ESTADÍSTICA GENERAL:

MODELO DE REGRESIÓN LOGÍSTICA EN LA PREDICCIÓN DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS

Ciencia Latina Internacional
CID editorial
REDALAC
INDEXA
AGENCIA ISBN
Crossref

Septiembre 2023 – CID - Centro de Investigación y Desarrollo

Copyright © CID - Centro de Investigación y Desarrollo

Copyright del texto © 2023 de Autores

biblioteca.ciencialatina.org

editorial@ciencialatina.org

Atención por WhatsApp al +52 22 26903834

Datos Técnicos de Publicación Internacional
Título: Estadística General: Modelo de regresión logística en la predicción del rendimiento académico.
Autor: Raúl Alfredo Sánchez Ancajima, Wilfredo Barrientos Farias, Washington Fernando Maliza Muñoz, Roger Álvaro Fernández Villarroel, Erlith Tafur Huaman
Editor: CID - Centro de Investigación y Desarrollo
Diseño de tapa: CID - Centro de Investigación y Desarrollo
Corrección de Estilo: CID - Centro de Investigación y Desarrollo
Formato: PDF
Páginas: 96 pág**.**
Tamaño: A4
Requisitos de sistema: Adobe Acrobat Reader
Modo de acceso: World Wide Web
ISBN: 978-99989-64-81-5
DOI: https://doi.org/10.37811/cli_w964

1ª. Edición. Año 2023. Editorial CID - Centro de Investigación y Desarrollo.

El contenido del libro y sus datos en su forma, corrección y fiabilidad son responsabilidad exclusiva de los autores. Permite la descarga de la obra y compartir siempre que los créditos se atribuyan a los autores, pero sin la posibilidad de cambiarlo de cualquier forma o utilizarlo con fines comerciales

Prohibida su reproducción por cualquier medio.

Distribución gratuita

ÍNDICE

CAPÍTULO 1: 5

ESTADÍSTICA GENERAL 5

MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN. 23

CÁLCULO DE LA MEDIANA 25

INTERVALO DE CONFIANZA 49

PRUEBA DE HIPÓTESIS (PH). 57

CONCLUSIONES 73

BIBLIOGRAFÍA 74

CAPÍTULO 2 75

PROYECTO DE INVESTIGACIÓN 75

RESUMEN 76

INTRODUCCIÓN 78

MATERIALES Y MÉTODOS 80

RESULTADOS 81

DISCUSIÓN 90

CONCLUSIONES 92

REFERENCIAS BIBLIOGRÁFICAS 93

Estadística General: Modelo de regresión logística en la predicción del rendimiento académico de estudiantes universitarios

Autores:
Raúl Alfredo Sánchez Ancajima
Willy Barrientos
Washington Fernando Maliza Muñoz
Roger Álvaro Fernández Villarroel
Erlith Tafur Huaman

Capítulo 1:

ESTADÍSTICA GENERAL

La estadística que muchos creen como algo nuevo, ya que hoy en día todo se fundamenta sobre datos o información estadística, para sombro de todos es tan antigua como la propia humanidad, ya que hay vestigios de su inicio antes de Cristo, como en China donde el emperador Yao (2,230 a. c) dispuso la realización de un censo en su imperio; algo similar ocurrió en la época del rey Herodes, relacionado con un empadronamiento cuando se habló del nacimiento de Cristo.

La estadística como disciplina tuvo su inicio en Alemania a mediados del siglo XVII, en el reinado de Godofredo de Achenwall (1719 – 1772), utilizó la palabra estadística y la separó de la sociología. Hoy en día alcanza un alto desarrollo, con la ayuda de tantos investigadores dedicados a perfeccionar métodos estadísticos modernos, que nos permiten seleccionar aquellas técnicas más adecuadas para analizar una determinada cantidad de información acelerando su proceso de cálculo con la ayuda de los programas de computación (Software estadísticos), disminuyendo las probabilidades de error.

En la vida diaria los diversos fenómenos de orden económico, social, político, educacional, e incluso biológico, aparecen, se transforman y finalmente desaparecen.

Para tan abundante y complejo material, es preciso tener un registro ordenado y continuo a fin de conseguir en un momento dado, los datos necesarios para el estudio de lo que ha sucedido, sucede o puede suceder.

Para ello se requiere contar con un método, o conjunto de reglas o principios, que nos permita la observación, el ordenamiento, la cuantificación y el análisis de dichos fenómenos. Ese método se denomina *Estadística*.

1.1 Definición de la Estadística.

La palabra *estadística* se refiere a un sistema o método usado en la recopilación, organización, análisis y descripción numérica de la información.

Los administradores aplican algunas técnicas estadísticas a virtualmente todas las ramas de las empresas públicas y privadas. Estas técnicas son tan diversas que los estadísticos, por lo general, consideran dos fases en el campo de la Estadística. En primer lugar, está la fase que sólo se limita a la descripción de una serie de datos sin llegar a conclusiones o a generalizar con respecto a un grupo mayor. Esta (fase) se conoce como *Estadística deductiva o descriptiva.* En segundo lugar, está la fase de análisis que trata de

llegar a conclusiones acerca de un grupo mayor basado en la información de un grupo menor o muestra: es esta la *Estadística inductiva o de inferencia*.

En un principio se consideraba que la función de la estadística era la descripción de las características de grupo, actividad que la hacía confundir con el papel que cumple la *historia* de observar y describir el hecho. En su origen, las estadísticas eran historias; hoy en día, la estadística, además de ser *descriptiva*, es *analítica*, considerándose esta última como la función más importante que realiza, ya que permite obtener conclusiones para un grupo mayor, denominado *población*, partiendo de una investigación realizada en un grupo menor, conocido como *muestra*, cuyos elementos, en la mayoría de casos, se seleccionan aleatoriamente.

La estadística descriptiva o deductiva tiene como finalidad colocar en evidencia aspectos característicos (promedios, variabilidad de los datos, etc.), que sirven para efectuar comparaciones sin pretender sacar conclusiones de tipo más general.

Esta descripción se realiza a través de la elaboración de cuadros, gráficos, cálculo de promedios, varianzas, proporciones de una o más variables, que deben estar relacionadas.

La estadística inferencial o inductiva busca dar explicaciones al conjunto de observaciones, probar la significación o validez de los resultados; intenta descubrir las causas que lo originan, con gran aplicación en el campo del muestreo, lográndose de esta manera, conclusiones que se extienden más allá de las estadísticas mismas.

Uno de los principales objetivos de la estadística es hacer inferencias acerca de los valores estadísticos de la población, denominados parámetros, a través de la información obtenida en una muestra, que permite el cálculo de estimadores.

La estadística, de acuerdo con todo lo anterior, se podría definir como aquel método que permite no sólo describir el hecho o fenómeno, sino deducir y evaluar conclusiones acerca de una población, utilizando resultados proporcionados por una muestra.

1.2. Algunos términos y conceptos que se deben conocer y manejar.

A continuación, se definen algunos términos que serán utilizados en el desarrollo de los diferentes capítulos, en los cuales se ha dividido esta asignatura.

Universo. Se denomina universo, a un conjunto de elementos (que consiste de personas, objetos, etc.) que contienen una o más *características* observables de naturaleza cualitativa o cuantitativa que se pueden medir en ellos.

Unidad elemental. A cada elemento de un universo se denomina unidad elemental.

Por ejemplo, los empleados de una empresa en un día laborable, constituyen un Universo en la que cada empleado (unidad elemental), tiene muchas características a ser observadas, como, por ejemplo: sexo, estado civil, lugar de procedencia, grado de instrucción, etc. (característica cualitativa), o número de hijos, ingresos mensuales, etc. (características cuantitativas).

El resultado de medir una característica observable de una unidad elemental, se denomina dato estadístico o valor observado o simplemente observación.

Población. Por otra parte, el universo viene definido por la investigación estadística a realizarse. Y como la medición de la característica especificada por la investigación se hace a cada unidad elemental, se puede considerar a la *población* como el conjunto de medidas o el recuento de todos los elementos que presentan una característica común especificada por la investigación estadística. El termino población se usa para denotar el conjunto de elementos del cual se extrae la muestra.

Marco muestral. Se denomina marco muestral o marco de referencia, a la lista, mapa o cualquier otro material aceptable, que contenga todas las *unidades* o *elementos* perfectamente identificadas y actualizadas, de la cual se selecciona la muestra.

Características. Es importante resaltar el hecho de que, a pesar de encontrarse una población constituida por un grupo de elementos o unidades, a la estadística no le interesa el elemento o la unidad en sí, sino su característica.

Estos son algunos ejemplos de lo que podrían denominarse elementos de una población y características de la misma.

Elementos	Características
Personas	Salarios
Personas	Horas trabajadas
Bananos	Número de nematodos
Parcelas de cultivo de papa	Rendimiento
Hogar	Consumo

Las características (o caracteres), corresponden a ciertos rasgos, cualidades o propiedades que poseen los elementos que constituyen la población o la muestra.

Muestra. Es un conjunto de medidas de una parte de los elementos pertenecientes a la población, con el fin de obtener información acerca de la población de la cual proviene. La muestra debe ser seleccionada de manera que sea representativa de la población. Para que una muestra sea representativa de la población se requiere que las unidades sean *seleccionadas aleatoriamente o al azar* (es decir, todos los elementos que componen la población tienen la misma posibilidad de ser seleccionados); ya sea utilizando métodos tales como: el sorteo, las tablas de números aleatorios, la selección sistemática o cualquier otro método que sea al azar.

Una muestra *no probabilística* corresponde a la selección de las unidades o elementos, según el criterio o conveniencia del investigador. En estos casos, algunas unidades tienen mayor posibilidad que otras de ser seleccionadas; por tal razón no se puede determinar la validez, ni la confianza que merecen dichos resultados. Casi por lo general, las unidades se seleccionan en forma caprichosa, voluntaria, por cuotas, tal como sucede en las encuestas de opinión, que en la mayoría de los casos no tienen credibilidad, de ahí que se le conozca como *Muestra circunstancial*.

Parámetro. Es una medida que describe numéricamente una característica de una población, tal como la media (μ) o la varianza (σ^2). También se le denomina valor verdadero, ya que una característica poblacional tendrá un solo parámetro (media, varianza, etc.). Sin embargo, una población puede tener varias características y, por tanto, varios parámetros. Algunos lo denominan como *valor estadístico de la población*.

Estadística o estadígrafo: La descripción numérica de una característica correspondiente a los elementos de una muestra aleatoria, se denomina *estadígrafo*, tal como la media ($\overline{\overline{X}}$) o la varianza (S^2) muestral. De una población se pueden obtener M

número de muestras posibles y en cada una de ellas se puede cuantificar la característica, obteniéndose por lo general, valores diferentes para cada muestra, a pesar de ser utilizado el mismo estadígrafo o medida.

También se le conoce como *Estimador Puntual* si se trata de un promedio, varianza, proporción, etc. Como por lo general, existe una diferencia entre el estimador y el parámetro, denominado *error*, es aconsejable utilizar el *estimador por intervalos*, dentro del cual deberá estar contenido el parámetro con cierto margen de seguridad, establecido por el investigador.

La diferencia entre el estimador y el parámetro, como ya se dijo, se le denomina como *error de muestreo*, y lo fija el investigador para determinar el tamaño de la muestra, con base en la experiencia y el conocimiento que tenga sobre el comportamiento y, la variación que tenga la característica o características principales relacionadas con el objetivo mismo de la investigación.

Errores ajenos al muestreo; no son cuantificables como el anterior y son resultados de errores sistemáticos, que se producen a lo largo de la investigación.

1.3. Variables estadísticas.

Una variable estadística es una característica de las personas u objetos que puede tomar diferentes valores. Las variables estadísticas generalmente se denotan con las letras X, Y, Z, etc.

Las variables se pueden clasificar de la forma siguiente:

1.3.1. Variable Cuantitativa. - Son aquellas variables que se obtienen como resultado de mediciones y conteos. Son variables cuantitativas: estatura, peso, ingreso, valor, producción, ventas, el número de personas por hogar, el número de piezas ensambladas que se han encontrado defectuosos, etc.

Las variables cuantitativas se clasifican en Discretas y continuas.

a) Una *variable discreta* sólo puede tomar unos determinados valores, siendo imposible que llegue a tomar valores intermedios entre dos consecutivos, es decir, toma únicamente valores enteros: 1, 2, 3, etc., tal es el caso del número de hijos por familia, el número de gusanos rosado por plantas de algodón, número de nematodos en 100 gramos suelo en el cultivo de cítricos, etc.

b) Una *variable continua* puede tomar cualquier valor en un intervalo especificado, es decir, se admiten valores fraccionarios, como el número de años de una persona, el

rendimiento de una nueva variedad de trigo, la presión sanguínea, el tiempo transcurrido antes de que falle un dispositivo y el número promedio de personas por hogar en una comunidad grande.

1.3.2. Variable Cualitativa. - Son variables cuyos valores consisten en categorías de clasificación, o sea se refiere a la cualidad que presenta la población, que no lleva clasificación numérica. Por ejemplo, la variable "especies de nematodos" puede adoptar las modalidades: Tylenchulus semiprenetrans (ataca a los cítricos) y el radopholus cimiles (ataca a los bananos). También son ejemplos de variables cualitativas: Estado civil, clase social, profesión, variedades de papa, marcas, calidad, etc.

A su vez esta variable puede clasificarse en: Nominal y Ordinal.

a) **Variable Cualitativa Nominal.** - Son aquellas que surgen cuando se definen categorías y se cuenta el número de observaciones pertenecientes a cada categoría y no lleva ninguna ordenación en las posibles modalidades.

- La variable cualitativa “color de la bebida gaseosa” con las posibles modalidades: Negro, Rojo, Naranja, es una variable cualitativa nominal.

- Son ejemplos de variables cualitativas: “estado civil”, “variedades de arroz”, “Nacionalidad”, entre otras.

b) Variable Cualitativa Ordinal. - Cuando se va más allá del nivel de clasificación y se busca ordenar los casos en términos del grado que posee una determinada característica, entonces se está trabajando con una variable cualitativa ordinal.

Ejemplos: la variable “status socioeconómico” con sus modalidades: clase baja, media y alta es una variable cualitativa ordinal. La variable “orden de mérito” cuyas modalidades son 1º, 2º, 3º, etc. mide las calificaciones de las unidades estadísticas en escala ordinal.

1.4. Obtención de datos a través experimentos y encuestas.

Una manera de obtener datos es a través de la observación directa. Un *experimento estadístico* es una forma de observación directa en la que se controlan algunos o todos los factores que pueden influir sobre la variable que se estudia.

En algunas situaciones, no es posible obtener datos en forma directa, sino que, más bien, la información debe obtenerse a partir de respuestas individuales. Una *encuesta estadística* es el proceso de recopilar datos pidiendo a personas que proporcionen

información. Los datos pueden obtenerse con métodos como la entrevista personal o telefónica, o través de un cuestionario escrito.

1.5. Presentación de los datos.

Después de obtener un conjunto de datos, es necesario presentarlos en forma tal, que facilite su compresión y su posterior análisis y utilización. No servirá de nada que estas medidas se presenten en un simple listado. Lo mejor será ordenarlos en tablas o cuadros y luego representarlo en gráficos.

Si se tienen pocos datos, y estos son valores discretos, entonces conviene presentar una distribución de frecuencias sin intervalos; pero si se tiene valores continuos o muchos valores discretos, conviene presentar una distribución de frecuencia por intervalos.

1.5.1. Distribución de frecuencias sin intervalos:

Ejemplo 1.1. Se pretende describir la gravedad de un ataque de mosca del mediterráneo a partir del número de picadas (variable X) constatado en 20 naranjas:

2, 1, 2, 4, 1, 3, 2, 3, 2, 0, 3, 2, 1, 3, 2, 3, 3, 1, 2, 4.

Al ordenar estos datos en forma ascendente, se obtienen cinco valores distintos 0, 1, 2, 3, 4 que se repiten respectivamente 1, 4, 7, 6, 2 veces. La distribución de frecuencias de X se da en la tabla 1.1.

Tabla 1.1. Distribución de frecuencias del número de picadas de la mosca del mediterráneo por naranja.

Número de picadas X_i	Número de naranjas n_i	Frecuencias Relativas h_i	Frecuencias Porcentajes h_i (%)
0	1	0.05	5
1	4	0.20	20
2	7	0.35	35
3	6	0.30	30
4	2	0.10	10
Total	20	1.00	100

Fuente. - Datos obtenidos en el laboratorio de biología de la UNP.

Gráfico.

La representación gráfica más común para este tipo distribución de frecuencias es el diagrama de barras que consiste en trazar en cada valor distinto de la variable, segmentos de líneas proporcionales a su frecuencia.

Figura 1.1. Diagrama de barras para los datos de la tabla 1.1.

1.5.2. Distribución de frecuencias por intervalos.

La distribución de frecuencia por intervalos o clases se usa cuando la variable estadística es continúa o cuando el número de valores distintos de una variable discreta es grande.

Elaboración de las tablas de frecuencias.

Para esclarecer la construcción de la tabla de frecuencias para datos agrupados en intervalos de clase, desarrollaremos un ejemplo.

Ejemplo 1.2. Se dispone de 50 parcelas de igual tamaño y fertilidad ubicados en la provincia de Huancabamba, durante el mes de marzo de 2007. Los datos de las observaciones del rendimiento (en kilogramos) de cierta variedad de papa por parcela se dan a continuación:

56 42 59 50 38 46 45 49 57 48 47 53 48 39 55 51 50 51

64 48 45 40 56 51 52 46 51 47 48 31 49 42 54 50 51 47

56 46 53 58 41 50 49 52 48 55 45 35 36 43.

Se pide representar los datos en una tabla de frecuencias.

Solución.

1º. Debemos de determinar el rango (R) de variación de los datos que se define por:

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

En nuestro ejemplo, tenemos:

Valor mínimo $X_{min} = 31$

Valor máximo $X_{máx} = 64$

Luego el rango es R= 64 - 31= 33

2º. Tenemos que calcular ¿cuántas clases deben formarse? Para eso se usa la fórmula de Sturges:

$$K = 1 + 3.322 \log n$$

donde:

K = número de clases

n = número de elementos en la muestra (tamaño de la muestra).

Está formula es muy útil porque orienta al principiante. Sin embargo, es un poco conservador y tiende a dar un número de clases un poco menor del que se utiliza en la práctica.

En nuestro ejemplo: K =?, n = 50

Aplicando la fórmula tenemos:

$$K = 1 + 3.322 \log 50 = 6.644$$

redondeando al entero inmediato mayor porque, como ya se indicó, la fórmula es un poco conservadora, obtenemos:

K = 7.

Nota. - por razones extraídas de la práctica, se adoptan los siguientes límites para K.

$$5 \leq K \leq 15$$

3º. La idea es resumir los valores con el fin de percibir algunas características o propiedades de los datos que no aparecen a simple vista. Para esto vamos a clasificar los 50 autobuses en 7 clases, de acuerdo con la fórmula de Sturges.

La amplitud de cada intervalo de clase que representamos con la letra c, se obtiene por medio de la fórmula:

$$c = \frac{R}{K}$$

Aplicando los datos a la fórmula anterior, se tiene: $c = \frac{33}{7} = 4.7143$

Redondeamos a 5.

Luego c = 5 y, por tanto, el nuevo recorrido será:

$$R' = c \times K = 5 \times 7 = 35$$

4º. Teniendo en cuenta que este recorrido es mayor que el recorrido original, buscaremos el exceso:

Exceso = 35 - 33= 2 kilogramos

Debemos repartir este exceso a los dos extremos del recorrido original, mitad a cada lado, en este caso 1 a cada extremo. Sumamos el número 1 al valor máximo: 64 + 1 = 65 y restamos el número 1 al valor mínimo: 31 - 1 = 30.

A partir de este valor 30 agregamos sucesivamente la amplitud 5 y obtenemos los puntos de división que determinan los 7 intervalos (ver figura 1.2).

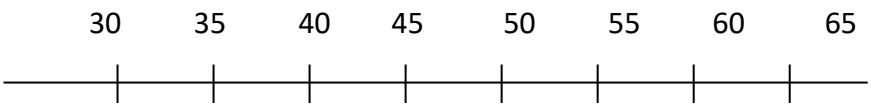


Figura 1.2

sin embargo, una dificultad se presenta cuando algunos de los datos coinciden con cualquiera de los puntos de división: 35, 40, 45, 50, 55 y 60.

5º. Supongamos que un dato es 35 ¿dónde lo colocamos? - ¿en el primer intervalo cuyos extremos son 30 y 35? - (ver Fig. 2.1) ó ¿en el segundo intervalo que tiene por extremos 35 y 40? -

Para aclarar está ambigüedad adoptamos, el intervalo abierto por la izquierda que no incluye al valor 30 (límite inferior del intervalo) y cerrado por la derecha que incluye al valor 35 (límite superior del intervalo). Este tipo de intervalo se representa matemáticamente por: $(Y'_{i-1} - Y'_i]$, donde Y'_{i-1} representa el límite inferior del intervalo y Y'_i representa el límite superior.

Así, en nuestro ejemplo el primer intervalo lo escribiremos como $(30 - 35]$.

Por tanto, los intervalos de clase quedan definidos como:

Intervalo de clases	(30 - 35]	(35 - 40]	(40 - 45]	(45 - 50]	(50 - 55]	(55 - 60]	(60 - 65]

Nota. - También existen otros dos tipos de intervalos de clases, definidas como:

$[Y'_{i-1} - Y'_i)$ Intervalo semicerrado. Este intervalo es cerrado por la izquierda (incluye al límite inferior del intervalo) y abierto por la derecha (no incluye al límite superior del intervalo).

$[Y'_{i-1} - Y']$ Intervalo cerrado. Este intervalo incluye a ambos límites del intervalo.

6º. Es conveniente que todos y cada uno de los datos que se hallen dentro de un mismo intervalo, estén representados por un mismo valor. Este valor caracteriza a la clase y por eso se llama marca de clase, se obtiene promediando los límites de cada intervalo. Una fórmula para calcular la marca de clase de un intervalo es:

$$Y_i = \frac{Y'_{i-1} + Y'_i}{2}$$

7º. A continuación debemos realizar la clasificación y conteo de los datos (ver tabla 1.2), es decir, colocar cada uno de ellos dentro de su clase, todos representados por un mismo signo: una tarja.

Tabla 1.2.

Intervalo de clase $(Y'_{i-1} - Y'_i]$	Marca de clase Y_i	Conteo	Frecuencia n_i
30 - 35	32.5		
35 - 40	37.5		
40 - 45	42.5		
45 - 50	47.5		
50 - 55	52.5		
55 - 60	57.5		
60 - 65	62.5		

A continuación, presentamos la tabla 1.3, conocida como tabla de distribución de frecuencias absolutas.

Tabla 1.3. Distribución de frecuencias de 50 parcelas según el rendimiento (en kgrs) de la variedad de papa V_1. Parcelas ubicadas en la Provincia de Huancabamba. Marzo de 2007.

Intervalo de clase $(Y'_{i-1} - Y'_i]$	Marca de clase Y_i	Frecuencia n_i	Frecuencia relativa	Frecuencia acumulada	Frecuencia acumulada
30 - 35	32.5	2	0.04	2	0.04
35 - 40	37.5	4	0.08	6	0.12
40 - 45	42.5	7	0.14	13	0.26
45 - 50	47.5	18	0.36	31	0.62
50 - 55	52.5	12	0.24	43	0.86
55 - 60	57.5	6	0.12	49	0.98
60 - 65	62.5	1	0.02	50	1.00
Total		50	1.00		

Fuente. Observaciones registradas por el propietario de las parcelas, ubicados en la provincia de Huancabamba, durante el mes de marzo de 2007.

- **Frecuencia absoluta: (n_i)**

Llamaremos frecuencia absoluta de un valor x_i de la variable estadística X, al número de veces aparece repetido dicho valor en el conjunto de las observaciones realizadas.

Propiedad. - $n_1 + n_2 + \cdots + n_k = \sum_{i=1}^{k} n_i = n$

- **Frecuencia absoluta acumulada: (N_i)**

La frecuencia absoluta de un valor x_i de la variable X es igual a la suma de los valores inferiores o iguales a dicho valor evidentemente. Así, la frecuencia absoluta acumulada del ultimo valor será n (donde n = número de observaciones realizadas).

$$N_i = n_1 + n_2 + \cdots + n_i = \sum_{j=1}^{i} n_j$$

Propiedad. - $N_i = N_{i-1} + n_i$, de donde $n_i = N_1 - N_{i-1}$

- **Frecuencia relativa: (h_i)**

Llamaremos frecuencia relativa de un valor observado x_i de la variable X, al cociente entre su frecuencia absoluta y el número de observaciones realizadas (n), y se denota por:

$$h_i = \frac{n_i}{n}, \quad 0 \le h_i \le 1 \ , \ \forall \ i = 1, 2, ..., k..$$

Propiedad. - $h_1 + h_2 + \cdots + h_k = \sum_{i=1}^{k} h_i = 1$

- **Frecuencia relativa acumulada: (H_i)**

Se llama frecuencia relativa acumulada de un valor x_i de la variable X, al cociente entre su frecuencia absoluta acumulada y el número de observaciones realizadas (n), y se denota por:

$$H_i = \frac{N_i}{n} = \frac{\sum_{j=1}^{i} n_j}{n} = \sum_{j=1}^{i} \frac{n_j}{n} = \sum_{j=1}^{i} h_j$$

Propiedad. - $H_i = H_{i-1} + h_i$

Interpretación. -

$n_2 = 4$, significa que 4 parcelas tuvieron un rendimiento de papa mayor que 35 kilogramos y menor o igual a 40 kilogramos.

$N_4 = 31$, significa que 31 parcelas tuvieron un rendimiento mayor que 30 kilogramos y menor o igual que 50 kilogramos; o también significa que 31 parcelas tuvieron un rendimiento menor o igual a que 50 kilogramos.

$h_3 = 0.14$, significa que el 14% de las parcelas tuvieron a un rendimiento mayor que 40 kilogramos y menor o igual a 45 kilogramos.

$H_5 = 0.86$, significa que el 86% del total de las parcelas tuvieron un rendimiento mayor que 30 kilogramos y menor o igual que 55 kilogramos; o también significa que 86% del total de parcelas tuvieron un rendimiento menor o igual a que 55 kilogramos.

$N_6 - N_2 = 49 - 6 = 43$ parcelas tuvieron un rendimiento mayor que 40 kilogramos y menor o igual que 60 kilogramos.

Ejercicio. - Abrir la ventana Descripción, escoger la opción Datos numéricos $\rightarrow$ Análisis unidimensional, enseguida seleccionar el icono Opciones tabulares $\rightarrow$ Tablas de frecuencias en el programa estadístico STATGRAPHICS Plus para Windows 5.1, para obtener la tabla de distribución de frecuencias de nuestro ejemplo ilustrativo que se viene considerando.

Gráfico de la distribución por intervalos.

Los gráficos más usados son: Histograma, Polígono de frecuencias y polígono de frecuencias acumuladas u ojiva.

a) Histograma.

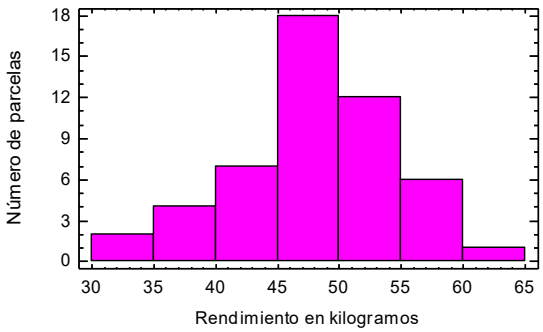
Es una representación gráfica de una distribución de frecuencias agrupadas en intervalos de clase, mediante una serie de rectángulos contiguos que tienen:

- sus bases sobre un eje horizontal y cuya longitud serán igual al tamaño de los intervalos de clase.

- Las alturas proporcionales a la frecuencia (absoluta o relativa).

Ejemplo 1.3. Abriendo la ventana *Descripción* y luego escogiendo la opción *Datos numéricos* $\rightarrow$ *Análisis unidimensional* $\rightarrow$ *Opciones gráficas* y finalmente seleccionar *Histograma de frecuencias* en el programa estadístico STAGRAPHICS, obtener el gráfico de histograma de frecuencias absolutas para el ejemplo 2.1.

Figura 1.3. Histograma de frecuencias de la distribución del rendimiento de papa por parcela de la tabla 1.3.

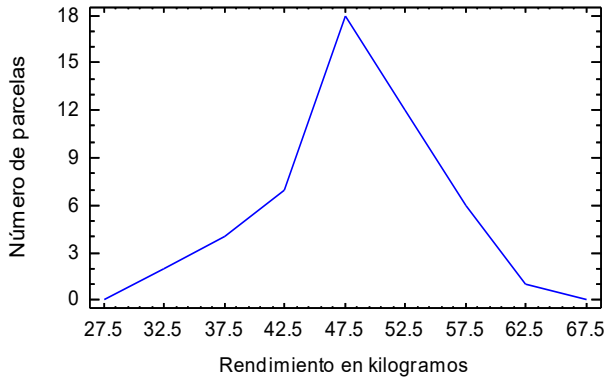


b) Polígono de frecuencias.

Cuando la variable está agrupada en intervalo de clase, el polígono de frecuencia se obtiene uniendo los puntos medios de las bases superiores de cada rectángulo en el histograma.

Ejemplo 1.4. Abriendo la ventana *Descripción* y luego escogiendo la opción *Datos numéricos* → *Análisis unidimensional* → *Opciones gráficas* → *Histograma* y haciendo clic con el botón derecho del *mouse* para que aparezca la ventana de dialogo *Opciones de ventana*, seleccionar en *Tipo de gráfico* la opción *Polígono* dentro del programa estadístico STATGRAPHICS, para obtener el polígono de frecuencias absolutas para el ejemplo 1.2.

Figura 1.4. Polígono de frecuencias del rendimiento en kilogramos por parcela de la tabla 1.3.

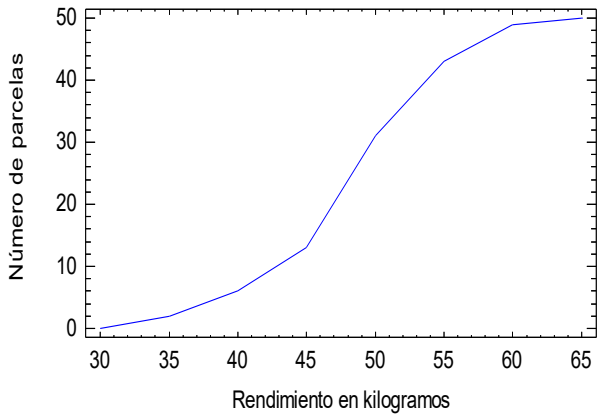


c) Polígono de frecuencias acumuladas u ojivas.

Está representación es válida para variables estadísticas agrupadas en intervalos de clase. En el eje de las abscisas representamos los distintos intervalos de clase que han de estar naturalmente traslapados. En el extremo superior de cada intervalo se levanta una vertical con altura igual a la frecuencia (absoluta o relativa) acumulada, luego se unen los extremos superiores de las verticales con segmentos rectilíneos. Así el polígono de frecuencias acumuladas absolutas alcanzará su máxima altura en el último intervalo.

Abriendo la ventana *Descripción* → *Datos numéricos* → *Análisis unidimensional* → *Opciones gráficas* → *Histograma* y haciendo clic con el botón derecho del *mouse* para que aparezca la ventana de dialogo *Opciones de ventana,* seleccionar en *Tipo de gráfico* la opción *Polígono* y finalmente seleccionar en *Frecuencia* la opción *Acumulada*, para obtener el polígono de frecuencias acumuladas absolutas para el ejemplo 1.2.

Figura 1.5. Polígono de frecuencias acumuladas del número de pasajeros por autobús de la tabla 1.3.



1.5.3. Distribución de frecuencias: Variable cualitativa

Consideremos el siguiente ejemplo:

Ejemplo 1.5. En una encuesta de opinión acerca de las preferencias de una marca de bebidas gaseosas por sus colores: Negro (N), Blanco (B), Rojo (R), 20 consumidores dieron las siguientes respuestas:

B, N, N, B, R, N, N, B, B, N, B, N, N, R, B, N, B, R, B, N.

Construir la distribución de frecuencias.

Solución.

La tabulación de estos datos, donde la variable cualitativa es X: Color de bebida gaseosa, es la distribución de frecuencias de la tabla 1.4.

Tabla 1.4. Distribución de personas por su color preferido de una marca de bebida gaseosa.

Color de Bebida	Número de Consumidores: n_i	Frecuencias relativas: h_i	Frecuencias Porcentajes: h_i(%)
Blanco (B)	8	0.40	40
Negro (N)	9	0.45	45
Rojo (R)	3	0.15	15
Total	20	1.00	100

Fuente. - Datos obtenidos de la encuesta realizada.

Ejercicio. - Mediante el programa STATGRAPHICS, el Abriendo la ventana *Descripción* y luego escogiendo la opción *Datos cualitativos* → *Tabulación - Entrada de datos*, construir la tabla de frecuencias de nuestro ejemplo ilustrativo sobre la preferencia de los colores de las bebidas gaseosas.

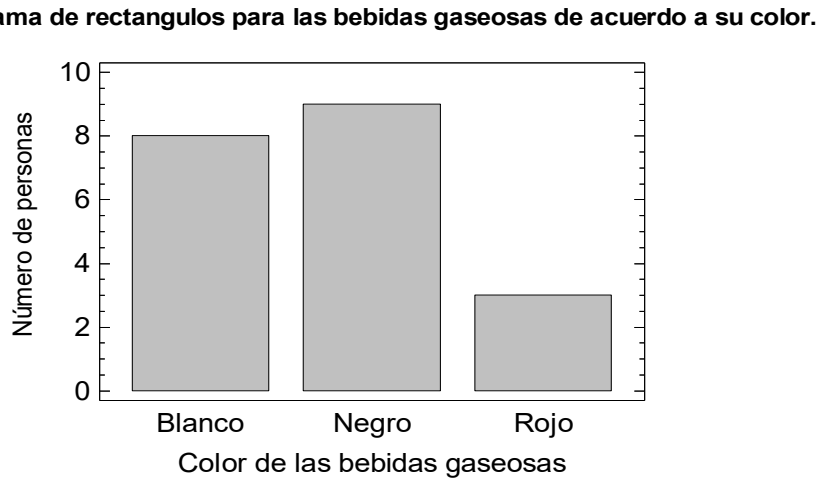
Gráficos.

Los gráficos más comunes para la distribución de frecuencias de variable cualitativa son el *de diagrama de rectángulos* y el *de sectores circulares*.

En un Diagrama de rectángulos los datos de cada una de las modalidades (caracteres cualitativos) se representa por un rectángulo vertical (u horizontal), cuya altura (o largo) es proporcional a su frecuencia (absoluta o relativa). Los rectángulos se dibujan dejando un espacio entre ellos.

Ejemplo 1.6. Abriendo la ventana *Descripción*, escogiendo *Datos cualitativos* → *Tabulación* y finalmente en el icono *Opciones gráficas* seleccionar *Diagrama de barras* para obtener el diagrama de rectángulos para los datos de la tabla 1.4.

Figura 1.6. Diagrama de rectángulos para los datos de la tabla 1.4.



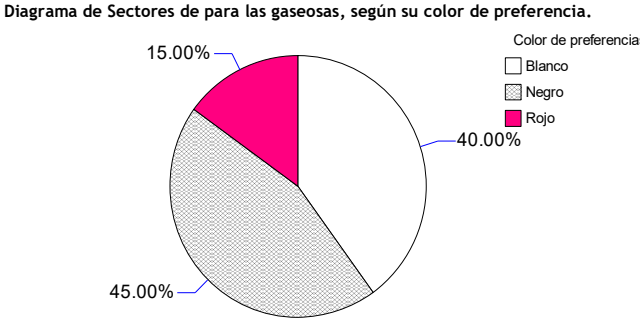
En un gráfico circular, los datos de cada categoría se representan por un sector circular. Es utilizado principalmente cuando se pretende comparar cada valor de la variable con el total. Para construir se divide el circulo en sectores, cuyas áreas serán proporcionales a los valores de la variable. Está división es obtenida a través de la regla de 3 simples.

Total $\rightarrow$ 360^o

Parte $\rightarrow$ x^o

Ejemplo 1.7. Construir el diagrama de sectores para la información contenida en la tabla 1.4. Abriendo la ventana *descripción*, escogiendo *Datos cualitativos* $\rightarrow$ *Tabulación* y finalmente en el icono *Opciones gráficas* seleccionar *Diagrama de sectores.*

Figura 1.7. Diagrama de sectores circulares para los datos de la tabla 1.4.



MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN.

2.1.- Medidas de tendencia central.

En el capítulo anterior estudiamos de que manera los datos podrían ser presentados en forma compacta, comprensible mediante tablas y gráficos. Sin embargo, con frecuencia necesitamos resumir aún más para facilitar el análisis e interpretación de la información. Cuando la variable en estudio es cuantitativa, el investigador puede estar interesado en encontrar un solo valor, que pueda caracterizar más nítidamente la naturaleza de los datos que se están midiendo.

Un valor que refleje la tendencia de los datos puede darse mediante las medidas de tendencia central o de posición. Las más importantes y muy usadas son: la media aritmética o media, la mediana, la media geométrica y la media armónica. También podemos mencionar a la moda, los cuartiles, los percentiles, etc. Estas medidas o estadígrafos son considerados como medidas de localización, puesto que señalan la localización de los valores más frecuentes o de valores extremos.

2.1.1- La Media Aritmética.

a) Datos no agrupados.

Sea x_1, x_2,..., x_n valores de la variable X. La media aritmética simple de X representada por $\overline{X}$ es dado por:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$ donde n= es el tamaño de la muestra.

b) Datos agrupados.

Sean x_1, x_2,..., x_k valores de la variable X ponderada por sus respectivas frecuencias absolutas: n_1, n_2, ..., n_k. La media aritmética de la variable X es dado por:

$$\overline{X} = \frac{\sum_{i=1}^{K} x_i\, n_i}{n}$$, donde $n = \sum_{i=1}^{k} n_i$.

Ejemplo 2.1. Considerando la información contenida en la tabla 2.1, determinar el número medio de pasajeros a bordo por autobús.

Tabla 2.1. Distribución de frecuencias de 50 parcelas de cultivo de papa, según el rendimiento en kilogramos por parcela.

$(Y'_{i-1} - Y'_i]$	Y_i	n_i	$Y_i n_i$
30 - 35	32.5	2	65
35 - 40	37.5	4	150
40 - 45	42.5	7	295
45 - 50	47.5	18	855
50 - 55	52.5	12	630
55 - 60	57.5	6	345
60 - 65	62.5	1	62.5
Total		50	2405

Luego la media aritmética de estos datos será:

$$\overline{Y} = \frac{\sum_{i=1}^{K} y_i\, n_i}{n} = \frac{2405}{50} = 48.1$$ Kilogramos por parcela.

Observación

- Si consideramos muestras de tamaño $n_1, n_2,...,n_r$ de una población, a los cuales le corresponden medias aritméticas $\overline{X}_1$, $\overline{X}_2$, ..., $\overline{X}_r$ respectivamente, entonces la media asociada a la muestra de tamaño $n_1 + n_2 +...+ n_r$ está dado por:

$$\overline{\overline{X}} = \frac{\overline{X}_1 n_1 + \overline{X}_2 n_2 + ... + \overline{X}_r n_r}{n_1 + n_2 ... + \overline{n}_r} = \frac{\sum_{i=1}^{r} \overline{X}_i\, n_i}{n}$$

donde $n = \sum_{i=1}^{r} n_i$.

- Si $p_1, p_2, ...,p_r$ son los pesos o ponderaciones asociados a los valores de la variable X: $x_1, x_2, ...,x_r$ respectivamente, entonces la media aritmética ponderada será:

Ejemplo 2.2. La empresa agroindustrial A tiene 100 empleados, con un sueldo promedio mensual por empleado de S/. 1500. La empresa agroindustrial B tiene 200 empleados con un sueldo promedio mensual de S/. 1400.

a) ¿Cuál es el sueldo promedio mensual de las dos empresas agroindustriales en conjunto?

b) si a las dos empresas agroindustriales se agrega una tercera con 50 empleados y un sueldo promedio mensual por empleado de S/. 1600, ¿Cuál es el sueldo promedio para las tres empresas en conjunto?

Solución.

a) $n_A = 100$, $\overline{X}_A = 1500$, $n_B = 200$, $\overline{X}_B = 1400$

si $\overline{X}_{A+B}$ es el promedio mensual de las dos empresas en conjunto, entonces,

$$\overline{X} = \frac{\overline{X}_A n_A + \overline{X}_B n_B}{n_A + n_B} = \frac{1500 \times 100 + 1400 \times 200}{100 + 200} = S/.1433.33$$

b) Ahora tenemos que:

$n_{A+B} = 300$ $\overline{X}_{A+B} = 1433.33$, $n_c = 50$ $\overline{X}_C = 1600$

Entonces, $\overline{X}_{(A+B+C)} = \frac{300 \times 1433.33 + 50 \times 1600}{300 + 50} = S/.1457.14$

Desventajas de la media aritmética.

1.- La media aritmética puede verse afectado por los valores extremos que no son representativos del resto de las observaciones.

2.- No se puede calcular la media aritmética en las distribuciones que tienen intervalos de clase abierto en los extremos.

2.1.2.- La Mediana.

La mediana es un valor que divide a un conjunto de observaciones ordenadas en forma ascendente o descendente en dos grupos de igual número de observaciones. La notación que vamos a emplear será: $\widetilde{X} = \text{Med}(X) = \text{mediana}$

Cálculo de la mediana

a) Datos no agrupados.

Para calcular el valor de la mediana de los datos $x_1, x_2, ..., x_n$ se tendrá en cuenta el siguiente procedimiento:

1) Se ordenan los datos en forma ascendente o descendente.

2) Si n es impar, el valor de la mediana es el valor del centro, es decir,

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

donde $\frac{n+1}{2}$ es la posición de la mediana.

3) si n es par, el valor de la mediana va a estar dado por:

$$\mathrm{Me} = \frac{X_{n/2} + X_{n/2+1}}{2}$$

Esto quiere decir, que el valor de la mediana se encuentra entre los valores cuya posición son: n/2 y (n/2+1).

Ejemplo 2.3. Las siguientes cifras son el rendimiento (en Toneladas de forraje) de sorgo forrajero de cierta variedad en 12 parcelas:

68, 55, 72, 91, 67, 83, 94, 74, 64, 105, 85, 58.

Determinar la mediana de estos importes.

Solución.

Ordenando la información en forma ascendente, tenemos:

55, 58, 64, 67, 68, 72, 74, 83, 85, 91, 94, 105.

Como el número de datos es par (n = 12), se tiene que la posición de la se localiza entre los valores centrales X_6 y X_7 , es decir, entre los valores 72 y 74. Por lo tanto el valor mediano

$$Me = \frac{72+74}{2} = 73 \text{ TM.}$$

Esto significa que el 50% de las parcelas de sorgo forrajero (es decir, 6 de ellas) tienen un rendimiento menor o igual que 73 toneladas de forraje y el 50% restante de las parcelas tienen un rendimiento mayor que 73 toneladas de forraje.

Ejemplo 2.4. En un estudio se analiza la presión sanguínea en una muestra de 9 mujeres, obteniéndose los siguientes resultados en orden:

35, 38, 40, 45, 50, 54, 55, 62 y 70.

Determine presión sanguínea mediana para este grupo de mujeres.

Solución.

En este caso, n es impar, y como los valores están ordenados en forma ascendente, el valor de la mediana se localiza en la posición (n + 1) / 2 = 10/2 = 5. Por tanto, el valor mediano es:

$$Me = X_{(5)} = 50$$

b) Datos agrupados.

En este caso el problema consiste en determinar un punto dentro del intervalo en que está comprendida la mediana.

Procedimiento:

1º. Calcular la posición de orden $\frac{n}{2}$.

2º. Por las frecuencias acumuladas se identifica la clase que contiene a la mediana, esto es, la clase para el cual se cumple:

$$N_{j-1} \leq \frac{n}{2} \leq N_j,$$

Con lo cual la mediana estará en la clase que tiene como frecuencia acumulada N_i.

3º. Utilizar la formula:

$$\tilde{X} = X'_{j-1} + c \cdot \left[\frac{\frac{n}{2} - N_{j-1}}{N_j - N_{j-1}} \right]$$

donde:

X'_{j-1} = límite inferior de la clase que contiene a la mediana.

n = tamaño de la muestra.

c = amplitud de la clase que contiene a la mediana.

N_j = frecuencia acumulada de la clase que contiene a la mediana.

N_{j-1} = frecuencia acumulada de la clase inmediatamente anterior a la clase que contiene a la mediana.

Ejemplo 2.5. A partir de la distribución de frecuencia de las 50 parcelas de papa, según su rendimiento en kilogramos (ver tabla 2.1), se pide determinar el rendimiento mediano por parcelas.

Solución.

$(Y'_{i-1} - Y'_i]$	Y_i	n_i	N_i
30 - 35	32.5	2	2
35 - 40	37.5	4	6
40 - 45	42.5	7	13
45 - 50	47.5	18	31
50 - 55	52.5	12	43
55 - 60	57.5	6	49
60 - 65	62.5	1	50
Total		50	

1º. $\frac{n}{2} = \frac{50}{2} = 25^{vo}$ posición

2º. Se identifica la clase que contiene a la mediana por la frecuencia acumulada, esto es, a través de la desigualdad:

$$N_3 = 13 < 25^{vo} < N_4 = 31$$

En este caso, la clase que contiene a la mediana es el cuarto.

3º. Reemplazar los datos en la formula obtenemos:

$$\tilde{X} = 45 + 5\left[\frac{25 - 13}{31 - 13}\right] = 48.33$$ Kilogramos.

Interpretación. - Este valor mediano significa, que el 50% de las parcelas de papa tuvieron un rendimiento menor o igual que 48.33 kilogramos, en tanto que el otro 50% tuvieron un rendimiento mayor que 48.33 kilogramos.

Ventajas de la mediana.

1. Algunas veces es un valor más representativo de un conjunto de datos que otros promedios (por ejemplo, que la media aritmética), gracias a su independencia, a sus valores extremos.

2. La mediana se puede calcular aún cuando los intervalos de clase de la distribución de frecuencias son de límites abiertos.

2.1.3.- Los Cuartiles.

Son valores que dividen a un conjunto de datos ordenados en forma ascendente o descendente en cuatro partes iguales (figura 2.1).

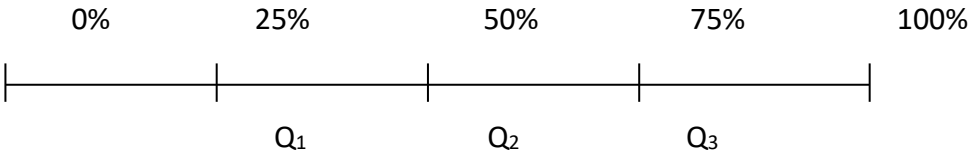


Figura 2.1

Q_1 = 1er cuartil, deja 25% de las observaciones menores o iguales a él y el 75% superiores a él.

Q_2 = 2do cuartil, coincide con la mediana.

Q_3 = 3er cuartil, deja 75% de las observaciones inferiores o iguales a él y el 25% de éstas superiores a él.

Es importante notar que entre dos cuartiles consecutivos se halla el 25% del número de valores.

Las fórmulas para calcular los cuartiles se derivan de la fórmula utilizada para calcular la mediana y los pasos para el cálculo son los mismos:

Procedimiento.

1º . Se calcula $\frac{r\,n}{4}$, para r = 1, 2, 3.

2º . Se identifica la clase que contiene a Q_r por medio de las frecuencias acumuladas, esto es, por la desigualdad:

$$N_{j-1} \leq \frac{r\,n}{4} \leq N_j$$

3º . Se aplica la fórmula:

$$Q_r = X'_{j-1} + c\cdot\left[\frac{\frac{r\,n}{4} - N_{j-1}}{N_j - N_{j-1}}\right] \quad r = 1, 2, 3.$$

Donde:

X'_{j-1} = límite inferior de la clase que contiene al cuartil Q_r.

n = tamaño de la muestra.

c = amplitud de la clase que contiene a Q_r.

N_j = frecuencia acumulada de la clase que contiene a Q_r.

N_{j-1} = frecuencia acumulada de la clase inmediatamente anterior a la clase que contiene a Q_r.

Ejemplo 2.6. Dada la siguiente distribución, determinar los cuartiles Q_1 y Q_3.

$(Y'_{i-1} - Y'_i]$	Y_i	n_i	N_i
30 - 35	32.5	2	2
35 - 40	37.5	4	6
40 - 45	42.5	7	13
45 - 50	47.5	18	31
50 - 55	52.5	12	43
55 - 60	57.5	6	49
60 - 65	62.5	1	50
Total		50	

Solución. -

1º. $\frac{n}{4} = \frac{50}{4} = 12.5^{vo}$; $\frac{3n}{4} = \frac{150}{4} = 37.5^{vo}$

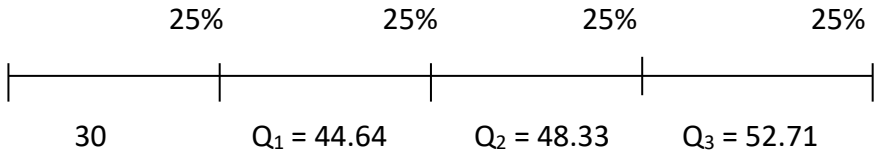
2º. Por las frecuencias acumuladas identificamos las clases que contienen a Q_1 y Q_3.Como $N_2 = 6 < \frac{n}{4} = 12.5 < 13 = N_3$, entonces el intervalo de clases que contiene a Q_1 es (40 - 45] y como $N_4 = 31 < \frac{3n}{4} = 37.5 < 43 = N_5$, entonces el intervalo de clases que contiene a Q_3 es (50 - 55].

3º. Usando las fórmulas dada para calcular Q_1 y Q_3 tenemos:

$$Q_1 = X'_{j-1} + c\cdot\left[\frac{\frac{n}{4} - N_2}{N_3 - N_2}\right] = 40 + 5\cdot\left[\frac{12.5 - 6}{13 - 6}\right] = 44.64 \text{ kilogramos.}$$

$$Q_3 = X'_{j-1} + c\cdot\left[\frac{\frac{3n}{4} - N_4}{N_5 - N_4}\right] = 50 + 5\cdot\left[\frac{37.5 - 31}{43 - 31}\right] = 52.71 \text{ kilogramos.}$$

De acuerdo a estos resultados, podemos afirmar que, en está distribución tenemos:



2.1.4. Percentiles.

Son valores que dividen la muestra ordenada en forma ascendente o descendente en 100 partes iguales.

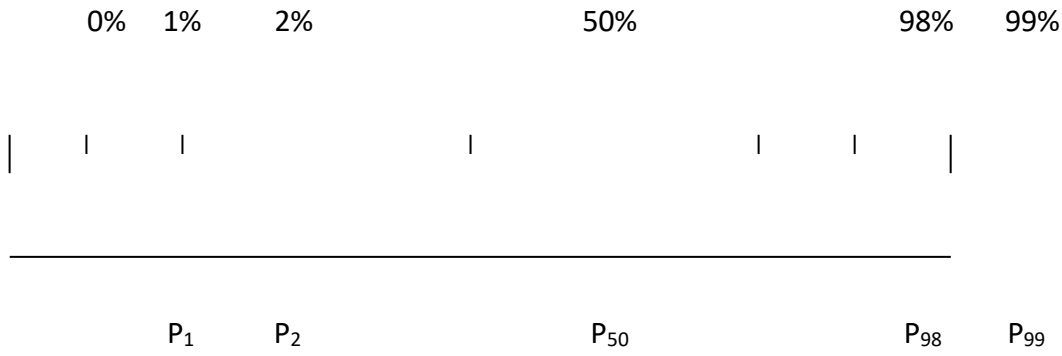


Figura 2.2

P_1 = 1er percentil, deja 1% de las observaciones menores o iguales a él y el 99% superiores a él.

P_{99} = 99vo percentil, deja 99% de las observaciones menores o iguales a él y el 1% superiores a él.

Las fórmulas para determinar los percentiles, son parecidos a los cuartiles, así:

$$P_r = X'_{j-1} + c \times \left[\frac{\frac{r\,n}{100} - N_{j-1}}{N_j - N_{j-1}} \right] \quad \forall \; r = 1, 2, \ldots, 99$$

donde:

X'_{j-1} = límite inferior de la clase que contiene a P_r , r = 1, 2, ..., 99.

n = tamaño de la muestra.

c = amplitud de la clase que contiene a P_r.

N_j = frecuencia acumulada de la clase que contiene a P_r.

N_{j-1} = frecuencia acumulada de la clase anterior a la clase que contiene a P_r.

Ejercicio. Determinar los valores de los puntos percentiles 10 y 90, para los datos del número de pasajeros a bordo, de la tabla 2.1 (página 16).

2.1.4. La Moda. -

La moda denotada por $\overset{\mu}{X} = M_0$, es un valor de la variable que tiene la más alta frecuencia, esto es, es el valor más frecuente en un conjunto de datos. La moda puede no existe, incluso si existe puede no ser única.

Ejemplo 2.7. En mes, 8 vendedores de artículos electrónicos vendieron los siguientes números de aparatos: 8, 11, 5, 14, 8, 11, 16 y 11. Considerando a este mes como a la población estadística que interesa, el número modal de unidades vendidas es:

$$\overset{\mu}{X} = M_0 = ...$$

Ejemplo 2.8. Considere la distribución de los pesos (en kilos) de 15 adultos:

63, 67, 70, 69, 81, 57, 63, 73, 68, 63, 71, 71, 71, 83.

La moda de estas observaciones es:

$$\overset{\mu}{X} = ...\text{kilos} \quad \text{y} \quad \overset{\mu}{X} = ...\text{kilos}$$

En este caso la distribución se llamará bimodal.

Ejemplo 2.9. Las notas promociónales de 7 alumnos son: 15, 16, 17, 14, 13, 12, 11.

Estos datos No tienen moda. Las distribuciones de este tipo se llaman uniforme.

En general, se tiene lo siguiente:

i) La distribución que tiene una sola moda se llama *unimodal*.

ii) La distribución que tiene dos modas se llama *bimodal*.

iii) La distribución que tiene más de dos modas se llama *multimodal*.

Cálculo de la Moda para datos agrupados.

Para datos agrupados en intervalos de clase, aplicaremos el siguiente procedimiento para el cálculo de la moda.

1º. Se identifica la clase modal (la clase con mayor frecuencia).

2º. Se aplica la fórmula:

$$\overset{\mu}{X} = X'_{j-1} + c\cdot\left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right]$$

Donde:

X'_{j-1} = límite inferior de la clase modal.

n_{Mo} = frecuencia de la clase modal.

$\Delta_1 = n_{Mo} - n_1$ (n_1 = frecuencia de la clase inmediatamente anterior a la clase modal).

$\Delta_2 = n_{Mo} - n_2$ (n_2 = frecuencia de la clase inmediatamente posterior a la clase modal).

c = amplitud de la clase que contiene a la mediana.

Ejemplo 2.10. Determinar la moda para la siguiente distribución (tabla 2.1).

$(Y'_{i-1} - Y'_i]$	Y_i	n_i
30 - 35	32.5	2
35 - 40	37.5	4
40 - 45	42.5	7
45 - 50	47.5	18
50 - 55	52.5	12
55 - 60	57.5	6
60 - 65	62.5	1
Total		50

Solución.

1º. El intervalo de clase de mayor frecuencia absoluta (18) es el cuarto intervalo: $(45 - 50]$

$\Delta_1 = 18 - 7 = 11$

$\Delta_2 = 18 - 12 = 6$

c = 5.

2º. Aplicando la formula tenemos:

$$\overset{\mu}{X} = 45 + 5\left[\frac{11}{11 + 6}\right] = 48.235 \text{ kilogramos.}$$

Este valor modal significa que el rendimiento más frecuente en las parcelas de cultivo de papa es de 48 kilogramos; o también que la mayoría de las parcelas tienen un rendimiento igual a 48 kilogramos.

2.2. Relación entre la Media, Mediana y Moda.

a) Distribuciones simétricas.

Se dice que una distribución de frecuencia es simétrica cuando valores de la variable equidistantes de un valor central tienen las mismas frecuencias. Es importante destacar en este caso que:

$$\overline{X} = \widetilde{X} = \overset{\mu}{X}$$

b) Para una distribución sesgada hacia la derecha (si la cola mayor se presenta a la derecha de la distribución) se tiene que:

$$\overline{X} > \widetilde{X} > \overset{\mu}{X}$$

Para una distribución sesgada a la izquierda tenemos:

$$\overline{X} < \widetilde{X} < \overset{\mu}{X}$$

Ejercicio. Considerando la distribución de la tabla 2.1 (página 16), se pide analizar la asimetría de la distribución.

2.3. Medida de dispersión.

Todos los valores representativos discutidos en las secciones precedentes han sido una especie de promedio o medida de posición. Sin embargo, el uso de un solo valor para describir una distribución oculta muchos fenómenos importantes. Por ejemplo, dos grupos separados de datos pueden contener la misma media, pero un grupo puede estar más disperso o esparcido alrededor del valor promedio que el otro.

Por lo que es necesario una medida de la dispersión, esparcimiento o variación para ayudar más completamente la distribución. Mientras menor es la dispersión, más típico es el valor de la media para toda la distribución.

Las medidas de dispersión que se utilizan con mayor frecuencia son: la varianza, la desviación estándar y el coeficiente de variación.

2.3.1. La varianza. -

Definición 1. (Para datos no agrupados). La varianza de una muestra $x_1, x_2, ..., x_n$ de la variable X, es dado por:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

Definición 2. (Para datos agrupados). La varianza de los valores una muestra y_1, y_2, ..., y_k de Y con frecuencias absolutas n_1, n_2, ..., n_k respectivamente, es dado por:

$$V(Y) = S^2 = \frac{\sum_{i=1}^{k}(y_i - \bar{y})^2 \times n_i}{n-1}$$

Observaciones.

1. La varianza poblacional se define en términos de la media poblacional μ , esto es:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \mu)^2}{N}$$

donde N es el tamaño de la población.

2.- Con la finalidad de reducir el volumen de operaciones en el cálculo de la varianza, podemos usar las fórmulas:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$ para datos no agrupados.

y

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 n_i - \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{n}\right]$$ para datos agrupados.

2.3.2.- La desviación estándar. -

Definición. - La desviación estándar o típica de los valores de la variable X se define como la raíz cuadrada positiva de la varianza, y se denota por $S = \sqrt{V(X)}$.

El valor numérico de S cuantifica el grado de dispersión de los valores de una variable con respecto a su media. Mientras mayor es la dispersión de las observaciones, mayor es la magnitud de sus desviaciones respecto a la media y por ende, más alto el valor numérico de la desviación estándar.

Ejemplo 2.11. Se determinó los pesos de una muestra de cartas procesadas en una oficina postal, pesadas hasta el gramo más próximo, son: 21, 18, 30, 12, 14, 17, 28, 10, 16 y 25.

Determine a) la varianza y b) la desviación estándar.

Solución. -

a) Cálculo de la varianza.

i) Aplicaremos la fórmula de definición de la varianza para datos no agrupados, donde $\overline{x}$ es la media aritmética del conjunto de observaciones.

En efecto, tenemos que la media aritmética es:

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{21 + 18 + \cdots + 25}{10} = \frac{191}{10} = 19.1 \text{ gramos.}$$

Luego aplicando fórmula para varianza:

$$S^2 = \frac{\sum_{i=1}^{10} (x_i - \overline{x})^2}{10-1} = \frac{(21-19.1)^2 + (18-19.1)^2 + \cdots + (25-19.1)^2}{9} = 45.6555 \quad \text{gramos}$$

al cuadrado.

ii) aplicando el método abreviado, tenemos:

$$\sum_{i=1}^{10} x_i = 21+18+\cdots+25 = 191 \quad \text{y} \quad \sum_{i=1}^{10} x_i^2 = 21^2 + 18^2 + \cdots + 25^2 = 4059$$

luego:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{9}\left[4059 - \frac{(191)^2}{10}\right] = 45.6555$$

b) cálculo de la desviación estándar.

La desviación estándar será la raíz cuadrada del valor obtenido en la varianza, es decir:

$$S = \sqrt{45.6555} = 6.7569 \text{ gramos.}$$

Ejemplo 2.12. Determine la desviación estándar de la siguiente distribución de frecuencias (ver tabla 2.1).

$(Y'_{i-1} - Y'_i]$	Y_i	n_i	$(y_i - \bar{y})^2$	$\frac{(y_i - \bar{y})^2}{n_i}$
30 - 35	32.5	2	243.36	486.72
35 - 40	37.5	4	112.36	449.44
40 - 45	42.5	7	31.36	219.52
45 - 50	47.5	18	0.36	6.48
50 - 55	52.5	12	19.36	232.32
55 - 60	57.5	6	88.36	530.16
60 - 65	62.5	1	207.36	207.36
Total		50		2132

Solución.

a) Aplicando la fórmula de la definición.

Se sabe que la media aritmética de esta distribución es:

$\overline{Y} = 48.1$ kilogramos

Luego aplicando la fórmula de definición la varianza, tenemos:

$$S^2 = \frac{\sum_{i=1}^{k} n_i (y_i - \bar{y})^2}{n-1} = \frac{2132}{49} = 43.5102$$ kilogramos al cuadrado.

Por tanto, la desviación estándar es: S = 6.596226 kilogramos.

b) Aplicando el método abreviado.

$(Y'_{i-1} - Y'_i]$	Y_i	n_i	$y_i n_i$	$y_i^2 n_i$
30 - 35	32.5	2	65	2112.5
35 - 40	37.5	4	150	5625
40 - 45	42.5	7	297.5	12643.75
45 - 50	47.5	18	855	40612.5
50 - 55	52.5	12	630	33075
55 - 60	57.5	6	345	19837.5
60 - 65	62.5	1	62.5	3906.25
Total		50	2405	117812.5

Luego tenemos:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 n_i - \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{n}\right] = \frac{1}{49}\left[117812.5 - \frac{(2405)^2}{50}\right] = 43.5102$$

Por tanto, la desviación estándar es: S = 6.59622 kilogramos.

2.3.3.- Medidas de Dispersión relativa.

Las medidas de dispersión que se han descrito en la sección anterior se expresan en las unidades de medidas originales, tales como altura, salarios, kilogramos, puntajes, etc. Se pueden utilizar para comparar la variación de dos distribuciones siempre que las variables se expresen en las mismas unidades de medida y sean aproximadamente del mismo tamaño promedio. Sin embargo, a veces es necesario comparar dos o más conjuntos de datos expresados en unidades diferentes. Es esta situación hay que utilizar una medida relativa de dispersión que sería el coeficiente de variación.

Definición. El coeficiente de dispersión, es una medida de dispersión relativa de la desviación estándar con respecto a la media, es decir:

Coeficiente de variación = $C.V = \frac{S}{\bar{x}}$

Generalmente el coeficiente de variación se expresa como un porcentaje. El C.V es un número puro independiente de la unidad de medición.

Observación:

- Si el C.V es menor del 10% se dice que hay poca dispersión;
- Si el C.V oscila entre el 10% y el 33% la dispersión existente es aceptable;
- Si el C.V oscila entre el 33% y el 50% se dice que hay alta dispersión;
- Pero si el C.V es mayor del 50% se dice que la dispersión es muy alta.

Ejemplo 2.13. Calcular la dispersión relativa para los datos de la tabla 2.1 (página 18).

En efecto tenemos que $\overline{Y} = 48.1$ y S = 6.596226, entonces el C.V es:

$$C.V = \frac{6.596226}{48.1} \times 100 = 13.71\%$$

Como el valor de C.V = 13.71% se encuentra entre 10% y 33%, indica que la dispersión es aceptable.

Ejemplo 2.14. Para 2 acciones comunes de empresas de la industria electrónica, el precio promedio de cierre en el mercado de valores durante un mes fue, para la acción A, de $1500, con una desviación estándar de $500. Para la acción B, el precio promedio fue de $5000, con una desviación estándar de $300.

a) Efectuar Hacer una comparación de la variabilidad absoluta de las acciones.

b) Realizar una comparación de la variabilidad con respecto al nivel medio de los precios.

Solución. -

a) Al hacer una comparación absoluta, resultó ser superior la variabilidad en el precio de la acción A debido a que muestra una mayor desviación estándar.

b) Con respecto a la variabilidad relativa, deben compararse los respectivos coeficientes de variación:

$$C.V(A) = \frac{500}{1500} = 0.33 \quad \text{y} \quad C.V(B) = \frac{300}{5000} = 0.060$$

Por ello, puede concluirse que el precio de la acción A ha sido casi 5 veces más variable que el precio de la acción B (con respecto al precio promedio para cada una de las dos).

EJERCICIOS DE AUTOEXAMEN Nº 01

1. *Los datos que se presentan a continuación, son el resultado de una encuesta: Se pide:*

ENCUESTA A 50 DISTINTOS PROFESIONALES QUE LABORAN EN EL DEPARTAMENTO DE TUMBES (MARZO 2010)

ORDEN	PROFESIÓN	INGRESOS MENSUALES (S/.)	CENTRO DE TRABAJO	HORAS DIARIAS	EDAD (años)	ESTADO CIVIL	NÚMERO DE HIJOS
1	ING. PESQUERO	1500	EMP. PRIVADA	12	32	SOLTERO	0
2	DOCENTE	300	UNIVERSIDAD	4	35	SOLTERO	0
3	OBSTETRA	850	HOSPITAL	8	27	CASADO	1
4	MÉDICO	2010	CLINICA	8	35	SOLTERO	1
5	MÉDICO	1800	CLINICA	8	40	CASADO	5
6	DOCENTE	320	UNIVERSIDAD	4	48	CASADO	3
7	ENFERMERA	700	HOSPITAL	12	35	CASADO	2
8	ING. PESQUERO	1000	EMP. PRIVADA	8	38	DIVORC.	3
9	DOCENTE	850	UNIVERSIDAD	5	35	CASADO	2
10	ENFERMERA	1100	HOSPITAL	8	38	CASADO	2
11	DOCENTE	435	UNIVERSIDAD	12	47	VIUDO	4
12	ABOGADO	2000	EMP. PRIVADA	8	49	SOLTERO	0
13	MÉDICO	1900	CLINICA	10	50	CASADO	3
14	ING. PESQUERO	900	EMP. PRIVADA	6	40	CASADO	6
15	ENFERMERA	600	HOSPITAL	6	25	SOLTERO	0
16	ENFERMERA	600	HOSPITAL	6	25	SOLTERO	0
17	ING. PESQUERO	500	EMP. PRIVADA	8	39	CASADO	6
18	ENFERMERA	1000	HOSPITAL	5	31	DIVORC.	1

19	ABOGADO	2100	EMP. PRIVADA	6	50	DIVORC.	8
20	MÉDICO	1100	HOSPITAL	5	49	CASADO	2
21	ING. PESQUERO	1500	EMP. PRIVADA	8	35	SOLTERO	0
22	MÉDICO	1900	HOSPITAL	8	44	CASADO	3
23	DOCENTE	1300	UNIVERSIDAD	10	48	CASADO	3
24	ENFERMERA	750	HOSPITAL	12	39	VIUDO	2
25	OBSTETRA	900	HOSPITAL	10	28	SOLTERO	0
26	MÉDICO	2050	CLINICA	12	47	CASADO	2
27	MÉDICO	1000	CLINICA	12	53	CASADO	4
28	DOCENTE	1100	UNIVERSIDAD	8	49	SOLTERO	2
29	ENFERMERA	1500	HOSPITAL	12	42	SOLTERO	1
30	MÉDICO	1500	HOSPITAL	6	45	CASADO	3
31	MÉDICO	1000	CLINICA	8	47	DIVORC.	5
32	ENFERMERA	840	HOSPITAL	8	48	DIVORC.	5
33	DOCENTE	1100	UNIVERSIDAD	8	47	CASADO	2
34	ENFERMERA	950	HOSPITAL	12	40	SOLTERO	1
35	ABOGADO	980	EMP. PRIVADA	6	55	CASADO	3
36	OBSTETRA	1100	CLINICA	10	32	CASADO	2
37	ENFERMERA	500	HOSPITAL	10	42	VIUDO	4
38	ENFERMERA	1200	HOSPITAL	10	47	SOLTERO	1
39	ENFERMERA	1059	HOSPITAL	6	33	CASADO	3
40	ING. PESQUERO	1800	EMP. PRIVADA	8	50	CASADO	7
41	MÉDICO	2000	CLINICA	10	50	SOLTERO	2

42	OBSTETRA	950	HOSPITAL	8	28	SOLTERO	0
43	DOCENTE	1250	UNIVERSIDAD	2	39	CASADO	2
44	ENFERMERA	750	HOSPITAL	12	38	DIVORC.	3
45	ABOGADO	1100	EMP. PRIVADA	8	32	DIVORC.	1
46	ING. PESQUERO	800	EMP. PRIVADA	6	44	CASADO	5
47	DOCENTE	1000	UNIVERSIDAD	6	28	SOLTERO	0
48	DOCENTE	1100	UNIVERSIDAD	6	24	SOLTERO	0
49	ENFERMERA	600	EMP. PRIVADA	6	33	CASADO	1
50	OBSTETRA	1000	HOSPITAL	8	30	CASADO	0

a) Identificar las variables y a que tipo pertenecen.
b) Representar los datos en distribuciones de frecuencias para datos no agrupados. Considerar las frecuencias. Absolutas simples (f_i), absolutas acumuladas (F_i), relativas simples (hi) y relativas acumuladas (Hi).
c) Interpretar que representan las frecuencias de la tercera fila, para cada variable.
d) Represente las distribuciones de frecuencia de datos agrupados para las variables que lo requieran. Considerar el número de intervalos y la amplitud del intervalo valor entero, utilice para ello la fórmula de Sturges.
e) Interpretar que representan las frecuencias de la cuarta fila, para cada variable.
f) Cuáles serían los intervalos si se utiliza otro valor de m cercano al utilizado en el caso anterior (pruebe para un valor de m menor en 1 y otro mayor en 1). ¿A su criterio cuál representa mejor a los datos?

2. El servicio de atención al paciente comprueba que el número de enfermos que acuden al servicio de urgencias de un hospital durante 100 días elegidos al azar es el siguiente:

87	80	109	91	83	70	91	93	86	95
71	86	98	104	107	89	109	16	102	83
86	107	85	115	103	69	74	70	86	92
97	118	102	122	97	78	91	108	101	85
100	80	93	90	90	79	93	102	107	107
83	103	99	87	92	107	80	90	72	100
118	81	115	88	70	87	99	105	99	103
99	98	88	90	89	73	93	94	82	80
84	115	87	91	113	113	108	75	76	85
115	103	88	99	92	93	71	89	112	119

Se pide:

a) Tabular la variable número de enfermos ingresados. Utilice la fórmula de Sturges para encontrar el número de intervalos.

b) Calcular e interpretar las frecuencias porcentuales simples y acumuladas del cuarto intervalo.

3. Se desea estudiar las alturas de un grupo de 20 alumnos. Realizar el estudio:

a) Con los datos sin agrupar.

b) Con los datos agrupados en intervalos de amplitud 10 cm.

Las alturas fueron expresadas en cm: 162-166-168-170-172-174-180-164-166-168-168-*172-178-182-164-166-168-170-176-188.*

Se pesaron a 50 alumnos de la Facultad de Ingeniería Civil y se obtuvieron los siguientes datos:

50.5	67.2	63.5	68.3	70.2	72.5	57.5	57.8	66.1	72.5
61.2	62.3	62.0	67.4	64.3	60.3	69.3	68.2	69.2	63.3
70.0	68.0	65.5	51.3	68.7	57.3	71.3	67.2	56.2	71.3
73.6	54.6	74.3	53.8	69.2	63.4	51.0	72.0	72.3	59.3
64.3	60.5	65.3	66.6	73.5	72.5	75.0	65.0	68.0	65.2

Se pide:

a) Construir la tabla de frecuencias para la variable peso, utilizando una amplitud entera.

b) Construir la tabla de frecuencias para la variable peso, utilizando una amplitud redondeada a 1 decimal.

c) Interpretar las frecuencias simples y acumuladas de la 1º y 2º tabla de frecuencias para el tercer intervalo.

5. Los tiempos de servicio en minutos obtenido de un servicentro de vehículos correspondiente a 45 clientes que fueron atendidos son:

58.2	73.5	46.8	65.1	37.4	74.2	63.7	73.6	538	56
69.5	48.6	64.5	79.4	25.5	72.6	50.6	53.4	59.7	68.2
43.2	81.9	44.1	69.3	69.8	43.2	50.6	53.4	59.7	68.2
77.6	61.8	70.7	80.8	74.8	77.6	61.8	70.7	80.8	41.4
60.5	46.7	53.6	40.3	42.4	60.5	46.7	53.6	40.3	42.4
55.2	63.8	66.3	47.7	71.4	68.6	85.7	72	57.8	68.6

a) Construir una distribución de frecuenta que tenga 7 intervalos y una amplitud entera.

b) Calcule e interprete las frecuencias porcentuales simples y acumuladas de la tercera fila.

6. Una muestra de los tiempos, en minutos, de llamadas telefónicas que recibe una secretaria son:

3	6	2	4	3	1	4	5	2	4	3	1	2	4	4
1	3	7	3	2	1	3	1	2	3	4	5	2	7	2
6	2	1	3	6	4	5	3	2	1	3	4	3	4	3
5	3	4	3	5	2	1	3	5	2	4	4	1	3	4

a) Organizar los datos en una distribución de frecuencias para datos no agrupados.

b) Interprete las frecuencias acumuladas simples, los porcentajes simples y acumulados de la tercera clase.

7. Considerando la amplitud del intervalo entera, construya los intervalos para la variable pago por consumo telefónico mensual en los siguientes casos:

a) Xmin = S/. 49,9 X max = S/. 385,8 intervalos= 6 (amplitud redondeada a décimos)

b) Xmin = S/. 153,55 X max = S/. 595,75 intervalos= 8 (amplitud redondeada a centésimos)

8. Calcular los datos que faltan en la siguiente tabla:

$L_{i-1} - L_i$	fi	Hi	Fi
0 – 10	60	f1	60
10 – 20	n2	0,4	N2
20 – 30	30	f3	170
30 – 100	n4	0,1	N4
100 – 200	n5	f5	200

9. La distribución de edades de cien personas se había agrupado en cuatro intervalos recordándose solo los siguientes datos de la distribución:

i) El primer intervalo tiene 6 años como extremo superior, una frecuencia relativa de 0.2 y una amplitud de 4 años.

ii) En el segundo intervalo se acumulan 60 personas.

iii) Las marcas de clase del segundo y cuarto intervalos son 8 y 50 años respectivamente.

iv) El tercer intervalo presenta una frecuencia de 30 personas y una amplitud de 30 años.

Con esta información se pide:

a) Reconstruya la distribución de frecuencias

b) Interprete el 2 intervalo con sus respectivas frecuencias simples y acumuladas.

10. El aplazamiento en el cobro de las últimas cien ventas facturadas por un establecimiento se ha agrupado en cuatro intervalos, recordándose sólo los siguientes datos de la distribución:

a) El primer intervalo tiene 6 semanas como extremo superior, una frecuencia relativa de 0,2 y una amplitud de 4 semanas.

b) En el segundo intervalo se acumulan 60 ventas.

c) Las marcas de clase del segundo y cuarto intervalos son 8 y 50 semanas, respectivamente.

d) El tercer intervalo presenta una frecuencia de 30 ventas y una amplitud de 30 semanas.

Con esta información reconstruya la distribución de frecuencias e interprete las frecuencias simples y acumuladas del 3 intervalo.

EJERCICIOS DE AUTO EXAMEN Nº 02

1. A 50 clientes de una institución financiera se les ha preguntado por el tiempo que han tenido que esperar en la cola de la caja para realizar una gestión. Sus respuestas se han organizado en la siguiente tabla.

Tiempo	Clientes
0 – 5	20
5 – 10	10
10 – 15	15
15 – 20	3
20 – 25	2

a) Calcular las frecuencias absolutas y relativas, así como la marca de clase

b) Interprete los valores de las frecuencias del cuarto intervalo.

c) Represente los datos tabulados a través del histograma de frecuencias, el polígono y la ojiva menor que.

2. En la tabla siguiente tienes los datos del número de seguidores de diferentes religiones en el mundo, según una estimación de www.adherents.com

a) Calcular las frecuencias absolutas y relativas, así como la marca de clase

b) Represente los datos tabulados utilizando un tipo de gráfico que sea más conveniente.

c) Interprete los datos que se presentan en el gráfico.

Religión	Millones de personas
Cristianismo	2 000
Islam	1 300
Hinduismo	900
Ateo – agnóstico – sin religión	850
Budismo	360
Confucianismo/Taoismo	225
Animismo y religiones tradicionales africanas	245
Otras religiones	93

3. Se quiere realizar un estudio acerca del número de habitaciones que tienen los hoteles de la ciudad de Lima. Para ello se han obtenido los datos que figuran en la tabla adjunta.

Nº de Habitaciones	Nº de Hoteles
0 – 50	6
50 – 100	8
100 – 150	4
150 – 200	3
Más de 200	2

a) Calcular las frecuencias absolutas y relativas, así como la marca de clase

b) Interprete los valores de las frecuencias del cuarto intervalo.

c) Represente los datos tabulados a través del histograma de frecuencias, el polígono.

4. Los datos que se muestran a continuación representan las unidades producidas por un empleado durante 20 días de trabajo.

160	170	181	156	176
148	198	179	162	150
162	156	179	178	151
157	154	179	148	156

Resuma estos datos mediante:

a. Una distribución de frecuencias absolutas y relativas (simples y acumuladas).

b. Un histograma, un polígono de frecuencias, una ojiva menor que.

5. Prepare un gráfico con utilizando los datos que se encuentran en la Página Web del INEI (www.inei.gob.pe) acerca de la inflación, la población del Perú, la población de Tumbes, alumnos titulados por universidades públicas y privadas, y población en edad para trabajar.

INTERVALO DE CONFIANZA

1. INTERVALOS DE CONFIANZA (IC)

Los estimadores puntuales usados son:

Estadístico muestral Parámetro poblacional

$\bar{X} \longrightarrow \mu$

$\bar{P} \longrightarrow P$

$S \longrightarrow \sigma^2$

La media muestral es una estimación puntual de la media poblacional. *p* es una puntual de la proporción poblacional y s, la desviación estándar muestral, es una estimación puntual de σ, la desviación estándar poblacional.

El intervalo de confianza Es el conjunto de valores obtenido a partir de los datos muestrales en el que hay una determinada probabilidad de que se encuentre el parámetro poblacional. Esta probabilidad se le conoce como el nivel de confianza.

$$IC = (P_1 - P_2) \pm Z(1-\propto) \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_1(1-P_2)}{n_2}}}$$

La distribución muestral de X, permite localizar un intervalo que tenga una determinada probabilidad de contener a la media poblacional.

Si el tamaño de la muestra es razonablemente grande, el teorema del límite central permite establecer lo siguiente:

- El noventa y cinco por ciento de las medias muéstrales obtenidas de una. población se encuentran a no más de 1.96 desviaciones estándar de la media poblacional $\boldsymbol{\mu}$.
- El noventa y nueve por ciento de las medias poblacionales se encuentran a no más de 2.58 desviaciones estándar de la media poblacional.

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina nivel de confianza, y se denota $\mathbf{1} - \boldsymbol{\alpha}$. La probabilidad de equivocarnos se llama nivel de significancia y se simboliza con $\boldsymbol{\alpha}$. Generalmente se construyen intervalos con confianza $\mathbf{1} - \boldsymbol{\alpha} = \mathbf{95}\%$ (o significancia =5%). Menos frecuentes son los intervalos con $\boldsymbol{\alpha} = \mathbf{10}\%$ $\boldsymbol{o}$ $\boldsymbol{\alpha} = \mathbf{1}\%$.

1.1 Distribución normal

$\boldsymbol{P(1.96 < Z < 1.96) = 0.95}$, lo que indica que si una variable tiene distribución normal $\boldsymbol{N(\mu;\sigma)}$, entonces el 95% de las muestras.

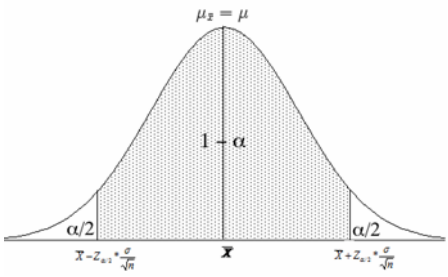
cumplen que $-\mathbf{1.96} \leq \frac{\overline{\boldsymbol{X}}-\boldsymbol{\mu}}{\boldsymbol{\sigma}_{\overline{\boldsymbol{X}}}} \leq \mathbf{1.96}$, donde $\boldsymbol{\sigma}_{\overline{x}} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$, error estándar del muestreo

Despejando $\overline{\boldsymbol{X}} - \mathbf{1.96} * \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} \leq \boldsymbol{\mu} \leq \overline{\boldsymbol{X}} + \mathbf{1.96} * \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$ con una probabilidad de $\mathbf{1} - \boldsymbol{\alpha}$

Esto quiere decir que de 100 muestras de tamaño n, 95 contienen la media poblacional , cuando la variable X es normal y se conoce $\boldsymbol{\sigma}$.

1.2 Estimación para la media poblacional. Muestras grandes.

Cuando el tamaño de la muestra es grande o la variable tiene distribución normal, el intervalo de confianza está dado por:



$$\overline{X} - Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}}$$

$$P\left[\overline{X} - Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Como el error estándar está afectado por el tamaño de la muestra, este a su vez afecta el intervalo de confianza. Conforme aumenta el tamaño de la muestra, el error estándar disminuye, indicando esto que hay menos variabilidad en la distribución muestral de a media muestral. La estimación obtenida de una muestra grande será más precisa que una estimación obtenida de una muestra pequeña.

Cuando el tamaño de la muestra n es mayor o igual a 30, se aplica el teorema de límite central asegura que la media muestral sigue la distribución normal. Si la media muestral tiene una distribución normal, se puede usar la distribución normal estándar, es decir, z, para los cálculos.

Ejemplo. En una encuesta aplicada a 1600 colombiano se encontró que en promedio ven 14.6 horas de televisión a la semana con una desviación estándar de 5 horas. Realice una estimación de la variable con para un nivel de confianza del 80%.

Datos

N=1600

$\overline{X} = 14.6$

$\sigma = 5h$

$1 - \alpha = 0.80$

$z_{0.80} = 1.28$

$$\boldsymbol{P\left[14.6 - 1.28 * \frac{5}{\sqrt{1600}} \leq \mu \leq 14.6 + 1.28 * \frac{5}{\sqrt{1600}}\right] = 0.80}$$

$$\boldsymbol{P[14.44 \leq \mu \leq 14.76] = 0.80}$$

Los colombianos ven televisión en promedio entre 14.4 y 14.76 horas con un nivel de confianza del 80%.

1.3 Estimación para la media poblacional. Muestras pequeñas.

Cuando el número de observaciones es menor de 30, la estimación del intervalo se basa en las suposiciones que si la población es normal o que si se conoce la desviación estándar de la población.

En caso que la muestra sea pequeña, menor de 30, y se conozca la varianza de la población σ^2, el intervalo de confianza es.

$$\bar{X} - \qquad Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\alpha_{/2} * \frac{\sigma}{\sqrt{n}}$$

Cuando el tamaño de la muestra sea pequeño y no se conozca la varianza poblacional σ^2, se utiliza la desviación estándar de la muestra *s*, y la distribución de probabilidad t.

La distribución de probabilidad t o distribución t de Student es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño. Ésta es la base del popular test de la t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones.

La distribución t surge, en la mayoría de los estudios estadísticos prácticos, cuando la desviación típica de una población se desconoce y debe ser estimada a partir de los datos de una muestra.

Cuando el tamaño de la muestra sea pequeño y no se conozca la varianza poblacional σ^2, se utiliza la desviación estándar de la muestra *s*, y la distribución de probabilidad t.

La distribución de probabilidad t o distribución t de Student es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño. Ésta es la base del popular test de la t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones.

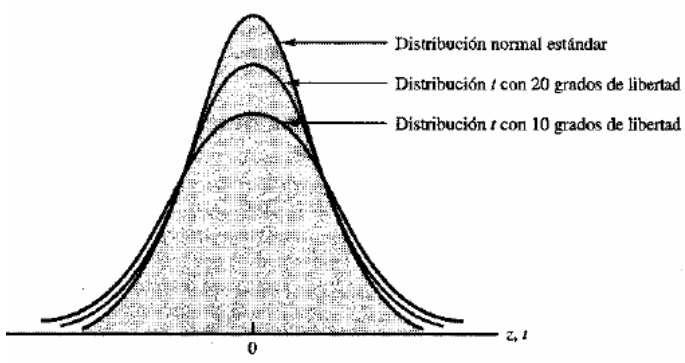
La distribución t surge, en la mayoría de los estudios estadísticos prácticos, cuando la desviación típica de una población se desconoce y debe ser estimada a partir de los datos de una muestra.

La distribución te tiene las siguientes características:

- Al igual que la distribución Z, es una distribución continua.
- La distribución t tiene una media de cero, es simétrica respecto de la media y se extiende de $-\propto$ a $+\propto$ la varianza de t es $\frac{V}{V+2}$ para $\boldsymbol{V > 2}$.

Cuando los grados de libertad son suficientemente grandes la varianza de la distribución t tiende a 1.

- Tiene forma acampanada y simétrica.
- No hay una distribución t, sino una "familia" de distribuciones t. todas con la misma media cero, pero con su respectiva desviación estándar diferente de acuerdo con el tamaño de la muestra n. Existe una distribución t para una muestra de 20, otra para una muestra de 22, y así sucesivamente.
- La distribución t es más ancha y más plana en el centro que la distribución normal estándar como resultado de ello se tiene una mayor variabilidad en las medias de muestra calculadas a partir de muestras más pequeñas. Sin embargo, a medida que aumenta el tamaño de la muestra, la distribución t se aproxima a la distribución normal estándar.



GL	Estimación intervalos de confianza									
	70%	75%	80%	85%	90%	95%	96%	98%	98%	99.00%
1	1.96261	2.41421	3.07768	4.16530	6.31375	12.70620	15.89454	31.82052	31.82052	63.65674
2	1.38621	1.60357	1.88562	2.28193	2.91999	4.30265	4.84873	6.96456	6.96456	9.92484
3	1.24978	1.42263	1.63774	1.92432	2.35336	3.18245	3.48191	4.54070	4.54070	5.84091
4	1.18957	1.34440	1.53321	1.77819	2.13185	2.77645	2.99853	3.74695	3.74695	4.60409
5	1.15577	1.30095	1.47588	1.69936	2.01505	2.57058	2.75651	3.36493	3.36493	4.03214
6	1.13416	1.27335	1.43976	1.65017	1.94318	2.44691	2.61224	3.14267	3.14267	3.70743
7	1.11916	1.25428	1.41492	1.61659	1.89458	2.36462	2.51675	2.99795	2.99795	3.49948
8	1.10815	1.24032	1.39682	1.59222	1.85955	2.30600	2.44898	2.89646	2.89646	3.35539
9	1.09972	1.22966	1.38303	1.57374	1.83311	2.26216	2.39844	2.82144	2.82144	3.24984
10	1.09306	1.22126	1.37218	1.55924	1.81246	2.22814	2.35931	2.76377	2.76377	3.16927

Para la estimación del intervalo de confianza, el valor de t depende de los grados de libertad, n-1, (fila) y del nivel de confianza.

Por ejemplo, para un nivel de confianza de 90% y 5 grados de libertad, t = 2.01505

GL	Estimación interval				
	70%	75%	80%	85%	90%
1	1.96261	2.41421	3.07768	4.16530	6.31375
2	1.38621	1.60357	1.88562	2.28193	2.91999
3	1.24978	1.42263	1.63774	1.92432	2.35336
4	1.18957	1.34440	1.53321	1.77819	2.13185
5	1.15577	1.30095	1.47588	1.69936	2.01505

Intervalo de confianza para muestras pequeñas donde no se conoce la varianza poblacional.

$$\bar{X} - t * \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t * \frac{s}{\sqrt{n}}$$

$$P\left[\bar{X} - t * \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t * \frac{s}{\sqrt{n}}\right] = 1 - \alpha$$

Ejemplo. En una encuesta aplicada a 10 personas de la calle se encontró que en promedio ven 2.5 horas de televisión al día con una desviación estándar de 0.8 horas. Con un nivel de confianza del 95% estime el intervalo d confianza para la media

poblacional.

En ese caso:

N=10;$\overline{\boldsymbol{X}} = \mathbf{2.5}; \boldsymbol{S} = \mathbf{0.8}\boldsymbol{h}$; $\mathbf{1} - \boldsymbol{\alpha} = \mathbf{0.95}$; por lo tanto $\boldsymbol{t}_{\mathbf{9.095}} = \mathbf{2.262}$

$$\boldsymbol{P}\left[\mathbf{2.5} - \mathbf{2.262} * \frac{\mathbf{0.8}}{\sqrt{\mathbf{10}}} \leq \boldsymbol{\mu} \leq \mathbf{2.5} + \mathbf{2.262} * \frac{\mathbf{0.8}}{\sqrt{\mathbf{10}}}\right] = \mathbf{0.95}$$

$$\boldsymbol{P}[\mathbf{1.92775} \leq \boldsymbol{\mu} \leq \mathbf{3.0722}] = \mathbf{0.95}$$

Se estima que la población debe ver en promedio entre 1.93h y 3.07h de televisión al día con un nivel de confianza del 95%.

1.4 Estimación para la proporción poblacional. Muestras grandes

Todo el análisis para la estimación de la media poblacional para muestras grandes se aplica para determinar la proporción poblacional.

$$\overline{\boldsymbol{p}} - \boldsymbol{Z} * \sqrt{\frac{\overline{\boldsymbol{p}} * (\mathbf{1} - \overline{\boldsymbol{p}})}{\boldsymbol{n}}} \leq \boldsymbol{\mu} \leq \overline{\boldsymbol{p}} - \boldsymbol{Z} * \sqrt{\frac{\overline{\boldsymbol{p}} * (\mathbf{1} - \overline{\boldsymbol{p}})}{\boldsymbol{n}}}$$

$$\boldsymbol{P}\left[\overline{\boldsymbol{p}} - \boldsymbol{Z} * \sqrt{\frac{\overline{\boldsymbol{p}} * (\mathbf{1} - \overline{\boldsymbol{p}})}{\boldsymbol{n}}} \leq \boldsymbol{\mu} \leq \overline{\boldsymbol{p}} - \boldsymbol{Z} * \sqrt{\frac{\overline{\boldsymbol{p}} * (\mathbf{1} - \overline{\boldsymbol{p}})}{\boldsymbol{n}}}\right] = \mathbf{1} - \boldsymbol{\alpha}$$

Recuerde que $\boldsymbol{p} = \frac{\boldsymbol{N° \, de \, exitos}}{\boldsymbol{n}}$

Ejemplo. La cadena de televisión TVK está considerando la posibilidad de sustituir una de sus series policiales por una serie de comedia con orientación familiar. Antes de tomar una decisión, los directores de la red toman una muestra de 400 televidentes. Después de ver la nueva serie, 250 indican que sí la verían y sugieren que reemplace a la serie policíaca. Estime el valor de la proporción poblacional con un nivel de 99%.

Datos:

N=400

x=250

p=400/250=0.625

q=1-0.625=0.375

1-$\boldsymbol{\alpha}$=0.99;$\boldsymbol{Z}_{\mathbf{0.995}} = \mathbf{2.58}$

$$P\left[0.625 - 2.58 * \sqrt{\frac{0.625 * 0.375}{400}} \leq \mu \leq 0.625 + 2.58 * \sqrt{\frac{0.625 * 0.375}{400}}\right] = 0.99$$

$$P[0.6008 \leq \mu \leq 0.6492] = 0.99$$

Entre el 60.08% y el 64.92% de los televidentes aceptan el cambio de programa

1.5 Estimación para la proporción poblacional. Muestras pequeñas

Todo el análisis para la estimación de la proporción poblacional para muestras grandes se aplica para determinar la proporción poblacional par muestras pequeñas.

$$\overline{p} - Z * \sqrt{\frac{\overline{p} * (1 - \overline{p})}{n}} \leq \mu \leq \overline{p} - Z * \sqrt{\frac{\overline{p} * (1 - \overline{p})}{n}}$$

$$P\left[\overline{p} - Z * \sqrt{\frac{\overline{p} * (1 - \overline{p})}{n}} \leq \mu \leq \overline{p} - Z * \sqrt{\frac{\overline{p} * (1 - \overline{p})}{n}}\right] = 1 - \alpha$$

Ejemplo. Un periodista estaba preguntando en la calle sobre la opinión que tienen los transeúntes sobre la pena de muerte. De las 25 persona a las que se les preguntó, 20 estuvieron de acuerdo. Encontrar el intervalo de confianza con un nivel de confianza del 95%.

N=425;x=20;p=20/25=0.80; q=1-0.8=0.20;1-α=0.95;$Z_{24;0.95} = 2.0639$

$$P\left[0.80 - 2.0639 * \sqrt{\frac{08 * 0.2}{25}} \leq \mu \leq 0.80 + 2.0639 * \sqrt{\frac{08 * 0.2}{25}}\right] = 0.95$$

$$P[0.634888 \leq \mu \leq 96.5112] = 0.95$$

Según los transeúntes, entre el 63.49% y el 96.51% de los transeúntes están de acuerdo con la pena de muerte

PRUEBA DE HIPÓTESIS (PH).

La estadística inferencial es el proceso de usar la información de una muestra para describir el estado de una población. Sin embargo, es frecuente que usemos la información de una muestra para probar un reclamo o conjetura sobre la población. El reclamo o conjetura se refiere a una hipótesis. El proceso que corrobora si la información de una muestra sostiene o refuta el reclamo se llama prueba de hipótesis.

En la prueba de hipótesis se pone a prueba un reclamo hecho sobra la naturaleza de una población a base de la información de una muestra. El reclamo se llama hipótesis estadística.

Hipótesis es una aseveración acerca de una población.

Hipótesis Estadística: Una hipótesis estadística es una afirmación acerca de un parámetro poblacional, hecho con el propósito de ponerlo a prueba.

Por ejemplo, la premisa formulada por un productor de baterías para autos de que su batería dura en promedio 48 meses, es una hipótesis estadística porque el manufacturero no inspecciona la vida de cada batería que él produce. Si surgieran quejas de parte de los clientes, entonces se pone a prueba el reclamo del manufacturero.

El procedimiento para mostrar si la afirmación es válida o no, se conoce como prueba de hipótesis. Ese procedimiento está basado sobre los resultados de una muestra realizada para tal fin.

La hipótesis estadística sometida a prueba se llama la hipótesis nula, y se denota como $\boldsymbol{H_0}$.

- La hipótesis Nula ($\boldsymbol{H_0}$):

Es la afirmación o conjetura que se hace sobre el parámetro poblacional.

En caso que no se logre tomar como válida la hipótesis nula se debe tener una afirmación alterna que se debe dar como válidas y conocida cono la hipótesis alterna que se denota como $\boldsymbol{H_1}$.

- Hipótesis Alterna ($\boldsymbol{H_1}$):

Una premisa que es cierta cuando la hipótesis nula es falsa.

Por ejemplo, para el productor de baterías

Por ejemplo, para aprobar o desaprobar el reclamo pronunciado por el productor de baterías debemos probar la hipótesis estadística de que $\boldsymbol{\mu \geq 48}$. Por lo tanto, la hipótesis nula es:

$\boldsymbol{H_0: \geq 40\ H_1: < 48}$

Luego de tener $\boldsymbol{H_0}$ y $\boldsymbol{H_1}$ se procede a tomar una muestra aleatoria de baterías y medir su vida media.

Para probar si la hipótesis nula es cierta, se toma una muestra aleatoria y se calcula la información, como el promedio, la proporción, etc. Esta información muestral se llama estadística de prueba.

Tipos de Errores

A base de la información de una muestra nosotros podemos cometer dos tipos de errores en nuestra decisión.

- Rechazar$\boldsymbol{H_0}$ siendo que es cierta. Error tipo 1
- Aceptar $\boldsymbol{H_0}$ siendo que es falsa. Error tipo 2

El error Tipo 1 se da cuando se rechaza la Hipótesis Nula siendo que es válida.

El error Tipo 2 se da cuando aceptamos la Hipótesis Nula siendo que es falsa.

Prueba estadística. Prueba Z para la diferencia entre dos proporciones, divida ésta por el error estándar de la diferencia.

$$\boldsymbol{Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_1(1-P_2)}{n_2}}}}$$

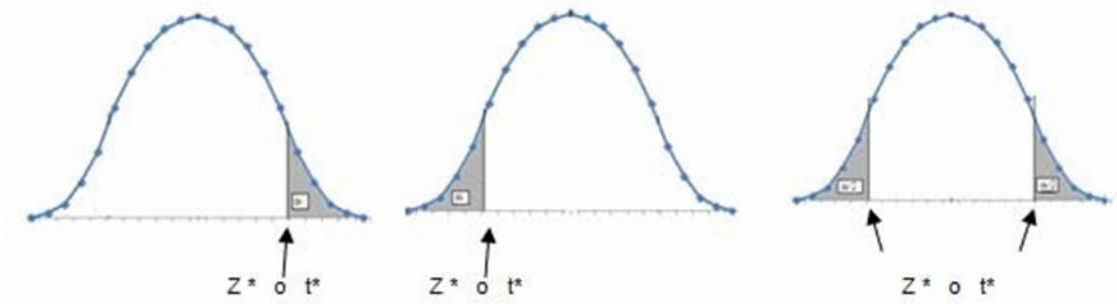
- Nivel de Significancia ($\boldsymbol{\alpha}$)

Para ser muy cuidadosos en no cometer el error tipo 1, debemos especificar la probabilidad de rechazar $\boldsymbol{H_0}$, denotada por $\boldsymbol{\alpha}$.

El nivel de significancia es la probabilidad de cometer el error tipo I. Este valor debe ser pequeño.

Los valores más usados para hacer prueba de hipótesis es 1%, 5% y 10%.

Usando un valor pre-asignado de $\boldsymbol{\alpha}$ se construye una región de rechazo o región crítica en la curva normal estándar o en la curva t que indica si debemos rechazar $\boldsymbol{H_0}$ o aceptarla.



- Región Crítica o de Rechazo. Una región crítica o de rechazo es una parte de la curva de *z* o de la curva t donde se rechaza $\boldsymbol{H_0}$. La región puede ser de una cola o de dos dependiendo de la Hipótesis Alterna. Si la hipótesis alterna es $\boldsymbol{H_1}: \boldsymbol{\mu} > \boldsymbol{k}$, la cola es hacia la derecha, si $\boldsymbol{H_1}: \boldsymbol{\mu} < \boldsymbol{k_1}$, la cola es hacia la derecha, o si $\boldsymbol{H_1}: \boldsymbol{\mu} = \boldsymbol{k}$, son dos colas.

2.1 Prueba de Hipótesis Unilaterales y Bilaterales sobre la Media

En primer lugar, veremos cómo se escribirían las hipótesis que queremos contrastar:

$\boldsymbol{H_0}$ se llama hipótesis nula y es lo contrario de lo que sospechamos que va a ocurrir (suele llevar los signos igual, mayor o igual y menor o igual)

$\boldsymbol{H_1}$ se llama hipótesis alternativa y es lo que sospechamos que va a ser cierto (suele llevar los signos distinto, mayor y menor). Los contrastes de hipótesis pueden ser de dos tipos:

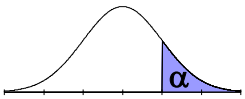
- Bilateral: En la hipótesis alternativa aparece el signo =.
- Unilateral: En la hipótesis alternativa aparece o el signo > o el signo <.

Nivel de significancia $\boldsymbol{\alpha}$: Es la probabilidad de cometer un error de tipo I. $\boldsymbol{\alpha}$ es la probabilidad de cometer un error de tipo II. De los dos, el más importante es $\boldsymbol{\alpha}$.

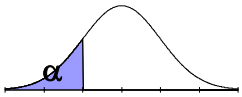
Debido a que los dos errores anteriores a la vez son imposibles de controlar, vamos a fijarnos solamente en el nivel de significación, este es el que nos interesa ya que la hipótesis alternativa que estamos interesados en probar y no queremos aceptarla si en realidad no es cierta, es decir, si aceptamos la hipótesis alternativa queremos equivocarnos con un margen de error muy pequeño.

El nivel de significación lo marcamos nosotros. Si es grande es más fácil aceptar la hipótesis alternativa cuando en realidad es falsa. El valor del nivel de significación suele ser un 5%, lo que significa que 5 de cada 100 veces aceptamos la hipótesis alternativa cuando la cierta es la nula.

Ejemplos



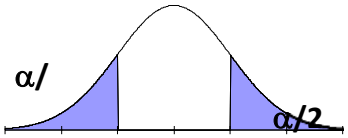
Para H_1: μ > valor aceptado, la región de rechazo está dada por:
(cola derecha, z* ó t*)



Para H_1 : $\mu \neq k$

Para H_1 : μ < valor aceptado, la región de rechazo está dada por :
(cola izquierda, z* ó t*)

Muestras Grandes



Para H_1: μ = valor aceptado, la región de rechazo es de dos colas y está dada por:
(2-colas, Z* ó t*)

En las pruebas de hipótesis para la media (μ), cuando se conoce la desviación estándar (σ) poblacional, o cuando el valor de la muestra es grande (30 o más), el valor estadístico de prueba es z y se determina a partir de:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

El valor estadístico z, para muestra grande y desviación estándar poblacional desconocida se determina por la ecuación:

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

2.2 Muestras Pequeñas

En la prueba para una media poblacional con muestra pequeña y desviación estándar poblacional desconocida se utiliza el valor estadístico t.

$$\boldsymbol{t = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}$$

Procedimiento para una prueba de hipótesis

Los pasos a seguir son:

1. Formular la hipótesis nula $\boldsymbol{H_0}$ y la alternativa $\boldsymbol{H_1}$, de acuerdo al problema.
2. Escoger un nivel de significación o riesgos $\boldsymbol{\alpha}$.
3. Elegir la estadística de prueba apropiada, cuya distribución por muestreo sea conocida en el supuesto de que $\boldsymbol{H_0}$ es cierta.
4. En base a $\boldsymbol{\alpha}$ y $\boldsymbol{H_1}$, determinar el valor (o los valores) críticos y con ello se establecen las regiones de aceptación o rechazo.
5. Calcular los valores de la prueba estadística a partir de una muestra aleatoria de tamaño n, $\boldsymbol{H_0}$ y reemplazarlos en la estadística de prueba elegida en el paso 3, para hallar el valor experimental.
6. Tomar la decisión de aceptar Ho si el valor experimental cae en la región de aceptación y rechazarla si dicho valor cae en la región crítica o de rechazo.
7. Opcional: Si se rechaza $\boldsymbol{H_0}$, se puede hallar un intervalo de confianza para el parámetro de interés.

2.3 Prueba de hipótesis sobre la media poblacional.

Caso A: Cuando la varianza poblacional es conocida.

La hipótesis de que el parámetro poblacional μ = X toma un determinado valor K Conocemos que la población se distribuye normalmente y conocemos también su varianza, o bien si nos es desconocida, el tamaño muestral es lo suficientemente grande como para poder utilizar la muestral cómo poblacional.

Determina un nivel de significación para la realización del contraste y plantea el supuesto de una muestra aleatoria de tamaño n.

Así conocemos que $\boldsymbol{\overline{X} \Rightarrow N\left[\mu, \sigma/\sqrt{n}\right]}$ de lo que deducimos que $\boldsymbol{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \Longrightarrow N[0,1]}$ de forma que la hipótesis nula es $\boldsymbol{H_0: \mu = k}$

El estadístico está dado por $\boldsymbol{Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}$

Ejemplo Nº 1. De 100 observaciones de una población normal se obtiene que $\boldsymbol{\bar{X} = 5}$, $\boldsymbol{\sigma = 2}$.Contrastar con un nivel de significación del 5% la hipótesis de que la media de la población sea 7.

Aplicando el procedimiento para probar una hipótesis tenemos:

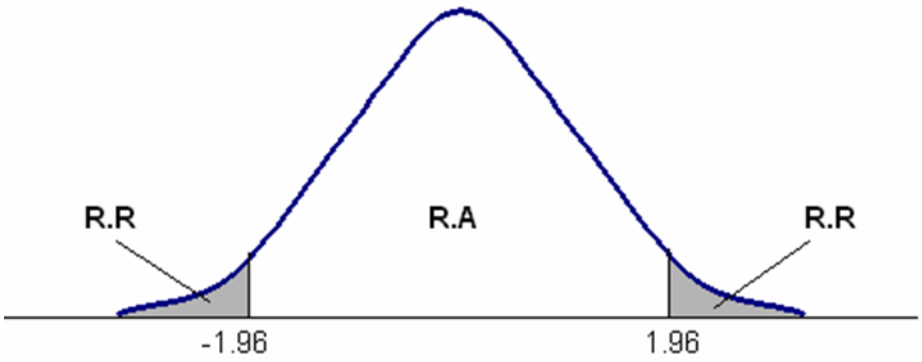
1. $\boldsymbol{H_0: \mu = 7}$

   $\boldsymbol{H_1: \mu \neq 7}$

2. El nivel de significancia es del 5% $(\boldsymbol{\alpha = 5\%})$

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

4. Establecemos la región de aceptación y de rechazo:



5. Realizamos la prueba estadística: $\boldsymbol{Z = \frac{5-7}{2/\sqrt{100}}}$

6. $\boldsymbol{\mu_0 \neq 7}$

Cuando no se conoce la varianza poblacional y para una muestra pequeña.

Deseamos contrastar la hipótesis de que el parámetro poblacional toma un determinado valor.

Desconocemos la varianza de la población y, dado que el tamaño muestral es pequeño (menor o igual a 30), podemos utilizar varianza en su lugar.

Hemos determinado un nivel de significación para la realización del contraste y vamos a plantearlo en el supuesto de realizar una muestra aleatoria de tamaño n menor o igual a 30.

En este caso el estadístico de prueba será t:

$$\boldsymbol{t = \frac{\bar{X} - \mu}{s/\sqrt{n}}}$$

Ejemplo. Se escoge a 17 individuos al azar y se les mide, resultando que su estatura media es de 1,71 metros con desviación típica de 0,02. Contrastar la hipótesis de que la estatura media nacional sea de 1.75 metros si utilizamos un nivel de la significación del 5%. Se supone normalidad

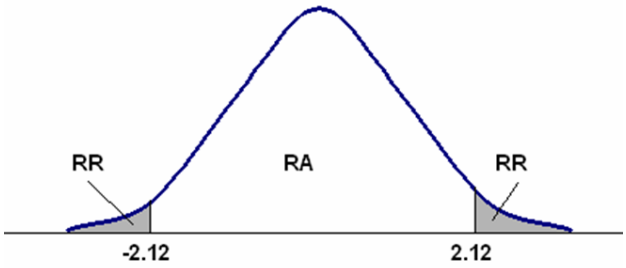
Aplicando el procedimiento para probar una hipótesis tenemos:

1. $\boldsymbol{H_0: \mu = 7}$

   $\boldsymbol{H_1: \mu \neq 7}$

2. El nivel de significancia es del 5% $(\boldsymbol{\alpha = 5\%})$

3. Como n es menor de 30 se busca en la tabla para un nivel de caso $\boldsymbol{t} *= {}^{+}/_{-}\mathbf{2.11991}$



4. Establecemos la región de aceptación y de rechazo: Utilizando la tabla T.

5. Se calcula el estadístico de prueba : $\boldsymbol{t = \frac{1.71-1.75}{0.02/\sqrt{17}} = -8.25}$

6. Dado que t=-8.25 y no pertenece a la región de aceptación estamos en condiciones de rechazar la hipótesis nula, luego aceptar la alternativa:$\boldsymbol{\mu \neq 1.75}$

2.4 Prueba de Hipótesis sobre la Proporción de una Población p

Se trata de efectuar una prueba de hipótesis acerca de la proporción de elementos con cierto atributo en una población, hipótesis de la forma:

$$\boldsymbol{H_0: p = p_0}$$
$$\boldsymbol{H_1: p \neq p_0}$$
$$\boldsymbol{H_0: p \leq p_0}$$
$$\boldsymbol{H_1: p > p_0}$$
$$\boldsymbol{H_0: p \geq p_0}$$
$$\boldsymbol{H_1: p < p_0}$$

$$Z = \frac{p - p_0}{\sqrt{\frac{p_{0(1-p_0)}}{n}}}$$

Dónde: $p = \frac{x}{n}$ (proporción muestral)

Ejemplo 1. Una empresa de publicidad desea comprobar si un determinado programa de televisión es visto por el 30% de la audiencia potencial. Para ello se escoge al azar una muestra de 200 familias resultando que de ellas 50 lo ven asiduamente. Contrastar la hipótesis con un nivel de significación del 5%.

Aplicando el procedimiento para probar una hipótesis tenemos:

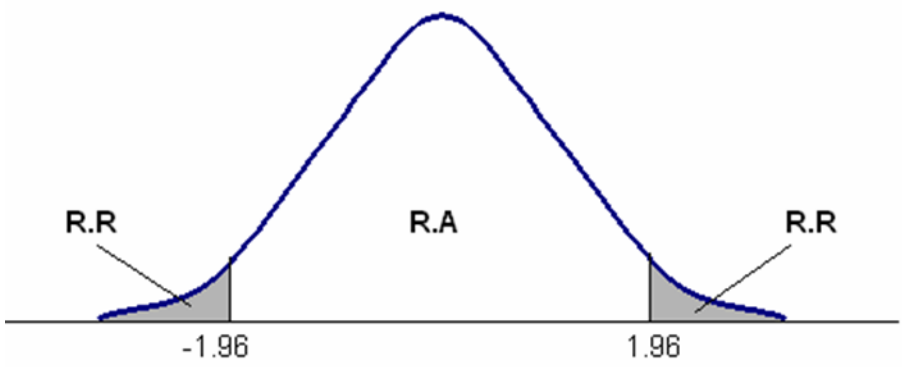
1. $H_0: p = 0.3$

$H_1: p \neq 0.30$

2. El nivel de significancia es del 5% $(\alpha = 5\%)$

3. $z * = {}^{+}/_{-} 1.96$

4. Establecemos la región de aceptación y de rechazo



5. Realizamos la prueba estadística:

$$Z = \frac{p - p_0}{\sqrt{\frac{p_{0(1-p_0)}}{n}}} = \frac{0.25 - 0.30}{\sqrt{\frac{0.3(1 - 0.3)}{200}}} = -1.54$$

6. Dado que Z=-1.54 y pertenece a la región de aceptación estamos en condiciones de acepta la hipótesis nula, es decir: $p = 0,3$

2.5 Cálculo de la Probabilidad para el error tipo II

La probabilidad del error Tipo I es siempre igual al nivel de significancia que se utiliza al probar las hipótesis nulas. Dicho de otra manera, hay todavía una probabilidad (valor de $\boldsymbol{\alpha}$) de que se pudiera elegir una muestra al azar que diera una media comprendida dentro de la región de rechazo ó las regiones de rechazo. Ello haría que al rechazar la hipótesis nula siendo verdadera cometiéramos el error Tipo I.

La única forma en que se puede determinar la probabilidad del error Tipo II ($\boldsymbol{\beta}$) es con respecto a un valor específico incluido dentro del rango de la hipótesis alternativa. Ejemplo:

News & World Report publicó un artículo sobre la carrera de éxitos de Wal- Mart. Actualmente es la mayor cadena de ventas al por menor de la nación. Empezó con una sola tienda de descuento en la pequeña localidad de Rogers, Arkansas, y a crecido hasta poseer 1300 tiendas en 25 estados. Este éxito le ha valido a Sam Walton, fundador y mayor accionista, el título del hombre más rico de América. Las ventas anuales se cifran en 15 millones de dólares por tienda.

a) Si se elige al azar una muestra de 120 tiendas y se hallan unas ventas medias de 15.39 millones de dólares, con una desviación estándar de 2.9 millones de dólares. Pruebe la hipótesis 15 millones con un nivel de significancia del 10%

b)Si la μ es en realidad 14.8 millones de dólares, ¿Cuál es la probabilidad de cometer el error Tipo II.

Datos

N= 120

$\boldsymbol{\bar{X} = 15.39\ millones}$

$\boldsymbol{\sigma = 2.9\ millones}$

$\boldsymbol{\alpha = 10}$

1. Establecer la hipótesis

$\boldsymbol{H_0: \mu = 15}$

$\boldsymbol{H_1: \mu \neq 15}$

2. Establecer la estadística de prueba

$$\boldsymbol{Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}}$$

3. Definir el nivel de significancia y la zona de rechazo

4. Nivel de significancia : $\boldsymbol{\beta = 0.10}$ zona de rechazo $= \{z/z < 1.64\ o\ z/z > 1.64\}$

5. Calcular la estadística de prueba:

$$\boldsymbol{Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}}$$

el valor de la media poblacional es el que está a prueba en la hipótesis, por lo tanto:

$$\boldsymbol{Z = \frac{15.39 - 15}{2.9 / \sqrt{120}}}$$

Como 1.47 < 1.64, No se rechaza $\boldsymbol{H_0}$

6. Conclusión. Existe evidencia para decir que las ganancias anuales por tienda son de 15 millones de dólares por tienda con un nivel de significancia de 0.10.

Pasos para calcular el error Tipo II

1. Plantear la hipótesis nula y alternativa para la prueba

$$H_0: \mu = 15$$

$$H_1: \mu \neq 15$$

2. Determinar el valor crítico de la media muestral que debe utilizarse para probar la hipótesis nula con un nivel de significancia dado.

$$\overline{X_C = \mu \pm Z\,\sigma\,/\sqrt{n}}$$

$$\bar{x} = 15 \pm (1.64)2.9/\sqrt{120} = 14.5639$$

$$\bar{x} = 15 \pm (1.64)2.9/\sqrt{140} = 15.4342$$

3. Identificar la probabilidad del error Tipo I correspondiente al valor crítico que se calculó antes, como base para la regla de decisión. = 0 .10.

4. Identificar la probabilidad del error Tipo II correspondiente a la regla de decisión, dada una media alternativa especificada.

$$P(error\ de\ tipo\ II) = P(14.56 < \bar{X} < 15.44)$$

$$Z = \frac{15.4841 - 14.8}{2.9/\sqrt{120}}, Z = \frac{15.5659 - 14.8}{2.9/\sqrt{10}} = 0.88$$

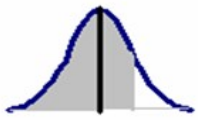
$$P(error\ de\ tipo\ II) = P(-0.88 < \bar{X} < 2.40) = 0.8024$$

t

=DISTR.T.INV(α;GL)

GL	Intervalos de confianza (1- α)													
	80%	85%	86%	88%	90%	92%	94%	92%	95%	96.00%	97%	98%	99%	99.90%
	Nivel de significancia para pruebas de una cola													
	0.10	0.075	0.07	0.06	0.05	0.04	0.03	0.04	0.025	0.02	0.015	0.01	0.005	0.0005
	Nivel de significancia para pruebas de dos colas													
	0.2000	0.1500	0.1400	0.1200	0.1000	0.0800	0.0600	0.0800	0.0500	0.0400	0.0300	0.0200	0.0100	0.0010
1	3.07768	4.16530	4.47374	5.24218	6.31375	7.91582	10.57889	7.91582	12.70620	15.89454	21.20495	31.82052	63.65674	636.61925
2	1.88562	2.28193	2.38338	2.62016	2.91999	3.31976	3.89643	3.31976	4.30265	4.84873	5.64278	6.96456	9.92484	31.59905
3	1.63774	1.92432	1.99502	2.15624	2.35336	2.60543	2.95051	2.60543	3.18245	3.48191	3.89605	4.54070	5.84091	12.92398
4	1.53321	1.77819	1.83752	1.97123	2.13185	2.33287	2.60076	2.33287	2.77645	2.99853	3.29763	3.74695	4.60409	8.61030
5	1.47588	1.69936	1.75289	1.87268	2.01505	2.19096	2.42158	2.19096	2.57058	2.75651	3.00287	3.36493	4.03214	6.86883
6	1.43976	1.65017	1.70021	1.81165	1.94318	2.10431	2.31326	2.10431	2.44691	2.61224	2.82893	3.14267	3.70743	5.95882
7	1.41492	1.61659	1.66430	1.77021	1.89458	2.04601	2.24088	2.04601	2.36462	2.51675	2.71457	2.99795	3.49948	5.40788
8	1.39682	1.59222	1.63827	1.74024	1.85955	2.00415	2.18915	2.00415	2.30600	2.44898	2.63381	2.89646	3.35539	5.04131
9	1.38303	1.57374	1.61854	1.71758	1.83311	1.97265	2.15038	1.97265	2.26216	2.39844	2.57380	2.82144	3.24984	4.78091
10	1.37218	1.55924	1.60308	1.69984	1.81246	1.94810	2.12023	1.94810	2.22814	2.35931	2.52748	2.76377	3.16927	4.58689
11	1.36343	1.54756	1.59063	1.68558	1.79588	1.92843	2.09614	1.92843	2.20099	2.32814	2.49066	2.71808	3.10581	4.43698
12	1.35622	1.53796	1.58040	1.67387	1.78229	1.91231	2.07644	1.91231	2.17881	2.30272	2.46070	2.68100	3.05454	4.31779
13	1.35017	1.52992	1.57184	1.66408	1.77093	1.89887	2.06004	1.89887	2.16037	2.28160	2.43585	2.65031	3.01228	4.22083
14	1.34503	1.52310	1.56458	1.65578	1.76131	1.88750	2.04617	1.88750	2.14479	2.26378	2.41490	2.62449	2.97684	4.14045
15	1.34061	1.51723	1.55833	1.64865	1.75305	1.87774	2.03429	1.87774	2.13145	2.24854	2.39701	2.60248	2.94671	4.07277
16	1.33676	1.51213	1.55291	1.64246	1.74588	1.86928	2.02400	1.86928	2.11991	2.23536	2.38155	2.58349	2.92078	4.01500
17	1.33338	1.50766	1.54815	1.63703	1.73961	1.86187	2.01500	1.86187	2.10982	2.22385	2.36805	2.56693	2.89823	3.96513
18	1.33039	1.50371	1.54395	1.63224	1.73406	1.85534	2.00707	1.85534	2.10092	2.21370	2.35618	2.55238	2.87844	3.92165
19	1.32773	1.50019	1.54020	1.62797	1.72913	1.84953	2.00002	1.84953	2.09302	2.20470	2.34565	2.53948	2.86093	3.88341
20	1.32534	1.49704	1.53685	1.62415	1.72472	1.84433	1.99371	1.84433	2.08596	2.19666	2.33624	2.52798	2.84534	3.84952
21	1.32319	1.49419	1.53383	1.62071	1.72074	1.83965	1.98804	1.83965	2.07961	2.18943	2.32779	2.51765	2.83136	3.81928
22	1.32124	1.49162	1.53109	1.61759	1.71714	1.83542	1.98291	1.83542	2.07387	2.18289	2.32016	2.50832	2.81876	3.79213
23	1.31946	1.48928	1.52860	1.61476	1.71387	1.83157	1.97825	1.83157	2.06866	2.17696	2.31323	2.49987	2.80734	3.76763
24	1.31784	1.48714	1.52633	1.61217	1.71088	1.82805	1.97399	1.82805	2.06390	2.17154	2.30691	2.49216	2.79694	3.74540
25	1.31635	1.48517	1.52424	1.60979	1.70814	1.82483	1.97010	1.82483	2.05954	2.16659	2.30113	2.48511	2.78744	3.72514
26	1.31497	1.48336	1.52232	1.60760	1.70562	1.82186	1.96651	1.82186	2.05553	2.16203	2.29581	2.47863	2.77871	3.70661
27	1.31370	1.48169	1.52054	1.60558	1.70329	1.81913	1.96320	1.81913	2.05183	2.15782	2.29091	2.47266	2.77068	3.68959
28	1.31253	1.48014	1.51890	1.60371	1.70113	1.81659	1.96014	1.81659	2.04841	2.15393	2.28638	2.46714	2.76326	3.67391
29	1.31143	1.47870	1.51737	1.60197	1.69913	1.81424	1.95729	1.81424	2.04523	2.15033	2.28217	2.46202	2.75639	3.65941
30	1.31042	1.47736	1.51595	1.60035	1.69726	1.81205	1.95465	1.81205	2.04227	2.14697	2.27826	2.45726	2.75000	3.64596
40	1.30308	1.46772	1.50570	1.58871	1.68385	1.79631	1.93566	1.79631	2.02108	2.12291	2.25027	2.42326	2.70446	3.55097
60	1.29582	1.45820	1.49560	1.57723	1.67065	1.78085	1.91703	1.78085	2.00030	2.09936	2.22292	2.39012	2.66028	3.46020
100	1.29007	1.45067	1.48761	1.56817	1.66023	1.76866	1.90237	1.76866	1.98397	2.08088	2.20150	2.36422	2.62589	3.39049
200	1.28580	1.44508	1.48168	1.56144	1.65251	1.75963	1.89152	1.75963	1.97190	2.06723	2.18568	2.34514	2.60063	3.33984
500	1.28325	1.44175	1.47814	1.55743	1.64791	1.75425	1.88507	1.75425	1.96472	2.05912	2.17630	2.33383	2.58570	3.31009
1000	1.28240	1.44064	1.47696	1.55610	1.64638	1.75247	1.88293	1.75247	1.96234	2.05643	2.17319	2.33008	2.58075	3.30028

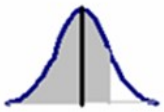
TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \overline{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.5000000	0.5039894	0.5079783	0.5119665	0.5159534	0.5199388	0.5239222	0.5279032	0.5318814	0.5358564	0.0
0.1	0.5398278	0.5437953	0.5477584	0.5517168	0.5556700	0.5596177	0.5635595	0.5674949	0.5714237	0.5753454	0.1
0.2	0.5792597	0.5831662	0.5870644	0.5909541	0.5948349	0.5987063	0.6025681	0.6064199	0.6102612	0.6140919	0.2
0.3	0.6179114	0.6217195	0.6255158	0.6293000	0.6330717	0.6368307	0.6405764	0.6443088	0.6480273	0.6517317	0.3
0.4	0.6554217	0.6590970	0.6627573	0.6664022	0.6700314	0.6736448	0.6772419	0.6808225	0.6843863	0.6879331	0.4
0.5	0.6914625	0.6949743	0.6984682	0.7019440	0.7054015	0.7088403	0.7122603	0.7156612	0.7190427	0.7224047	0.5
0.6	0.7257469	0.7290691	0.7323711	0.7356527	0.7389137	0.7421539	0.7453731	0.7485711	0.7517478	0.7549029	0.6
0.7	0.7580363	0.7611479	0.7642375	0.7673049	0.7703500	0.7733726	0.7763727	0.7793501	0.7823046	0.7852361	0.7
0.8	0.7881446	0.7910299	0.7938919	0.7967306	0.7995458	0.8023375	0.8051055	0.8078498	0.8105703	0.8132671	0.8
0.9	0.8159399	0.8185887	0.8212136	0.8238145	0.8263912	0.8289439	0.8314724	0.8339768	0.8364569	0.8389129	0.9
1.0	0.8413447	0.8437524	0.8461358	0.8484950	0.8508300	0.8531409	0.8554277	0.8576903	0.8599289	0.8621434	1.0
1.1	0.8643339	0.8665005	0.8686431	0.8707619	0.8728568	0.8749281	0.8769756	0.8789995	0.8809999	0.8829768	1.1
1.2	0.8849303	0.8868606	0.8887676	0.8906514	0.8925123	0.8943502	0.8961653	0.8979577	0.8997274	0.9014747	1.2
1.3	0.9031995	0.9049021	0.9065825	0.9082409	0.9098773	0.9114920	0.9130850	0.9146565	0.9162067	0.9177356	1.3
1.4	0.9192433	0.9207302	0.9221962	0.9236415	0.9250663	0.9264707	0.9278550	0.9292191	0.9305634	0.9318879	1.4
1.5	0.9331928	0.9344783	0.9357445	0.9369916	0.9382198	0.9394292	0.9406201	0.9417924	0.9429466	0.9440826	1.5
1.6	0.9452007	0.9463011	0.9473839	0.9484493	0.9494974	0.9505285	0.9515428	0.9525403	0.9535213	0.9544860	1.6
1.7	0.9554345	0.9563671	0.9572838	0.9581849	0.9590705	0.9599408	0.9607961	0.9616364	0.9624620	0.9632730	1.7
1.8	0.9640697	0.9648521	0.9656205	0.9663750	0.9671159	0.9678432	0.9685572	0.9692581	0.9699460	0.9706210	1.8
1.9	0.9712834	0.9719334	0.9725711	0.9731966	0.9738102	0.9744119	0.9750021	0.9755808	0.9761482	0.9767045	1.9
2.0	0.9772499	0.9777844	0.9783083	0.9788217	0.9793248	0.9798178	0.9803007	0.9807738	0.9812372	0.9816911	2.0
2.1	0.9821356	0.9825708	0.9829970	0.9834142	0.9838226	0.9842224	0.9846137	0.9849966	0.9853713	0.9857379	2.1
2.2	0.9860966	0.9864474	0.9867906	0.9871263	0.9874545	0.9877755	0.9880894	0.9883962	0.9886962	0.9889893	2.2
2.3	0.9892759	0.9895559	0.9898296	0.9900969	0.9903581	0.9906133	0.9908625	0.9911060	0.9913437	0.9915758	2.3
2.4	0.9918025	0.9920237	0.9922397	0.9924506	0.9926564	0.9928572	0.9930531	0.9932443	0.9934309	0.9936128	2.4
2.5	0.9937903	0.9939634	0.9941323	0.9942969	0.9944574	0.9946139	0.9947664	0.9949151	0.9950600	0.9952012	2.5
2.6	0.9953388	0.9954729	0.9956035	0.9957308	0.9958547	0.9959754	0.9960930	0.9962074	0.9963189	0.9964274	2.6
2.7	0.9965330	0.9966358	0.9967359	0.9968333	0.9969280	0.9970202	0.9971099	0.9971972	0.9972821	0.9973646	2.7
2.8	0.9974449	0.9975229	0.9975988	0.9976726	0.9977443	0.9978140	0.9978818	0.9979476	0.9980116	0.9980738	2.8
2.9	0.9981342	0.9981929	0.9982498	0.9983052	0.9983589	0.9984111	0.9984618	0.9985110	0.9985588	0.9986051	2.9
3.0	0.9986501	0.9986938	0.9987361	0.9987772	0.9988171	0.9988558	0.9988933	0.9989297	0.9989650	0.9989992	3.0
3.1	0.9990324	0.9990646	0.9990957	0.9991260	0.9991553	0.9991836	0.9992112	0.9992378	0.9992636	0.9992886	3.1
3.2	0.9993129	0.9993363	0.9993590	0.9993810	0.9994024	0.9994230	0.9994429	0.9994623	0.9994810	0.9994991	3.2
3.3	0.9995166	0.9995335	0.9995499	0.9995658	0.9995811	0.9995959	0.9996103	0.9996242	0.9996376	0.9996505	3.3
3.4	0.9996631	0.9996752	0.9996869	0.9996982	0.9997091	0.9997197	0.9997299	0.9997398	0.9997493	0.9997585	3.4
3.5	0.9997674	0.9997759	0.9997842	0.9997922	0.9997999	0.9998074	0.9998146	0.9998215	0.9998282	0.9998347	3.5
3.6	0.9998409	0.9998469	0.9998527	0.9998583	0.9998637	0.9998689	0.9998739	0.9998787	0.9998834	0.9998879	3.6
3.7	0.9998922	0.9998964	0.9999004	0.9999043	0.9999080	0.9999116	0.9999150	0.9999184	0.9999216	0.9999247	3.7
3.8	0.9999277	0.9999305	0.9999333	0.9999359	0.9999385	0.9999409	0.9999433	0.9999456	0.9999478	0.9999499	3.8
3.9	0.9999519	0.9999539	0.9999557	0.9999575	0.9999593	0.9999609	0.9999625	0.9999641	0.9999655	0.9999670	3.9
4.0	0.9999683	0.9999696	0.9999709	0.9999721	0.9999733	0.9999744	0.9999755	0.9999765	0.9999775	0.9999784	4.0
4.1	0.9999793	0.9999802	0.9999811	0.9999819	0.9999826	0.9999834	0.9999841	0.9999848	0.9999854	0.9999861	4.1
4.2	0.9999867	0.9999872	0.9999878	0.9999883	0.9999888	0.9999893	0.9999898	0.9999902	0.9999907	0.9999911	4.2
4.3	0.9999915	0.9999918	0.9999922	0.9999925	0.9999929	0.9999932	0.9999935	0.9999938	0.9999941	0.9999943	4.3
4.4	0.9999946	0.9999948	0.9999951	0.9999953	0.9999955	0.9999957	0.9999959	0.9999961	0.9999963	0.9999964	4.4
4.5	0.9999966	0.9999968	0.9999969	0.9999971	0.9999972	0.9999973	0.9999974	0.9999976	0.9999977	0.9999978	4.5
4.6	0.9999979	0.9999980	0.9999981	0.9999982	0.9999983	0.9999983	0.9999984	0.9999985	0.9999986	0.9999986	4.6
4.7	0.9999987	0.9999988	0.9999988	0.9999989	0.9999989	0.9999990	0.9999990	0.9999991	0.9999991	0.9999992	4.7
4.8	0.9999992	0.9999992	0.9999993	0.9999993	0.9999994	0.9999994	0.9999994	0.9999994	0.9999995	0.9999995	4.8
4.9	0.9999995	0.9999995	0.9999996	0.9999996	0.9999996	0.9999996	0.9999996	0.9999997	0.9999997	0.9999997	4.9
5.0	0.9999997	0.9999997	0.9999997	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	5.0
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

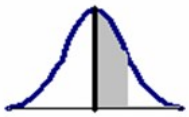
TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \overline{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.0
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.1
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.2
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.3
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.4
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.5
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	0.6
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	0.7
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	0.8
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	0.9
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	1.0
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	1.1
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	1.2
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	1.3
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	1.4
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	1.5
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	1.6
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	1.7
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	1.8
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	1.9
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	2.1
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	2.2
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	2.3
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	2.4
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	2.5
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	2.6
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	2.7
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	2.8
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	2.9
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900	3.0
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929	3.1
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	3.2
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965	3.3
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	3.4
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983	3.5
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989	3.6
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992	3.7
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995	3.8
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997	3.9
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998	4.0
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999	4.1
4.2	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	4.2
4.3	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	4.3
4.4	0.99999	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.4
4.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.5
4.6	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.6
4.7	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.7
4.8	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.8
4.9	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.9
5.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	5.0
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \overline{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586	0.0
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535	0.1
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409	0.2
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173	0.3
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793	0.4
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240	0.5
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490	0.6
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524	0.7
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327	0.8
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891	0.9
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214	1.0
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298	1.1
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147	1.2
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774	1.3
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189	1.4
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408	1.5
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449	1.6
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327	1.7
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062	1.8
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670	1.9
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169	2.0
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574	2.1
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899	2.2
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158	2.3
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361	2.4
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520	2.5
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643	2.6
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736	2.7
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807	2.8
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861	2.9
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900	3.0
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929	3.1
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950	3.2
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965	3.3
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976	3.4
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983	3.5
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989	3.6
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992	3.7
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995	3.8
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997	3.9
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998	4.0
4.1	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49999	0.49999	4.1
4.2	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	4.2
4.3	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	4.3
4.4	0.49999	0.49999	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.4
4.5	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.5
4.6	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.6
4.7	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.7
4.8	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.8
4.9	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	4.9
5.0	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	5.0
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

=DISTR.T.INV(α;GL)

GL	Intervalos de confianza (1- α)													
	80%	85%	86%	88%	90%	92%	94%	92%	95%	96.00%	97%	98%	99%	99.90%
	Nivel de significancia para pruebas de una cola													
	0.10	0.075	0.07	0.06	0.05	0.04	0.03	0.04	0.025	0.02	0.015	0.01	0.005	0.0005
	Nivel de significancia para pruebas de dos colas													
	0.2000	0.1500	0.1400	0.1200	0.1000	0.0800	0.0600	0.0800	0.0500	0.0400	0.0300	0.0200	0.0100	0.0010
1	3.07768	4.16530	4.47374	5.24218	6.31375	7.91582	10.57889	7.91582	12.70620	15.89454	21.20495	31.82052	63.65674	636.61925
2	1.88562	2.28193	2.38338	2.62016	2.91999	3.31976	3.89643	3.31976	4.30265	4.84873	5.64278	6.96456	9.92484	31.59905
3	1.63774	1.92432	1.99502	2.15624	2.35336	2.60543	2.95051	2.60543	3.18245	3.48191	3.89605	4.54070	5.84091	12.92398
4	1.53321	1.77819	1.83752	1.97123	2.13185	2.33287	2.60076	2.33287	2.77645	2.99853	3.29763	3.74695	4.60409	8.61030
5	1.47588	1.69936	1.75289	1.87268	2.01505	2.19096	2.42158	2.19096	2.57058	2.75651	3.00287	3.36493	4.03214	6.86883
6	1.43976	1.65017	1.70021	1.81165	1.94318	2.10431	2.31326	2.10431	2.44691	2.61224	2.82893	3.14267	3.70743	5.95882
7	1.41492	1.61659	1.66430	1.77021	1.89458	2.04601	2.24088	2.04601	2.36462	2.51675	2.71457	2.99795	3.49948	5.40788
8	1.39682	1.59222	1.63827	1.74024	1.85955	2.00415	2.18915	2.00415	2.30600	2.44898	2.63381	2.89646	3.35539	5.04131
9	1.38303	1.57374	1.61854	1.71758	1.83311	1.97265	2.15038	1.97265	2.26216	2.39844	2.57380	2.82144	3.24984	4.78091
10	1.37218	1.55924	1.60308	1.69984	1.81246	1.94810	2.12023	1.94810	2.22814	2.35931	2.52748	2.76377	3.16927	4.58689
11	1.36343	1.54756	1.59063	1.68558	1.79588	1.92843	2.09614	1.92843	2.20099	2.32814	2.49066	2.71808	3.10581	4.43698
12	1.35622	1.53796	1.58040	1.67387	1.78229	1.91231	2.07644	1.91231	2.17881	2.30272	2.46070	2.68100	3.05454	4.31779
13	1.35017	1.52992	1.57184	1.66408	1.77093	1.89887	2.06004	1.89887	2.16037	2.28160	2.43585	2.65031	3.01228	4.22083
14	1.34503	1.52310	1.56458	1.65578	1.76131	1.88750	2.04617	1.88750	2.14479	2.26378	2.41490	2.62449	2.97684	4.14045
15	1.34061	1.51723	1.55833	1.64865	1.75305	1.87774	2.03429	1.87774	2.13145	2.24854	2.39701	2.60248	2.94671	4.07277
16	1.33676	1.51213	1.55291	1.64246	1.74588	1.86928	2.02400	1.86928	2.11991	2.23536	2.38155	2.58349	2.92078	4.01500
17	1.33338	1.50766	1.54815	1.63703	1.73961	1.86187	2.01500	1.86187	2.10982	2.22385	2.36805	2.56693	2.89823	3.96513
18	1.33039	1.50371	1.54395	1.63224	1.73406	1.85534	2.00707	1.85534	2.10092	2.21370	2.35618	2.55238	2.87844	3.92165
19	1.32773	1.50019	1.54020	1.62797	1.72913	1.84953	2.00002	1.84953	2.09302	2.20470	2.34565	2.53948	2.86093	3.88341
20	1.32534	1.49704	1.53685	1.62415	1.72472	1.84433	1.99371	1.84433	2.08596	2.19666	2.33624	2.52798	2.84534	3.84952
21	1.32319	1.49419	1.53383	1.62071	1.72074	1.83965	1.98804	1.83965	2.07961	2.18943	2.32779	2.51765	2.83136	3.81928
22	1.32124	1.49162	1.53109	1.61759	1.71714	1.83542	1.98291	1.83542	2.07387	2.18289	2.32016	2.50832	2.81876	3.79213
23	1.31946	1.48928	1.52860	1.61476	1.71387	1.83157	1.97825	1.83157	2.06866	2.17696	2.31323	2.49987	2.80734	3.76763
24	1.31784	1.48714	1.52633	1.61217	1.71088	1.82805	1.97399	1.82805	2.06390	2.17154	2.30691	2.49216	2.79694	3.74540
25	1.31635	1.48517	1.52424	1.60979	1.70814	1.82483	1.97010	1.82483	2.05954	2.16659	2.30113	2.48511	2.78744	3.72514
26	1.31497	1.48336	1.52232	1.60760	1.70562	1.82186	1.96651	1.82186	2.05553	2.16203	2.29581	2.47863	2.77871	3.70661
27	1.31370	1.48169	1.52054	1.60558	1.70329	1.81913	1.96320	1.81913	2.05183	2.15782	2.29091	2.47266	2.77068	3.68959
28	1.31253	1.48014	1.51890	1.60371	1.70113	1.81659	1.96014	1.81659	2.04841	2.15393	2.28638	2.46714	2.76326	3.67391
29	1.31143	1.47870	1.51737	1.60197	1.69913	1.81424	1.95729	1.81424	2.04523	2.15033	2.28217	2.46202	2.75639	3.65941
30	1.31042	1.47736	1.51595	1.60035	1.69726	1.81205	1.95465	1.81205	2.04227	2.14697	2.27826	2.45726	2.75000	3.64596
40	1.30308	1.46772	1.50570	1.58871	1.68385	1.79631	1.93566	1.79631	2.02108	2.12291	2.25027	2.42326	2.70446	3.55097
60	1.29582	1.45820	1.49560	1.57723	1.67065	1.78085	1.91703	1.78085	2.00030	2.09936	2.22292	2.39012	2.66028	3.46020
100	1.29007	1.45067	1.48761	1.56817	1.66023	1.76866	1.90237	1.76866	1.98397	2.08088	2.20150	2.36422	2.62589	3.39049
200	1.28580	1.44508	1.48168	1.56144	1.65251	1.75963	1.89152	1.75963	1.97190	2.06723	2.18568	2.34514	2.60063	3.33984
500	1.28325	1.44175	1.47814	1.55743	1.64791	1.75425	1.88507	1.75425	1.96472	2.05912	2.17630	2.33383	2.58570	3.31009
1000	1.28240	1.44064	1.47696	1.55610	1.64638	1.75247	1.88293	1.75247	1.96234	2.05643	2.17319	2.33008	2.58075	3.30028

Conclusiones

Podemos concluir que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que, para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error. Dada una variable aleatoria con distribución Normal N(μ, σ), el objetivo es la construcción de un intervalo de confianza para el parámetro μ, basado en una muestra de tamaño n de la variable.

La construcción de intervalos específicos depende de las características de la población (normal o no, etc.), de los parámetros o combinaciones de parámetros a los que se les construye (media, varianza, proporción, coeficiente de correlación) tamaño muestral y parámetros poblacionales conocidos

En la inferencia estadística existen dos aproximaciones complementarias: pruebas de hipótesis y estimación. Las pruebas de hipótesis evalúan la probabilidad asociada a la hipótesis nula (H0) de que no hay efecto o diferencia. El valor de p obtenido refleja la probabilidad de rechazar la H0 siendo esta verdadera; en ningún caso prueba que la hipótesis alternativa, de que, si hay efecto o diferencia, sea verdadera. Error tipo I (α) es un falso negativo: rechazar H0 cuando esta es verdadera. Error tipo II (β) es un falso negativo, aceptar H0 la cuando esta es falsa.

La prueba de hipótesis para una media con varianza conocida, las suposiciones para esta prueba son mínimas. La población tiene una media $\boldsymbol{\mu}$ y una varianza $\boldsymbol{\sigma^2}$ conocida. El estadístico de prueba se basa principalmente en la media muestral. En cambio, sino se conocen las varianzas se estiman los grados de libertad de la estadística son k. Si $\boldsymbol{k \geq 30}$ el estadístico t tiene una distribución aproximadamente normal y su procedimiento es similar al caso donde se conocen las varianzas.

Bibliografía

Marco Antonio Triana (2009). Contraste de Hipótesis. UAO

Carlos Escalante Angulo (2010). Prueba de hipótesis frente a intervalos de confianza. Universidad Nacional.

Rafael Vargas Barrera (2008). Estadística II. Bogotá D.C: Escuela Superior de Administración Pública.

J. Humberto Mayorga A. (2004). Inferencia Estadística. Bogotá: Universidad Nacional de Colombia.

Erik Cobo, Belchin Kostov, Jordi Cortés, José Antonio González y Pilar Muñoz. (2014). Intervalos de confianza. UPC

Bibliografía Electrónica

https://sisbib.unmsm.edu.pe/bibvirtualdata/libros/Matematicas/inferencia/pdf/a03.pdf

https://moodle2.unid.edu.mx/dts_cursos_mdl/pos/ME/TD/AM/10/Pruebas_Media_Poblacional.pdf

http://terra.geociencias.unam.mx/~ramon/EstInf/Clase13.pdf

file:///C:/Users/admin/Downloads/Dialnet-PruebaDeHipotesisFrenteAIntervalosDeConfianza-5599153.pdf

http://www.mat.uda.cl/hsalinas/cursos/2008/spss/apuntes/practica4.pdf

http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/prueba_hipotesis_2012.pdf

http://www.esap.edu.co/portal/wp-content/uploads/2017/10/6-Estadstica-ii.pdf

https://www.ugr.es/~eues/webgrupo/Docencia/MonteroAlonso/estadisticaII/tema4.pdf

http://bdigital.unal.edu.co/48058/2/9587013743.PDF

http://verso.mat.uam.es/~pablo.fernandez/6-intervalos-curso17-18.pdf

Capítulo 2

PROYECTO DE INVESTIGACIÓN

"Aplicación del modelo de regresión logística en la predicción del rendimiento académico de estudiantes de la Universidad Nacional de Tumbes en el primer ciclo de estudios 2012"

Application of logistic regression model in predicting academic performance of students at National University of Tumbes in the first cycle of studies 2012.

RESUMEN

El rendimiento académico de los estudiantes de la Universidad Nacional de Tumbes es un problema que necesita ser estudiado debido a que la mayoría de los estudiantes ingresan con muchas deficiencias académicas que son mostradas en el examen de admisión (bajo rendimiento); así en el presente trabajo se elaboró un modelo estadístico que nos ayude a determinar la probabilidad de encontrar alumnos académicamente desaprobados en el primer ciclo académico y para esto se consideró un criterio de exigencia (promedio ponderado menor a once en escala de cero a veinte puntos) para clasificar como desaprobado a un estudiante en el primer ciclo de su carrera; esto es aplicando un modelo de regresión logística binaria calificado como modelo econométrico, se pudo observar que son cuatro variables las que influyen en el rendimiento académico, tales variables son: sexo, financiamiento de sus estudios, la escuela profesional, temas de matemática y economía llevados en la etapa escolar, todas son significativas al nivel del 5%. Las variables que incrementan la probabilidad de que un estudiante este desaprobado son: Sexo y Financiamiento de sus Estudios; y las que disminuyen este probabilidad son: Temas de Economía Llevados y la escuela profesional, además se puede observar que el punto de corte es de 0,5189 y es el mejor resultado de capacidad predictiva al obtenerse una Sensibilidad de 92,68% y Especificidad de 62,12%, teniendo en cuenta una efectividad del 83,5%, esto es calificación bueno. De otro lado, la variable relacionada con los estilos de aprendizaje fue la gran ausente en el modelo, lo cual posibilita rechazar la afirmación de que puedan tener una influencia específica sobre el rendimiento del estudiante de la Universidad Nacional de Tumbes durante su primer año de estudios.

Palabras clave: *rendimiento académico; regresión logística; modelo econométrico*

ABSTRACT

The academic performance of students at the Universidad Nacional de Tumbes is a problem that needs to be studied because most students enter with many academic deficiencies that are displayed in the entrance exam, so at work developed a statistical model to help determine the likelihood of finding academic students who failed the first academic year and this is considered a demanding criterion for classifying a student as deprecated in the first cycle of his career, this is by applying a binary logistic regression model that is an econometric model, it was observed that four variables are the variable influencing academic performance variables such as sex, financing their education, career, mathematics and economics issues brought into the school years, all significant at level of 5%. The variables that increase the likelihood that a student is deprecated are 2: Sex and finance their studies, and that this probability decreases are: Driven Economic Issues and representing the Specialty Code also shows that the cutoff is 0,5189 and the best result of predictive and obtained a sensitivity of 92,68% and 62,12% specificity, for which on aggregate its got a 83,5% rating this is good. On the other hand, the variable related to learning styles was largely absent in the model, making it possible to reject the claim that may have a specific influence on student performance of Tumbes National University during his freshman year .

Keywords: *academic performance,; logistic regression; econometric model*

Introducción

El rendimiento académico se constituye en un indicador del nivel de aprendizaje alcanzado por el estudiante, representa el nivel de eficacia en la consecución de los objetivos curriculares para las diversas asignaturas, y en nuestro país se expresa mediante un calificativo o promedio ponderado basado en el sistema vigesimal; es decir, las notas varían de 0 a 20 puntos, donde el puntaje de menor o igual a 10 es desaprobatorio; por ello, en el sistema educativo nacional e internacional es importante este indicador y existe la necesidad de estudiar los determinantes del rendimiento académico de los estudiantes de parte de las instituciones educativas; al igual que de las universidades, para de esa manera se logren diseñar políticas educativas que permita alcanzar, en los estudiantes, el más alto rendimiento académico y bajo nivel de deserción estudiantil.

Esquivel y Rojas (2005), muestran evidencia que en situaciones de aprendizaje y rendimiento; el elemento central es la motivación, está les permite alcanzar una meta y será sus éxitos o fracasos de su conducta como estudiantes.

Jadue (2003), indica que la familia tiene participación activa y un rol importante como mediador del aprendizaje, siendo la madre un fuerte predictor del rendimiento académico. Sin embargo, las familias de nivel socioeconómico bajo, deben lidiar con una serie de situaciones que dificultan su rol en el apoyo temprano del niño en el plano académico como la inestabilidad laboral, recursos materiales y financieros inadecuados, problemas maritales y familiares.

Abarca y Sánchez (2005), han encontrado, según lo expresado por los estudiantes, que un obstáculo clave en el bajo rendimiento académico son sus deficiencias en el perfil cognitivo, además de conocimientos disciplinarios insuficientes. Definen dos dificultades primordiales: no saben estudiar y no saben aprender. En el terreno de las habilidades de aprendizaje se localizan dos tipos: Competencias para pensar y comprender vs. Repetir y memorizar en las diversas disciplinas, y el conocimiento de técnicas de estudio, que sobre todo apuntan a cómo revisar la bibliografía y comprenderla.

Arlette y La Serna (2009), señalan entre sus principales resultados que, provenir de un colegio de “enseñanza tradicional”, es decir, que no es preuniversitario, incrementa el rendimiento académico en casi un punto. Ello se explicaría dado que los estudiantes de colegios tradicionales reciben una educación mucho más integral, cuyo objetivo no es solo asegurar el ingreso a la universidad, sino también desarrollar en ellos otras competencias

que terminarían contribuyendo indirectamente a un mejor rendimiento en la educación superior, como el arte, el deporte o los valores religioso-humanistas. No obstante, convendría realizar investigaciones más profundas para descubrir cada una de las raíces que explican la brecha entre los colegios preuniversitarios y los de enseñanza tradicional.

Ramirez (2004), refiere que los encargados de realizar la actividad docente asumen que las altas tasas de desaprobados se deben a fallas de formación de los estudiantes lograda en enseñanza media (etapa escolar), carencia de responsabilidad y de perseverancia del estudiante, problemas de inteligencia y carencia de aptitudes verbales y/o matemáticas.

En nuestra experiencia como docentes debemos decir que en la Universidad Nacional de Tumbes el rendimiento académico de los estudiantes en el primer ciclo afronta muchas deficiencias y mucho más en las áreas básicas como son matemática y lenguaje, además sabemos que la mayoría de los estudiantes ingresan con nota desaprobatoria, si se hace la equivalencia al sistema vigesimal los puntajes que obtienen en el examen de admisión.

Si bien el número de universidades ha aumentado, nada garantiza que la calidad de la oferta en educación superior haya crecido a la par o, al menos, se haya mantenido. Además, la brecha que existe entre la educación básica (escolar) y la universitaria es un problema evidente en Tumbes y en todo el país. Los jóvenes suelen llegar a los estudios superiores sin los conocimientos previos, las habilidades y las actitudes necesarias para aprovechar al máximo el proceso de aprendizaje. “Ello genera severas distorsiones en la actividad lectiva, sobre todo durante los primeros ciclos académicos, en los cuales se tiene que reforzar materias y competencias que deberían haber sido desarrolladas en la educación básica. Las repercusiones de estas deficiencias se observan en la universidad a través de indicadores tales como el alto número de repeticiones en los cursos básicos y el número de ciclos que emplean los alumnos para terminar su carrera” (MINEDU 2006, p18)

Por las razones mencionadas se ha elaborado un modelo de regresión logística que permite predecir el rendimiento académico antes de que este ocurra, para poder tomar las medidas preventivas, proponer cambios en el sistema de admisión a la universidad y realizar mejores políticas educativas en las instituciones de formación escolar y superior de la Región Tumbes.

Materiales y Métodos

La población corresponde a los ingresantes a la Universidad Nacional de Tumbes durante el primer semestre del año 2012 con un total de 460 estudiantes. Debido a que la universidad ofrece doce carreras profesionales, se decidió realizar un muestreo estratificado con lo cual se obtuvo el 50% de la población equivalente a 230 estudiantes. El único criterio de inclusión es el período de ingreso de los estudiantes, dado que este grupo presenta información académica, psicológica y socio-económica referida a las variables de interés. Para la construcción de la base de datos se aplicó una encuesta a los alumnos acerca de diversos aspectos de su currículo escolar mediante una ficha de información. Mediante este instrumento se hicieron preguntas orientadas a identificar el tipo de contenidos de los cursos de matemática y economía de la educación secundaria, la cantidad de alumnos en cada salón de clases, así como información respecto a lo de enseñanza preuniversitaria, información sobre la situación del hogar y su nivel socio-económico, financiamiento de sus estudios y de opinión familiar con respecto a los estudios universitarios.

Luego con la ayuda del programa estadístico Statical Package for the Social Sciences – SPSS en la versión número 20 de la IBM, se elabora un modelo de regresión logística que permite identificar los principales determinantes del rendimiento del primer ciclo de estudios de los mencionados alumnos. La codificación que por defecto efectúa el SPSS no ha tenido que realizar cambios internos dado que en la matriz de datos la variable dependiente se ha encontrado codificada como: Desaprobado (0), Aprobado (1).

El modelo incorpora un conjunto de veintiocho variables explicativas entre las que destacan aquellas referidas al rendimiento escolar del alumno y las características académicas de su colegio de procedencia (número de alumnos en aula, conocimientos básicos en matemática), sus indicadores psicológicos (estilos de aprendizaje), así como otras variables socio-familiares y de identificación (ingreso económico familiar entre otras).

Asimismo, se utilizó el método por pasos hacia adelante (razón de verosimilitud). Este método es automático ya que deja al programa que vaya introduciendo variables en el modelo, empezando por aquellas que tienen coeficientes de regresión más grandes, estadísticamente más significativos (Reyes 2007)

Resultados

1. Modelo de Regresión Logística

Se consideró la totalidad de los registros utilizados en la base de datos, en este caso contamos con 230 estudiantes lo cual nos dio una lectura de ningún dato perdido. A través de la regresión logística, se ha determinado que en el modelo siempre predice la categoría más común (Aprobado) es correcto para el 71.3% de la muestra (Cuadro 1). El modelo se incluye una constante, el valor de corte fue 0,500. Así se predijo correctamente a todos los estudiantes en situación académica Aprobado, pero ignora a todos los que presentan Desaprobado. Esto generó una línea base que se evaluó posteriormente.

Cuadro 1. Proporción de categoría más común de estudiantes Aprobado ingresantes a la Universidad Nacional de Tumbes durante el primer semestre del año 2012.

Observado			Pronosticado		
			Situación académica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 0	Situación académica	Desaprobado	0	66	0,0
		Aprobado	0	164	100,0
	Porcentaje global				71,3

2. Bloque 1: Por pasos hacia adelante (Razón de Verosimilitudes)

Los algoritmos por pasos encuentran un subgrupo de variables que maximicen la verosimilitud, en nuestro caso el análisis por pasos, con la selección hacia adelante y el estadístico razón de verosimilitud, tomó cinco pasos o iteraciones.

En el cuadro 03, se observa que en el modelo analizado en el Cuadro 01, el valor de -2LL es 275,726, este valor comparado con el obtenido en el paso cinco del bloque 1 del Cuadro 2, se reduce en 63 unidades, lo cual aumenta la efectividad del modelo.

Cuadro 2. Pruebas sobre los coeficientes del modelo.

		Chi cuadrado	gl	Sig.
Paso 1	Paso	23,834	1	0,000
	Bloque	23,834	1	0,000
	Modelo	23,834	1	0,000
Paso 2	Paso	13,738	1	0,000
	Bloque	37,572	2	0,000
	Modelo	37,572	2	0,000
Paso 3	Paso	14,541	1	0,000
	Bloque	52,114	3	0,000
	Modelo	52,114	3	0,000
Paso 4	Paso	7,537	1	0,006
	Bloque	59,650	4	0,000
	Modelo	59,650	4	0,000
Paso 5	Paso	4,003	1	0,045
	Bloque	63,653	5	0,000
	Modelo	63,653	5	0,000

La primera fila (PASO) es la que corresponde al cambio de verosimilitud. La segunda fila (Bloque) es el cambio en -2LL entre bloques de entrada sucesivos durante las construcciones del modelo. La tercera fila (Modelo) es la diferencia entre el valor de -2LL para el modelo sólo con la constante y el valor de -2LL para el modelo actual.

Los resultados de la prueba de bondad de ajuste de Hosmer y Lemeshow, en el cual se observa que la bondad de ajuste en el paso 5 es 13,508 y se distribuye como un valor de chi-cuadrada con una significancia de 0.096.

Cuadro 3. Historial de iteraciones

Iteración		-2 log de la verosimilitud	Coeficientes Constante
Paso 0	1	275,886	0,852
	2	275,726	0,910
	3	275,726	0,910

Cuadro 4. Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	43,111	7	0,000
2	21,815	8	0,005
3	12,664	7	0,081
4	22,793	8	0,004
5	13,508	8	0,096

Cuadro 5. Clasificación de proporción de categorías en cada iteración.

Observado			Pronosticado		
			Situación académica		Porcentaje correcto
			Desaprobado	Aprobado	
Paso 1	Situación académica	Desaprobado	3	63	4,5
		Aprobado	12	152	92,7
	Porcentaje global				67,4
Paso 2	Situación académica	Desaprobado	29	37	43,9
		Aprobado	13	151	92,1
	Porcentaje global				78,3
Paso 3	Situación académica	Desaprobado	34	32	51,5
		Aprobado	11	153	93,3
	Porcentaje global				81,3
Paso 4	Situación Académica	Desaprobado	40	26	60,6
		Aprobado	12	152	92,7
	Porcentaje global				83,5
Paso 5	Situación académica	Desaprobado	40	26	60,6
		Aprobado	12	152	92,7
	Porcentaje global				83,5

El cuadro 5, muestra que el porcentaje global de predicciones correctas es 83,5%, se observa también que del total de Estudiantes Desaprobados el 60,60% (Especificidad) son pronosticados correctamente y del total de Estudiantes Aprobados el 92,7% (Sensibilidad) se pronostican correctamente y en el global es de 83.5%, lo cual es de tener en cuenta porque el porcentaje global es altamente influenciado por la sensibilidad del modelo, esto traduce cuanta efectividad tiene el modelo para la investigación; asumir más del 80% es de clasificación bueno, por tal motivo se infiere que el modelo puede empezar a trabajar. Finalmente se puede observar en el cuadro 06 que el modelo se ha ajustado con un total de 4 variables de las 28 iniciales, todas ellas son significativas al nivel del 5%. Las variables que incrementan la probabilidad de encontrar un alumno en situación académica Desaprobado son dos: sexo y financiamiento de sus estudios con un factor de oportunidad de 3.294 y 3.067 respectivamente. Además, las variables que disminuyen la probabilidad de encontrar un alumno en situación académica desaprobado son dos: Temas de Economía Llevados en secundaria y la carrera profesional con un factor de oportunidad 0.687 y 0.828 respectivamente. Se consideró Código de carrera para efecto de procesar resultados, sin embargo, es importante decir que se refiera a la Carrera profesional.

Cuadro 06: Variables en la ecuación.

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1a	Código carrera	-0,194	0,042	21,791	1	0,000	0,823	0,759	0,893
	Constante	2,164	0,327	43,688	1	0,000	8,706		
Paso 2b	Sexo	1,174	0,322	13,272	1	0,000	3,233	1,720	6,079
	Código carrera	-0,172	0,042	17,014	1	0,000	0,842	0,775	0,913
	Constante	1,442	0,365	15,638	1	0,000	4,229		
Paso 3c	Sexo	1,280	0,338	14,304	1	0,000	3,597	1,853	6,984
	Financiamiento estudios	1,137	0,322	12,462	1	0,000	3,118	1,658	5,863
	Código carrera	-0,190	0,044	18,728	1	0,000	0,827	0,759	0,901
	Constante	-1,877	1,002	3,511	1	0,061	0,153		
Paso 4d	Sexo	1,135	0,348	10,614	1	0,001	3,110	1,571	6,154
	Financiamiento estudios	1,112	0,328	11,488	1	0,001	3,039	1,598	5,779
	Temas economía llevados	-0,360	0,132	7,466	1	0,006	0,698	0,539	0,903
	Código carrera	-0,189	0,044	18,342	1	0,000	0,828	0,759	0,902
	Constante	-0,967	1,071	0,816	1	0,366	0,380		
Paso 5e	Sexo	1,192	0,353	11,415		0,001	3,294	1,650	6,577

	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
Financiamiento estudios	1,121	0,329	11,598	1	0,001	3,067	1,609	5,846
Hijos del alumno	-23,099	40192,97	0,000	1	1,000	0,000	0,000	.
Temas economía llevados	-0,376	0,133	8,011	1	0,005	0,687	0,530	0,891
Código carrera	-0,189	0,044	18,044	1	0,000	0,828	0,759	0,903
Constante	-0,961	1,075	0,799	1	0,371	0,382		

a. Variable(s) introducida(s) en el paso 1: Código Carrera o Carrera profesional.

b. Variable(s) introducida(s) en el paso 2: Sexo.

c. Variable(s) introducida(s) en el paso 3: Financiamiento de sus Estudios.

d. Variable(s) introducida(s) en el paso 4: Temas de Economía Llevados.

e. Variable(s) introducida(s) en el paso 5: Hijos del Alumno.

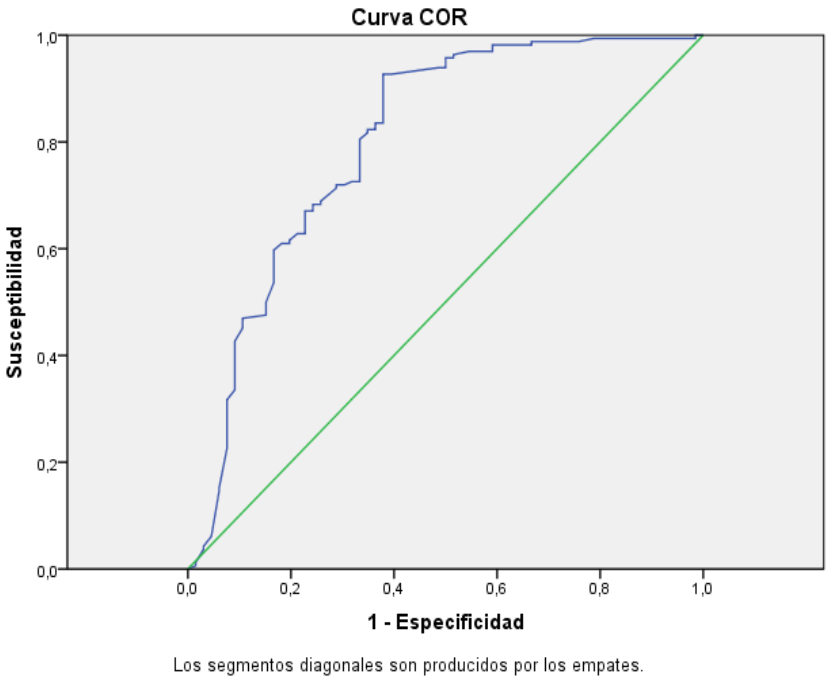
4.2.1 Curva COR (Curva Operativa de Rendimiento)

La validez del modelo se presenta en forma gráfica en la figura 1, donde la Curva COR está bastante alejada de la diagonal principal, por lo que podemos afirmar que el modelo brinda una buena calidad predictiva.

En el Cuadro 7 se aprecia que el área bajo la curva es 0,802 muy por encima del mínimo exigible para un método de diagnóstico.

El error estándar para esa estimación es 0,036 que multiplicado por 1,96 (nivel de confianza del 95,0%) que sumando y restando nos da el intervalo de confianza con un límite inferior de 0,732 y límite superior 0,872.

Figura 1. Curva Operativa de Rendimiento



De acuerdo a los resultados obtenidos se afirma que el área bajo la Curva COR es significativamente mayor que el mínimo exigible 0.5.

Cuadro 7. Variables resultado de contraste: Probabilidad pronosticada

La variable (o variables) de resultado de contraste: Probabilidad pronosticada tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

4.3 Modelo de Regresión Logística

De acuerdo a los resultados del modelamiento con regresión logística en SPSS mediante el método "Adelante", con el criterio de Razón de Verosimilitudes (método de contraste de los parámetros) se ha obtenido el siguiente modelo:

$$P(Estado = Desaprobado) = \frac{1}{1+e^{-(-0.961+1.192\times S+1.121\times F-0.376\times T-0.189\times C)}}$$

Donde:

S=Sexo del estudiante,

F=Financiamiento de sus estudios,

Área	Error tipa	Sig. Asintótica b	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
0,802	0,036	0,000	0,732	0,872

T=Temas llevados en economía durante la secundaria,

C=Código de su carrera o especialidad.

El modelo se ajusta correctamente de acuerdo a la prueba de Bondad de Ajuste de Hosmer y Lemeshow y además tiene un buen nivel de capacidad predictiva.

Con el modelo ya justificado podemos realizar algunos cálculos:

a) Un estudiante de contabilidad que financia sus estudios por medio de una beca y que llevo curso de economía con herramientas algebraicas tiene una probabilidad de desaprobar en su primer ciclo igual a 0,274

b) Una estudiante de Ing. Agrícola que financia sus estudios por medio de una beca y que llevo curso de economía con modelos gráficos tiene una probabilidad de desaprobar en su primer ciclo igual a 0,369

c) Un estudiante de hotelería y turismo que tiene trabajo propio y que llevo curso de economía con modelos gráficos tiene una probabilidad de desaprobar en su primer ciclo igual a 0,584

Discusión

Luego de analizar la relación entre sexo y rendimiento escolar, los estudios realizados en los países desarrollados suelen concluir que las mujeres rinden mejor que los hombres. En cambio, en los países latinoamericanos se ha encontrado que los varones tienen un mejor rendimiento (Arlette y La Serna 2009), la desventaja de las mujeres también es recurrente en las investigaciones desarrolladas en el Perú, existe deserción por motivos matrimoniales o de gestación; Cueto (2004) concluyó, que el atraso y la deserción escolar eran más probables en estudiantes de sexo femenino. En el modelo obtenido la variable sexo influye aumentando la probabilidad de que un estudiante tenga situación académica desaprobado, no se analiza si las mujeres tienen un mejor rendimiento académico.

De manera parcial se coincide con Jadue (2003), al constatar que las familias de nivel socioeconómico bajo afrontan una serie de situaciones que dificultan su rol en pro del apoyo en el plano académico, pues en nuestro estudio el financiamiento de los estudios influye en el rendimiento académico del estudiante, aumentando la probabilidad de encontrarse en el grupo de Desaprobados.

Los contenidos de las asignaturas de economía (materia que utiliza herramientas matemáticas) enseñados durante la etapa escolar también son de especial importancia (Castellanos 1998, Valdivieso 2002), ya que si el estudiante tiene conceptos definidos en algebra, geometría (modelos gráficos en planos cartesianos), sistema de ecuaciones, cálculo diferencial, bien definidos aumenta la probabilidad de tener un mejor rendimiento académico y posicionarse en el grupo de alumnos Aprobados.

Con respecto a la especialidad (código de la carrera), influye en el rendimiento académico (Brunner 2003) disminuyendo la probabilidad de estar en situación desaprobado, esto se puede entender en la diferencia de cursos de cada especialidad, como por ejemplo, no se dicta la asignatura de matemáticas en las escuelas académico profesional de derecho y obstetricia, esto permite que los estudiantes tengan un mayor promedio ponderado en comparación a los estudiantes con materias de matemáticas en su plan curricular. En el Perú, también se ha encontrado que los hombres tienden a rendir mejor que las mujeres en matemáticas (MINEDU 2004).

Se ha confirmado que el hecho de que un alumno posea un estilo de aprendizaje no acorde con la carrera elegida puede constituir un factor importante en el fracaso

académico (Tejedor 2003, Rojas 2006), en nuestro estudio la variable relacionada con los estilos de aprendizaje fue la gran ausente, por carecer de significancia para este modelo, se recomienda realizar otros estudios más específicos sobre esta variable.

Nuestra universidad debe considerar muy importante investigar el rendimiento académico de sus estudiantes; analizar los resultados del examen de admisión (Reyes, 2003) y (Torres, 2007), el origen de su institución educativa y estimar su rendimiento académico con el fin de crear planes de asesoría académica que les permita obtener un mejor rendimiento académico evitando que llegan a la cuarta matrícula y su fracaso universitario.

La variable estilos de aprendizaje fue la gran ausente en nuestro modelo, sin embargo, se recomienda crear una unidad académica de psicología donde se oriente al estudiante sobre la importancia de saber estudiar.

BASE DE DATOS: Puede revisar la base de datos y resultados en el siguiente enlace drive:

https://drive.google.com/drive/folders/16JjaiyrqoKRjAbI-BcTkU5cX-m4BeK16?usp=sharing

Conclusiones

El modelo de regresión logística es un buen procedimiento para predecir el éxito en el primer ciclo, pues el modelo ajustado a sólo cuatro variables de las 28 iniciales, se ha estimado que todas ellas son significativas al nivel del 5%.

Las variables que incrementan la probabilidad de tener un estudiante en situación académica desaprobado son dos: Sexo con un factor de oportunidad de 3,294 y Financiamiento de sus Estudios con un factor de oportunidad de 3,067. Las variables que disminuyen la probabilidad tener un estudiante en situación académica desaprobado son: Temas de Economía llevados en secundaria con un factor de oportunidad de 0,687 y carrera profesional con un factor de oportunidad de 0,828.

Los contenidos de los cursos de matemáticas y economía enseñados durante su etapa escolar también son de especial importancia, ya que, si el alumno llevó durante sus estudios secundarios cálculo diferencial y/o geometría analítica, su probabilidad de estar en situación académica desaprobatoria disminuye en alrededor de medio punto.

Referencias bibliográficas

Abarca, Alex, y Mario Sánchez. «Factores Generales que inciden en el rendimiento académico.» Revista electrónica Actualidades Investigativas en Educación. Universidad de Costa Rica, 2005.

Arlette, Barco, y Karlos La Serna. ¿Qué explica el rendimiento académico en el primer año de estudios universitarios? Un estudio de caso en la Universidad del Pacífico. Documento de Discusión, Lima: Centro de Investigación, Universidad del Pacifico, 2009.

Brunner, J, y G Elacqua. «Factores que inciden en una educación efectiva: Evidencia internacional.» Editado por Ediciones Universidad Diego Portales. Hevia, 2003.

Castellanos, Luis, Concepción González, Antonia González, y Isabel Manzano. Asociación española de profesores universitarios. 1998. http://www.uv.es/asepuma/VI/17.PDF.

Cueto, Santiago. «Factores Predictivos del Rendimiento Escolar Deserción e Ingreso a Educación Secundaria en una Muestra de Estudiantes de Zonas Rurales del Perú.» Archivos Analíticos de Políticas Educativas 12, nº 35 (2004).

Diaz, Mario. Modalidades de enseñanza centradas en el desarrollo de competencias orientaciones para promover el cambio metodológico en el marco del Espacio Europeo de Educación Superior. Oviedo: Ediciones Universidad de Oviedo, 2006.

Esquivel, Mar, y Roger Rojas. «Motivos de los estudiantes de nuevo ingreso para estudiar Post Grado en Educación.» Revista Iberoamericana de Educación, 2005.

Fernandez, Mario. «Factores de personalidad y rendimiento académico. Un estudio comparativo.» Cuadernos de Psicologia: Universidad de Lima, 1990: 127-200.

Gonzales, Mar, Pol Álvarez, Luis Cabreba, y Bethencourt J. «EL abandono de los estudiantes universitarios: factores determinantes y medidas preventivas.» Revsita española de pedagogía 65, nº 236 (2007): 71-86.

Jadue, George. «Transformaciones familiares en Chile: Riesgo creciente para el desarrollo emocional, psicosocial y la educación.» Estudios Pedagógicos, 2003.

Jimenez, César. «Educación familiar y alumnos con alto rendimiento.» Revista española de pedagogía 64, nº 234 (2006): 273-29.

MINEDU. Unidad de Medición de la Calidad Educativa. Informe, Ministerio de Educación, 2004.

MINEDU. Unidad de Medición de la Calidad Educativa. Lima: Ministerio de Educación, 2006.

Ministerio de Educación del Perú. «La universidad en el Perú. Razones para una reforma universitaria.» Documento de discusión., 2006.

Perez, Poll. Psicología Educativa. Lima: Editorial San Marcos, 2008.

Ramirez, P. Motivación y rendimiento académico en cuatro carreras del área biológica de la U. de Antofagasta. Seminario de título, Antofagasta- Chile: Facultad de Educación y Ciencias Humanas, 2004.

Reyes, Jimmy, Carlos Flores, Juan Duarte, y Pedro Ramírez. «Una aplicación del modelo de regresión logistica en la predicción del rendimiento estudiantil.» Estudios Pedagógicos 32, nº 2 (2007): 101-120.

Reyes, Juan. Relación entre el rendimiento académico, la ansiedad ante los examenes, los rasgos de personalidad, el autoconcepto y el asertividad en estudiantes del primer año de psicología de la UNMSM. Tesis, Lima: Editorial San Marcos, 2003.

Rodriguez, Abarca, y Mario. Sánchez. «Actualidades Investigativas en Educación.» Actualidades Investigativas en Educación, 2005: Vol. 5.

Rojas, G, R Salas, y C Jimenez. «Estilos de aprendizaje y estilos de pensamiento entre estudiantes universitarios.» Estudios Pedagógicos 1, nº 32 (2006): 49-75.

Sánchez, Oviedo. «Factores que influyen en el rendimiento académico del estudiante universitario.» TECNOCIENCIAChiguagua, 2011: 90-97.

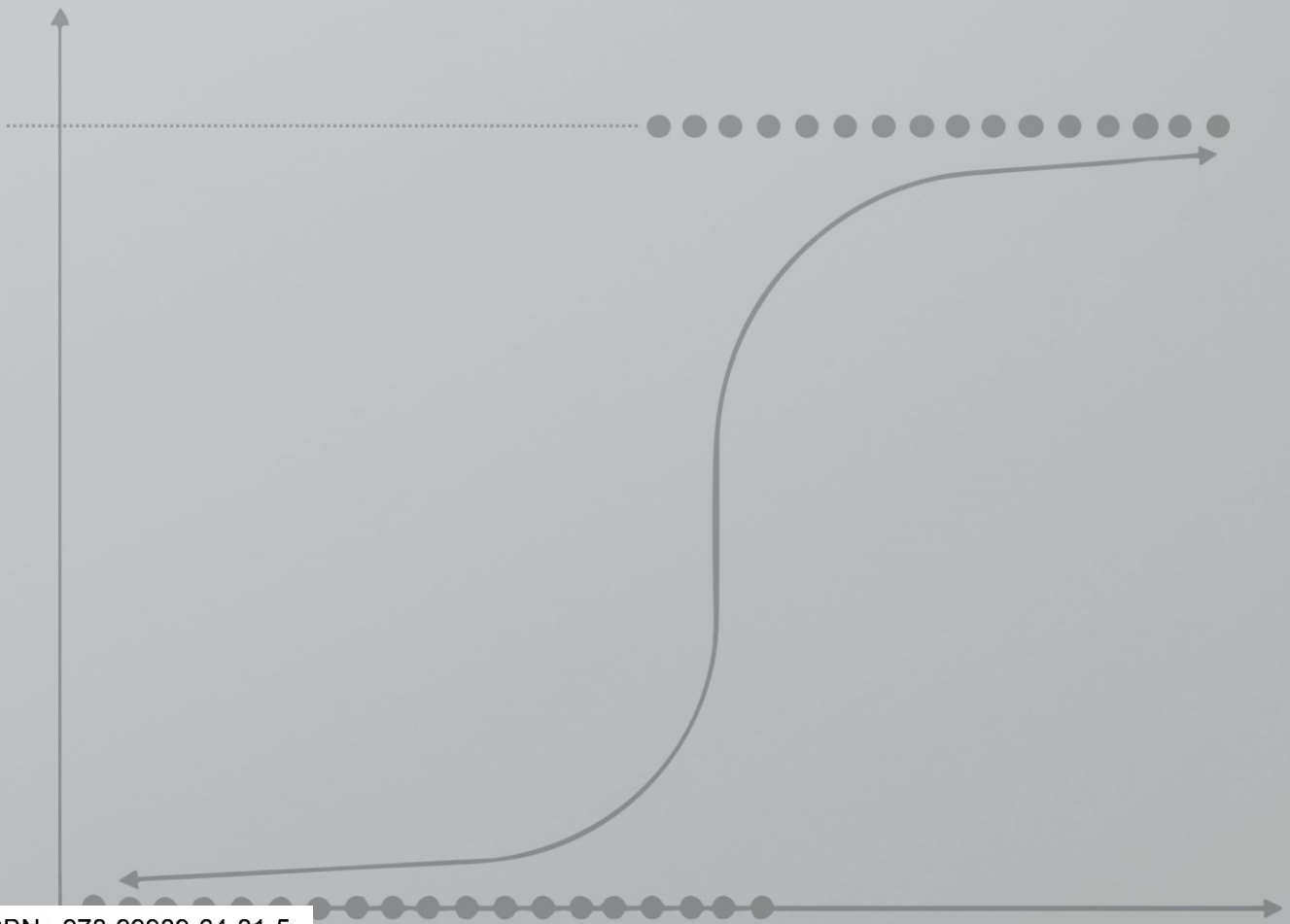
Tejedor, F. «El poder explicativo de algunos determinantes del rendimiento en los estudios universitarios.» Revista Española de Pedagogía 61, nº 224 (2003): 5-32.

Tejedor, F., y A. García. «Causas del bajo rendimiento del estudiante universitario.» Revista Educación, 2007: 443-473.

Torres, Marcos, Lajo Raúl, Eduardo Campos, Riveros, y Manuel. «Rendimiento académico de los alumnos de una facultad de educación de una universidad pública y su percepción de la calidad académica de los docentes.» Revsita de Investigación en Psicología 1, nº 10 (2007): 71-89.

Valdivieso, Marco, K Monar, y Mario Granda. «Análisis de los determinantes del rendimiento de los estudiantes de ESPOL.» Revista Tecnológica 17, nº 1 (2002): 213-218.

Septiembre 2023 - CID - Centro de Investigación y Desarrollo
Copyright © - CID - Centro de Investigación y Desarrollo
Copyright del texto © 2023 de Autores
Formato: PDF
Tamaño: A4 210 x 297 mm
Requisitos de sistema: Adobe Acrobat Reader
Modo de acceso: World Wide Web
biblioteca.ciencialatina.org
editorial@ciencialatina.org
Atención por WhatsApp al +52 22 2690 3834



ISBN: 978-99989-64-81-5