \bar{X}_2 = 85 - 5 = 80 .$$

$$\sigma_{\bar{D}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{240}{30} + \frac{340}{40}} = 4,06.$$

Límites de confianza: $\bar{D} \pm Z_{\frac{\alpha}{2}} \times \sigma_{\bar{D}} = 5 \pm 1,96 \times 4,06 = 5 \pm 7,95.$

Límite inferior: $L_i = \bar{D} - Z_{\frac{\alpha}{2}} \times \sigma_{\bar{D}} = 5 - 1,96 \times 4,06 = -2,95.$

Límite superior: $L_s = \bar{D} + Z_{\frac{\alpha}{2}} \times \sigma_{\bar{D}} = 5 + 1,96 \times 4,06 = 12,95.$

$$P\left(\bar{D} - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \delta \leq \bar{D} + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha \Rightarrow$$

$$P(-2,95 < \delta < 12,95) = 95 \%.$$

Para los casos en los que el límite inferior es negativo y el superior positivo, no existe diferencia significativa entre los promedios, hecho que se puede verificar al realizar la prueba de hipótesis correspondiente, tema que se estudiará más adelante, en la sección de prueba de hipótesis.

R/ Con una probabilidad del 95 %, se puede afirmar que la verdadera diferencia de promedios está comprendida entre $-2,95$ y $12,95$.

7.2.4 Límites de confianza para diferencia de proporciones ($\Delta p = p_1 - p_2$)

$$\bar{p}_1 = \frac{X_1}{n_1} \text{ y } \bar{p}_2 = \frac{X_2}{n_2}.$$

$$\Delta \bar{p} = \bar{p}_1 - \bar{p}_2.$$

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}.$$

El nuevo error estándar se determina así:

$$\sigma_{\Delta \bar{p}} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

Con este error se tiene mayor precisión en la estimación, es decir, la diferencia entre L_s y L_i es menor.

Ejemplo:

Mediante un intervalo de confianza del 98 %, se desea conocer la verdadera diferencia de porcentajes de alumnos que aprueban la materia de Estadística según género, sabiendo que, de 30 hombres, aprobaron 25, y de 20 mujeres, aprobaron 15.

Solución:

$$\bar{p}_1 = \frac{X_1}{n_1} \text{ y } \bar{p}_2 = \frac{X_2}{n_2}.$$

$$\Delta\bar{p} = \bar{p}_1 - \bar{p}_2$$

$$\bar{p}_1 = \frac{X_1}{n_1} = \frac{25}{30} = 0,83 \text{ y } \bar{p}_2 = \frac{X_2}{n_2} = \frac{15}{20} = 0,75.$$

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{25 + 15}{30 + 20} = \frac{40}{50} = 0,8.$$

$$\Delta\bar{p} = \bar{p}_1 - \bar{p}_2 = 0,83 - 0,75 = 0,08.$$

$$\sigma_{\Delta\bar{p}} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

Límites de confianza:

$$(\bar{p}_1 - \bar{p}_2) \pm Z_{\frac{\alpha}{2}} \times \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0,08 \pm 2,33 \times 0,115 = 0,08 \pm 0,268.$$

$$\text{Límite inferior: } L_i = 0,08 - 0,268 = -0,188.$$

$$\text{Límite superior: } L_s = 0,08 + 0,268 = 0,348.$$

De la relación de probabilidad se tiene lo siguiente:

$$P\left(\left(\bar{p}_1 - \bar{p}_2\right) - Z_{\frac{\alpha}{2}} \times \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} < \Delta P < \left(\bar{p}_1 - \bar{p}_2\right) + Z_{\frac{\alpha}{2}} \times \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right) = 1 - \alpha.$$

De aquí se sigue que: $P(-0,188 < \Delta P < 0,348) = 98\%$.

R/ La verdadera diferencia de proporciones está entre $-0,188$ y $0,348$; por lo tanto, no existe diferencia significativa entre los porcentajes de alumnos que aprueban la materia.

7.3 PRUEBAS DE HIPÓTESIS

Las hipótesis son supuestos que el investigador hace acerca de los parámetros de una población para decidir si un parámetro ha cambiado. Por ejemplo, para determinar si un nuevo medicamento es más eficiente que otro; sin método de enseñanza produce mejores resultados que otro.

7.3.1 PROCEDIMIENTO

Para realizar una prueba de hipótesis se deben aplicar los siguientes pasos:

Determinar la distribución de probabilidad adecuada

La distribución de probabilidad puede ser normal, t de Student, Ji-Cuadrado, F (Fisher).

La *Distribución Normal* se utiliza para pruebas de hipótesis en poblaciones con distribución normal, en los siguientes casos: a) Si la muestra es pequeña ($n \leq 30$) y se conoce la desviación estándar de la población (σ); b) la muestra es grande ($n > 30$) sin importar si se conoce o no el valor de la desviación estándar (σ).

La *Distribución t de Student* se utiliza para muestras pequeñas ($n \leq 30$) y no se conoce la desviación estándar (σ) de la población. Se aplica para realizar pruebas de hipótesis de la media y en diferencia de medias en muestras dependientes e independientes.

La *Distribución Ji-Cuadrado* se utiliza para realizar prueba de hipótesis de la varianza poblacional en muestras de tamaño menor o igual que 100. También se utiliza en pruebas no paramétricas como la prueba de independencia, de homogeneidad y de bondad de ajuste. La prueba de independencia permite comprobar si existe relación entre dos variables cualitativas; y la prueba de bondad de ajuste se utiliza para comparar si una distribución de frecuencias se aproxima a una distribución de probabilidades.

La *Distribución F* se utiliza para comparar dos varianzas, en análisis de varianza (también llamado anova) y en otras pruebas no paramétricas.

Plantear las hipótesis.

Se plantea dos hipótesis: hipótesis nula, que se representa por H_0 , y la hipótesis alterna o de trabajo, que se representa por H_1 .

Ejemplo:

Si las hipótesis se refieren a la *media de la población*, se pueden presentar tres casos: prueba unilateral derecha, izquierda y bilateral, como se indica en la Tabla 6. El sentido de la prueba lo determina la hipótesis alterna, tal como se indica en la Tabla 6.

Tabla 6. Ejemplo de planteamiento de hipótesis respecto a la media poblacional

| <i>Prueba unilateral derecha</i> | <i>Prueba unilateral izquierda</i> | <i>Prueba bilateral</i> |
|----------------------------------|------------------------------------|-------------------------|
| $H_0: \mu \leq \mu_0$ | $H_0: \mu \geq \mu_0$ | $H_0: \mu = \mu_0$ |
| $H_1: \mu > \mu_0$ | $H_1: \mu < \mu_0$ | $H_1: \mu \neq \mu_0$ |

Escoger el nivel de significancia α

α = *probabilidad de rechazar la hipótesis H_0 cuando es verdadera* (Error tipo I).

Generalmente, este valor se escoge entre 1 % y 5 %. Cuando se trabaja con el 5% y se rechaza la hipótesis nula, se puede concluir que existe una diferencia significativa entre el estadístico de prueba y el parámetro; y si se rechaza la

hipótesis nula con un nivel de significancia del 1%, la diferencia es altamente significativa. Al ubicar el valor de significancia en la curva de probabilidad, quedan definidas las regiones de *aceptación* y de *rechazo* de H_0 .

Representar gráficamente la situación

Sin importar el tipo de distribución el valor de α divide la curva en dos secciones: región de aceptación y región rechazo de la hipótesis nula.

Si la prueba es unilateral derecha, el valor de α se ubica en la cola derecha de la curva; si es unilateral izquierda, α se ubica en la cola izquierda; y si es bilateral, el valor de α se ubica $\frac{\alpha}{2}$ en cola derecha y $\frac{\alpha}{2}$ en la izquierda. El valor de α determina un valor crítico $Z_\alpha, t_\alpha, \chi^2_\alpha, F_\alpha$.

El Gráfico 12, corresponde a una representación de una prueba unilateral derecha.

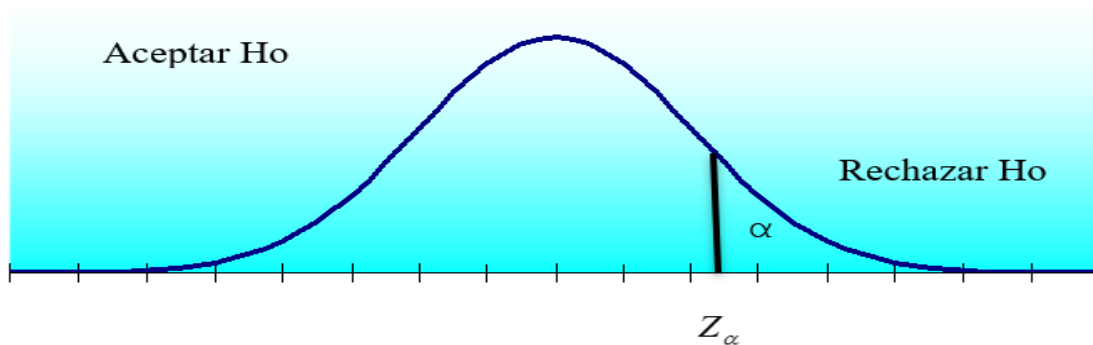


Gráfico 12. Ejemplo de zona de aceptación y rechazo de la hipótesis nula

Calcular el estadígrafo de prueba Z, t, χ^2 o F

Se determina con la fórmula correspondiente a la distribución.

Plantear las reglas de decisión, y, comparar los valores y tomar la decisión

Existen tres métodos para tomar la decisión. A continuación, se describen los métodos para una distribución normal.

Primer método

Reglas de decisión

Si $Z_{calculado} > Z_{\alpha}$, se rechaza H_0 .

Si $Z_{calculado} \leq Z_{\alpha}$, se acepta H_0 .

Comparar y tomar la decisión.

Segundo método

La decisión también se puede tomar comparando el valor de α con $p - value$, el cual corresponde a $P(Z > Z_{calculado})$, en el caso de una prueba unilateral derecha;

y $P(Z < Z_{calculado})$ en prueba unilateral izquierda.

Reglas de decisión

Si $p - value \geq \alpha$, se acepta la hipótesis nula.

Si $p - value < \alpha$, se rechaza la hipótesis nula.

Comparar y tomar la decisión.

Tercer método

Calcular el punto crítico K y comparar con el estadígrafo de prueba:

$$K = \text{parámetro} + Z_{\alpha} \times \text{error estándar.}$$

Reglas de decisión

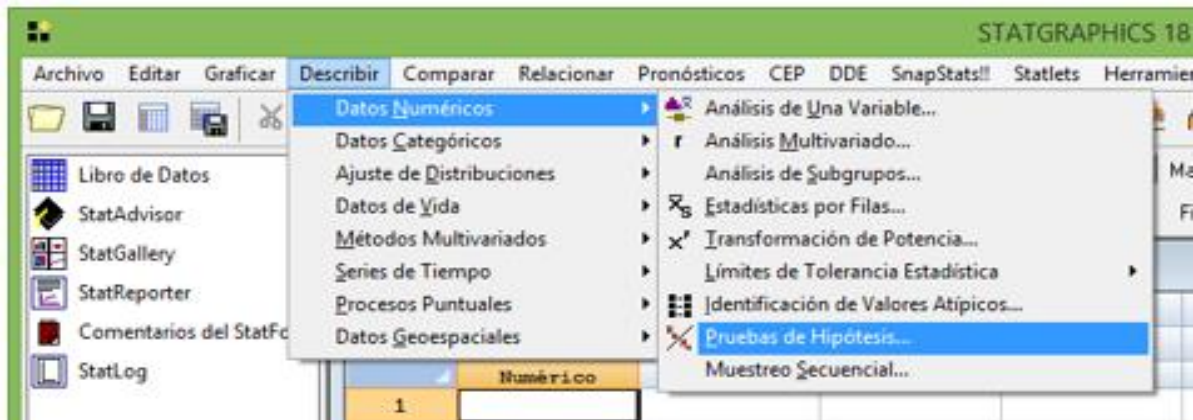
Si el estadígrafo de prueba es mayor que K , se rechaza H_0 .

Si el estadígrafo de prueba es menor que K , se acepta H_0

Comparar y tomar la decisión.

Prueba de hipótesis con Statgraphics

En el menú Seleccionar la opción *DESCRIBIR* datos numéricos, luego prueba de hipótesis para la media, proporciones o desviación estándar, según el caso, y digitar los valores que solicite en cada casilla; seleccionar mayor que, menor que o diferente; Aceptar y se obtiene como resultado los límites de confianza y la prueba de hipótesis (ver Imagen 12).



Pruebas de Hipótesis

Parámetro

- Media Normal
- Sigma Normal
- Proporción Binomial
- Tasa de Poisson
- Coeficiente de Correlación

Hipótesis Nula: 0,5

Media Muestral: 0,0 Sigma Muestral: 1,0

Proporción de la Muestra: 0,5 Tasa de la Muestra: 1,0

Correlación de la Muestra: 0,0

Tamaño de Muestra: 100

Aceptar
Cancelar
Ayuda

Opciones Prueba de Hipótesis

Hipótesis Alterna

- No Igual
- Menor que
- Mayor que

Alpha: 5,0 % Usar Prueba-Z

Aceptar
Cancelar
Ayuda

Imagen 12. Prueba de hipótesis con Statgraphics

Para diferencia de promedios o proporciones, en el menú se debe escoger la opción COMPARAR dos muestras pruebas de hipótesis, y teclear los datos solicitados. Si son muestras dependientes o independientes, opción COMPARAR dos muestras, dependientes o independientes.

Se presentan ejemplos sobre pruebas de hipótesis de la media, proporciones, diferencia de medias, diferencia de proporciones y se trabaja el tema de potencia de una prueba.

7.3.1.1 Prueba de hipótesis para la media (μ)

Ejemplo:

Una compañía comercializa un programa de adiestramiento basado en la motivación; asegura que las tasas de rendimiento de los trabajadores se elevan después de realizar un curso. La compañía interesada en comprar el programa realiza un muestreo aleatorio de 80 empleados y encuentra que su promedio es $\bar{X} = 77$ puntos de rendimiento en una escala de 0 a 100. La compañía que ofrece el programa, asegura que el promedio de rendimiento es $\mu = 75$ puntos por trabajador, con una desviación estándar $\sigma = 13$ puntos. Se desea comprobar la afirmación del oferente del programa, con un nivel de significancia del 5 %.

Solución:

Los resultados del muestreo son los siguientes:

Promedio: $\bar{X} = 77$; desviación estándar $\sigma = 13$ y tamaño de la muestra $n = 80$.

Planteamiento de hipótesis:

$H_0: \mu \leq 75 \rightarrow$ El curso no aumenta el rendimiento.

$H_1: \mu > 75 \rightarrow$ El curso aumenta el rendimiento.

Nivel de significancia: $\alpha = 5\%$ (ver Gráfico 13).

$Z_\alpha = 1,64$.

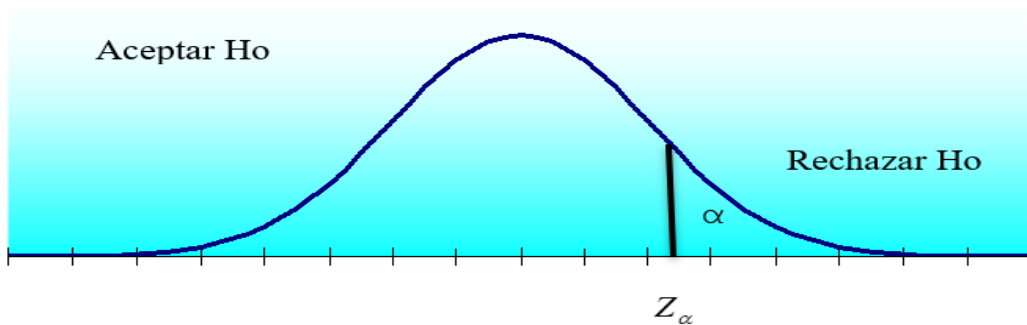


Gráfico 13. Prueba de hipótesis para la media

Se calcula el valor de Z ($Z_{calculado}$) de la siguiente manera:

$$Z_{calculado} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{77 - 75}{\frac{13}{\sqrt{80}}} = 1,38.$$

Reglas de decisión:

Si $Z_{calculado} \geq Z_{\alpha}$, se rechaza H_0 .

Si $Z_{calculado} \leq Z_{\alpha}$, se acepta H_0 .

Dado que $Z_{calculado} = 1,38 < Z_{\alpha} = 1,64$, se acepta H_0 .

R/ No existe evidencia para concluir que el curso aumente el rendimiento.

Segundo método

Determinar $p - value$.

$$p - value = P(Z > 1,38) = 0,0838$$

Reglas de decisión:

Si $p - value \geq \alpha$, se acepta H_0 .

Si $p - value < \alpha$, se rechaza H_0 .

Como $p - value = 0,0808 > 0,05$, se acepta H_0 .

R/ No existe evidencia para concluir que el curso aumente el rendimiento.

Tercer método

Calcular el punto crítico K .

$$K = \mu_0 + Z_{\alpha} \times \frac{\sigma}{\sqrt{n}} = 75 + 1,64 \times \frac{13}{\sqrt{80}} = 77,38$$

Como $\bar{X} < K$, se acepta H_0 .

R/ No existe evidencia para concluir que el curso aumente el rendimiento.

7.3.1.2. Prueba de hipótesis para la proporción (p)

Ejemplo:

Un fabricante afirma que al menos el 30% del público prefiere su producto. Se toma una muestra aleatoria de 100 personas para verificar su afirmación, 25 de ellas expresaron su preferencia por el producto. ¿Este dato es suficiente evidencia para comprobar la información del fabricante a un nivel de significancia del 5%?

Solución:

Planteamiento de hipótesis:

$H_0: p \geq 0,30$ (El fabricante tiene razón).

$H_1: p < 0,30$.

$\alpha = 5\%$

$n = 100$

$x = 25$

$$\bar{p} = \frac{x}{n} = \frac{25}{100} = 0,25$$

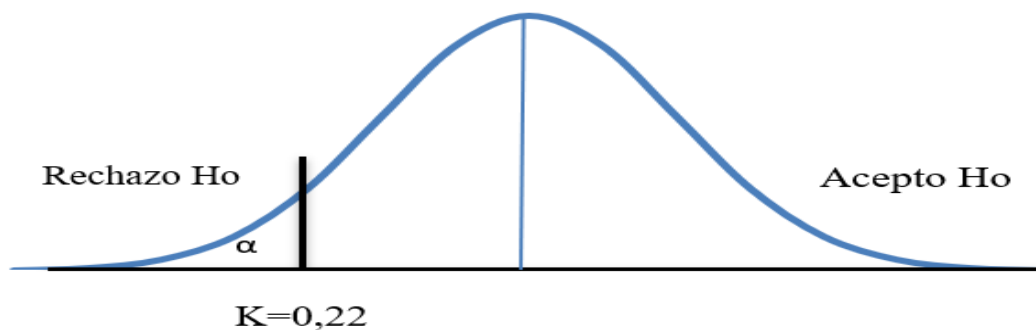


Gráfico 14. Prueba de hipótesis para la proporción

Primer método

Calcular Z.

$$Z_{\text{calculado}} = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0,25 - 0,30}{\sqrt{\frac{0,30 \times 0,70}{100}}} = \frac{-0,05}{\sqrt{0,00021}} = \frac{-0,05}{0,0458} = -1,09.$$

$$Z_{\alpha} = -1,64$$

Reglas de decisión

Si $Z_{\text{calculado}} < Z_{\alpha}$ rechazo H_0 .

Si $Z_{\text{calculado}} \geq Z_{\alpha}$ acepto H_0 .

Como $Z_{\text{calculado}} > Z_{\alpha}$ se acepta H_0 .

Segundo método

$$p - \text{value} = P(\bar{p} < 0,25) = P(Z < -1,09) = 0,1369 .$$

Como $p - \text{value} > \alpha$, se acepta H_0 .

Tercer método

$k = \text{parámetro} + Z \times \text{error estándar} .$

$$\text{Punto crítico: } k = p_o + \sqrt{\frac{pq}{n}} = 0,30 - 1,64 \times \sqrt{\frac{0,30 \times 0,70}{100}} = 0,22 .$$

Reglas de decisión:

Si $\bar{p} > k$, entonces acepto H_0 .

Si $\bar{p} < k$, entonces rechazo H_0 .

Dado que $\bar{p} = 0,25 > k = 0,22$ entonces se acepta H_0 (ver Gráfico 14).

R/ El fabricante tiene la razón cuando afirma que al menos el 30% del público prefiere su producto.

7.3.1.3 Prueba de hipótesis para la diferencia de medias

$$\delta = \mu_1 - \mu_2$$

Ejemplo:

Un diseñador de productos está interesado en reducir el tiempo de secado de una pintura. Se prueban dos fórmulas de pintura: la fórmula 1 tiene el contenido químico estándar, mientras que la fórmula 2 tiene un nuevo ingrediente, que tiende a reducir el tiempo de secado. De la experiencia se sabe que la desviación estándar del tiempo de secado es ocho minutos, y esta variabilidad inherente no debe verse afectada por la adición del nuevo ingrediente. Se pintan 35 placas con la fórmula 1 y otras 35 con la fórmula 2. Los dos tiempos muestrales promedio de secado son 116 minutos para la fórmula 1 y 112 minutos para la fórmula 2. ¿A qué conclusión puede llegar el diseñador del producto sobre la eficacia del nuevo ingrediente, al nivel de significancia 0,01?

Solución:

Planteamiento de hipótesis:

$$H_0: \mu_1 - \mu_2 = \delta = 0.$$

$$H_1: \mu_1 - \mu_2 = \delta > 0.$$

$$\alpha = 1\%$$

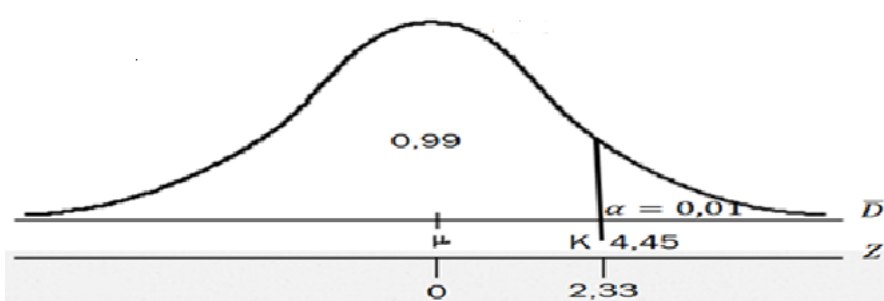


Gráfico 15. Prueba de hipótesis para la diferencia de medias

| | n | σ | \bar{X} |
|-----------|-----|----------|-----------|
| Fórmula 1 | 35 | 8 | 116 |
| Fórmula 2 | 35 | 8 | 112 |

$$k = \delta \pm Z_{\alpha} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} .$$

$$k = 0 + 2,33 \times \sqrt{\frac{8^2}{35} + \frac{8^2}{35}} = 4,45 .$$

$$\bar{D} = \bar{X}_1 - \bar{X}_2 = 116 - 112 = 4 .$$

Reglas de decisión

Si $\bar{D} < k$, se acepta H_0 .

Si $\bar{D} > k$, se rechaza H_0 .

Dado que $\bar{D} = 4 < 4,45$, entonces acepta H_0 (ver Gráfico 15).

R/ NO hay evidencia suficiente para demostrar que el tiempo de secado disminuye significativamente.

$$p - value = P(\bar{D} > 4) = P(Z > 1,91) = 0,0281 .$$

Como $p - value > \alpha$, se acepta H_0 .

7.3.1.4 Prueba de hipótesis para la diferencia de proporciones (Δp)

Ejemplo:

Una compañía asegura que el mercado para su producto X tiene una aceptación de iguales proporciones (porcentaje de éxitos) en la ciudad A que en la ciudad B. Un especialista en mercado pone en duda dicha afirmación y, para tal fin, tomó una muestra aleatoria de 500 amas de casa en la ciudad A y encontró que el 59.6 % de las

mismas prefería el artículo X. Por otra parte, tomó una muestra aleatoria de 300 amas de casa en la ciudad B y encontró que el 50 % de las mismas preferían el artículo X. ¿Existe una diferencia real entre las dos ciudades a un nivel de significancia 5 %?

Solución:

Planteamiento de hipótesis:

$$H_0: p_1 = p_2 \text{ o } p_1 - p_2 = 0.$$

$$H_1: p_1 - p_2 \neq 0.$$

$$\alpha = 5\%$$

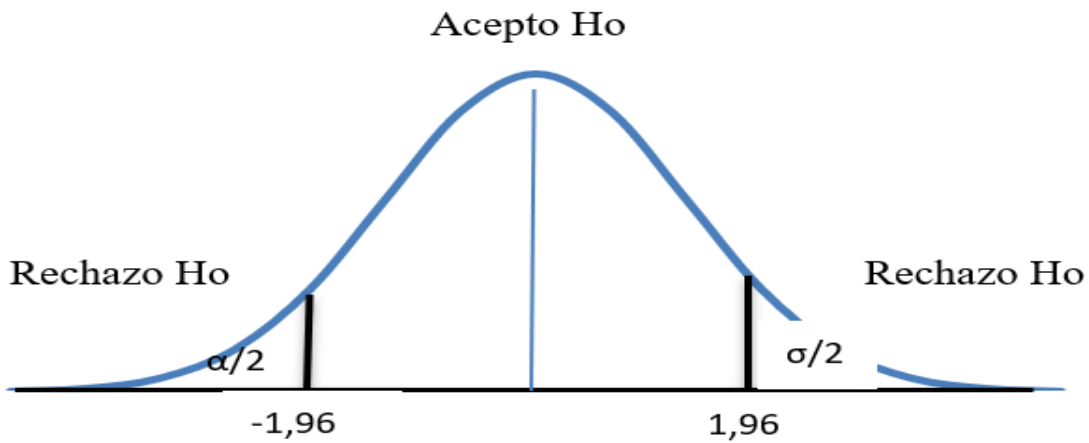


Gráfico 16. Prueba de hipótesis para la diferencia de proporciones

| Muestra 1 | Muestra 2 |
|-----------------------|---------------------|
| $n_1 = 500$ | $n_2 = 300$ |
| $\bar{p}_1 = 59,6 \%$ | $\bar{p}_2 = 50 \%$ |
| $\bar{q}_1 = 40,4 \%$ | $\bar{q}_2 = 50 \%$ |

$$\Delta_{\bar{p}} = 0,596 - 0,50 = 0,096.$$

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{500 \times 0,596 + 300 \times 0,50}{500 + 300} = \frac{298 + 150}{800} = 0,56.$$

$$\sigma_{\Delta\bar{p}} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0,56 \times 0,44 \left(\frac{1}{500} + \frac{1}{300}\right)} = 0,036.$$

$Z_{\text{critico}} = 1,96$ (ver Gráfico 16).

$$k_1 = \Delta_p - Z \times \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0 - 1,96 \times 0,036 = -0,07056.$$

$$k_2 = \Delta_p + Z \times \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0 + 1,96 \times 0,036 = 0,07056.$$

Reglas de decisión:

Si $k_1 < \Delta_{\bar{p}} < k_2$, entonces se acepta H_0 .

Si $\Delta_{\bar{p}} < k_1$, entonces se rechaza H_0 .

Si $\Delta_{\bar{p}} > k_2$, entonces se rechaza H_0 .

Dado que $\Delta_{\bar{p}} = \bar{p}_1 - \bar{p}_2 = 0,596 - 0,50 = 0,096 > k_2 = 0,070$, entonces se rechaza H_0 .

$$p - \text{value} = P(\Delta_{\bar{p}} > 0,096) = P(Z > 2,69) = 0,036.$$

Como $p - \text{value} < \alpha$, se rechaza H_0 .

R/ A la luz de la información muestral, se concluye con una baja probabilidad de error, de 0,0036, que la proporción de aceptación del producto es diferente en las ciudades A y B.

7.3.1.5 Potencia de una prueba

La Potencia de una prueba estadística o el Poder Estadístico es la probabilidad de que la hipótesis nula sea rechazada cuando esta es falsa (es decir, la probabilidad de no cometer un error del tipo II). La Potencia es, en general, una función de las distribuciones posibles, a menudo determinada por un parámetro, bajo la hipótesis

alternativa. A medida que aumenta la Potencia, las posibilidades de que se presente un error del tipo II, disminuyen. La probabilidad de que ocurra un error de tipo II, se conoce como la Tasa de Falsos Negativos (β), que es la probabilidad de aceptar la H_0 cuando ésta es falsa. Por lo tanto, la Potencia es igual a $1 - \beta$, que también se conoce como Sensibilidad.

| Estado de naturaleza | Aceptar | Rechazar |
|-----------------------|------------------------------------|---|
| Si H_0 es verdadera | Decisión correcta ($1 - \alpha$) | Error tipo I (α) |
| Si H_0 es falsa | Error tipo II (β) | Decisión correcta ($1 - \beta$) Potencia |

Ejemplo:

Un fabricante de autos sostiene que sus vehículos consumen en promedio 55,5 galones de combustible cada 100 km. Un vendedor de la compañía comprueba el consumo de gasolina de 60 autos y encuentra que el consumo medio de este grupo es de 55,65 galones por cada 100 km. Si la desviación estándar del consumo es 70 galones, determinar si es cierta la afirmación del fabricante a un nivel de significancia del 1 %.

Solución:

Datos:

$$\alpha = 1 \%$$

$$n = 60$$

$$\bar{X} = 55,65$$

$$\sigma = 70$$

Planteamiento de hipótesis:

$H_0: \mu = 55,5$ galones (El fabricante tiene razón).

$H_1: \mu > 55,5$ galones (El vendedor tiene razón).

Punto crítico k :

$$K = \mu_0 + Z \times \frac{\sigma}{\sqrt{n}}$$

Remplazamos los datos en la fórmula:

$$K = \mu_0 + Z_\alpha \times \frac{\sigma}{\sqrt{n}} = 55,5 + 2,33 \times \frac{70}{\sqrt{60}} = 76,5.$$

Reglas de decisión:

Si $\bar{X} < K$ se acepta H_0 .

Si $\bar{X} > K$ se rechaza H_0 .

Dado que $\bar{X} = 55,65 < K = 76,5$, entonces se acepta H_0 .

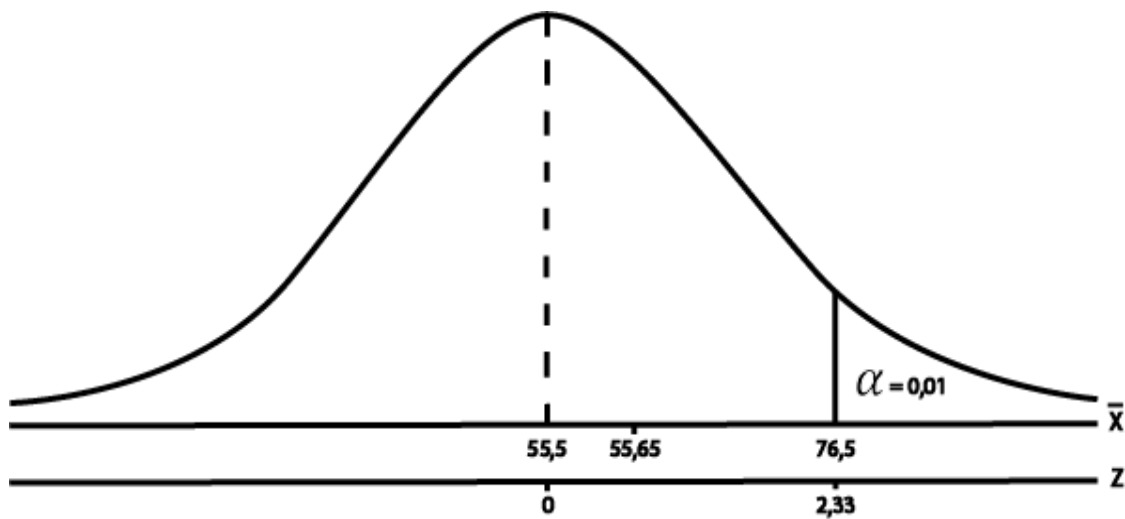


Gráfico 17. Punto crítico $k= 76,5$ para calcular la potencia de la prueba

Potencia:

Encontrar la potencia de la prueba utilizando el punto K y diversos valores de la media (ver Gráfico 17).

μ_1 : 80; 85; 90 y 95.

a) μ_1 : 80; b) μ_1 : 85; c) μ_1 : 90; d) μ_1 : 95

Solución:

a) El valor de β para la media de μ_1 : 80

$$Z = \frac{76,5 - 80}{\frac{70}{\sqrt{60}}} = -0,3872 .$$

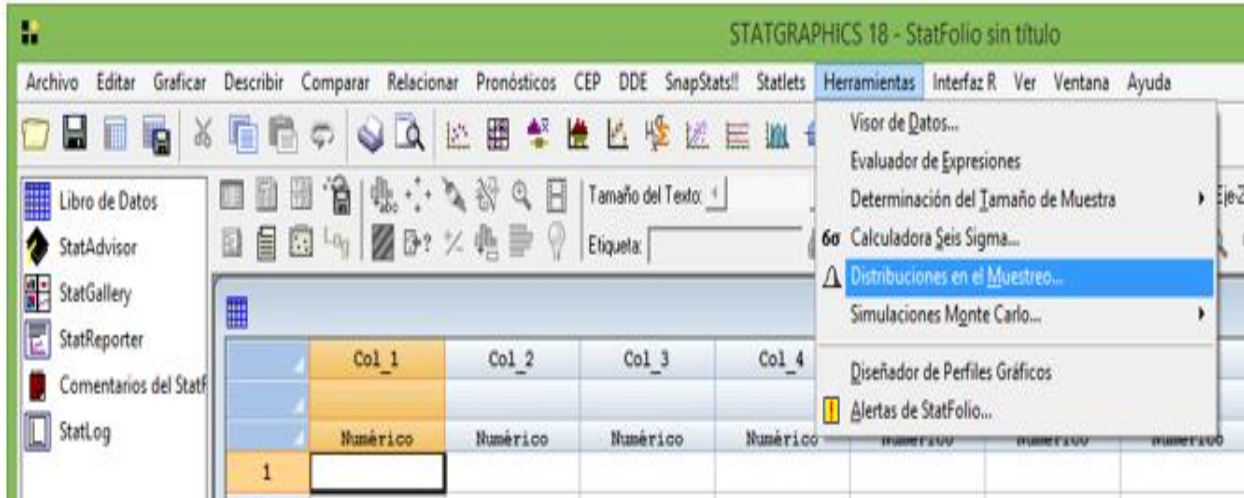
Se busca en la tabla normal el valor $Z = -0,3872$ y se obtiene el valor 0,3493 .

Entonces, $\beta = P(\bar{X} < K/\mu_1 = 80) = P(\bar{X} < 76,5/\mu_1 = 80) = P(Z < -0,3872) = 0,3493 .$

Media 80, error estándar 9,037.

Cálculo de la potencia de la prueba ($1 - \beta$) con Statgraphics

En Statgraphics, para encontrar la Potencia, en el menú se escoge la opción herramientas, distribuciones muestrales, seleccionar la distribución normal, t, Ji-Cuadrado o F, según el caso, y digitar la información solicitada (ver Imagen 13).



Opciones de distribuciones en el muestreo ✖

Límite Inferior: Límite Superior: Área:

Calcular

- Área límite inferior
- Área límite superior
- Área entre límites
- Área fuera de límites
- Valor crítico inferior para un área dada
- Valor crítico superior para un área dada
- Valor crítico conteniendo un área dada
- Valor crítico excluyendo un área dada

Distribuciones en el muestreo ✖

Distribución

- Normal

| | |
|--|---|
| Media: <input type="text" value="80.0"/> | Desviación típica: <input type="text" value="9.036"/> |
|--|---|
- t de Student

| | |
|---|--|
| Grados de libertad: <input type="text" value="10.0"/> | |
|---|--|
- Chi-Cuadrada

| | |
|---|--|
| Grados de libertad: <input type="text" value="10.0"/> | |
|---|--|
- F (Razón de Varianzas)

| | |
|--|---|
| G.I. numerador: <input type="text" value="1.0"/> | G.I. denominador: <input type="text" value="10.0"/> |
|--|---|

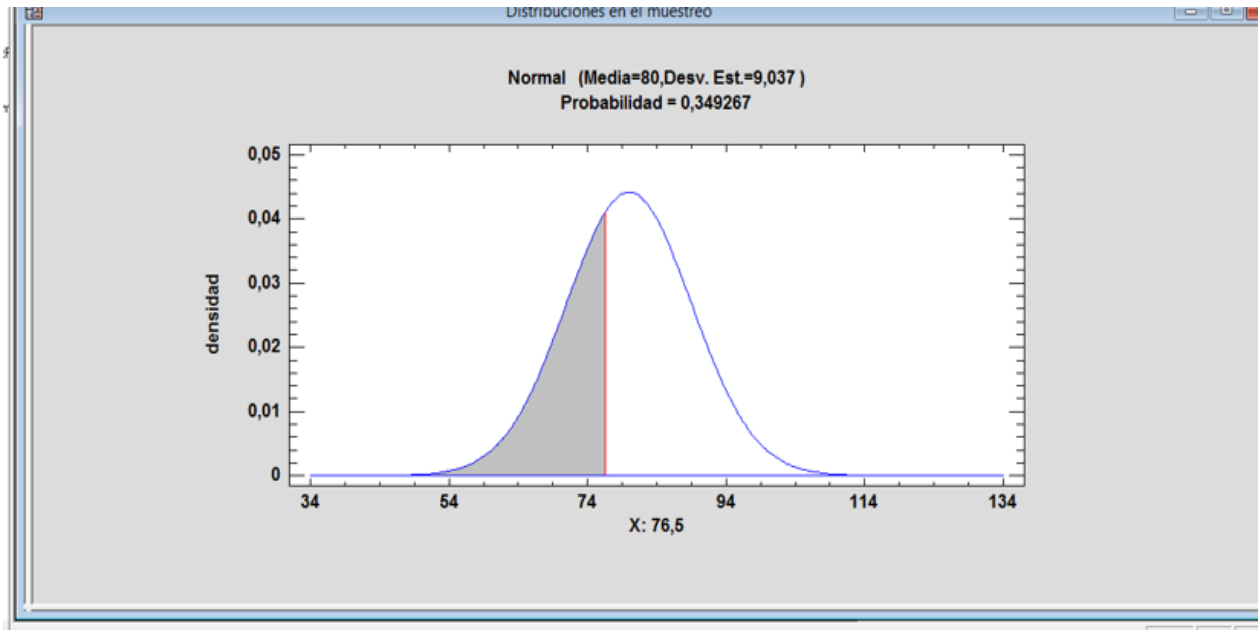


Imagen 13. Cálculo de potencia con Statgraphics: DISTRIBUCIONES EN EL MUESTREO

b) El valor de β para la media de $\mu_1: 85$ (ver Imagen 14).

$$Z = \frac{76,5 - 85}{\frac{70}{\sqrt{60}}} = \frac{-8,5}{9,037} = -0,94.$$

Se busca en la tabla normal el valor $Z = -0,94$ y se obtiene el valor 0,1736.

Entonces, $\beta = P(\bar{X} < 76,5 / \mu_1 = 85) = P(Z < -0,94) = 0,1736$.

c) El valor de β para la media de $\mu_1: 90$

$$Z = \frac{76,5 - 90}{\frac{70}{\sqrt{60}}} = \frac{-13,5}{9,037} = -1,49.$$

Se busca en la tabla normal el valor $Z = -1,49$ y se obtiene el valor 0,0681.

Entonces, $\beta = P(\bar{X} < 76,5 / \mu_1 = 90) = P(Z < -1,49) = 0,0681$.

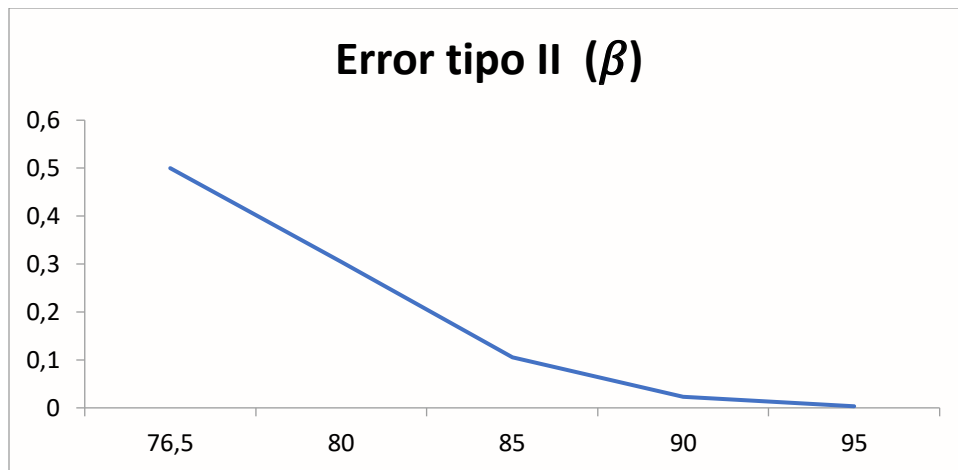
d) El valor de β para la media de $\mu_1: 95$

$$Z = \frac{76,5 - 95}{\frac{70}{\sqrt{60}}} = \frac{-18,5}{9,037} = -2,05 .$$

Se busca en la tabla normal el valor $Z = -2,05$ y se obtiene el valor 0,0202 .

Entonces, $\beta = P(\bar{X} < 76,5 / \mu_1 = 95) = P(Z < -2,05) = 0,0202 .$

| μ_1 | β | $1 - \beta$ |
|---------|---------|-------------|
| 76,5 | 0,50 | 0,5 |
| 80 | 0,3483 | 0,6517 |
| 85 | 0,1736 | 0,8264 |
| 90 | 0,0681 | 0,9319 |
| 95 | 0,0202 | 0,9798 |



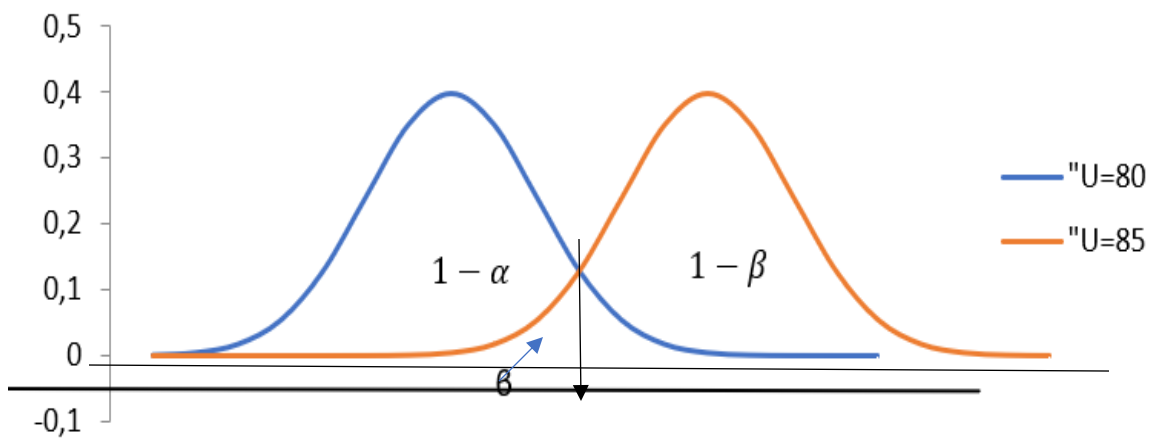
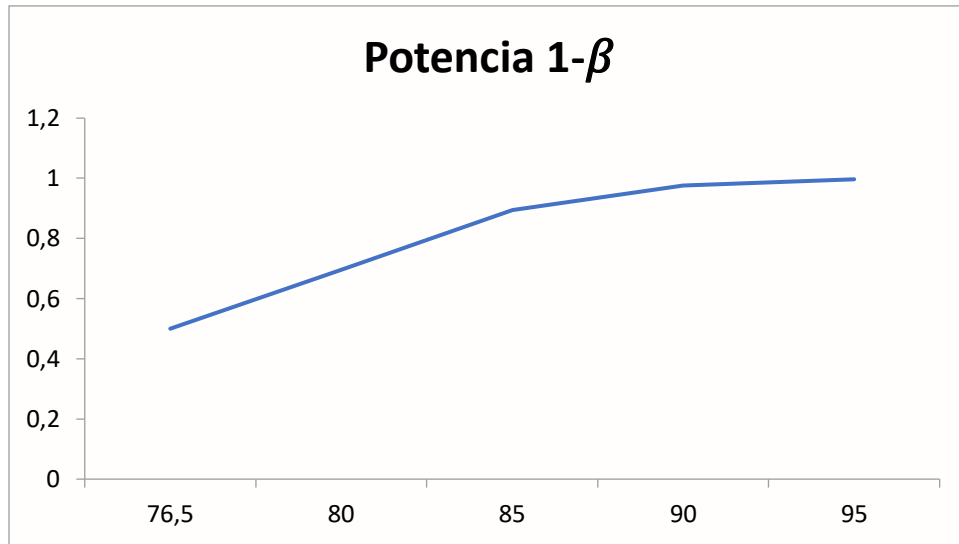


Imagen 14. Representación de la potencia de la prueba

Potencia de una prueba de hipótesis con Microsoft Excel

Para encontrar la Potencia en Excel, insertar función estadística DISTR.NORM.N, en X digitar el estadígrafo de prueba, la media, la desviación estándar y en acumulado 1, con lo cual se obtiene el área de la izquierda y si necesita el área de la derecha, se procede por complemento.

Se calcula β cuando la media es 80, error estándar 9,037 y el estadígrafo de prueba 76,5.

Se presenta una captura de pantalla, con los datos del punto a) (ver Imagen 15).

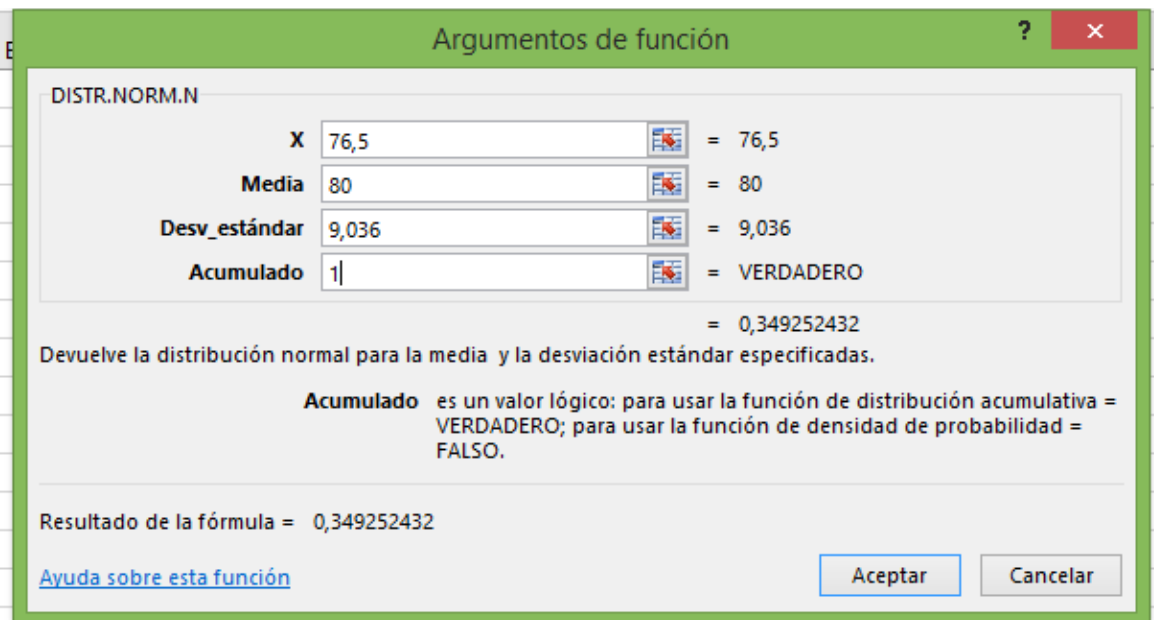
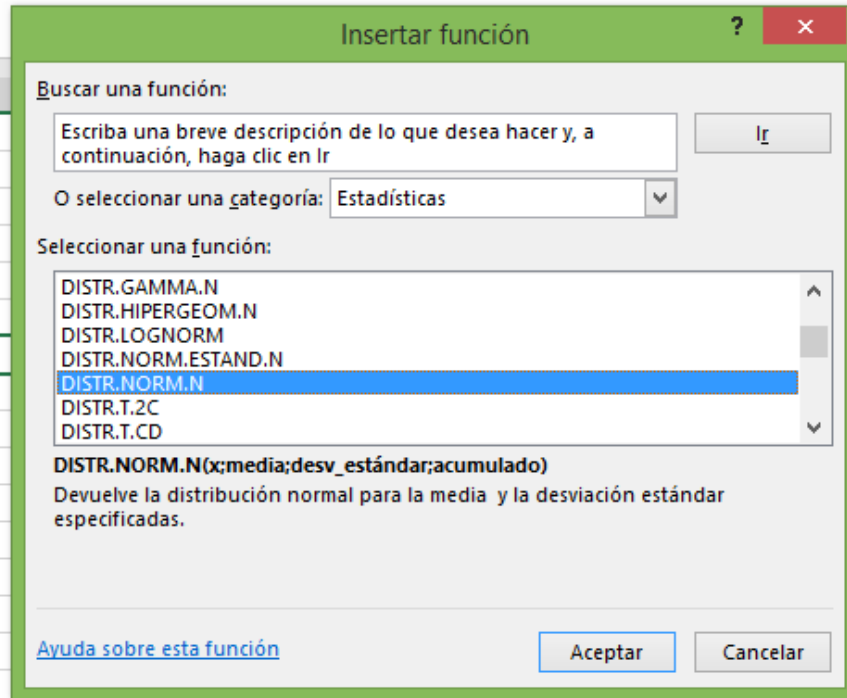


Imagen 15. Microsoft Excel: argumentos de función DISTR.NORM.N

7.3.2 Distribución t de Student

La Distribución Normal se utilizó para hacer inferencias sobre muestras grandes en distribución muestral de medias y en diferencia de medias, tanto para estimación por intervalos como para prueba de hipótesis.

Cuando las muestras provienen de una distribución normal y son pequeñas ($n \leq 30$), y, además, se desconoce la desviación estándar de la población (σ), se utiliza la distribución t de Student para prueba de hipótesis de la media y para diferencia de medias en muestras dependientes e independientes.

Se sabe que las variables Z y t se expresan como sigue:

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}} \text{ con desviación estándar sesgada.}$$

$$t = \frac{\bar{X} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \text{ con desviación estándar insesgada.}$$

Observando las fórmulas de Z y t , se encuentra que la diferencia está en el denominador, así: cuando se aplica la *distribución normal*, se utiliza el valor de σ , que corresponde a la desviación estándar de la población y cuando se aplica la *distribución t de Student*, se utiliza el estadígrafo s , que representa la desviación estándar de una muestra de tamaño n . El valor $\nu = n - 1$, representa el número de grados de libertad de la distribución t de Student.

Concepto de grados de libertad, es el número de valores que se asignan libremente para obtener un resultado, por ejemplo, si la suma de 5 números debe dar 50 se puede asignar libremente cuatro valores, y el quinto se encuentra por diferencia. En este caso el número de grados de libertad es $u = n - 1 = 5 - 1 = 4$

Ejemplo:

Si una muestra es de tamaño $n = 20$ y el área $(1 - \alpha)$ a la izquierda es 95%, entonces $t_{(0,95;\nu=19)} = 1,73$.

7.3.2.1 Estimación por intervalos para la media de la población

Para encontrar los límites de confianza, es necesario tomar una muestra de tamaño n , definir el grado de confiabilidad $1 - \alpha$, y calcular el promedio y la desviación estándar de la muestra. En relación a los límites, estos se determinan de manera similar a la distribución normal.

Ejemplo:

Encontrar los límites del 95% de confianza para la verdadera media de la población, si en una muestra de 10 personas, la variable edad tiene los siguientes datos:

X (edad): 10 12 18 18 15 20 21 14 13 20.

Calcular promedio y desviación estándar de la muestra.

El promedio es $\bar{X} = 16,1$ y la desviación estándar $S = 3,62$. Para buscar el valor de t en la tabla, se ubica en $\nu = 9$ grados de libertad, y si el área a la izquierda es 0,975, se busca $t = 0,975$. En este caso, el valor correspondiente es $t = 2,26$, con grado de confianza 95% y $\nu = 9$ grados de libertad.

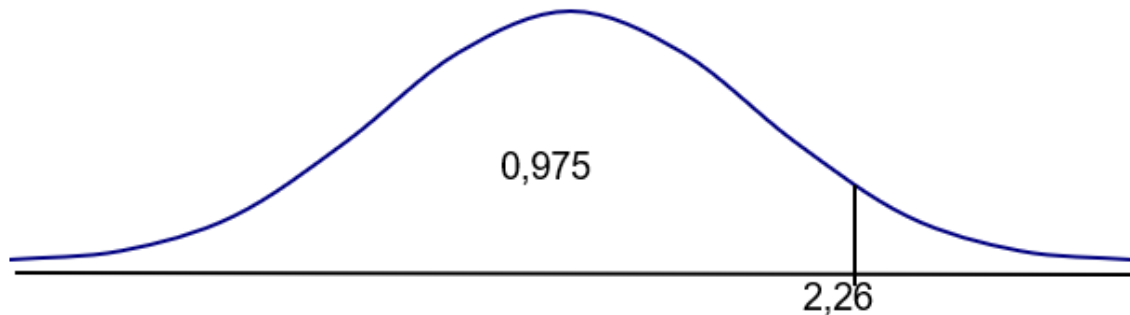


Gráfico 18. Estimación por intervalos para la media de la población

Límites de confianza para poblaciones infinitas: $\bar{X} \pm t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n-1}}$.

Límites de confianza para poblaciones finitas: $\bar{X} \pm t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n-1}} \sqrt{\frac{N-n}{N-1}}$.

Solución:

El valor de $t_{\frac{\alpha}{2}}$ para 9 grados de libertad y con un área a la izquierda de 0,975 es 2,26 (ver Gráfico 18); el promedio de los datos es 16,1 y la desviación estándar 3,62, entonces los límites de confianza son:

$$\text{Límites de confianza: } \bar{X} \pm t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n-1}} \sqrt{\frac{N-n}{N-1}} = 16,1 \pm 2,26 \times \frac{3,62}{\sqrt{9}} = 16,1 \pm 2,73 .$$

$$\text{Límite inferior: } L_i = 16,1 - 2,73 = 13,37 .$$

$$\text{Límite superior: } L_s = 16,1 + 2,73 = 18,83 .$$

R/ Con un 95 % de confiabilidad, la media de la población está comprendida entre 13,37 y 18,83; es decir: $13,37 < \mu < 18,83$.

7.3.2.2 Límites de confianza para diferencia de medias poblacionales

Para encontrar los límites de la diferencia de medias, es necesario considerar si las muestras son dependientes o independientes.

Caso 1: muestras independientes

De manera similar a la determinación del valor de Z en la distribución normal, el valor t se puede calcular de la siguiente manera:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} .$$

Los límites de confianza para la diferencia de medias son:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} .$$

Si las varianzas son iguales, para obtener una mejor aproximación de la varianza muestral, es posible agrupar los datos de las dos muestras, como si se tratara de una sola.

$$\text{Varianza 1: } S_1^2 = \frac{\sum(X - \bar{X})^2}{n_1}; \quad \text{Varianza 2: } S_2^2 = \frac{\sum(X - \bar{X})^2}{n_2} \text{ (varianza sesgada).}$$

$$\text{Varianza 1: } \hat{S}_1^2 = \frac{\sum(X - \bar{X})^2}{n_1 - 1}; \quad \text{Varianza 2: } \hat{S}_2^2 = \frac{\sum(X - \bar{X})^2}{n_2 - 1} \text{ (varianza insesgada).}$$

De lo anterior se tiene que:

$$S_1^2 = \frac{\sum(X - \bar{X})^2}{n_1} \rightarrow S_1^2 \times n_1 = \sum(X - \bar{X})^2.$$

$$\hat{S}_1^2 = \frac{\sum(X - \bar{X})^2}{n_1 - 1} \rightarrow \hat{S}_1^2 \times (n_1 - 1) = \sum(X - \bar{X})^2.$$

De aquí se deduce que:

$$S_1^2 \times n_1 = \hat{S}_1^2 \times (n_1 - 1).$$

Para el caso de la varianza sesgada, se tiene lo siguiente:

$$n_1 S_1^2 = \sum(X - \bar{X})^2; \quad n_2 S_2^2 = \sum(X - \bar{X})^2.$$

La varianza por agrupación queda expresada de la siguiente manera:

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}.$$

Ahora bien, si se asume que las varianzas son iguales, se tiene las siguientes relaciones:

$$S_1^2 = S_2^2 = S^2.$$

$$S_{\bar{D}} = \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}} = S \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = S \times \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}}$$

$$S_{\bar{D}} = S \times \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}}$$

$$S_{\bar{D}} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}}$$

$$t = \frac{\bar{D} - \delta}{S_{\bar{D}}}$$

Límites de confianza: $\bar{D} \pm t_{\frac{\alpha}{2}} \times S_{\bar{D}}$.

Ejemplo:

Se desea estimar la diferencia en la productividad de los empleados de dos fábricas A y B; para lo cual, se toma dos muestras de 6 empleados de cada fábrica y se mide su productividad, así:

| | Muestra fábrica A | Muestra fábrica B |
|----------------------|-------------------|-------------------|
| | 20 | 17 |
| | 18 | 15 |
| | 19 | 15 |
| | 15 | 14 |
| | 14 | 13 |
| | 13 | 13 |
| Medidas estadísticas | Muestra fábrica A | Muestra fábrica B |
| Promedio | 16,5 | 14,5 |
| Varianza | 6,92 | 1,92 |

$$S_{\bar{D}} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}} = 1,33.$$

$$\bar{D} = \bar{X}_1 - \bar{X}_2 = 16,5 - 14,5 = 2.$$

$$t(10 \text{ grados de libertad}; 0,95) = 1,81.$$

$$\text{Límites de confianza: } \bar{D} \pm t_{\frac{\alpha}{2}} * S_{\bar{D}} = 2 \pm 1,81 * 1,33 = 2 \pm 2,41.$$

$$\text{Límite inferior: } L_i = 2 - 2,41 = -0,41.$$

$$\text{Límite superior: } L_s = 2 + 2,41 = 4,41.$$

$$\text{Entonces, } P\left(\bar{D} - t_{\frac{\alpha}{2}} * S_{\bar{D}} < \delta < \bar{D} + t_{\frac{\alpha}{2}} * S_{\bar{D}}\right) = P(-0,41 < \delta < 4,41) = 95\%.$$

R/ Con una confiabilidad del 95%, se puede concluir que, la verdadera diferencia de medias $\delta = \mu_1 - \mu_2$ está comprendida entre -0,41 y 4,41.

Caso 2: muestras dependientes

En este caso, las muestras son del mismo tamaño y el procedimiento para determinar los límites de confianza, es el siguiente:

$$\text{Límites de confianza: } \bar{D} \pm t_{\frac{\alpha}{2}} \times S_{\bar{D}}.$$

El error estándar de la diferencia de medias es:

$$S_{\bar{D}} = \frac{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n-1}}}{\sqrt{n}}.$$

Ejemplo:

La tabla que sigue, muestra los puntajes de 7 estudiantes correspondientes a dos pruebas en una escala de 0 a 100.

| No. | X_1 | X_2 |
|-----|-------|-------|
| 1 | 70 | 68 |
| 2 | 75 | 73 |
| 3 | 83 | 80 |
| 4 | 83 | 84 |
| 5 | 78 | 75 |
| 6 | 69 | 70 |
| 7 | 78 | 72 |

Solución:

Para cada estudiante se determina la diferencia $D = X_1 - X_2$ y sus cuadrados.

| No. | X_1 | X_2 | D | D^2 |
|-------|-------|-------|-----|-------|
| 1 | 70 | 68 | 2 | 4 |
| 2 | 75 | 73 | 2 | 4 |
| 3 | 83 | 80 | 3 | 9 |
| 4 | 83 | 84 | -1 | 1 |
| 5 | 78 | 75 | 3 | 9 |
| 6 | 69 | 70 | -1 | 1 |
| 7 | 78 | 72 | 6 | 36 |
| Total | | | 14 | 64 |

Si se toma 2 muestras dependientes de 7 datos, determinar los límites del 98 % de confianza para la verdadera diferencia de medias.

$$\text{Promedio: } \bar{D} = \frac{\sum D}{n} = \frac{14}{7} = 2.$$

$$\overline{D^2} = \frac{\sum D^2}{n} = \frac{64}{7} = 9,14.$$

Valor de t con $n - 1$ grados de libertad y 98% de confiabilidad: $t_{(0,99;6)} = 3,14$.

$$S_{\bar{D}} = \frac{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n-1}}}{\sqrt{n}} = \frac{\sqrt{\frac{64 - \frac{14^2}{7}}{6}}}{\sqrt{7}} = \frac{\sqrt{6}}{\sqrt{7}} = \frac{2,45}{2,64} = 0,92.$$

Límites de confianza: $\bar{D} \pm t_{\frac{\alpha}{2}} * S_{\bar{D}} = 2 \pm 3,14 * 0,92 = 2 \pm 2,88$.

Límite inferior: $L_i = 2 - 2,88 = -0,88$.

Límite superior: $L_s = 2 + 2,88 = 4,88$.

Entonces, $P\left(\bar{D} - t_{\frac{\alpha}{2}} * S_{\bar{D}} < \delta < \bar{D} + t_{\frac{\alpha}{2}} * S_{\bar{D}}\right) = P(-0,88 < \delta < 4,88) = 98\%$.

R/ La verdadera diferencia de medias $\delta = \mu_1 - \mu_2$, con un 98% de confiabilidad, se encuentra entre $-0,88$ y $4,88$.

Cuando el límite inferior es negativo y el límite superior, no existe diferencia entre los promedios de la población.

7.3.2.3 Prueba de hipótesis para la media de población con t de Student

La prueba t de Student fue desarrollada en 1908 por el químico inglés William Sealey Gosset 1876-1937, mientras trabajaba en técnicas de control de calidad para las destilerías de Guinness en Dublín. Debido a que, en la destilería, su puesto de trabajo no era inicialmente de estadístico y su dedicación debía estar exclusivamente encaminada a mejorar los costos de producción, publicó sus hallazgos anónimamente firmando sus artículos con el seudónimo de Student.

Caso 1: contraste unilateral derecho

$$H_0: \mu = \mu_0 .$$

$$H_1: \mu > \mu_0 .$$

Si el nivel de significancia se fija en 5 %, entonces, en la tabla de la distribución t , se debe buscar el t crítico utilizando el valor de α y los grados de libertad correspondientes.

Con los datos de la muestra, se obtiene el promedio aritmético y la desviación estándar; luego se calcula el estadígrafo de prueba t , utilizando la siguiente fórmula:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}} \text{ con desviación sesgada.}$$

$$t = \frac{\bar{X} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \text{ con desviación insesgada.}$$

Decisión:

Si $t_{calculado} > t_{crítico}$, se rechaza la hipótesis H_0

Si $t_{calculado} \leq t_{crítico}$, se acepta la hipótesis H_0

Ejemplo:

Se midió el nivel colesterol total a 12 pacientes varones adultos escogidos al azar de una población normal. Los resultados obtenidos arrojan una media muestral de 407 mg/dl y una desviación estándar de 6 mg/dl. Analizar la hipótesis de que el resultado encontrado es significativamente superior al valor límite de referencia de 400 mg/dl. Utilice un nivel de significancia de 1 %.

Solución:

$$\bar{X} = 407 \text{ (media de la muestra) .}$$

$$\mu_0 = 400 \text{ (media poblacional hipotética) .}$$

$$s = 6 \text{ (desviación estándar sesgada) .}$$

$$n = 12 \text{ (tamaño de la muestra).}$$

$$\nu = n - 1 = 11 \text{ (grados de libertad).}$$

$$\alpha = 1\% .$$

Planteamiento de hipótesis:

$$H_0: \mu = \mu_0 .$$

$$H_1: \mu > \mu_0 .$$

Reglas de decisión:

Si $t_{calculado} \geq t_{crítico}$, se rechaza H_0 .

Si $t_{calculado} \leq t_{crítico}$, se acepta H_0 .

$$t_{crítico} (\alpha = 0,01, \nu = 11) = 2,72$$

$$t_{calculado} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{407 - 400}{\frac{6}{\sqrt{11}}} = 3,87 .$$

Dado que $t_{calculado} = 3,87 > t_{crítico} = 2,72$, entonces se rechaza H_0 .

R/ Debido a que se rechaza la hipótesis H_0 , se concluye que el resultado de la muestra de 407 mg/dl es significativamente superior a 400 mg/dl.

$$p - \text{value} = P(t > t_{calculado}) = P(t > 3,87) = 0,0013 .$$

Como $p - \text{value} < \alpha$, entonces, se rechaza H_0 .

7.3.2.4 Prueba de hipótesis con t de Student para dos muestras independientes

Ejemplo:

Una tienda de artículos de deportes se encuentra en un centro comercial de mediano tamaño. Para planificar el volumen del personal, el director pide que le ayuden a averiguar si existen pruebas contundentes de que las ventas son menores los sábados que los lunes, con un nivel de significancia del 5%. Para el efecto, se recogen muestras aleatorias de 15 sábados y 15 lunes de una población normal, obteniendo los siguientes resultados:

Lunes:

$$\bar{x}_1 = 908,2; s_1 = 469,8; n_1 = 15.$$

Sábados:

$$\bar{x}_2 = 1.078; s_2 = 633; n_2 = 15.$$

Solución:

Planteamiento de hipótesis:

$H_0: \delta = \mu_1 - \mu_2 = 0$ (No hay diferencia en las ventas).

$H_1: \delta = \mu_1 - \mu_2 < 0$ (Las ventas son menores los días lunes).

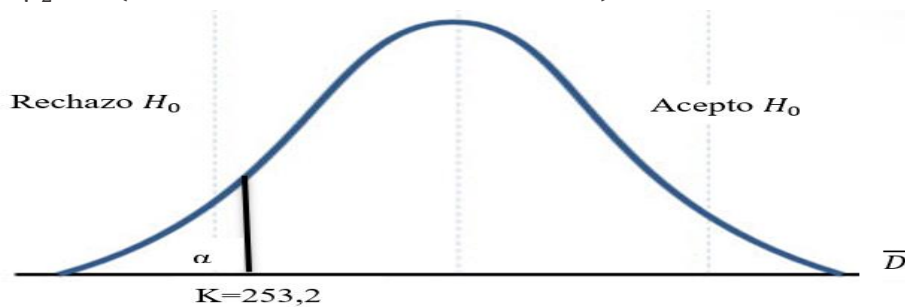


Gráfico 19. Prueba de hipótesis con t de Student para dos muestras independientes

$$S_{\bar{D}} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}}.$$

$$S_{\bar{D}} = \sqrt{\frac{15 \times (469,8)^2 + 15 \times (633)^2}{28}} \times \sqrt{\frac{15 + 15}{225}} = 576,96 \times 0,25 = 148,97.$$

$$t_{\text{crítico}}(0,05; 28) = -1,70 .$$

$$\text{Punto crítico } k = \delta + t \times S_{\bar{D}} .$$

Punto crítico $k = 0 - 1,70(148,97) = -253,2$ (ver Gráfico 19).

Reglas de decisión:

Si $\bar{D} > k$ se acepta H_0 .

Si $\bar{D} < k$ se rechaza H_0 .

$$\bar{D} = \bar{X}_1 - \bar{X}_2 = 908,2 - 1078 = -169,8 .$$

Dado que $\bar{D} = -169,8 > k = -253,2$ entonces se acepta H_0 .

R/ No existen pruebas contundentes que indiquen que las ventas de los sábados sean menores a las ventas de los días lunes.

$$p - \text{value} = P(\bar{D} < -169,8) = P(t < -1.14) = 0,1320 .$$

Como $p - \text{value} > \alpha$, se acepta H_0 .

7.3.2.5 Prueba de hipótesis con t de Student para dos muestras dependientes

Ejemplo:

Se toma una muestra de 7 empleados de una empresa y se les realiza una prueba antes de iniciar un curso de capacitación y otra después del curso. Comprobar si el curso fue efectivo con un nivel de significancia del 5%.

| No. | X_1 | X_2 | D | D^2 |
|-----|-------|-------|-----|-------|
| 1 | 70 | 68 | 2 | 4 |
| 2 | 75 | 73 | 2 | 4 |
| 3 | 83 | 80 | 3 | 9 |
| 4 | 83 | 84 | -1 | 1 |
| 5 | 78 | 75 | 3 | 9 |
| 6 | 69 | 70 | -1 | 1 |
| 7 | 78 | 72 | 6 | 36 |

| | | |
|-------|----|----|
| Total | 14 | 64 |
|-------|----|----|

Solución:

$H_0: \delta = \mu_2 - \mu_1 = 0$ (El curso **NO** es efetivo).

$H_1: \delta = \mu_2 - \mu_1 > 0$ (El curso **ES** efetivo).

Promedio: $\bar{D} = \frac{\sum D}{n} = \frac{14}{7} = 2.$

$\overline{D^2} = \frac{\sum D^2}{n} = \frac{64}{7} = 9,142.$

Valor de t con $n - 1$ grados de libertad y 98% de confiabilidad es: $t_{(0,99;6)} = 3,14.$

$$S_{\bar{D}} = \frac{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n-1}}}{\sqrt{n}} = \frac{\sqrt{\frac{64 - \frac{14^2}{7}}{6}}}{\sqrt{7}} = \frac{\sqrt{6}}{\sqrt{7}} = \frac{2,45}{2,64} = 0,92.$$

Punto crítico $K = \delta + t * S_{\bar{D}} = 0 + 3,14 * 0,92 = 2,89.$

Reglas de decisión:

Si $\bar{D} > K$ se rechaza $H_0.$

Si $\bar{D} < K$ se acepta $H_0.$

Decisión:

Dado que $\bar{D} = 2 < K = 2,89$ entonces se acepta $H_0.$

R/ El curso no es efetivo para aumentar los niveles de productividad.

p - value = $P(\bar{D} > 2) = P(t > 2,17) = 0,0365.$

Como $p - \text{value} > \alpha$, se acepta H_0 .

7.3.3 Distribución Ji-Cuadrado (χ^2)

La distribución Ji-Cuadrado se aplica para estimación y prueba de hipótesis de una varianza; además, se utiliza para analizar pruebas no paramétricas, tales como pruebas de homogeneidad, de independencia y de bondad de ajuste.

7.3.3.1 Estimación por intervalos para la varianza poblacional con Ji-Cuadrado (χ^2)

Los siguientes datos son la cantidad de vuelos (X) que una aerolínea despacha diariamente:

X : 23 25 10 26 14 23 14 21 16 11.

Estimar la varianza y la desviación estándar de la población con un 90% de confiabilidad.

Solución:

$$\bar{x} = 18,3.$$

$$s = 5,66.$$

$$s^2 = 32,0356. \text{ Varianza sesgada.}$$

$$\hat{s} = 5,9638.$$

$$\hat{s}^2 = 35,57. \text{ Varianza insesgada.}$$

$$n = 10.$$

$$1 - \alpha = 90\%.$$

$$v = n - 1 = 10 - 1 = 9.$$

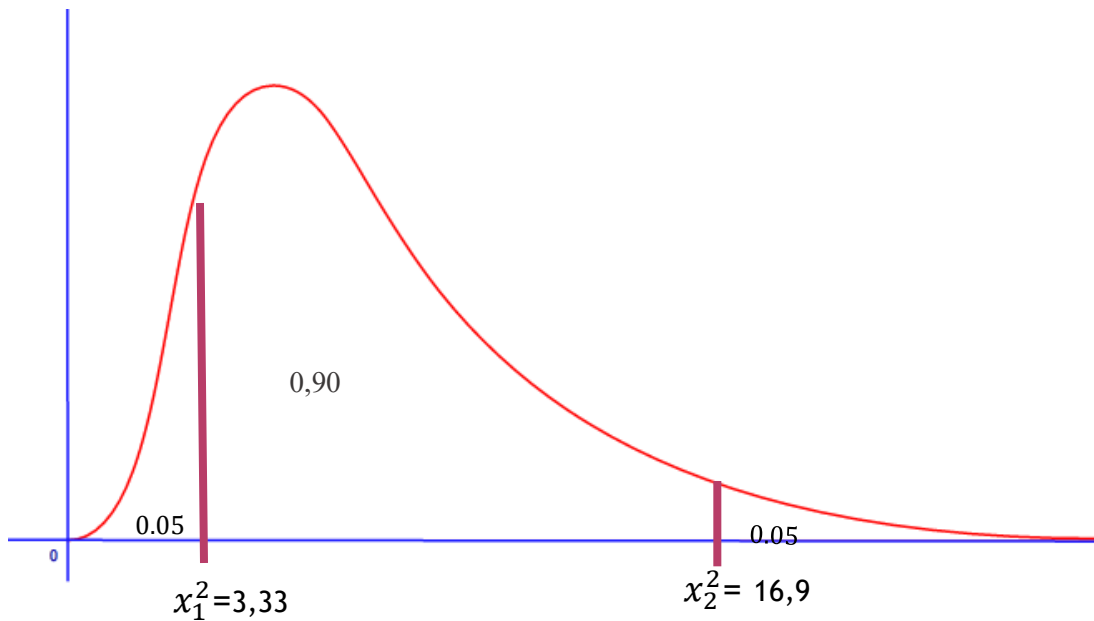


Gráfico 20. Estimación por intervalos para la varianza poblacional con Ji-Cuadrado

$$L_i = \frac{ns^2}{x_2^2} = \frac{10 \times 32,0356}{16,9} = 18,95. \quad \text{Con} \quad \text{varianza} \quad \text{sesgada.}$$

$$L_s = \frac{ns^2}{x_1^2} = \frac{10 \times 32,0356}{3,33} = 96,2. \quad \text{Con varianza sesgada.}$$

$$L_i = \frac{(n-1)\hat{s}^2}{x_2^2} = \frac{9 \times 35,57}{16,9} = 18,94. \quad \text{Con} \quad \text{varianza} \quad \text{inesgada.}$$

$$L_s = \frac{(n-1)\hat{s}^2}{x_1^2} = \frac{9 \times 35,57}{3,33} = 96,1. \quad \text{Con varianza in sesgada.}$$

$$P(18,95 < \sigma^2 < 96,2) = 90\% \rightarrow P(\sqrt{18,95} < \sqrt{\sigma^2} < \sqrt{96,2}) = 90\% .$$

Entonces, $P(4,3531 < \sigma < 9,81) = 90\%$ (ver Gráfico 20).

R/ Con un 90% de confiabilidad, se puede afirmar que la variación que tiene el promedio de los vuelos de la aerolínea está entre 4,35 y 9,81.

7.3.3.2 Prueba de hipótesis para la varianza (σ^2)

Ejemplo:

En una empresa dedicada a la producción de muebles ejecutivos, se decide tomar una muestra de 10 empleados, con el fin de saber cuáles son los niveles de productividad de cada uno. La productividad dentro de la organización se mide teniendo en cuenta la cantidad de productos fabricados por mes, y se la clasifica en una escala de 1 a 10, donde 10 es el grado de productividad más alto y 1 el más bajo. Los resultados encontrados para cada uno de los empleados escogidos son los siguientes:

| | | | | | | | | | | |
|---------------|----|---|---|---|---|---|----|---|---|----|
| Empleado | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Productividad | 10 | 7 | 9 | 8 | 6 | 7 | 10 | 9 | 8 | 7 |

Se desea comprobar si la varianza entre los grados de productividad de los empleados es diferente de 2, con un nivel de significancia del 5%.

Solución:

$$H_0: \sigma^2 = 2 .$$

$$H_1: \sigma^2 \neq 2 .$$

$$\alpha = 5\% .$$

$$n = 10 .$$

$$S^2 = 1,69 \text{ varianza sesgada.}$$

$$\hat{S}^2 = 1,87777 \text{ varianza insesgada.}$$

Solución:

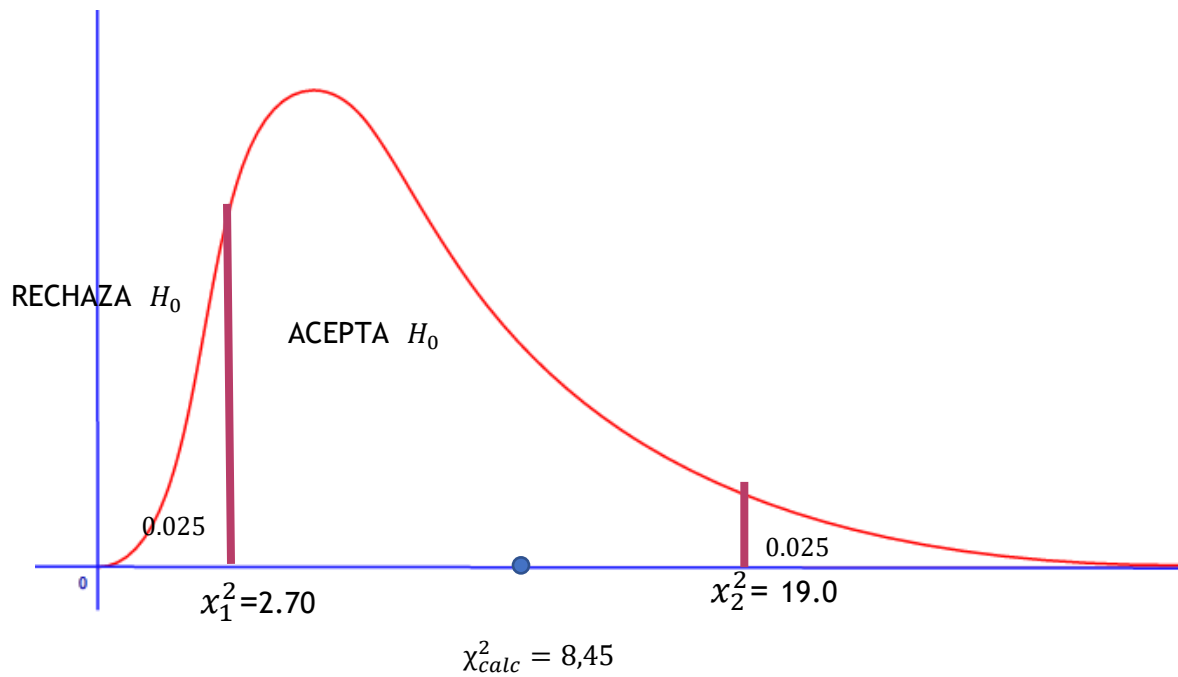


Gráfico 21. Prueba de hipótesis para la varianza

Ji – Cuadrado crítico 1: $\chi^2_1(0,025; 9) = 2,70$.

Ji – Cuadrado crítico 2: $\chi^2_2(0,0975; 9) = 19,0$.

Ji – Cuadrado calculado = $\frac{ns^2}{\sigma^2} = 8,45$ con varianza sesgada.

Ji – Cuadrado calculado = $\frac{(n-1)s^2}{\sigma^2} = 8,45$ con varianza insesgada.

Reglas de decisión:

Si $\chi_{calc}^2 < \chi_1^2$ se rechaza la hipótesis H_0 .

Si $\chi_{calc}^2 > \chi_2^2$ se rechaza la hipótesis H_0 .

Si $\chi_1^2 < \chi_{calc}^2 < \chi_2^2$ se acepta la hipótesis H_0 .

Dado que, $2,70 < 8,45 < 19$, entonces se acepta H_0 (ver Gráfico 21).

R/ Por lo tanto, se puede concluir que la varianza de productividad de todos los empleados, es 2.

7.3.3.3 Prueba de bondad de ajuste

La prueba de bondad de ajuste se aplica cuando se trata de comprobar si una distribución de frecuencias observadas se ajusta a un modelo de probabilidad.

Ejemplo:

La siguiente tabla muestra las frecuencias observadas y esperadas al lanzar un dado durante 120 veces. Comprobar la hipótesis de que el dado está perfectamente equilibrado al nivel de significancia del 5 %.

Solución:

H_0 : el dado está perfectamente equilibrado (los datos se ajustan a una distribución uniforme).

H_1 : el dado está cargado (los datos no se ajustan a una distribución uniforme).

| | | | | | | |
|----------------------|----|----|----|----|----|----|
| Cara del dado | 1 | 2 | 3 | 4 | 5 | 6 |
| Frecuencia observada | 25 | 17 | 15 | 23 | 24 | 16 |
| Frecuencia esperada | 20 | 20 | 20 | 20 | 20 | 20 |

El valor de Chi-Cuadrado calculado es:

$$\chi^2_{\text{calculado}} = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = \frac{(25 - 20)^2}{20} + \frac{(17 - 20)^2}{20} + \frac{(15 - 20)^2}{20} + \frac{(23 - 20)^2}{20} + \frac{(24 - 20)^2}{20} + \frac{(16 - 20)^2}{20} = 5,0.$$

Este valor se compara con el Ji-Cuadrado crítico que se busca en la tabla, según el nivel de significancia; en este caso 5% y los grados de libertad $v = (n - 1)(k - 1)$, es

decir, número de columnas menos uno, multiplicado por el número de filas menos uno, de la tabla anterior:

$$\nu = (6 - 1)(2 - 1) = 5.$$

Valor Ji – Cuadrado crítico: $\chi^2_{crítico}(\alpha = 0,05; \nu = 5) = 11,1$.

Reglas de decisión:

Si $\chi^2_{calculado} \leq \chi^2_{crítico}$ entonces se acepta H_0 .

Si $\chi^2_{calculado} > \chi^2_{crítico}$ entonces se rechaza H_0 .

Decisión:

Dado que $\chi^2_{calculado} = 5,0 < \chi^2_{crítico} = 11,1$ entonces, se acepta H_0 .

R/ El dado está perfectamente equilibrado, es decir, los datos se ajustan a una distribución uniforme.

$$p - \text{value} = P(\chi^2 > 5,0) = 0,4159.$$

Como $p - \text{value} > \alpha$, se acepta H_0 .

7.3.2.4 Prueba de independencia

La prueba Ji-Cuadrado se utiliza para determinar si dos variables categóricas (cualitativas) están relacionadas.

En general, las hipótesis son:

H_0 : las variables son *independientes*.

H_1 : las variables son *dependientes*.

Cuando se presenta una tabla 2×2 , es necesario hacer la corrección de *Yates* que consiste en restar 0,5 a la diferencia entre frecuencias observadas y frecuencias esperadas.

Ejemplo 1:

Dos grupos *A* y *B* formados por 100 individuos cada uno, padecen una enfermedad. Se administra un suero al grupo *A*, pero no al grupo *B* (que se llama grupo control o testigo); en todo lo demás, los dos grupos son tratados idénticamente. Los resultados del estudio se presentan en la siguiente tabla:

| Grupo | Se recuperan | No se recuperan | Total |
|-------------------------|--------------|-----------------|-------|
| A (utiliza suero) | 75 | 25 | 100 |
| B (no utiliza el suero) | 65 | 35 | 100 |
| Total | 140 | 60 | 200 |

Analizar si el suero utilizado en el tratamiento fue efectivo, es decir, si tuvo relación con el mejoramiento del paciente, con un nivel de significancia del 5%.

Solución:

Hipótesis:

H_0 : la recuperación del paciente es independiente de la aplicación del suero (el suero NO fue efectivo).

H_1 : la recuperación del paciente depende de la aplicación del suero (el suero fue efectivo).

Para aplicar la prueba Chi-Cuadrado, es necesario construir una tabla de frecuencias esperadas, bajo la hipótesis de que la recuperación es independiente del suero, así:

Frecuencias esperadas:

| Frecuencias observadas | Se recuperan | No se recuperan | Total |
|-------------------------|--------------|-----------------|-------|
| A (utiliza el suero) | 70 | 30 | 100 |
| B (no utiliza el suero) | 70 | 30 | 100 |
| Total | 140 | 60 | 200 |

Se trata de una tabla 2×2 , con un grado de libertad, por lo cual, basta encontrar solo una frecuencia esperada con la siguiente fórmula:

$$\frac{\text{total columna} \times \text{total fila}}{\text{gran total}}.$$

$$\frac{140 \times 100}{200} = 70.$$

Conociendo la primera frecuencia, se determina las demás, por lo cual, esta tabla sólo tiene un grado de libertad. El valor de Chi-Cuadrado calculado, es:

$$\chi^2_{\text{calculado}} = \sum \frac{(f_o - f_e)^2}{f_e}.$$

$$\chi^2_{\text{calculado}} = \frac{(75 - 70)^2}{70} + \frac{(65 - 70)^2}{70} + \frac{(25 - 70)^2}{70} + \frac{(35 - 70)^2}{70} = 2,38.$$

Con la corrección de *Yates* para tablas 2×2 , se tiene:

$$\chi^2_{\text{calculado}} = \frac{([75 - 70] - 0,5)^2}{70} + \frac{([65 - 70] - 0,5)^2}{70} + \frac{([25 - 70] - 0,5)^2}{70} + \frac{([35 - 70] - 0,5)^2}{70} = 1,93.$$

Entonces, $\chi^2_{\text{calculado}} = 1,93$.

El valor de Ji-Cuadrado crítico, con un 5% de significancia y 1 grado de libertad, es:

$$\chi^2_{\text{crítico}}(\alpha = 0,05; \nu = 1) = 3,84.$$

En ambos casos, es decir, sin la corrección de *Yates* y con la corrección de *Yates*, se obtiene que $\chi^2_{\text{calculado}} < \chi^2_{\text{crítico}}$, por lo tanto, se *acepta* la hipótesis nula, es decir, el suero no fue efectivo para la recuperación del paciente.

R/ El suero no fue efectivo para la recuperación del paciente.

$$p - \text{value} = P(\chi^2 > 1,93) = 0,1648.$$

Como $p - \text{value} > \alpha$, se *acepta* H_0 .

Ejemplo 2:

Se desea averiguar, si existe alguna asociación significativa entre la formación académica y el rendimiento laboral para un grupo de 200 empleados. El nivel de formación académica se clasifica en tres clases: primaria, secundaria y universitaria y el rendimiento académico, se clasifica en tres categorías: excelente, bueno y regular. Los resultados de la investigación se presentan en la siguiente tabla, con un nivel de significancia del 5%.

Tabla de frecuencias observadas:

| | Primaria | Secundaria | Universitaria | Total |
|-----------|----------|------------|---------------|-------|
| Excelente | 10 | 40 | 10 | 60 |
| Bueno | 30 | 30 | 20 | 80 |
| Regular | 10 | 30 | 20 | 60 |
| Total | 50 | 100 | 50 | 200 |

Solución:

Planteamiento de hipótesis:

H_0 : el rendimiento laboral es *independiente* de la preparación académica.

H_1 : el rendimiento laboral es *dependiente* de la preparación académica.

Tabla de frecuencias esperadas:

| | Primaria | Secundaria | Universitaria | Total |
|-----------|----------|------------|---------------|-------|
| Excelente | 15 | 30 | 15 | 60 |
| Bueno | 20 | 40 | 20 | 80 |
| Regular | 15 | 30 | 15 | 60 |
| Total | 50 | 100 | 50 | 200 |

| $\frac{(f_0 - f_e)^2}{f_e}$ | | |
|--|-------|-------|
| 1.667 | 3.333 | 1.667 |
| 5.000 | 2.500 | 0.000 |
| 1.667 | 0.000 | 1.667 |
| 0.000 | 0.000 | 0.000 |
| 8.333 | 5.833 | 3.333 |
| $\chi^2_{\text{calculado}} = \sum \frac{(f_0 - f_e)^2}{f_e} = 17,5.$ | | |

Los grados de libertad se encuentran multiplicando columnas menos uno, por filas menos 1, así: $\nu = (3 - 1)(3 - 1) = 4$.

$$\chi^2_{\text{crítico}}(\alpha = 0,01; \nu = 4) = 13,3.$$

Reglas de decisión:

Si $\chi^2_{\text{calculado}} \leq \chi^2_{\text{crítico}}$ entonces se acepta H_0 .

Si $\chi^2_{\text{calculado}} > \chi^2_{\text{crítico}}$ entonces se rechaza H_0 .

Dado que $\chi^2_{\text{calculado}} = 17,5 > \chi^2_{\text{crítico}} = 13,3$, se rechaza la hipótesis nula.

$$p - \text{value} = P(\chi^2 > 17,5) = 0,0015.$$

Como $p - \text{value} < \alpha$, se rechaza H_0 .

R/ El rendimiento laboral si *depende* de la preparación académica.

7.3.4 Distribución F (Fisher)

Una de las aplicaciones de la Distribución F es la prueba de hipótesis para comparar dos varianzas.

Ejemplo:

Se informa que la varianza de las notas de los estudiantes de grado décimo es igual a la varianza de los estudiantes de grado undécimo; no obstante, un profesor dice que las varianzas de las notas de los estudiantes son diferentes. Se toman dos muestras

aleatorias de 13 y 11 estudiantes. Comprobar si el profesor tiene la razón, con un nivel de significancia del 5 %.

Solución:

$$\alpha = 0,05; n_1 = 11; n_2 = 13.$$

Datos grado 10:

Calificaciones X_1 : 4,1; 3,6; 4,0; 4,0; 4,2; 3,9; 3,5; 4,4; 4,4; 4,3; 4,2 .

$$\bar{X}_1 = 4,06; S_1^2 = 0,082.$$

Datos grado 11:

Calificaciones X_2 : 4,0; 3,7; 3,9; 4,2; 4,0; 4,5; 3,5; 3,9; 3,8; 4,3; 4,1; 4,1; 3,9 .

$$\bar{X}_2 = 3,99; S_2^2 = 0,062 .$$

Planteamiento de hipótesis:

$H_0: \sigma_1^2 = \sigma_2^2$ (implica que $\frac{\sigma_1^2}{\sigma_2^2} = 1$). No tiene razón el profesor.

$H_1: \sigma_1^2 \neq \sigma_2^2$ (implica que $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$). Tiene razón el profesor

$$v_1 = 11 - 1 = 10.$$

$$v_2 = 13 - 1 = 12.$$

$$F_{crítico 1} (0,025; v_1 = 10; v_2 = 12) = 0,2762$$

$$F_{crítico 2} (0,975; v_1 = 10; v_2 = 12) = 3,3736$$

F calculado:

$$F_{calculado} = \frac{s_1^2}{s_2^2} = \frac{0,082}{0,062} = 1,32 .$$

Reglas de decisión:

Si $F_{calculado} < F_{crítico\ 1}$, entonces se rechaza H_0

Si $F_{calculado} > F_{crítico\ 2}$, entonces se rechaza H_0

Si $F_{crítico\ 1} < F_{calculado} < F_{crítico\ 2}$, entonces se acepta H_0

En el ejemplo, se cumple que:

$F_{crítico\ 1} = 0,2762 < F_{calculado} = 1,32 < F_{crítico\ 2} = 3,3736$, entonces se acepta H_0 .

Decisión con $p - value$:

$p - value = 0,6376$.

Dado que $p - value > \alpha$, entonces se acepta H_0 .

R/ Las varianzas entre las notas de los estudiantes de grado 10 y de grado 11 son iguales; por lo tanto, el profesor no tiene la razón.

Con Statgraphics

En el menú, seleccionar comparar dos muestras, pruebas de hipótesis.

Pruebas de Hipótesis (Comparación)

Comparar

- Medias Normales
- Sigmas Normales
- Proporciones Binomiales
- Tasas de Poisson
- Coeficientes de Correlación

Hipótesis Nula para Razón de Varianzas:
1.0

Media Muestra 1: 0,0 Media Muestra 2: 0,0

Sigma Muestra 1: 0,9055 Sigma Muestra 2: 0,7874

Proporción Muestra 1: 0,5 Proporción Muestra 2: 0,5

Tasa 1: 1,0 Tasa Muestra 2: 1,0

Correlación Muestra 1: 0,5 Correlación Muestra 2: 0,5

Tamaño Muestra 1: 11 Tamaño Muestra 2: 13

Aceptar
Cancelar
Ayuda

Opciones Prueba de Hipótesis

Hipótesis Alternativa

- No Igual
- Menor que
- Mayor que

Alpha: 5,0 %

Asumir Sigmas Iguales
 Usar Prueba-Z

Aceptar
Cancelar
Ayuda

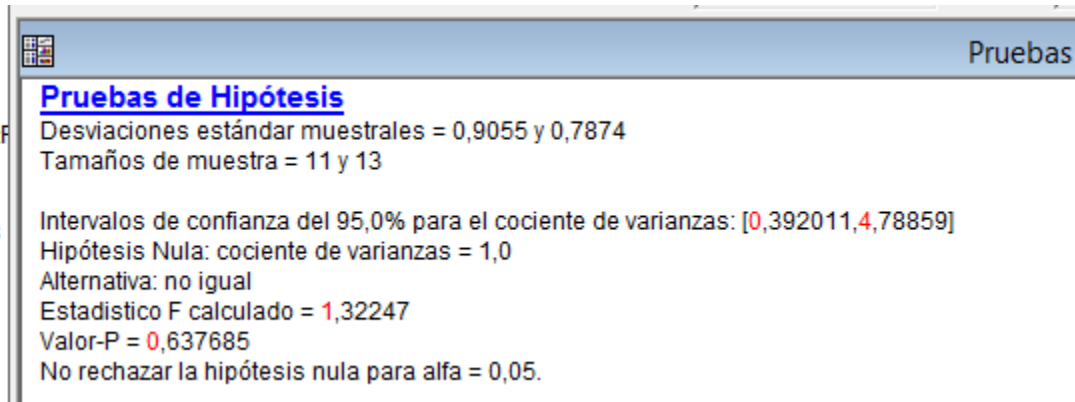


Imagen 16. Prueba de hipótesis para comparar dos varianzas con Statgraphics

Como se puede observar (ver Imagen 16), mediante Statgraphics, se confirma que las varianzas entre las notas de los estudiantes de grado 10 y de grado 11 son iguales; por lo tanto, el profesor no tiene la razón.



Capítulo 8

Problemas Propuestos

CAPÍTULO 8

PROBLEMAS PROPUESTOS

8.1 CÁLCULO DE PROBABILIDADES

Determinar el número de maneras en que se puede escoger un comité compuesto por 3 hombres y 2 mujeres, de un grupo de 7 hombres y 5 mujeres.

Todos los años se selecciona una delegación de 4 estudiantes de un colegio para asistir a la Asamblea Anual de la Asociación de Estudiantes.

¿De cuántas maneras puede escogerse la delegación, si hay 12 estudiantes elegibles?

¿De cuántas maneras, si dos de los estudiantes elegibles no asisten al mismo tiempo?

En un examen, un estudiante tiene que contestar 8 de 10 preguntas.

¿Cuántas maneras tiene para escoger?

¿Cuántas maneras, si las 3 primeras preguntas son obligatorias?

¿Cuántas si tiene que contestar 4 de las 5 primeras preguntas?

El jefe de personal de una compañía desea contratar dos agentes de ventas de un total de 4 solicitantes. Suponga que los solicitantes varían en cuanto a sus capacidades y habilidades; denote por 1, 2, 3, 4 a los solicitantes, donde 1 es el mejor, 2 el que sigue y así sucesivamente.

Defina el experimento aleatorio.

Cuántos puntos tiene el experimento.

Cuál es el espacio muestral.

Escriba los elementos de los sucesos: A= seleccionar los dos mejores; B= seleccionar al menos uno de los mejores.

Cuatro socios elegidos al azar deben expresar su opinión favorable o contraria a un proyecto determinado.

¿Cuántos puntos tiene el experimento?

¿Cuáles son los resultados?

Represente los resultados en un diagrama de árbol.

Escriba los resultados de los siguientes sucesos: A= todos están a favor; B= todos están en contra; C= por lo menos uno está a favor.

Un experimento, consiste en seleccionar tres piezas de un proceso manufacturero y observar si son defectuosos (D) o no defectuosos (D^c).

Determinar el número de puntos del espacio muestral.

Escriba todos los elementos del espacio.

Escriba los elementos de los sucesos: A= el número de piezas defectuosas es cero; B= hay exactamente dos defectuosas.

Un administrador desea implantar un nuevo sistema para la selección de personal en su empresa, el proyecto es presentado para su aprobación, a la junta directiva integrada por 5 miembros.

Determinar el número de puntos del espacio muestral.

¿Cuál es el espacio?

¿Cuál es la probabilidad de que sea aprobado el proyecto?

Encuentre la probabilidad para cada uno de los sucesos de los problemas anteriores.

Sean los eventos A y B con $P(A) = \frac{1}{4}$; $P(A \cup B) = \frac{1}{3}$. Determinar $P(B)$ si:

Si A y B son mutuamente excluyentes.

Si A y B son independientes.

Un ingeniero utiliza dos máquinas A y B, en la construcción de una obra. Las probabilidades de que las máquinas operen correctamente son: $P(A) = \frac{1}{2}$; $P(B) = \frac{2}{3}$ y la probabilidad de que ambas funcionen correctamente es $\frac{1}{4}$. Calcular la probabilidad que las máquinas funcionen correctamente:

Por lo menos una máquina

Ninguna

Sólo una

Sea el espacio muestral $S = \{a, b, c, d, e, f\}$ con las siguientes probabilidades: $P(a) = \frac{1}{16}$, $P(b) = \frac{1}{16}$, $P(c) = \frac{1}{8}$, $P(d) = \frac{3}{16}$, $P(e) = \frac{1}{4}$ y $P(f) = \frac{5}{16}$. Sean los sucesos $A = \{a, c, e\}$, $B = \{c, d, e, f\}$. Calcular:

$$P(A \cap B)$$

$$P(A/B)$$

$$P(A \cup B)$$

$$P(A - B)$$

Una sociedad está conformada por 3 economistas, 4 abogados y 2 ingenieros. Se desea elegir la junta directiva integrada por 3 miembros de la sociedad. Encontrar la probabilidad de que:

- Todos sus miembros sean de la misma profesión.
- De que todas las profesiones estén representadas en la junta.

Si no se permiten repeticiones:

¿Cuántos números de 3 dígitos se pueden formar con los seis dígitos 2, 3, 5, 6, 7 y 9?

¿Cuántos de éstos son menores que 400?

¿Cuántos son pares?

¿Cuántos son impares?

¿Cuántos son múltiplos de 5?

¿De cuántas maneras un profesor puede escoger 1 o más estudiantes de 6 elegibles?

Simplificar las siguientes Expresiones:

$$\frac{(n+1)!}{\frac{n!}{\frac{n!}{(n-2)!}}}$$

¿Cuántas placas para automóvil se pueden fabricar, si cada placa consta de 2 letras diferentes, seguidas de 3 dígitos diferentes?

Resolver el problema anterior, si el primer dígito debe ser diferente de *cero*.

Durante una semana determinada, la probabilidad de que unas acciones ordinarias aumenten de precio (*A*) es 0.30, la probabilidad de que permanezcan constantes (*C*) es 0.20 y la probabilidad de que disminuyan de precio (*D*) es 0.50.

¿Los sucesos *A*, *C*, *D* son excluyentes? ¿Por qué?

¿Cuál es la probabilidad de que esas acciones aumenten de precio o permanezcan sin cambio?

¿Cuál es la probabilidad de que el precio cambie durante la semana?

Un aparato electrónico consta de dos partes *A* y *B*. Con base en una serie de pruebas, se presuponen las siguientes probabilidades: la probabilidad de que *A* falle, es de 0.20; la probabilidad de que *B* falle, es 0.15 y la probabilidad de que ambas partes fallen, es 0.15. Calcular las siguientes probabilidades:

De que *A* o *B* fallen.

De que falle *B*.

De que falle solamente *A*.

De que falle *A* si se sabe que ha fallado *B*.

Por estudios de tránsito en una vía, se sabe que el 25% de los vehículos que transitan son de servicio público pesado, el 30% de servicios público liviano y el 45% de servicio particular. También se conoce, que la probabilidad de accidente en cada clase de vehículos son 0,2; 0,3 y 0,15, respectivamente. Se desea conocer la probabilidad de accidente en dicha vía.

El 80% de los obreros que ingresan a una planta electrónica, asisten a un curso de capacitación. El 86% de ellos cumplen con la cuota de producción, además, el 35% de los obreros que no asistieron al curso cumplen con la cuota de producción.

¿Qué probabilidad existe de que un obrero cumpla con la cuota?

¿Si cumple con la cuota, cuál es la probabilidad de que haya asistido al curso?

¿Cuál es la probabilidad de que no cumpla con la cuota?

Si no cumple con la cuota, ¿cuál es la probabilidad de que no haya asistido?

La siguiente tabla, presenta los resultados de un experimento para analizar la resistencia de un material, en 200 pruebas.

| Material | Resiste (C) | No resiste (D) | Total |
|----------|---------------|------------------|-------|
| Tipo (A) | 70 | 50 | 120 |
| Tipo (B) | 40 | 40 | 80 |
| Total | 110 | 90 | 200 |

$P(A), P(B), P(C), P(D), P(A \cap C), P(A \cap D), P(A/C), P(B/D), P(A \cup C), P(A^c/D^c), P(B^c/C)$.

Sean A, B y C eventos mutuamente excluyentes, con las siguientes probabilidades:

$$P(A) = 0,2, P(B) = 0,3 \text{ y } P(C) = 0,2$$

Calcular: $P(A \cup B \cup C); P(A^c \cap (B \cup C))$

Se lanza un par de dados. Encuentre la probabilidad de obtener:

Una suma de 8.

Cuando más una suma de 5.

Si se eligen al azar 3 libros de un estante que contiene 5 novelas, 3 libros de poemas y un diccionario, determinar la probabilidad de que:

Se seleccione el diccionario.

Se elijan dos novelas y un libro de poemas.

Suponga que, en un grupo de 500 estudiantes universitarios, se encuentra que 210 fuman, 258 ingieren bebidas alcohólicas, 216 comen entre comidas, 122 fuman e ingieren bebidas alcohólicas, 83 comen entre comidas e ingieren bebidas alcohólicas, 97 fuman y comen entre comidas y 52 participan en estas 3 malas prácticas para la salud. Si se elige al azar un miembro de ese grupo, encuentre la probabilidad de que el estudiante:

Fume pero no ingiera bebidas alcohólicas;

Coma entre comidas e ingiera bebidas alcohólicas, pero no fume;

No fume ni coma entre comidas.

A partir de experiencias previas, un corredor de acciones considera que, bajo las condiciones económicas actuales, un cliente invertirá en bonos libres de impuestos con una probabilidad de 0.6, en fondos Mutualistas con una probabilidad de 0.3 y tanto en bonos libres de impuestos como en fondos Mutualistas con una probabilidad de 0.15. Encuentre la probabilidad de que el cliente invierta:

En bonos libres de impuestos o en fondos mutualistas.

Que no inviertan en bonos libres de impuestos ni en fondos mutualistas.

La probabilidad de que una estación de servicio sirva gasolina a 0, 1, 2, 3, 4, 5 o más automóviles durante un periodo de 30 minutos, son: 0.03, 0.18, 0.24, 0.28, 0.10 y 0.17, respectivamente. Encuentre la probabilidad de que:

Más de 2 automóviles reciban gasolina.

Máximo 4 reciban gasolina.

Por lo menos 4 reciban gasolina.

Si R es el evento de que un convicto haya cometido un robo armado y D es el evento de que este convicto haya vendido droga, plantee en palabras, qué probabilidades están expresadas en los siguientes casos:

$$P(R/D)$$

$$P(D^c/R)$$

$$P(R^c/D)$$

En la siguiente tabla se clasifica una muestra aleatoria de 200 adultos, de acuerdo al género y nivel de educación.

| Educación | Masculino | Femenino |
|---------------|-----------|----------|
| Primaria | 38 | 45 |
| Secundaria | 28 | 50 |
| Universitaria | 22 | 17 |

Si se elige al azar una persona de este grupo, encuentre la probabilidad de que:

La persona sea hombre, dado que tiene educación secundaria.

Que la persona no tenga grado universitario si es mujer.

En el último año, de un grupo de 100 estudiantes de educación superior, 42 estudiaron matemáticas, 68 psicología, 54 historia, 22 matemáticas e historia, 25 matemáticas y psicología, 7 estudiaron historia, pero no matemáticas ni psicología, 10 estudiaron las 3 áreas y 8 no estudiaron ninguna de las 3. Si se elige al azar un estudiante determine la probabilidad de que:

- a) La persona estudie psicología, si no estudia matemáticas.
- b) Que una persona estudie psicología, si estudia historia y matemáticas.

La probabilidad de que un médico diagnostique en forma correcta una determinada enfermedad, es de 0.7. Dado que el médico hace un diagnóstico incorrecto, la probabilidad de que un paciente presente una demanda, es 0.9. Determinar la probabilidad de que el doctor haga un diagnóstico incorrecto y el paciente presente la demanda.

La policía planea hacer respetar los límites de velocidad, utilizando radares en 4 ubicaciones diferentes dentro de los límites de la ciudad. Se operan radares en cada una de las ubicaciones L1, L2, L3, L4 en 40%, 30%, 20% y 10% del tiempo. Una persona que rebasa los límites de velocidad en su camino al trabajo tiene, respectivamente, las probabilidades de 0.2, 0.1, 0.5 y 0.2, de pasar por estos lugares. Determinar la probabilidad de que reciba una multa.

A un sospechoso se le aplica un suero de la verdad, que se sabe es confiable en el 90% cuando la persona es culpable y en el 99% cuando es inocente. Si el sospechoso se escogió de un grupo del cual sólo el 5% han cometido alguna vez un crimen y el suero indica que la persona es culpable, ¿cuál es la probabilidad de que la persona sea inocente?

Un empresario cuenta con la opción de invertir en 2 de 5 proyectos en el próximo año; el empresario ignora que solo tres de esos 5 proyectos producirán ganancias. Si elige los 2 proyectos al azar:

Cuál es el experimento.

¿Cuántos puntos tiene el espacio muestral?

¿Determinar la probabilidad de que mínimo uno de los proyectos que producen ganancia sea escogido?

Encontrar la probabilidad de que máximo dos de los proyectos que no producen ganancia, sean escogidos.

Una empresa de servicios evalúa el funcionamiento de una podadora para sus clientes. El cliente dispone de muchas posibilidades para escoger, puesto que hay podadoras fáciles de podar, de dificultad mediana y de difícil operación, las hay caras o baratas, con reparación costosa, regular o barata. Determinar la probabilidad de que:

Un cliente solicite una podadora de fácil operación, barata y de reparación regular.

Solicite una podadora de dificultad mediana en la operación.

Solicite una podadora barata.

Un fabricante del producto A, ha conservado registros sobre la calidad de su producto y tiene la tabla que sobre el número de unidades defectuosas encontrados en 200 unidades del producto, tomadas en 6 grupos, así:

| Número de defectuosos | No. de unidades del producto A |
|-----------------------|--------------------------------|
| 0 | 100 |
| 1 | 60 |
| 2 | 20 |
| 3 | 5 |
| 4 | 5 |
| 5 | 10 |
| Total | 200 |

Elaborar una distribución de probabilidades y representarla gráficamente.

Calcular el valor esperado, la varianza y la desviación estándar.

Encontrar la función acumulativa y calcular la probabilidad de que haya por lo menos 2 defectuosos.

Si A y B son sucesos independientes, demostrar que:

A y B^c son independientes.

A^c y B son independientes.

Los compradores de volúmenes grandes de mercancía, utilizan con frecuencia, esquemas de muestreo de inspección para controlar la calidad de las mercancías que arriban. Los lotes de mercancías, son rechazados o aceptados sobre la base de los resultados obtenidos al inspeccionar algunos artículos seleccionados del lote. Suponga que un inspector de una planta procesadora de alimentos, ha aceptado el 98% de los lotes que son de *buena calidad*. Además, el inspector acepta el 94% de *todos* los lotes. El 5% de los lotes son de mala calidad. Elaborar el diagrama de árbol y la tabla de probabilidad correspondiente. Además, calcular el porcentaje de lotes que son:

Rechazados si son de mala calidad

Rechazados o de mala calidad.

Rechazados y de mala calidad

Aceptados, si son de buena calidad.

(Elabore un diagrama de árbol y el cuadro correspondiente)

8.2 MODELOS DE PROBABILIDAD

Debido a las altas tasas de interés, una firma consultora, indica que el 30% de sus cuentas por cobrar están vencidas. Si un contador escoge aleatoriamente 5 de esas cuentas, determinar la probabilidad de que:

Ninguna de las cuentas esté vencida.

Exactamente dos estén vencidas.

La mayoría de las cuentas de la muestra, estén vencidas.

Se desea elegir una junta directiva formada por un presidente, un tesorero y un secretario.

¿Cuántas planchas diferentes se pueden postular, si la asamblea está compuesta por 20 miembros y todos tienen igualdad de derechos para pertenecer a la junta?

Si su candidato para presidente es la persona A, ¿cuál es la probabilidad de que él quede elegido en ese cargo?

Las calificaciones de una evaluación, se distribuyen normalmente, con media $\mu = 3,8$ y desviación estándar $\sigma = 0,3$. Determinar el porcentaje de estudiantes que obtuvieron calificaciones:

Por debajo de 4.

Por encima de 3,5.

Entre 3,5 y 4,0.

La probabilidad de que un potencial cliente haga una compra es del 20%.

¿Cuál es la probabilidad de que un vendedor que visita a 10 clientes potenciales, menos de 3 hagan una compra?

¿Cuál es el valor esperado y la varianza asociada a los 10 clientes?

Suponga que, en una plantación de café, el 40% de las matas están infectados con Broca. Para detectar la presencia del insecto se toma una muestra aleatoria de 100 matas:

¿Cuál es la variable aleatoria para medir la infección?

¿Qué distribución de probabilidad será adecuada?

Calcule la probabilidad de que más del 50% de las matas examinadas tengan Broca.

El 10% de las semillas de cierta planta *no germinan*. Las semillas se empaquetan en cajas de 10 unidades y se venden con la garantía de que por lo menos 9 de ellas germinarán. Si un cliente compra una caja, ¿cuál es la probabilidad de que ésta cumpla la garantía?

Una compañía de seguros considera que solamente el 0,1% de la población le ocurre cierto tipo de accidente cada año. ¿Cuál debe ser el costo (X) de una póliza de *diez millones de pesos* para que la compañía tenga una utilidad (valor esperado) de \$10.000 anuales por cada póliza vendida?

Nota: denotar por X al costo o valor que paga el cliente por adquirir la póliza. Observe que, si al cliente no le ocurre accidente, este valor sería una utilidad para la compañía.

Una empresa de correo distribuye la correspondencia de tarjetas de crédito, así: un 10% son de la zona norte, de ellos el 30% están vencidos; el resto son de la zona sur y de ellos el 10% están vencidos. Si un recibo está vencido:

¿Cuál es la probabilidad de que sea de la zona norte?

¿Cuál es la totalidad de recibos vencidos?

Suponga que los 4 motores de un avión comercial operan independientemente y que la probabilidad de que un motor falle durante un vuelo, es 0.01. El avión puede llegar a su destino si por lo menos un motor está en buenas condiciones. ¿Cuál es la probabilidad de que el avión no llegue a su destino?

Debido a las altas tasas de interés, una compañía informa que el 30% de sus cuentas por cobrar están vencidas. Si un contador escoge aleatoriamente 5 de esas cuentas, qué modelo de probabilidad utilizaría para encontrar las siguientes probabilidades:

Ninguna cuenta vencida.

Exactamente 2 vencidas.

La mayoría vencidas.

¿Cuál es la variable?

Al sistema de contabilidad de una empresa ingresan diariamente 15 registros contables (N), de los cuales, 5 se digitan con error. Se toma una muestra de 4 registros. Calcular la probabilidad de que:

En la muestra no haya registros con error.

De que por lo menos uno de los cuatro registros tenga error

Un comerciante recibe un pedido de 20 televisores, de los cuales 4 son defectuosos. Si toma al azar una muestra de 3 aparatos, calcular la probabilidad que:

No haya televisores defectuosos en la muestra.

Exactamente uno sea defectuoso.

El 40% de los empleados de una compañía tienen seguro de vida. Si se toma una muestra aleatoria de 10 empleados:

¿Cuál es la función de probabilidad para las personas aseguradas?

¿Qué indica la variable aleatoria?

¿Cuál es la probabilidad de que todas las personas estén aseguradas?

¿Cuál es la probabilidad de que por lo menos una persona esté asegurada?

¿Cuál es y cómo se interpreta el valor esperado y la desviación estándar?

El Hemacitómetro es un aparato que se utiliza para medir la densidad de células por cuadrícula. Según experimentos anteriores, se conoce que el promedio de células por cuadrícula es igual a 4. Calcule la probabilidad de que en las próximas mediciones se encuentre por lo menos 3 células por cuadrícula.

El salario promedio de los trabajadores de una empresa es de \$ 3.800 por hora y la desviación estándar es de \$250.

¿Qué porcentaje de empleados ganan menos de \$ 4.000 la hora?

Si se toma una muestra de 25 empleados, ¿cuál es la probabilidad de que el promedio sea superior a \$ 3.600?

Al examinar la longitud del tronco madre del *Pemphigus Populi Transversus*, se encontró que la variable se distribuye normalmente con media 4,4 mm y desviación estándar 0,12 mm. Determinar el porcentaje de troncos que tendrán las siguientes longitudes:

Superior a 4,50 mm.

Entre 4,20 y 4,65 mm.

Suponga que en una cosecha de tomate de la variedad Hamstead, solamente el 60% cumplen con la especificación de frutos de buena calidad. La mercancía se empaca en cajas de 30 unidades y se despachan al mercado nacional. En el departamento de compras, el encargado de control de calidad, revisa aleatoriamente una caja de cada envío: si más de la mitad de los frutos cumplen con las medidas de buena calidad, acepta el envío de lo contrario lo rechaza.

¿Cuál es la probabilidad de rechazar el envío?

¿Cuál es la probabilidad de aceptarlo?

Calcular el promedio aritmético, la varianza y la desviación estándar.

El peso medio de las frutas de un gran cargamento, es de 15 onzas y la desviación estándar 1,62 onzas. Si su peso está distribuido normalmente, ¿qué porcentaje de frutas tendrán un peso entre 18 y 20 onzas?

Los recaudos diarios del impuesto predial, se distribuyen normalmente con un promedio de 35 millones diarios y desviación estándar 5 millones de pesos. Encuentre la probabilidad de que un día cualquiera se recaude:

Más de 48 millones.

Menos de 40 millones.

Entre 30 y 40 millones.

La distribución salarial de los empleados de una empresa es normal con media \$800.000

y desviación típica \$60.000. Determinar el porcentaje de empleados que gana:

Menos de \$900.000

Más de \$ 620.000

La vida útil de cierta marca de baterías es normal, con media 30 meses y desviación estándar 6 meses. Calcular el porcentaje de baterías que tendrán una duración:

Menor de 24 meses.

Entre 24 y 40 meses.

Superior a 40 meses.

Los agentes de aduanas de los Estados Unidos, chequean los documentos de las mercancías que entran al país para ver si cada envío se encuentra debidamente legalizado. Los registros del departamento muestran que el 50% de los envíos tienen su documentación correcta. Si se toma una muestra aleatoria de 8 envíos, utilizando el modelo Binomial, calcule la probabilidad de que por lo menos uno tenga su documentación debidamente legalizada.

La probabilidad de que un cierto tipo de componente se comporte adecuadamente bajo condiciones de alta temperatura es del 90 %. Si el dispositivo tiene 6 componentes, calcular la probabilidad de que:

Todos los componentes se comporten adecuadamente.

Falle por lo menos uno de los componentes.

El promedio de clientes que llegan a la ventanilla de un banco es 4 por minuto. Cuál será la función de probabilidad y la variable aleatoria para determinar la probabilidad de que durante el próximo minuto:

No lleguen clientes.

Lleguen máximo 3 clientes.

Lleguen por lo menos 4 clientes.

Determinar el área bajo la curva normal en cada uno de los siguientes casos:

$$P(0 < Z < 1,2)$$

$$P(-0,68 < Z < 0)$$

$$P(-0,46 < Z < 2,21)$$

$$P(Z < -1,28)$$

$$P(Z > 2,33)$$

Determinar el valor de K, en cada uno de los siguientes casos:

$$P(Z < K) = 0,05$$

$$P(Z > K) = 0,01$$

$$P(Z < K) = 0,95$$

$$P(K1 < Z < K2) = 0,98$$

$$P(Z > K) = 0,90$$

La empresa de energía A, empezará a promover la conservación de energía, ofreciendo tasas de descuento a los consumidores que mantengan su uso de energía

por debajo de ciertos estándares establecidos. Un reporte reciente de la empresa, afirma que el 70% de los residentes en la ciudad donde está la empresa A, ha reducido su uso de energía eléctrica, lo suficiente para ser tenidos en cuenta en tarifas de descuento. Supongamos que selecciona al azar 10 usuarios, entonces, determinar la probabilidad de que:

Por lo menos 7 reciban descuentos.

Máximo 4 reciban descuentos.

Todos reciban descuentos

8.3 ESTADÍSTICA INFERENCIAL

Un estudiante gasta mensualmente un promedio de \$ 25.000 en materiales, con desviación estándar \$ 4.000. Si se elige al azar una muestra de 25 estudiantes, ¿cuál es la probabilidad de que el promedio de la muestra sea inferior a \$ 26.000?

El contenido de nicotina de cierta marca de cigarrillos, tiene distribución normal, con media desconocida y desviación estándar 1 miligramo.

Construir un intervalo del 95% de confianza para la media con base a una muestra de 36 cigarrillos, cuyo contenido promedio de nicotina es de 30 miligramos.

¿Cuál es el margen de error?

¿Cuál es el error estándar del promedio?

Los diámetros interiores de tubos de acero producidos por una fábrica, tienen distribución normal, con media 10 pulgadas y desviación estándar 0.1 pulgada. Tubos con diámetros superiores a 10,17 e inferiores a 9,83 pulgadas, se consideran de mala calidad. ¿Qué porcentaje de la producción cumple las especificaciones de buena calidad?

La probabilidad de que un estudiante de Ingeniería de Sistemas apruebe la materia, es del 50%. Si en el curso hay 50 estudiantes, cuál es la probabilidad de que:

Más del 54% gane la materia.

Menos del 48% gane la materia.

Entre el 48% y el 54% aprueben la materia.

Cierto aparato electrónico tiene una duración media de 1500 horas y una desviación estándar de 120 horas. Si se toma una muestra aleatoria de 36 aparatos, determinar la probabilidad de que el promedio:

Sea inferior a 1540 horas.

Esté entre 1480 y 1540 horas.

Un fabricante de tubos fluorescentes, afirma que la vida media de este material es de 1600 horas y la desviación estándar 420 horas. Los consumidores creen que el fabricante exagera. Para comprobarlo, se toma una muestra de 64 tubos y se encuentra que la vida media es de 1500 horas. ¿Quién tiene la razón a un nivel de significancia del 1%?

La media y la desviación típica de la carga máxima soportada por 60 cables, son 11.09 y 0.73 toneladas, respectivamente. Calcular los límites del 95% de confianza para la media de todos los cables de este tipo.

Dos muestras aleatorias tomadas de poblaciones normales, con varianzas idénticas dan los siguientes resultados:

$$n_1 = 30, n_2 = 40, \bar{X}_1 = 10, \bar{X}_2 = 25, S_1 = 34,6, S_2 = 30$$

Contrastar la hipótesis, de que la verdadera diferencia de las medias es -10 , frente a la alternativa de que es menor a -10 , a un nivel de significancia del 5%.

Un fabricante de cierto shampoo para el cabello, distribuye el tamaño "profesional" en 100 salones de belleza de Bogotá. Se ha determinado que el consumo promedio de su producto es de 2800 cojines mensuales, con desviación estándar de 280 cojines. Si se toma una muestra probabilística de 36 salones, ¿cuál es la probabilidad de que el consumo promedio en un mes sea inferior a 2900?

Se desea conocer la resistencia media a la ruptura de cierta clase de material. Una muestra de 40 cables elegidos al azar, revela una tensión media de ruptura igual a 2400 libras y una desviación típica de 150 libras. Determinar el intervalo del 95% de confianza para la verdadera resistencia media a la ruptura de los cables.

Un fabricante afirma que al menos el 20% del público prefiere su producto. Se toma una muestra aleatoria de 100 personas para verificar su afirmación, 16 de las cuales expresaron su preferencia por el producto. ¿Esta información es suficiente evidencia para refutar la afirmación del fabricante a un nivel de significancia del 5%?

Los salarios diarios en cierta empresa están distribuidos normalmente, con una media de \$ 1320. Si el 9% de las medias de los salarios diarios en una muestra de 36 obreros,

es inferior a \$ 1250, ¿cuál es la desviación estándar de los salarios diarios de esa empresa?

De cada una de dos poblaciones normales e independientes, con iguales medias y desviaciones 6,40 y 7,20, respectivamente, se extraen muestras de 64 elementos. Encontrar la probabilidad de que la diferencia entre las medias muestrales sea superior a 0,60.

El 46% de los sindicatos del país, están en contra de comerciar con China Continental. ¿Cuál es la probabilidad de que una muestra de 100 sindicatos muestre que más del 52% tienen la misma posición?

Sea X la vida útil de cierto aparato electrónico, con media desconocida y desviación típica 4000 horas. Una muestra aleatoria de 100 observaciones, dio como resultado una media de 30000 horas de duración. Construir un intervalo del 95% de confianza para la verdadera media de duración.

Las A y B elaboran dos tipos de cables que tienen resistencia media a la rotura de 4500 y 4000 libras, con desviaciones 200 y 300 libras, respectivamente. Si se comprueban 50 cables de A y 100 de B, ¿cuál es la probabilidad de que la media de resistencia a la rotura de A sea al menos 600 libras más que B?

Un fabricante de autos sostiene que sus vehículos consumen, en promedio, 5,5 galones de gasolina cada 100 kilómetros. Un vendedor de la compañía comprueba el consumo de gasolina de 35 autos y encuentra que el consumo medio de este grupo es de 5,65 galones por cada 100 kms. Si la desviación estándar del consumo es 0,35 galones, ¿es cierta la afirmación del fabricante a un nivel de significancia del 1 %?

Una investigación en una universidad, realizada para determinar si el poseer un automóvil es perjudicial al rendimiento académico de los estudiantes, se basó en dos muestras aleatorias de 100 alumnos cada una.

El promedio de aprovechamiento y la varianza de los estudiantes fue:

Sin automóvil: $\bar{X}_1 = 2,70$; $S_1^2 = 0,36$

Con automóvil: $\bar{X}_2 = 2,54$; $S_2^2 = 0,40$

¿Presentan los datos suficiente evidencia que indiquen una diferencia entre el rendimiento medio de los dos grupos a un nivel de significancia del 5 %?

Las lámparas que fabrica cierta empresa, tienen una vida media de 800 horas y una desviación estándar de 60 h. Determinar la probabilidad de que en una muestra aleatoria de 406 lámparas, la vida media:

Esté entre 790 y 810 horas.

Sea superior a 820 horas.

Una muestra de 200 artículos producidos por una máquina, debe tener como especificaciones, un diámetro medio de 3.6 cms. Si la desviación estándar es de 0,21 cms, plantear una regla de decisión del 5% de significancia para determinar, a partir de qué valores la máquina no cumple con las especificaciones (Sugerencia $H_0: \mu=3,6$ cms \Rightarrow cumple las especificaciones).

El número de horas de duración de una pila para transistores, tiene distribución normal, con media 100 horas y desviación 20 horas:

¿Qué proporción de pilas durarán entre 100 y 120 horas?

Si se toma muestras de 36 pilas, ¿qué porcentaje de muestras tendrán promedio entre 97 y 105 horas?

Se sabe, que cierta marca de computadores tiene el 65% del mercado. Si se toma dos muestras de 200 usuarios cada una, ¿cuál es la probabilidad de que la diferencia entre la primera y la segunda muestra sea superior al 10% en las preferencias de los usuarios de esa clase de computadores?

Se ha determinado, que la capacidad de un puente es de 28 toneladas, con desviación estándar 5 toneladas. Una muestra de 30 camiones que transitan por esa vía, presenta un promedio de 30,5 toneladas. Comprobar si se está sobrepasando la capacidad del puente a un nivel de significancia del 5%.

Se selecciona una muestra aleatoria de 500 compradores de un centro comercial, con el fin de determinar la distancia promedio que recorren hasta allí. La muestra revela que el promedio es 23,5 millas y la desviación estándar 10.4 millas. ¿Cuáles son los límites del 95% de confianza para la verdadera media?

En el problema anterior, ¿cuál debería ser el tamaño de muestra, si se desea un error máximo de 1,5 millas y una confiabilidad del 98%?

Una compañía transnacional instituyó recientemente, un programa de seguridad en el trabajo para reducir el tiempo perdido debido a accidentes de trabajo. En los 48 meses siguientes a la implantación del programa, el tiempo perdido a causa de accidentes de trabajo, promedió 91 horas por mes, con una desviación estándar de 14 horas. En los 50 meses anteriores al programa de seguridad, el tiempo perdido era de 108 horas y la desviación estándar 12 horas. Estime la diferencia del tiempo perdido debido a accidentes de trabajo antes y después del programa de seguridad, usando un intervalo del 90% de confianza.

Contrastar la hipótesis, de que la variabilidad de las compras anuales de los dos tipos de clientes, son iguales frente a la alternativa de que no son iguales, a un nivel de significancia del 5%.

Se ha determinado que el consumo promedio de gaseosas en la ciudad es de 5300 botellas diarias, con desviación 250 botellas. Sin embargo, el gerente de la empresa Coca Cola asegura que el consumo ha disminuido. En este sentido, se toma una muestra durante los últimos 30 días y se encuentra que el promedio es de 4500 botellas. Determinar, a un nivel de significancia del 1%, si el consumo ha disminuido efectivamente.

Se toman dos muestras aleatorias de 25 empleados cada una, la primera de la empresa A y la segunda de la empresa B. Se obtiene los siguientes resultados: promedios 500 y 800, respectivamente y desviaciones $S_1 = 25, S_2 = 10$. Comprobar la hipótesis de que las varianzas son iguales o diferentes con $\alpha = 5\%$.

La media de una población es 375 y la desviación estándar 48. Si la probabilidad de que el promedio muestral esté comprendido entre 370 y 380 es del 95%, determinar el tamaño de la muestra.

La proporción de audiencia de TV, que ve cierto programa el sábado en la noche, fue del 50% según se ha encontrado previamente. Se cree que la proporción ha bajado. Para comprobarlo, se tomó una muestra de 100 televidentes y se encontró que el 45 veían el programa. Si el nivel de significancia es del 1%, ¿se puede concluir que la proporción ha bajado efectivamente?

Se desea comprobar las hipótesis, de que la productividad media de los empleados de la fábrica A es igual a la de la fábrica B, o es diferente a un nivel de significancia de 5%. Para tal fin, se toma muestras independientes de cada fábrica de 6 empleados y se mide su productividad, así:

Muestra1 (A): 20, 18, 19, 15, 14, 13

Muestra 2 (B): 17, 15, 15, 14, 13, 13

Determinar:

El promedio y la varianza en cada muestra.

El intervalo del 98% de confianza para la verdadera diferencia de medias

La tabla que sigue, indica el número de artículos producidos por tres máquinas. El jefe de control de calidad afirma, que todas están funcionando correctamente, si el porcentaje de defectuosos esperado es igual al 15% en cada una de las máquinas.

| Calidad | Máquina 1 | Maquina 2 | Máquina 3 | Total |
|-------------|-----------|-----------|-----------|-------|
| Buenos | 50 | 47 | 56 | 153 |
| Defectuosos | 5 | 14 | 8 | 27 |
| Total | 55 | 61 | 64 | 180 |

Calcular las frecuencias esperadas bajo la hipótesis del jefe

Comprobar, con un nivel de significancia del 5%, si las máquinas están operando correctamente

La variable X representa el número de vehículos vendidos en un concesionario de la ciudad de Pasto, en cada día de la semana pasada.

| Día: | lunes | martes | miércoles | jueves | viernes |
|------|-------|--------|-----------|--------|---------|
| X: | 3 | 5 | 7 | 8 | 9 |

Calcular:

El promedio aritmético y la desviación estándar de las ventas.

Determinar un intervalo del 95 % de confianza para el verdadero promedio de ventas.

Una empresa dedicada a la construcción, proporciona guantes a sus empleados para que desarrollen adecuadamente sus funciones. En estas condiciones, se recomienda un nuevo tipo de guantes de mayor duración, pero que evidentemente serán más costosos. La empresa comprará los nuevos guantes, si su vida media es significativamente superior a 120 días, de lo contrario, seguirá usando los actuales. La empresa toma una muestra de 36 pares de guantes, encontrando un promedio de

125 días. Si la desviación estándar es de 18 días y el nivel de significancia del 1%, ¿cuál será la decisión de la empresa? ($H_0: \mu = 120 \text{ días}$)

El porcentaje de alumnos que aprueban una materia, según se ha encontrado previamente, es del 80%; se cree que el rendimiento aumentará para el presente año. Para comprobarlo, se tomó una muestra de 30 alumnos y se encuentra que el 83% gana la materia. Si el nivel de significancia es del 5%, ¿se puede concluir que el rendimiento académico ha mejorado?

Un administrador afirma, que las utilidades de su empresa ascienden a la suma de 50 millones de pesos en promedio y que la desviación estándar es 6 millones. Los empleados de la compañía creen que el monto de utilidades es inferior. Para comprobarlo, toman una muestra de 10 empresas similares y encuentran un promedio de 46 millones de pesos:

¿Quién tiene la razón a un nivel de significancia del 1%?

Encuentre los límites del 95% de confiabilidad para las utilidades de la empresa.

Las siguientes son las cifras de ventas (X) que reportaron 12 vendedores a su empresa:

X: 10, 12, 13, 15, 20, 24, 22, 25, 20, 24, 25, 23.

Probar la hipótesis de que el promedio de ventas en la población es diferente de 24, a un nivel del 5%.

Probar la hipótesis de que la varianza σ^2 es menor que 32 con $\alpha = 1\%$

Encontrar los límites del 95% para la verdadera varianza y desviación estándar.

El director de una empresa quiere averiguar, si existe alguna relación entre la participación de los empleados en los programas de la empresa y el estrato según el área residencial, con α del 5%. Los datos se presentan en la siguiente tabla:

| Participación | Estrato bajo | Estrato medio | Estrato alto | Total |
|---------------|--------------|---------------|--------------|-------|
| Nunca | 20 | 32 | 40 | 92 |
| Ocasional | 15 | 28 | 44 | 87 |
| Regular | 10 | 18 | 23 | 51 |
| Total | 45 | 78 | 107 | 230 |

La siguiente tabla, muestra el número de horas perdidas por accidentes de trabajo en 6 plantas de una fábrica, antes y después de ponerse en marcha un programa de seguridad industrial. ¿Los datos proporcionan suficiente evidencia para concluir que el programa fue efectivo para reducir los accidentes que causan pérdidas de tiempo a un α del %?

| Planta | A | B | C | D | E | F |
|----------------------|----|----|----|----|----|----|
| Antes del programa | 40 | 64 | 42 | 70 | 58 | 30 |
| Después del programa | 37 | 58 | 40 | 65 | 52 | 29 |

La siguiente tabla, muestra la relación existente entre las notas de estudiantes en Matemáticas y en Estadística. Contrastar la hipótesis de que los resultados son independientes a un nivel de significancia α del 5%.

| Estadística | Matemáticas | | | Total |
|-------------|-------------|-------|------|-------|
| | Alta | Media | Baja | |
| Alta | 56 | 71 | 12 | 139 |
| Media | 47 | 163 | 38 | 248 |
| Baja | 14 | 42 | 85 | 141 |
| Total | 117 | 276 | 135 | 528 |

La media de una población, es 375 y la desviación estándar 48. Si la probabilidad de que el promedio muestral este comprendido entre 370 y 380 es del 95%, ¿cuál debe ser el tamaño de la muestra?

Suponga que dos grupos de estudiantes realizan una evaluación idéntica. El primer grupo está formado por 18 estudiantes del colegio1, con promedio 85 y varianza 240; el segundo grupo, está constituido por 12 estudiantes del colegio2, con promedio 80 y varianza 340. Determinar los límites de 98 % de confianza para la verdadera diferencia de promedios.

Se toman dos muestras aleatorias de 25 empleados cada una; la primera de la empresa A y la segunda de la empresa B. Se obtiene las siguientes resultados

promedios: 500 y 800 respectivamente; y $S_1 = 25, S_2 = 10$. Comprobar la hipótesis de que las varianzas son iguales o son diferentes con $\alpha = 2\%$.

El promedio y la varianza de las ventas de un supermercado en los últimos 15 días fueron \$ 340.000 y \$ 400.000 respectivamente. Determinar un intervalo del 98% de confianza para la verdadera varianza y la desviación estándar.

Se desea estimar el salario promedio (μ) de todos los empleados públicos del municipio. Para tal fin, se toma una muestra de 200 empleados y se encuentra que el promedio es \$280000 y la desviación estándar \$30000.

Determinar un intervalo del 95% de confiabilidad para el verdadero promedio.
¿Cuál sería el margen de error?

A continuación, se presenta el número de transistores que no satisfacen un requisito de calidad de producción en 20 muestras de 10 transistores cada una. Pruebe la hipótesis de que los datos se aproximan a la distribución Binomial con $n = 10$ y $p = 0,30$, a un nivel de significancia del 5%.

Artículos defectuosos (X): 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10
Cantidad de muestras (f_0): 0; 1; 2; 4; 5; 5; 2; 1; 0; 0; 0

Probar la hipótesis;

H_0 : El número de fallas mecánicas de una planta se aproxima a la distribución de Poisson con media 2,5; $n=40$ y nivel de significancia del 1%.

Número de fallas (X): 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13.
Frecuencia observada (f_0): 1; 6; 8; 10; 7; 4; 3; 1; 0; 0; 0; 0; 0; 0

La tabla que sigue, muestra el número de accidentes en una muestra de 56 fábricas. Compruebe la hipótesis:

H_0 : la distribución de frecuencias se ajusta a una distribución normal. Determinar el promedio y la desviación estándar; nivel de significancia 5 %.

| Número de accidentes | f_0 |
|----------------------|-------|
| 1,45 – 1.75 | 6 |
| 1,75 – 2,05 | 12 |
| 2,05 – 2,35 | 14 |
| 2,35- 2,65 | 9 |
| 2,65- 2,95 | 8 |
| 2,95 – 3,25 | 7 |
| Total | 56 |

BIBLIOGRAFÍA

- Canavos, G. C. (1984). *Probabilidad y Estadística*. Ma. Graw Hill.
- Chou, Y. L. (1977). *Análisis Estadístico*. Editorial Interamericana.
- Hanke J. E. & Reitsh, A. G. (1994). *Estadística para negocios*. Mc Graw Hill.
- Kazmier, L. (1987). *Estadística para la Administración y la Economía*. Serie Schaum. Mc Graw Hill.
- Levin, R. & Rubin, D. (1994). *Estadística para administradores*. Prentice Hall.
- Martinez, B. C. (1987). *Estadística*. Editorial ECOE.
- Mendenhall, W. (1979). *Introducción a la probabilidad y Estadística*. Wastworth Internacional.
- Miller, J. R. (1964). *Probabilidad y Estadística para Ingenieros*. 5a. Editorial Prentice May Hispanoamericana S.A.
- Spiegel, M. (1970). *Estadística. Teoría y 875 problemas resueltos*. Serie Schaum. McGraw-Hill.

ANEXOS

ANEXO A: PRUEBA DE HIPÓTESIS PARA μ

| Planteamiento | | Datos |
|-------------------------------------|-----------------|-------|
| Hipótesis nula | $H_0: \mu =$ | 3,6 |
| Hipótesis alterna | $H_1: \mu \neq$ | 3,6 |
| Nivel de significancia | α | 0,01 |
| Tamaño de muestra | n | 36 |
| Promedio muestral | \bar{x}_1 | 3,75 |
| Desviación estándar de la población | σ | 0,5 |

| Proceso | | |
|-----------------------------|---|-------|
| $Z_{\text{critico 1}}$ | Z_1 | -2,58 |
| $Z_{\text{critico 2}}$ | Z_2 | 2,58 |
| Error estándar del promedio | $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ | 0,08 |
| $Z_{\text{calculado}}$ | $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ | 1,80 |
| Punto crítico 1 | $t_{k_1} = \mu_0 + Z_1 \frac{\sigma}{\sqrt{n}}$ | 3,39 |
| Punto crítico 2 | $t_{k_2} = \mu_0 + Z_2 \frac{\sigma}{\sqrt{n}}$ | 3,81 |

| Reglas de decisión | Conclusión |
|--|--------------|
| Si $Z_1 < Z_{\text{calculado}} < Z_2$ entonces se acepta H_0 | Acepto H_0 |
| $t_{k_1} < \bar{x}_1 < t_{k_2}$ entonces se acepta H_0 | Acepto H_0 |

ANEXO B: TABLA PARA DISTRIBUCIÓN NORMAL

$P(Z < Z_0)$ corresponde a áreas a la izquierda de Z_0 .

Ejemplo:

Determinar la probabilidad que $Z < 2,54$

$$P(Z < 2,54) = 0,9945$$

| Z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 |
| 1 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 |

| | | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 |
| 2 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 |
| 3 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 |
| 3,1 | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 |
| 3,2 | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 |
| 3,3 | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 |
| 3,4 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 |

Fuente: elaboración propia con función de Excel

ANEXO C. TABLA PARA DISTRIBUCIÓN t

La primera fila corresponde a las áreas a la izquierda de un valor t y la primera columna representa los grados de libertad.

Ejemplo:

Determinar el valor t que deja a la izquierda un área de 0,995 con $\nu = 6$ grados de libertad.

$$t(0,995; 6) = 3,71$$

Eso significa que, la probabilidad de que $P(t \leq 3,71; \nu = 6) = 0,995$

| ν | 0,995 | 0,990 | 0,975 | 0,9 | 0,8 | 0,75 | 0,7 | 0,6 | 0,55 |
|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|
| 1 | 63,66 | 31,82 | 12,71 | 3,08 | 1,376 | 1,000 | 0,727 | 0,325 | 0,158 |
| 2 | 9,92 | 6,96 | 4,30 | 1,89 | 1,061 | 0,816 | 0,617 | 0,289 | 0,142 |
| 3 | 5,84 | 4,54 | 3,18 | 1,64 | 0,978 | 0,765 | 0,584 | 0,277 | 0,137 |
| 4 | 4,60 | 3,75 | 2,78 | 1,53 | 0,941 | 0,741 | 0,569 | 0,271 | 0,134 |
| 5 | 4,03 | 3,36 | 2,57 | 1,48 | 0,920 | 0,727 | 0,559 | 0,267 | 0,132 |
| 6 | 3,71 | 3,14 | 2,45 | 1,44 | 0,906 | 0,718 | 0,553 | 0,265 | 0,131 |
| 7 | 3,50 | 3,00 | 2,36 | 1,41 | 0,896 | 0,711 | 0,549 | 0,263 | 0,130 |
| 8 | 3,36 | 2,90 | 2,31 | 1,40 | 0,889 | 0,706 | 0,546 | 0,262 | 0,130 |
| 9 | 3,25 | 2,82 | 2,26 | 1,38 | 0,883 | 0,703 | 0,543 | 0,261 | 0,129 |
| 10 | 3,17 | 2,76 | 2,23 | 1,37 | 0,879 | 0,700 | 0,542 | 0,260 | 0,129 |
| 11 | 3,11 | 2,72 | 2,20 | 1,36 | 0,876 | 0,697 | 0,540 | 0,260 | 0,129 |
| 12 | 3,05 | 2,68 | 2,18 | 1,36 | 0,873 | 0,695 | 0,539 | 0,259 | 0,128 |
| 13 | 3,01 | 2,65 | 2,16 | 1,35 | 0,870 | 0,694 | 0,538 | 0,259 | 0,128 |
| 14 | 2,98 | 2,62 | 2,14 | 1,35 | 0,868 | 0,692 | 0,537 | 0,258 | 0,128 |
| 15 | 2,95 | 2,60 | 2,13 | 1,34 | 0,866 | 0,691 | 0,536 | 0,258 | 0,128 |
| 16 | 2,92 | 2,58 | 2,12 | 1,34 | 0,865 | 0,690 | 0,535 | 0,258 | 0,128 |
| 17 | 2,90 | 2,57 | 2,11 | 1,33 | 0,863 | 0,689 | 0,534 | 0,257 | 0,128 |

| ν | 0,995 | 0,990 | 0,975 | 0,9 | 0,8 | 0,75 | 0,7 | 0,6 | 0,55 |
|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|
| 18 | 2,88 | 2,55 | 2,10 | 1,33 | 0,862 | 0,688 | 0,534 | 0,257 | 0,127 |
| 19 | 2,86 | 2,54 | 2,09 | 1,33 | 0,861 | 0,688 | 0,533 | 0,257 | 0,127 |
| 20 | 2,85 | 2,53 | 2,09 | 1,33 | 0,860 | 0,687 | 0,533 | 0,257 | 0,127 |
| 21 | 2,83 | 2,52 | 2,08 | 1,32 | 0,859 | 0,686 | 0,532 | 0,257 | 0,127 |
| 22 | 2,82 | 2,51 | 2,07 | 1,32 | 0,858 | 0,686 | 0,532 | 0,256 | 0,127 |
| 23 | 2,81 | 2,50 | 2,07 | 1,32 | 0,858 | 0,685 | 0,532 | 0,256 | 0,127 |
| 24 | 2,80 | 2,49 | 2,06 | 1,32 | 0,857 | 0,685 | 0,531 | 0,256 | 0,127 |
| 25 | 2,79 | 2,49 | 2,06 | 1,32 | 0,856 | 0,684 | 0,531 | 0,256 | 0,127 |
| 26 | 2,78 | 2,48 | 2,06 | 1,31 | 0,856 | 0,684 | 0,531 | 0,256 | 0,127 |
| 27 | 2,77 | 2,47 | 2,05 | 1,31 | 0,855 | 0,684 | 0,531 | 0,256 | 0,127 |
| 28 | 2,76 | 2,47 | 2,05 | 1,31 | 0,855 | 0,683 | 0,530 | 0,256 | 0,127 |
| 29 | 2,76 | 2,46 | 2,05 | 1,31 | 0,854 | 0,683 | 0,530 | 0,256 | 0,127 |
| 30 | 2,75 | 2,46 | 2,04 | 1,31 | 0,854 | 0,683 | 0,530 | 0,256 | 0,127 |
| 40 | 2,70 | 2,42 | 2,02 | 1,30 | 0,851 | 0,681 | 0,529 | 0,255 | 0,126 |
| 60 | 2,66 | 2,39 | 2,00 | 1,30 | 0,848 | 0,679 | 0,527 | 0,254 | 0,126 |
| 120 | 2,62 | 2,36 | 1,98 | 1,29 | 0,845 | 0,677 | 0,526 | 0,254 | 0,126 |

ANEXO D. TABLA PARA DISTRIBUCIÓN JI-CUADRADO

La primera fila corresponde a las áreas a la izquierda de un valor χ^2 y la primera columna representa los grados de libertad.

Ejemplo:

Determinar el valor χ^2 que deja a la izquierda un área 0,975 con $\nu = 10$ grados de libertad.

$$\chi^2(0,975; 10) = 20,48$$

Eso significa que, la probabilidad de que $P(\chi^2 \leq 20,48; \nu = 10) = 0,975$

| ν | 0,995 | 0,99 | 0,975 | 0,9 | 0,75 | 0,50 | 0,25 | 0,1 | 0,025 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 7,88 | 6,63 | 5,02 | 2,71 | 1,32 | 0,45 | 0,10 | 0,02 | 0,00 |
| 2 | 10,60 | 9,21 | 7,38 | 4,61 | 2,77 | 1,39 | 0,58 | 0,21 | 0,05 |
| 3 | 12,84 | 11,34 | 9,35 | 6,25 | 4,11 | 2,37 | 1,21 | 0,58 | 0,22 |
| 4 | 14,86 | 13,28 | 11,14 | 7,78 | 5,39 | 3,36 | 1,92 | 1,06 | 0,48 |
| 5 | 16,75 | 15,09 | 12,83 | 9,24 | 6,63 | 4,35 | 2,67 | 1,61 | 0,83 |
| 6 | 18,55 | 16,81 | 14,45 | 10,64 | 7,84 | 5,35 | 3,45 | 2,20 | 1,24 |
| 7 | 20,28 | 18,48 | 16,01 | 12,02 | 9,04 | 6,35 | 4,25 | 2,83 | 1,69 |
| 8 | 21,95 | 20,09 | 17,53 | 13,36 | 10,22 | 7,34 | 5,07 | 3,49 | 2,18 |
| 9 | 23,59 | 21,67 | 19,02 | 14,68 | 11,39 | 8,34 | 5,90 | 4,17 | 2,70 |
| 10 | 25,19 | 23,21 | 20,48 | 15,99 | 12,55 | 9,34 | 6,74 | 4,87 | 3,25 |
| 11 | 26,76 | 24,72 | 21,92 | 17,28 | 13,70 | 10,34 | 7,58 | 5,58 | 3,82 |
| 12 | 28,30 | 26,22 | 23,34 | 18,55 | 14,85 | 11,34 | 8,44 | 6,30 | 4,40 |
| 13 | 29,82 | 27,69 | 24,74 | 19,81 | 15,98 | 12,34 | 9,30 | 7,04 | 5,01 |
| 14 | 31,32 | 29,14 | 26,12 | 21,06 | 17,12 | 13,34 | 10,17 | 7,79 | 5,63 |
| 15 | 32,80 | 30,58 | 27,49 | 22,31 | 18,25 | 14,34 | 11,04 | 8,55 | 6,26 |
| 16 | 34,27 | 32,00 | 28,85 | 23,54 | 19,37 | 15,34 | 11,91 | 9,31 | 6,91 |
| 17 | 35,72 | 33,41 | 30,19 | 24,77 | 20,49 | 16,34 | 12,79 | 10,09 | 7,56 |

| ν | 0,995 | 0,99 | 0,975 | 0,9 | 0,75 | 0,50 | 0,25 | 0,1 | 0,025 |
|-------|--------|--------|--------|--------|--------|-------|-------|-------|-------|
| 18 | 37,16 | 34,81 | 31,53 | 25,99 | 21,60 | 17,34 | 13,68 | 10,86 | 8,23 |
| 19 | 38,58 | 36,19 | 32,85 | 27,20 | 22,72 | 18,34 | 14,56 | 11,65 | 8,91 |
| 20 | 40,00 | 37,57 | 34,17 | 28,41 | 23,83 | 19,34 | 15,45 | 12,44 | 9,59 |
| 21 | 41,40 | 38,93 | 35,48 | 29,62 | 24,93 | 20,34 | 16,34 | 13,24 | 10,28 |
| 22 | 42,80 | 40,29 | 36,78 | 30,81 | 26,04 | 21,34 | 17,24 | 14,04 | 10,98 |
| 23 | 44,18 | 41,64 | 38,08 | 32,01 | 27,14 | 22,34 | 18,14 | 14,85 | 11,69 |
| 24 | 45,56 | 42,98 | 39,36 | 33,20 | 28,24 | 23,34 | 19,04 | 15,66 | 12,40 |
| 25 | 46,93 | 44,31 | 40,65 | 34,38 | 29,34 | 24,34 | 19,94 | 16,47 | 13,12 |
| 26 | 48,29 | 45,64 | 41,92 | 35,56 | 30,43 | 25,34 | 20,84 | 17,29 | 13,84 |
| 27 | 49,64 | 46,96 | 43,19 | 36,74 | 31,53 | 26,34 | 21,75 | 18,11 | 14,57 |
| 28 | 50,99 | 48,28 | 44,46 | 37,92 | 32,62 | 27,34 | 22,66 | 18,94 | 15,31 |
| 29 | 52,34 | 49,59 | 45,72 | 39,09 | 33,71 | 28,34 | 23,57 | 19,77 | 16,05 |
| 30 | 53,67 | 50,89 | 46,98 | 40,26 | 34,80 | 29,34 | 24,48 | 20,60 | 16,79 |
| 40 | 66,77 | 63,69 | 59,34 | 51,81 | 45,62 | 39,34 | 33,66 | 29,05 | 24,43 |
| 50 | 79,49 | 76,15 | 71,42 | 63,17 | 56,33 | 49,33 | 42,94 | 37,69 | 32,36 |
| 60 | 91,95 | 88,38 | 83,30 | 74,40 | 66,98 | 59,33 | 52,29 | 46,46 | 40,48 |
| 70 | 104,21 | 100,43 | 95,02 | 85,53 | 77,58 | 69,33 | 61,70 | 55,33 | 48,76 |
| 80 | 116,32 | 112,33 | 106,63 | 96,58 | 88,13 | 79,33 | 71,14 | 64,28 | 57,15 |
| 90 | 128,30 | 124,12 | 118,14 | 107,57 | 98,65 | 89,33 | 80,62 | 73,29 | 65,65 |
| 100 | 140,17 | 135,81 | 129,56 | 118,50 | 109,14 | 99,33 | 90,13 | 82,36 | 74,22 |

ANEXO E. TABLA DISTRIBUCIÓN F

La primera fila corresponde a los grados de libertad v_1 y la primera columna a los grados de libertad v_2 . Todos los valores de F que aparecen dentro de la tabla, dejan a la izquierda un área de 0,95 y a la derecha un área de 0,5.

Ejemplo:

Determinar el valor F que deja a la izquierda un área 0,95 con $v_1 = 10$ y $v_2 = 8$ grados de libertad.

$$F(0,95; v_1 = 10; v_2 = 8) = 3,35$$

Eso significa que, la probabilidad de que $P(F \leq 3,35; v_1 = 10, v_2 = 8) = 0,95$

Área a la izquierda: 0,95

| v_1/v_2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 30 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 161,4 | 199,5 | 215,7 | 224,6 | 230,2 | 234,0 | 236,8 | 238,9 | 240,5 | 241,9 | 243,9 | 245,9 | 248,0 | 250,1 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,35 | 19,37 | 19,38 | 19,40 | 19,41 | 19,4 | 19,4 | 19,5 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,89 | 8,85 | 8,81 | 8,79 | 8,74 | 8,7 | 8,7 | 8,6 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 | 5,91 | 5,9 | 5,8 | 5,7 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,77 | 4,74 | 4,68 | 4,6 | 4,6 | 4,5 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 | 4,00 | 3,9 | 3,9 | 3,8 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,64 | 3,57 | 3,5 | 3,4 | 3,4 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,35 | 3,28 | 3,2 | 3,2 | 3,1 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,14 | 3,07 | 3,0 | 2,9 | 2,9 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 | 2,91 | 2,8 | 2,8 | 2,7 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 | 2,79 | 2,7 | 2,6 | 2,6 |
| 12 | 4,75 | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 | 2,69 | 2,6 | 2,5 | 2,5 |
| 13 | 4,67 | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 | 2,60 | 2,5 | 2,5 | 2,4 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 | 2,53 | 2,5 | 2,4 | 2,3 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 | 2,48 | 2,4 | 2,3 | 2,2 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,42 | 2,4 | 2,3 | 2,2 |

| v_1/v_2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 30 |
|-----------|------|------|------|------|------|------|------|------|------|------|------|-----|-----|-----|
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 | 2,38 | 2,3 | 2,2 | 2,1 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,34 | 2,3 | 2,2 | 2,1 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 | 2,31 | 2,2 | 2,2 | 2,1 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 | 2,28 | 2,2 | 2,1 | 2,0 |
| 21 | 4,32 | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,49 | 2,42 | 2,37 | 2,32 | 2,25 | 2,2 | 2,1 | 2,0 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 | 2,23 | 2,2 | 2,1 | 2,0 |
| 23 | 4,28 | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,44 | 2,37 | 2,32 | 2,27 | 2,20 | 2,1 | 2,0 | 2,0 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 | 2,18 | 2,1 | 2,0 | 1,9 |
| 25 | 4,24 | 3,39 | 2,99 | 2,76 | 2,60 | 2,49 | 2,40 | 2,34 | 2,28 | 2,24 | 2,16 | 2,1 | 2,0 | 1,9 |
| 26 | 4,23 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 | 2,15 | 2,1 | 2,0 | 1,9 |
| 27 | 4,21 | 3,35 | 2,96 | 2,73 | 2,57 | 2,46 | 2,37 | 2,31 | 2,25 | 2,20 | 2,13 | 2,1 | 2,0 | 1,9 |
| 28 | 4,20 | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 | 2,12 | 2,0 | 2,0 | 1,9 |
| 29 | 4,18 | 3,33 | 2,93 | 2,70 | 2,55 | 2,43 | 2,35 | 2,28 | 2,22 | 2,18 | 2,10 | 2,0 | 1,9 | 1,9 |
| 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 | 2,09 | 2,0 | 1,9 | 1,8 |
| 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 | 2,00 | 1,9 | 1,8 | 1,7 |
| 60 | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 | 1,92 | 1,8 | 1,7 | 1,6 |
| 120 | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,18 | 2,09 | 2,02 | 1,96 | 1,91 | 1,83 | 1,8 | 1,7 | 1,6 |

Acerca de los Autores

Alberto Javier Mesa Guerrero. Docente adscrito al Departamento de Matemáticas y Estadística, Facultad de Ciencias Exactas y Naturales, Universidad de Nariño. Licenciado en Matemáticas y Física, Universidad de Nariño. Especialista en Computación para la Docencia, Universidad Mariana. Estadístico en Salud, Universidad de Antioquia. Profesor en categoría Asociado de la Universidad de Nariño.

Correo electrónico: soundmesa@yahoo.com.

Segundo Javier Caicedo-Zambrano. Docente adscrito al Departamento de Matemáticas y Estadística, Facultad de Ciencias Exactas y Naturales, Universidad de Nariño. Doctor en Ciencias de la Educación, Universidad del Tolima. Licenciado en Matemáticas y Física, Universidad de Nariño. Ingeniero de Sistemas, Universidad Antonio Nariño. Especialista en Computación para la Docencia, Universidad Mariana. Especialista en Multimedia Educativa, Universidad Antonio Nariño. Magister en Software Libre, Universidad Autónoma de Bucaramanga. Asesor de Desarrollo Académico, Universidad de Nariño. Profesor Tiempo Completo Universidad de Nariño.

Correo electrónico: jacaza1@gmail.com; jacaza1@udenar.edu.co.



Editorial

Universidad de **Nariño**

Fecha de publicación: 2022
San Juan de Pasto - Nariño - Colombia

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA INFERENCIAL

El libro contiene exactamente las temáticas que se desarrollan en los cursos de cálculo de probabilidades y estadística de la Universidad de Nariño, por lo cual, en dichos cursos se podría utilizar como texto guía. No se trata de un libro de texto en sí mismo, sino de unas notas de clase de un curso de Cálculo de Probabilidades y Estadística de la Universidad de Nariño.

El libro está organizado en ocho (8) capítulos. El primero, “Introducción a la teoría de conjuntos”, presenta las operaciones entre conjuntos y el concepto de producto cartesiano. El segundo capítulo, “Introducción al cálculo de probabilidades”, presenta los fundamentos de las técnicas de contar, los conceptos de espacio muestral, definición de probabilidad, las leyes de probabilidad y finaliza con el estudio de espacios finitos de probabilidad. El capítulo tercero, trata sobre probabilidad condicional, procesos estocásticos finitos, probabilidad total y probabilidad de Bayes. El Capítulo cuarto, “Variables aleatorias”, desarrolla el tema de variables aleatorias discretas y continuas. El capítulo quinto, “Modelos de probabilidades”, trata algunos modelos de distribución de probabilidad, tanto discretos como continuos. En el capítulo sexto, “Distribuciones muestrales”, se estudia la distribución muestral de la media, de la proporción de éxitos, de la diferencia de medias y de la diferencia de proporciones. En el capítulo séptimo, se trabaja la aplicación de las distribuciones en el cálculo de probabilidades, la estimación de parámetros, la prueba de hipótesis y el análisis de varianza. Finalmente, el capítulo ocho, “Problemas propuestos”, presenta diversos problemas sobre cálculo de probabilidades, modelos de probabilidad y aplicaciones de la estadística inferencial.



Universidad de **Nariño**
FUNDADA EN 1904

ai

Universidad de **Nariño**

ACREDITADA DE ALTA CALIDAD
RESOLUCIÓN MEN 10567 - MAYO 23 DE 2017

Editorial
Universidad de Nariño