UNIVERSIDAD NACIONAL DE COLOMBIA - SEDE PALMIRA

NOTAS DE ESTUDIO DE ESTADISTICA Y PROBABILIDAD

Este documento tiene un resumen de los temas a evaluar en la asignatura Estadística y Probabilidad

Dr. Luis Eduardo Orobio Andrade
01/11/2010

TABLA DE CONTENIDO

I. CONCEPTOS BASICOS

DEFINICIÓN DE ESTADÍSTICA

Estadística es la ciencia que contiene un conjunto de métodos para obtener, recolectar, clasificar, registrar, presentar y analizar datos con el fin de investigar las relaciones existentes para en lo posible establecer normas, leyes y principios

CLASIFICACION DE LA ESTADISTICA

Ejemplo se pretende encontrar el personal adecuado para ingresar a una empresa recién constituida: pasos para lograr este objetivo:

ESTADISTICA DESCRIPTIVA

- Obtener : Colocar un clasificado ; solución a mi problema
- Recolectar: Recibir hojas de vida para los cargos solicitado
- Clasificar: Evaluar el perfil de las personas para los cargos
- Registrar: Se definen las personas a ser contratadas
- Presentar: Presentarse en la empresa a entrevista
- Analizar: De acuerdo a la entrevista, analizar el cargo de la personas

ESTADISTICA INFERENCIAL

- Leyes: Los hombres deben tener libreta militar para emplearse
- Normas: Cumplir los horarios
- Principios: Ser sociable con sus compañeros

CONCEPTOS BÁSICOS DE ESTADISTICA

1. Fuentes de información
    i. *Primaria:* es aquella que se toma directamente para el estudio en cuestión. Ejemplo una encuesta, una entrevista etc.
    ii. *Secundaria*: es aquella que se toma en el estudio pero ha sido recolectada con otros propósitos. Ejemplo información del DANE, Cámara de Comercio etc.

2. Población: conjunto de elementos de interés en un estudio
    i. *Estudiada*: es aquella sobre la cual se van a extrapolar los resultados o sobre la cual se va hablar
    ii. *Accequible*: es aquella de fácil acceso al estudio

3. Muestra: es una parte representativa de la población

    i. *Muestreos probabilisticos* : es necesaria la estadística para tomar decisiones y se pueden generalizar los resultados
    ii. *Muestreos no probabilisticos* : es aquella en la cual no se pueden generalizar los resultados

4. Parámetro: Cualquier característica medible de una población

5. Estadístico: cualquier característica medible de una muestra

6. Variable: cualquier característica que se manifiesta a través de dos o mas modalidades

7. Constante: cualquier característica que se manifiesta a través de una sola modalidad

8. Confiabilidad: si podemos esperar en forma razonable que los resultados de dicha medición sean sistemáticamente precisos.

9. Validez: si mide lo que en realidad trata de medir.

10. Observación: es el conjunto de modalidades o valores de cada variable estadística medidos en un mismo individuo

11. Individuo: cada uno de los elementos que componen la población estadística

12. Caso: toda aquella información que se le puede extraer a un individuo, entidad o cosa

II. CLASIFICACION DE VARIABLES

La clasificación de variables es importante para la presentación y análisis de datos.

Esta clasificación de puede dar de acuerdo a su característica, nivel de medición y característica como se presenta a continuación:

VARIABLE CUALITATIVA

Variable cualitativa: es aquella característica que se manifiesta a través de atributos o cualidades y no de pueden ordenar. Ej. Estado civil, sexo, color, tipo de sangre etc.

Variable cualitativa nominal excluyente: es aquella característica que se manifiesta a través de atributos o cualidades, no se pueden ordenar y entre muchas opciones de respuesta solamente se pertenece a una. Consiste en categorías mutuamente excluyentes que no implican ningún orden. Ej. Estado civil, numero de documentos de identidad; etc.

Variable cualitativa nominal no excluyente: es aquella que se manifiesta a través de atributos o cualidades, no se pueden ordenar y tiene dos o más opciones de respuesta. Consiste en categorías mutuamente no excluyentes que no implican ningún orden. Ej. Tipos de gaseosas que acostumbra a tomar, razones de compra en un almacén etc.

Variable cualitativa nominal dicotomica: es aquella característica que se manifiesta a través de atributos o cualidades, no se pueden ordenar y solamente tiene dos opciones de respuesta. Ej. Genero, Si/No etc.

Variable cualitativa nominal dicotomizada: son aquellas características de cualquier tipo que se llevan a un éxito (Si cumple la condición buscada) o fracaso (Si no cumple la condición buscada). Ej. Encontrar personas mayores de edad, encontrar personas viudas, etc.

Variable cualitativa ordinal: es aquella característica que se manifiesta a través de atributos o cualidades y siempre se pueden ordenar. Ej. Estrato, rango docente etc.

VARIABLE CUANTITATIVA

Variable cuantitativa: Son aquellas características que se expresan mediante cantidades numéricas. Ej. Ingresos, gastos, peso etc.

Variable cuantitativa de razón: son aquellas características que se manifiestan numéricamente para las cuales las distancias entre números tienen tamaño constante y conocido. Y además la razón entre números tiene algún significado ya que existe un cero natural. Ej. Gastos, ingresos etc.

Variable cuantitativa de razón Discreta: son aquellas características que se manifiestan numéricamente y nunca admiten una modalidad intermedia entre dos características de sus modalidades. Ej. Numero de hijos, numero de objetos.

Variable cuantitativa de razón Continua: son aquellas características que se manifiestan numéricamente y siempre admiten una modalidad intermedia entre dos característica. Ej. Pesos, estaturas etc.

Variable cuantitativa de intervalos: son aquellas características que se manifiestan numéricamente y tienen un cero convencional; es decir puede tomar valores positivos y negativos. Ej. Temperatura, opinión, actitud, etc.

Variable cuantitativa de intervalos comparativa: es aquella en la cual se compara un producto con respecto a otro. Ejemplo. Que tal le parece la comida del restaurante A con respecto al restaurante B. (Mejor, Igual o Peor)

Variable cuantitativa de intervalos: es aquella en la cual se evalúa la opinión sobre alguna afirmación: (Muy de acuerdo, de acuerdo, Indiferente, En desacuerdo y Muy en desacuerdo)

Medición: es una asignación de números u otros signos a las características de los objetos de acuerdo con ciertas reglas especificadas con anterioridad

Escala**:** es una generación de un continuo en los que se colocan los objetos de medida

ETAPAS DE LAS CLASIFICACIÓN DE VARIABLES

De acuerdo a su característica: Cualitativas y cuantitativas

De acuerdo a su nivel de medición: Cualitativas nominales y ordinales o Cuantitativas de razón o intervalos

De acuerdo a su naturaleza: Cuantitativas de razón discretas o continuas, cuantitativas de intervalos comparativa y no comparativas

LAS ESCALAS DE MEDICIÓN SON ACUMULATIVAS:

$$\text{Razón} \subset \text{Intervalo} \subset \text{Ordinal} \subset \text{Nominal}$$

Ejemplo: Se desea tener una idea aproximada acerca del valor (dinero) del carro típico que poseen los profesores de una Universidad.

1. **Población**: conjunto de todos los carros de los profesores de la Universidad.
2. **Muestra**: es una proporción o parte de una población. Profesores del Depto. de Ingeniería
3. **Variable**: es el valor real de cada carro.
4. **Dato:** valor de un carro en particular. Carro del Profesor Ávila es de $15.000.000
5. **Datos:** conjunto de valores que corresponden a la muestra obtenida.
6. **Experimento:** método para obtener la información
7. **Parámetro:** es el valor promedio en la población.
8. **Valor Estadístico**: es el valor "promedio de la muestra".

EJERCICIOS DE COMPLEMENTO

1. Una fabrica actualmente cuenta con 400 empleados y desea ofrecer a los mismos un servicio de guardería, el cual posiblemente se instale a un kilómetro de distancia de la fábrica. Suponga que a usted lo encargan de realizar un estudio de las necesidades que los empleados tienen al respecto.

    a. Fuente de información a utilizar
    b. La población delimitándola claramente
    c. Trabajaría usted con una muestra o con una población. Justifique
    d. Cite cuatro características o variables pertinentes de investigar

2. Un fabricante de medicamentos veterinarios está interesado en la proporción de animales que padecen infecciones locales cuya condición puede ser controlada por un nuevo producto desarrollado por la empresa. Se condujo un estudio en el que participaron 5000 animales que padecen infecciones locales y se encontró que en el 80% de los animales se puede controlar la infección con el medicamento. Suponiendo que los 5000 animales son representativos del grupo de animales con infecciones locales, conteste las siguientes preguntas:

    a. Fuente de información utilizada
    b. Cuál es la población?
    c. Cuál es la muestra?
    d. Identifique el parámetro de interés
    e. Identifique el estadístico y proporcione su valor
    f. Se conoce el valor del parámetro?

3. Se ha hecho un estudio para determinar la preferencia de una marca especial de detergente por parte de las amas de casa. Entre las 50 amas de casa entrevistadas, 30 dijeron que preferían esta marca.

    a. ¿Qué constituye la muestra?.
    b. ¿Qué constituye la población?.
    c. ¿Cuál es la proporción, dentro de la muestra, de las amas de casa que prefieren la marca del detergente?

4. Una fabrica actualmente cuenta con 400 empleados y desea ofrecer a los mismos un servicio de guardería, el cual posiblemente se instale a un kilómetro de distancia de la fábrica. Suponga que a usted lo encargan de realizar un estudio de las necesidades que los empleados tienen al respecto.

    a. Fuente de información a utilizar
    b. La población delimitándola claramente
    c. Trabajaría usted con una muestra o con una población. Justifique
    d. Cite cuatro características o variables pertinentes de investigar

5. Comenta los errores cometidos al realizar los siguientes muéstreos:

a. Para hacer un estudio sobre el número de descendientes en las familias de una ciudad, tomo la muestra preguntando a los estudiantes de varios colegios,
b. Para realizar una encuesta sobre las novelas más vendidas pregunto a profesores de literatura, que son expertos en el tema,
c. Para realizar una encuesta sobre gustos del consumidor, una empresa de refrescos realizó las muestras dando a probar, a ciegas, entre el refresco de su marca y otro refresco sin determinar.

6. Clasifica las siguientes variables de acuerdo a su característica, nivel de medición y naturaleza:

    a. Carreras que se estudian en la U.S.B.
    b. Numero de cartas que se escriben en un mes
    c. Número de calzado
    d. Precio de un producto.
    e. Marcas de cerveza que conoces
    f. Numero de empleados de una empresa
    g. Altura
    h. Temperatura de un enfermo
    i. Distancia recorrida por cada estudiante para ir de su casa a la universidad
    j. Llamadas que llegan a central telefónica de la universidad en un día
    k. Preferencia por cierta marca de refresco
    l. Color del cabello de los estudiantes que toman el curso de estadística
    m. Talla de una camiseta Apolo
    n. Expectativa que tiene sobre la próxima capacitación Ecaes
    o. Porcentaje de estudiantes de la universidad que les gusta las asignaturas de economía
    p. Porcentaje de estudiantes que poseen moto

7. Se ha hecho un estudio para determinar la preferencia de una marca especial de detergente por parte de las amas de casa. Entre las 50 amas de casa entrevistadas, 30 dijeron que preferían esta marca.

    d. ¿Qué constituye la muestra?.
    e. ¿Qué constituye la población?.
    f. ¿Cuál es la proporción, dentro de la muestra, de las amas de casa que prefieren la marca del detergente?

8. En una fiesta, el 50% de los invitados son hombres. De todos los hombres de la fiesta, el 40% son calvos y de ellos el 50% habla inglés. Si 4 calvos hablan inglés. ¿Cuántas mujeres hay en la fiesta?.

9. Efectuar dos descuentos consecutivos, primero de un 10% y luego de un 20%, es equivalente a efectuar un solo descuento de...

10. Si Pedro tuviera un 15% menos de la edad que tiene, tendría 34 años. Hallar su edad actual.

11. En una población estudiada, hay 2000 mujeres y 8000 hombres. Si queremos seleccionar una muestra de 250 individuos en dicha población. ¿Cuántos deberán ser mujeres para que la muestra sea considerada representativa?

12. Indique, justificadamente, si las siguientes afirmaciones son verdaderas o falsas.

    a. Una población es una colección de algunos de los elementos en estudio.
    b. Los tamaños de los intervalos de una distribución de frecuencias deben ser siempre de igual tamaño.
    c. La mediana es un indicador que se ve afectado por la magnitud de los valores extremos.
    d. Cuando se trabaja con datos agrupados, podemos calcular una media aritmética aproximada si suponemos que cada valor de una clase dada es igual a su punto medio.
    e. Los percentiles que dividen los datos en 5 grupos de igual frecuencia se conocen como cuartiles.

III.SUMATORIA

DEFINICIÓN DE SUMATORIA

$$\sum_{i=1}^{n} X_i = X_1 + X_2 \quad + X_3 \quad + \ldots\ldots + X_n$$

Indica la suma de todos los X_i desde i = 1, 2,....., n

Donde: n ~ Limite superior de la sumatoria
$\sum$ ~ Símbolo de la sumatoria
X_i ~ Elemento genérico de la sumatoria
i ~ Limite inferior de la sumatoria

PROPIEDADES DE LA SUMATORIA

a. $\sum_{i=1}^{n} K$ = n K K es una constante

b. $\sum_{i=p}^{n} K$ = (n - p + 1) K p ≠ 1

c. $\sum_{i=1}^{n} KX_i$ = K $\sum_{i=1}^{n} X_i$

d. $\sum_{i=1}^{n} [X_i \pm Y_i] = \sum_{i=1}^{n} X_i \pm \sum_{i=1}^{n} Y_i$

e. $\sum_{i=1}^{n} {X_i}^2 \neq \left[\sum_{i=1}^{n} X_i\right]^2$

f. $\sum_{i=1}^{n} i = \left[\frac{n(n+1)}{2}\right]$

g. $\sum_{i=1}^{n} i^2 = \left[\frac{n(n+1)(2n+1)}{6}\right]$

h. $\sum_{i=1}^{n} i^3 = \left[\frac{n(n+1)}{2}\right]^2$

EJEMPLOS CON SUMATORIAS

Se tienen los valores en la siguiente tabla

Valor i	X_i	Y_i
1	5	3
2	2	7
3	12	8
4	9	10
5	14	18

Realice los siguientes ejercicios aplicando las propiedades de la sumatoria:

a. $\sum_{i=1}^{4} X_i = X_1 + X_2 + X_3 + X_4 = 5 + 2 + 12 + 9 = 28$

b. $\sum_{i=3}^{4} X_i = X_3 + X_4 = 12 + 9 = 21$

c. $\sum_{i=1}^{5} 2 = 5\ (2) = 10$

d. $\sum_{i=3}^{10} 2 = [\ 10 - 3 + 1\ ]\ (2) = 16$

e. $\sum_{i=1}^{5} 7X_i = 7\sum_{i=1}^{5} X_i = 7\ [X_1 + X_2 + X_3 + X_4 + X_5\ ]$

$= 7\ [\ 5 + 2 + 12 + 9 + 14\ ] = 7\ [42] = 294$

f. $\sum_{i=1}^{4} [X_i - Y_i - 2] = \sum_{i=1}^{4} X_i - \sum_{i=1}^{4} Y_i - \sum_{i=1}^{4} 2$

$= [\ X_1 + X_2 + X_3 + X_4\ ] - [\ Y_1 + Y_2 + Y_3 + Y_4\ ] - 4\ (2)$

$= [\ 5 + 2 + 12 + 9\ ] - [\ 3 + 7 + 8 + 10\ ] - 4\ (2)$

$= 28 - 28 - 8 = -\ 8$

g. $\sum_{i=3}^{4}[X_i - Y_i - 2]$ **=** $\sum_{i=3}^{4} X_i$ **-** $\sum_{i=3}^{4} Y_i$ **-** $\sum_{i=3}^{4} 2$

= [X_3 + X_4] - [Y_3 + Y_4] – [4 – 3 + 1] (2)

= [12 + 9] - [8 + 10] – 2 (2)

= 21 – 18 – 4 = - 1

h. $\sum_{i=1}^{5} X_i^{2} \neq [\sum_{i=1}^{5} X_i]^2$ **= [$X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2$] $\neq$ [$X_1 + X_2 + X_3 + X_4 + X_5$]^2**

= [$5^2 + 2^2 + 12^2 + 9^2 + 14^2$] $\neq$ [5 + 2 + 12 + 9 + 14]^2

= [$5^2 + 2^2 + 12^2 + 9^2 + 14^2$] $\neq$ [5 + 2 + 12 + 9 + 14]^2

= 450 $\neq$ [42]^2

= 450 $\neq$ 1764

i. $\sum_{i=5}^{8}[i-2]^3$ **=** $\sum_{i=5}^{8}[i^3 - 6i^2 + 12i - 4]$

= $\sum_{i=5}^{8} i^3$ **-** $\sum_{i=5}^{8} 6i^2$ **+** $\sum_{i=5}^{8} 12i$ **-** $\sum_{i=5}^{8} 4$

= [$\sum_{i=1}^{8} i^3$ **-** $\sum_{i=1}^{4} i^3$ **] - [** $\sum_{i=1}^{8} 6i^2$ **-** $\sum_{i=1}^{4} 6i^2$ **] + [** $\sum_{i=1}^{8} 12i$ **-** $\sum_{i=1}^{4} 12i$ **] - [** $\sum_{i=5}^{8} 4$ **]**

= [$\left[\frac{8(8+1)}{2}\right]^2$ **-** $\left[\frac{4(4+1)}{2}\right]^2$ **] - 6 [** $\left[\frac{8(8+1)(2*8+1))}{6}\right]$ **-** $\left[\frac{4(4+1)(2*4+1)}{6}\right]$ **]**

+ 12 [$\left[\frac{8(8+1)}{2}\right]$ **-** $\left[\frac{4(4+1)}{2}\right]$ **] - [8 – 5 + 1] (4)**

= [1296 – 324] – 6 [204 – 30] + 12 [36 – 10] – 16

= 972 – 1044 + 312 – 16

= 224

EJERCICIOS DE COMPLEMENTO

1. Si $X_1=4$; $X_2=8$; $X_3=10$; $X_4=12$; $X_5=15$; $X_6=5$, $X_7=4$; $X_8=14$; $X_9=2$. Resuelva la siguiente sumatoria, aplicando las diferentes propiedades de sumatorias:

a. $\left[\sum_{i=6}^{9}\left(X_9+X_i+3+i^2\right)\right]^2$

b. $\sum_{i=3}^{5}[2X_i-4Y_i]^2$

c. $\sum_{i=5}^{8}[3i+1]^2$

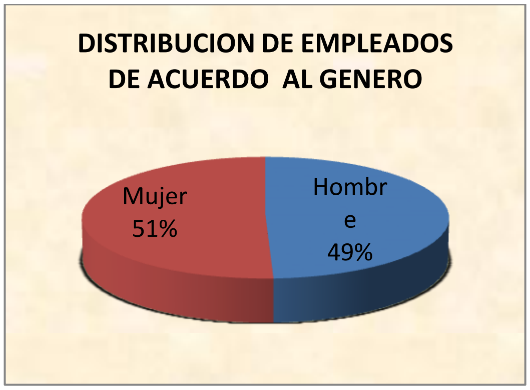
d. $\sum_{i=1}^{5}X_i[X_i-1]^2$

2. Simplifique la siguiente expresión utilizando las propiedades

$$\sum_{i=1}^{n}\left(X_i+a\right)^2-\sum_{i=1}^{n}\left(X_i-a\right)^2$$

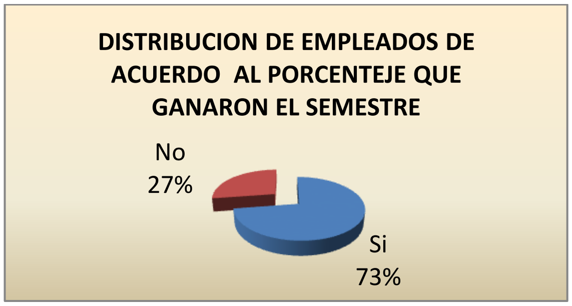
IV. PRESENTACION DE DATOS

PRESENTACIÓN DE DATOS CUALITATIVOS



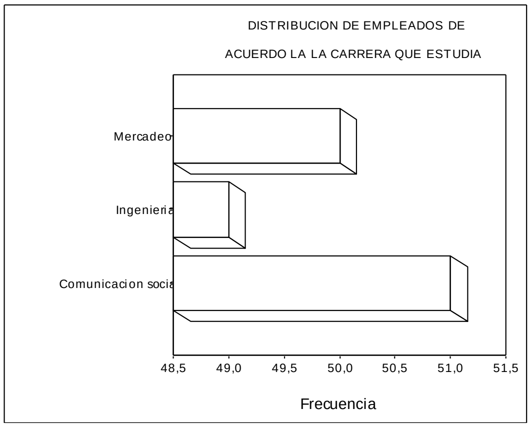
Variables cualitativa nominal dicotómica

	Genero	
	Recuento	%
Hombre	74	49,3%
Mujer	76	50,7%
Total	150	100,0%



Cuantitativa nominal dicotomizada

	Gano el semestre	
	Recuento	%
No	41	27,3%
Si	109	72,7%
Total	150	100,0%



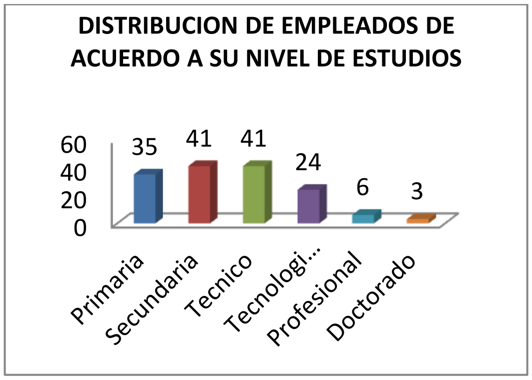
Variable cualitativa nominal excluyente

	Carrera que estudia	
	Recuento	%
Mercadeo	50	33,3%
Ingenieria	49	32,7%
Comunicacion social	51	34,0%
Total	150	100,0%

Cualitativa ordinal

Nivel de estudios

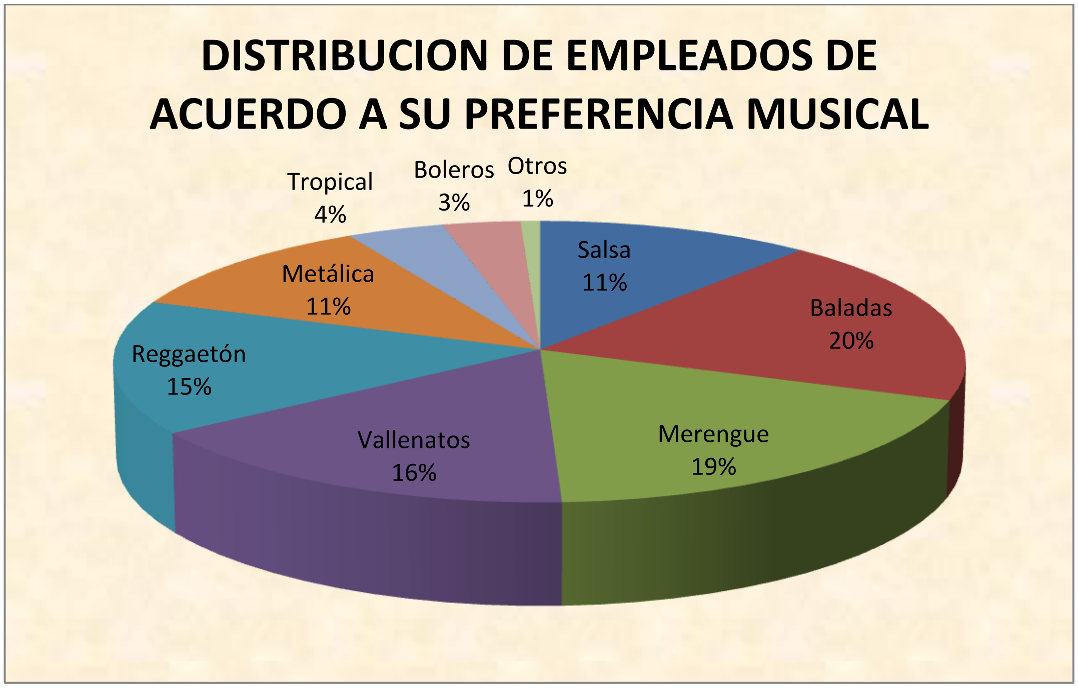
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Primaria	35	23,3	23,3	23,3
	Secundaria	41	27,3	27,3	50,7
	Tecnico	41	27,3	27,3	78,0
	Tecnologico	24	16,0	16,0	94,0
	Profesional	6	4,0	4,0	98,0
	Doctorado	3	2,0	2,0	100,0
	Total	150	100,0	100,0	

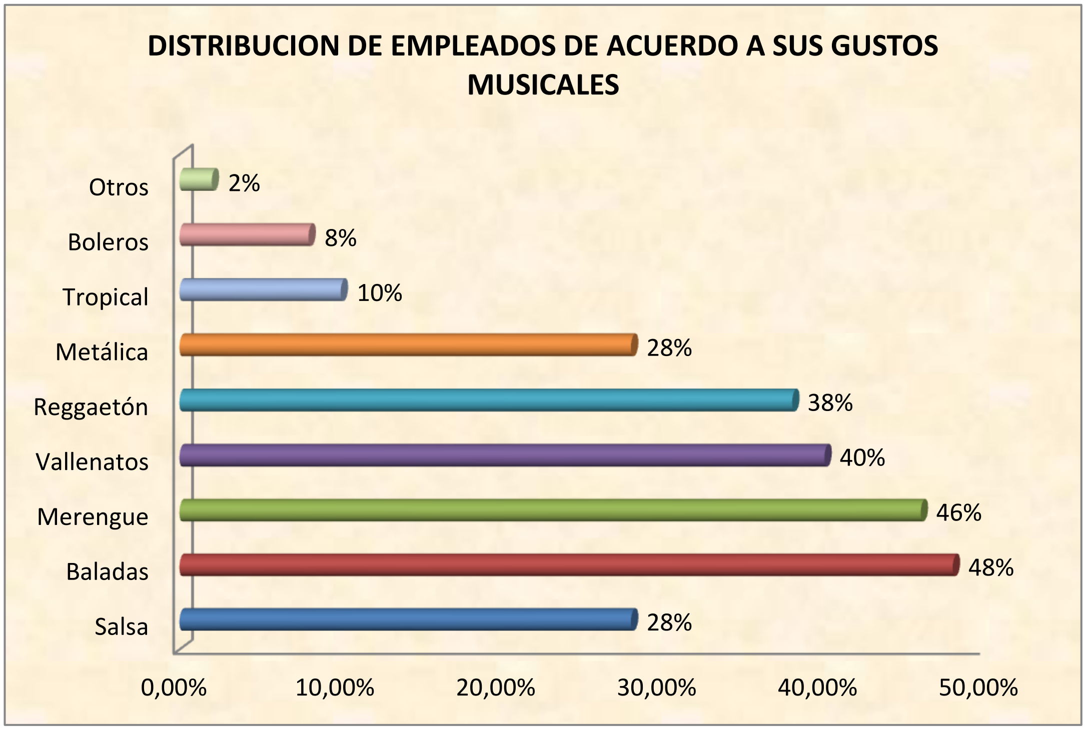


Cuantitativa nominal no excluyente

		Responses		Percent of Cases
		N	Percent	
TIPOS DE MUSICA QUE ACOSTUMBRA A ESCUCHAR	Salsa	14	11,3%	28,0%
	Baladas	24	19,4%	48,0%
	Merengue	23	18,5%	46,0%
	Vallenatos	20	16,1%	40,0%
	Reggaetón	19	15,3%	38,0%
	Metálica	14	11,3%	28,0%
	Tropical	5	4,0%	10,0%
	Boleros	4	3,2%	8,0%
	Otros	1	,8%	2,0%
Total		124	100,0%	248,0%

- La preferencia por las baladas de los empleados estudiados es de un 19.4%
- E1 40% de los empleados estudiados les gustan los vallenatos
- El numero esperado de preferencias musicales de los empleados estudiados es de aproximadamente dos tipos especialmente las baladas y el merengue como las dos principales


DISTRIBUCION DE EMPLEADOS DE ACUERDO A SU PREFERENCIA MUSICAL
Tropical 4%
Boleros 3%
Otros 1%
Salsa 11%
Metálica 11%
Baladas 20%
Reggaetón 15%
Vallenatos 16%
Merengue 19%

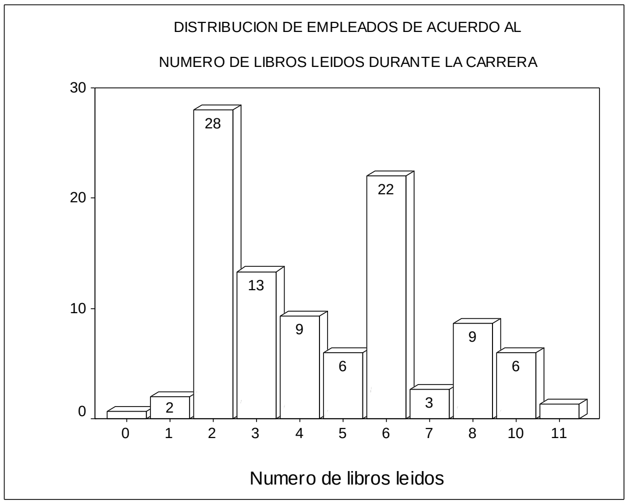

DISTRIBUCION DE EMPLEADOS DE ACUERDO A SUS GUSTOS MUSICALES
Otros 2%
Boleros 8%
Tropical 10%
Metálica 28%
Reggaetón 38%
Vallenatos 40%
Merengue 46%
Baladas 48%
Salsa 28%
0,00%
10,00%
20,00%
30,00%
40,00%
50,00%

PRESENTACIÓN DE DATOS CUANTITATIVOS

Cuantitativa de razón discreta

Numero de libros leidos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	1	,7	,7	,7
	1	3	2,0	2,0	2,7
	2	42	28,0	28,0	30,7
	3	20	13,3	13,3	44,0
	4	14	9,3	9,3	53,3
	5	9	6,0	6,0	59,3
	6	33	22,0	22,0	81,3
	7	4	2,7	2,7	84,0
	8	13	8,7	8,7	92,7
	10	9	6,0	6,0	98,7
	11	2	1,3	1,3	100,0
	Total	150	100,0	100,0	



Cuantitativa de razón datos brutos

Descriptivos

			Estadístico
Cuantas horas trabaja usted semanalmente	Media		43,03
	Intervalo de confianza para la media al 95%	Límite inferior	41,90
		Límite superior	44,15
	Media recortada al 5%		42,70
	Mediana		40,00
	Varianza		236,020
	Desv. típ.		15,36
	Mínimo		2
	Máximo		98
	Rango		96
	Amplitud intercuartil		9,00
	Asimetría		,532
	Curtosis		2,527

Construcción de una tabla de frecuencias de una variable cuantitativa de razón continúa datos

Pasos a seguir para la construcción de una tabla de frecuencias

.

a. Encontrar el rango de la serie de datos

   Rango = Máximo valor – Mínimo valor

b. Designar el número de intervalos a realizar: indicar en cuantas categorías se van a agrupar los datos.

   Estadísticamente la única limitante que se tiene es:

$$5 \le n \le 15$$

   Realizar muchos o pocos intervalos depende del estudio y nivel de precisión que se desea en las estimaciones

   Estos son formulas que me indican cuantos intervalos realizar, cuando no se tiene idea y depende del numero de datos.

   Método I.

   K : Numero de intervalos a realizar
   n : Numero de datos

$$K = \frac{Ln(n)}{Ln(2)}$$

   Método II

n	K
n < 50	5 a 7
50 < n < 100	6 a 10
100 < n < 250	7 a 12
n > 250	10 a 15

   Método III

$$K = 1 + 3.3 \text{ Log}(n)$$

c. Calcular la amplitud de cada intervalo

$$AI = \frac{(Rango + a)}{K}$$

El valor de “ a “ depende de los datos utilizados para el análisis

Datos	**Valor de a**
Enteros (10)	1
Si tienen un decimal (10.3)	0.1
Si tienen dos decimales (10.32)	0.01
Si tienen tres decimales (10.236)	0.001
Sucesivamente	……….

d. Calculo de los intervalos reales: son intervalos cerrados a la izquierda y abiertos a la derecha que solamente me sirven para realizar cálculos

[a – b)

e. Calculo de los intervalos de clase: Son intervalos cerrados la izquierda y cerrados a la derecha y solamente me sirven para interpretar

[a – b]

f. Calculo de la marca de clase: es el punto medio del intervalo real

$X' = \frac{(a+b)}{2}$; siendo “ a “ el limite inferior del intervalos y “ b “ el limite superior del intervalo

g. Método de tallos y hojas o diagrama de punto: es una forma grafica de organizar los datos para contar en numero de elementos que pertenecen a un intervalo determinado

h. Tabla de frecuencias

Ejemplo de construcción de una tabla de frecuencias.

Con base en la información de la edad de 50 empleados de una empresa caleña, construya la tabla de frecuencias utilizando los pasos anteriores:

32 38 26 29 32 41 28 31 45 26 45 35 40 30 31 40 37 33 28 30 30 41 39 38 33 35 31 36 37 32
25 39 38 46 26 32 27 23 37 34 23 28 25 28 41 37 29 32 33 38

Solución:

Pasos a seguir:

a. Encontrar el rango de la serie de datos

Rango = 46 – 23 = 23

b. *Designar el número de intervalos a realizar*

Método I.

K : Numero de intervalos a realizar
n : 50
K = (Ln 50 / Ln 2) = 5.64 ≅ 6

c. *Calcular la amplitud de cada intervalo*

$$AI = \frac{(Rango + a)}{K} = \frac{(23+1)}{6} = 4$$

d. *Calculo de los intervalos reales*

[23 – 27)
[27 – 31)
[31 – 35)
[35 – 39)
[39 – 43)
[43 – 47)

e. *Cálculo de los intervalos de clase*

[23 – 26]
[27 – 30]
[31 – 34]
[35 – 38]
[39 – 42]
[43 – 46]

f. *Calculo de la marca de clase*

$$X_1^{\prime} = \frac{(23+27)}{2} = 25$$

$$X_2^{\prime} = \frac{(27+31)}{2} = 29$$

$$X_3^{\prime} = \frac{(31+35)}{2} = 33$$

$$X_4^{\prime} = \frac{(35+39)}{2} = 37$$

$$X_5^{\prime} = \frac{(39+43)}{2} = 41$$

$$X_6^{\prime} = \frac{(43+47)}{2} = 45$$

g. Aplicación del método tallos y hojas

Edad del empleado en años: Grafico de tallos y hojas

Frecuencia	**Tallos &**	**Hojas**
14	2	33556667888899
28	3	0001112222233345567777888899
8	4	00111556

h. Tabla de frecuencias de una variable cuantitativa de razón continúa.

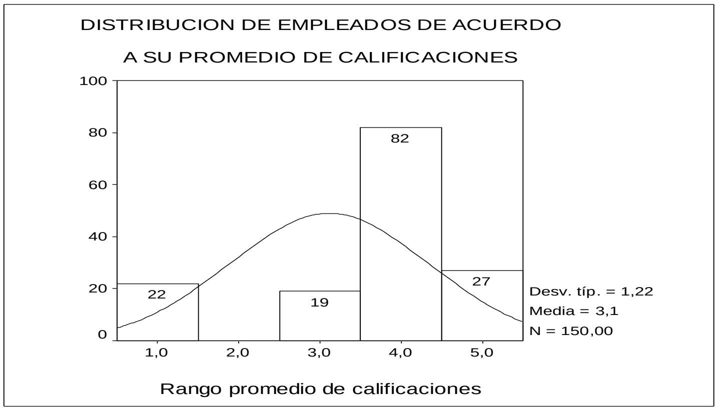
TABLA 1. DISTRIBUCIÓN DE FRECUENCIA DE EMPLEADOS DE LA EMPRESA DE ACUERDO A SU EDAD EN AÑOS

Intervalos reales	Intervalos de clase	Marca de clase X'_i	Frecuencia absoluta n_i	Frecuencia relativa h_i	Frecuencia absoluta acumulada N_i	Frecuencia relativa acumulada H_i
[23 – 27)	**[23 – 26]**	**25**	**8**	**0.16**	**8**	**0.16**
[27 – 31)	**[27 – 30]**	**29**	**12**	**0.24**	**20**	**0.40**
[31 – 35)	**[31 – 34]**	**33**	**11**	**0.22**	**31**	**0.62**
[35 – 39)	**[35 – 38]**	**37**	**11**	**0.22**	**42**	**0.84**
[39 – 43)	**[39 – 42]**	**41**	**5**	**0.10**	**47**	**0.94**
[43 – 47)	**[43 – 46]**	**45**	**3**	**0.06**	**50**	**1.00**
			50	**1.00**		

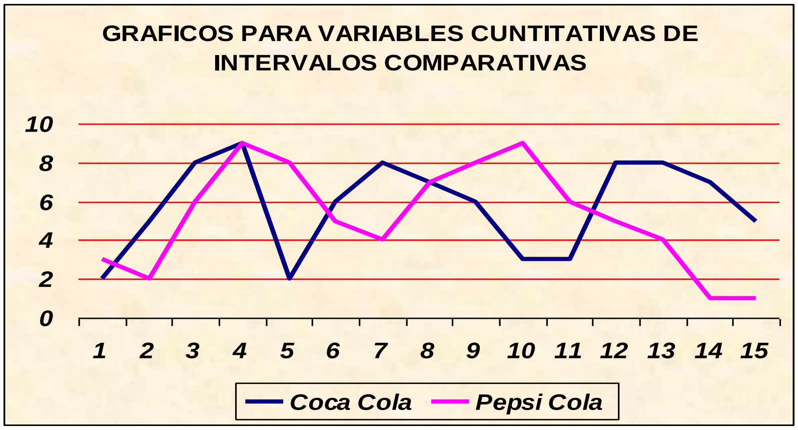
Cuantitativa de razón continúa datos agrupados

Rango promedio de calificaciones

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	[¨0 - 1.0]	22	14,7	14,7	14,7
	[¨2.1 - 3.0]	19	12,7	12,7	27,3
	[¨3.1 - 4.0]	82	54,7	54,7	82,0
	[¨4.1 - 5.0]	27	18,0	18,0	100,0
	Total	150	100,0	100,0	

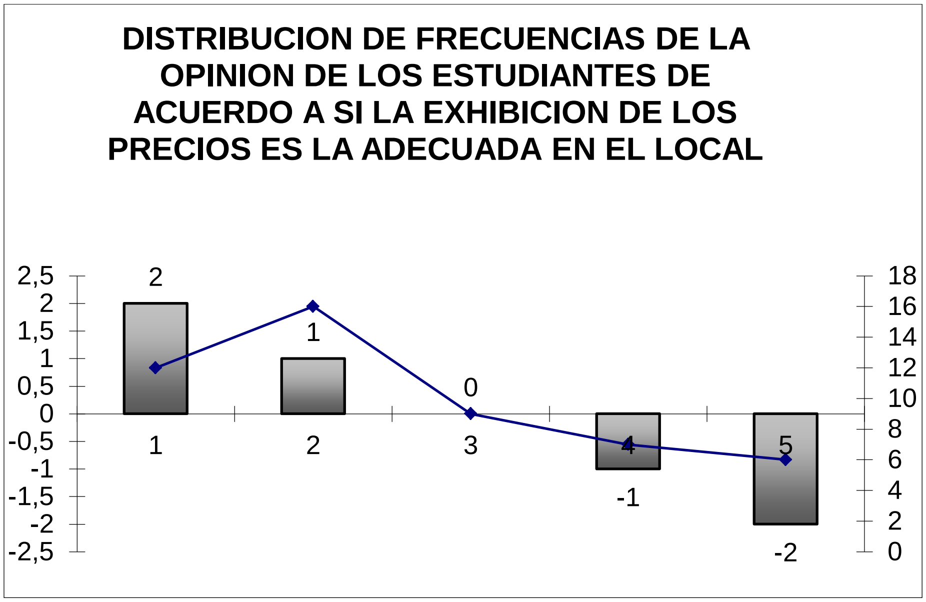


Cuantitativa de intervalos comparativa



Cuantitativa de intervalos no comparativa

Atributo	X	ni	hi	Ni	Hi
Completamente de acuerdo	2	12	0,24	12	0,24
De acuerdo	1	16	0,32	28	0,56
Neutral	0	9	0,18	37	0,74
En desacuerdo	-1	7	0,14	44	0,88
Completamente en desacuerdo	-2	6	0,12	50	1,00
		50	1,00		



EJERCICIOS DE COMPLEMENTO

1. En un centro de computación el número de veces que el sistema se detiene por saturación de este diariamente, fue recolectado para un periodo de 70 dias. Los datos obtenidos fueron los siguientes:

0	0	2	0	0	0	3	3	0	0	1	0	0	0
1	8	5	0	0	4	3	0	6	2	0	2	3	0
0	3	1	1	0	1	0	1	1	0	2	2	1	0
2	2	0	0	0	1	2	1	2	0	0	5	2	1
0	1	6	4	3	3	1	2	4	0	2	0	0	4

Con base a la información dada represente los datos a través de una tabla de frecuencias y gráficamente

2. Los siguientes datos corresponden a los salarios en miles de pasos de algunos empleados de una empresa

13,0	12,5	15,7	13,8	17,0	13,7	16,0	16,9	12,9
12,8	11,2	15,4	14,8	13,6	17,3	15,253	13,6	17,5
15,8	15,3	14,5	12,7	11,9	14,3	16,018	11,9	10,125
12,0	14,8	11,8	16,3	14,1	15,1	14,111	13,7	

Con base a la información dada represente los datos a través de una tabla de frecuencias y gráficamente

3. A continuación se indican las pérdidas y ganancias, en millones de dólares, de las 50 mayores empresas (por ventas) de la lista de 500 de Fortune en 1992. El valor más bajo es una pérdida de 4453 millones de dólares y el más alto una ganancia de 5600 millones. Construir una tabla de frecuencias con a = 1500, li = -4500.

-4453.00	-1484.00	-732.00	2056.00	2636.00	454.00	184.00	-387.00	17.00
5600.00	20.00	-617.00	97.00	120.00	-404.00	535.00	423.00	63.00
-2258.00	-1021.00	1154.00	939.00	311.00	3006.00	-142.00	1461.00	
-2827.00	1080.00	-1086.00	460.00	258.00	842.00	1403.00	308.00	
1294.00	454.00	601.00	73.00	1293.00	709.00	-273.00	97.00	
-795.00	-578.00	1681.00	505.00	1567.00	1773.00	368.00	755.00	

Con base a la información dada represente los datos a través de una tabla de frecuencias y gráficamente

4. En la fabricación de cierto tipo de clavos, aparecen un cierto número de defectuosos se han estudiado 200 lotes de 500 clavos cada uno obteniendo:

Clavos defectuosos	1	2	3	4	5	6	7	8
Numero de lotes	5	15	38	42	49	32	17	2

a. Completa la tabla de frecuencias
b. Interprete: $\mathbf{n_3}$, h_3, $\mathbf{N_3}$, $\mathbf{H_3}$
c. Calcula e interpreta la media, moda y mediana
d. Que porcentaje de lotes tienen mas de 6 clavos defectuosos

5. En Beverage Digest se informa que, con base en las ventas de 1998, las 5 marcas de refrescos que más se vendieron fueron Coke Classic, Diet Coke, Dr.Pepper, Pepsi Cola y Sprite. La lista siguiente proviene de una muestra de 50 compras de esas marcas fue:

Coke Classic	Dr.Pepper	Sprite	Coke	Classic	Pepsi Cola
Diet Coke	Diet Coke	Coke Classic	Diet Coke	Coke	Classic
Pepsi Cola	Pepsi Cola	Diet Coke	Coke Classic	Coke	Classic
Diet Coke	Pepsi Cola	Coke Classic	Diet Coke	Coke	Classic
Coke Classic	Coke Classic	Coke Classic	Coke	Classic	Pepsi Cola
Coke Classic	.Pepper	Sprite	Sprite	Coke	Classic
Sprite	Pepsi Cola	Pepsi Cola	Pepsi Cola	Coke	Classic
Dr.Pepper	Coke	Classic	Dr.Pepper	Pepsi Cola	Dr.Pepper
Pepsi Cola	Coke	Classic	Coke Classic	Pepsi Cola	Pepsi Cola
Diet Coke	Coke	Classic	Diet Coke	Pepsi Cola	Sprite

a. Construir la tabla de distribución de frecuencias.
b. Construir un gráfico.
c. ¿Qué porcentaje de las ventas tienen Pepsi Cola y Coke Classic?.

6. Los siguientes datos muestran la cantidad de pérdida de peso (en libras) para los 250 pacientes de una clínica para el control de peso durante el último año

Pesos (en libras)	[5 – 15)	[15 – 29)	[29 – 44)	[44 – 54)	[54 – 64)
h_i	$\frac{C}{4}$	C	0.18	$\frac{C}{2}$	0.12

a. Encuentre el valor de C
b. Completa la tabla de frecuencias
c. Realice una grafica

7. Representa gráficamente las ganancias medias de los trabajadores, según el sexo, en el cuarto trimestre de 2009, que se recogen en la siguiente tabla:

	Sueldo en ptas.	
Sector	Varones	Mujeres
Industria	284.363	206.204
Construcción	214.446	205.372
Servicios	263.554	195.447

8. Se ha realizado una encuesta en 30 hogares en la que se les pregunta el nº de individuos que conviven en el domicilio habitualmente. Las respuestas obtenidas han sido las siguientes:

4, 4, 1, 3, 5, 3, 2, 4, 1, 6, 2, 3, 4, 5, 5, 6, 2, 3, 3, 2, 2, 1, 8, 3, 5, 3, 4, 7, 2, 3.

a. Realice una tabla de frecuencias.
b. ¿Qué proporción de hogares está compuesto por tres o menos personas?
c. ¿Qué proporción de individuos vive en hogares con tres o menos miembros?
d. Represente gráficamente.

9. Represente gráficamente la procedencia de los extranjeros residentes en Colombia, en diciembre de 2009, recogidos en la siguiente tabla:

Procedencia	
Europa	353.556
América	166.709
Asia	66.340
África	213.012
Oceanía	1.013
Desconocida	699

10. Elaboremos un gráfico r, y los datos serán los referentes a los componentes de la totalidad de los Activos de la Compañía UHU.

Activo circulante: $ 300.000,00 = 25.00%
Fondos especiales: 60.000,00 = 5.00%
Inversiones: 120.000,00 = 10.00%
Activo fijo tangible 591.000,00 = 49.25%
Intangibles 90.000,00 = 7.50%
Cargos diferidos 24.000,00 = 2.00%
Otros activos 15.000,00 = 1.25%
TOTAL ACTIVOS 1.2000.000,00 =100.00%

11. Importaciones colombianas por países.(Miles de millones de pesos. 2009)

TOTAL	**574.518**
AMERICA (1)	*379.801*
USA (2)	270.722
M.C.C.(3)	382
ALADI (4)	72.854
GRUPO ANDINO (5)	22.461
CARICOM (6)	4.707
RESTO DE AMERICA	15.195
EUROPA (8)	*144.794*
C.E.E. (9)	129.703
ASIA (10)	*43.888*
JAPÓN (11)	28.152
AFRICA (12)	*2.579*
ARGELIA (13)	178
OCEANIA (14)	*4.454*
NUEVA ZELANDIA (1	3.780

V. TRANSFORMACIONES LINEALES

Se utilizan para encontrar el porcentaje de datos que no son fácilmente visibles en la tabla

Si se desea estimar el porcentaje de datos que son menores o iguales a al valor dado ¨a¨. Utilizamos la siguiente formula:

$$H(a) = h(X \le a) = H_{i-1} + \frac{h_i}{C_i}(a - L_{i-1})$$

Donde:

- Se encuentra el intervalo donde se encuentra el valor ä¨ buscado y a la posición le indicamos ¨i ¨
- H_{i-1} ~ indica la frecuencia relativa acumulada anterior de donde se encuentra el dato buscado
- h_i ~ indica la frecuencia relativa en donde se encuentra el dato buscado
- C_i ~ indica la amplitud del intervalo real donde se encuentra el dato buscado
- a ~ es el valor buscado
- L_{i-1} ~ Indica el limite inferior del intervalo real donde se encuentra el dato buscado

Si se desea estimar el porcentaje de datos que hay entre ¨a¨ y ¨b¨ .Utilizamos la siguiente formula**:**

$$h(a \le X \le b) = H(b) - H(a)$$

Ejemplo de aplicación de transformaciones lineales

TABLA 1. DISTRIBUCIÓN DE FRECUENCIA DE EMPLEADOS DE LA EMPRESA DE ACUERDO A SU EDAD EN AÑOS

Intervalos reales	Intervalos de clase	Marca de clase X'_i	Frecuencia absoluta n_i	Frecuencia relativa h_i	Frecuencia absoluta acumulada N_i	Frecuencia relativa acumulada H_i
[23 – 27)	**[23 – 26]**	**25**	**8**	**0.16**	**8**	**0.16**
[27 – 31)	**[27 – 30]**	**29**	**12**	**0.24**	**20**	**0.40**
[31 – 35)	**[31 – 34]**	**33**	**11**	**0.22**	**31**	**0.62**
[35 – 39)	**[35 – 38]**	**37**	**11**	**0.22**	**42**	**0.84**
[39 – 43)	**[39 – 42]**	**41**	**5**	**0.10**	**47**	**0.94**
[43 – 47)	**[43 – 46]**	**45**	**3**	**0.06**	**50**	**1.00**
			50	**1.00**		

a. Estimar el porcentaje de empleados menores de 30 años?

a = 30 y se encuentra en el intervalo real [27 – 31) i = 2

$$H(a) = h(X \le a) = H_{i-1} + \frac{h_i}{C_i}(a - L_{i-1})$$

$$H(30) = h(X \le 30) = 0.16 + \frac{0.24}{4}(30 - 27) = 0.34$$

El 34% de los empleados estudiados tienen edades son menores de 30 años

b. Estimar el porcentaje de empleados que tienen edades entre 30 y 40 años?

h (a ≤ X ≤ b) = H (b) – H (a)

h (30 ≤ X ≤ 40) = H (40) – H (30) = 0.865 – 0.34 = 0.525

El 52% de los empleados estudiados tienen edades entre 30 y 40 años

$$H(30) = h(X \le 30) = 0.16 + \frac{0.24}{4}(30 - 27) = 0.34$$

a = 40 y se encuentra en el intervalo real [39 – 43) i = 5

$$H(40) = h(X \le 40) = 0.84 + \frac{0.10}{4}(40 - 39) = 0.865$$

EJERCICIOS DE COMPLEMENTO

1. Se dispone del beneficio mensual anual obtenido el pasado año por 38 empresas caleñas:

Beneficio (millones de pesos)	Nº empresas
230-280	5
280-330	7
330-580	14
580-630	9
630-780	3

Se pide:

a. Calcular el beneficio medio de estas 38 empresas caleñas.
b. ¿Cuál es el beneficio mayor de la mitad de las empresas más modestas?
c. Determinar el beneficio más frecuente.
d. El 25% de las empresas mas rentables ¿qué nivel de beneficios tienen?

2. Se realiza un estudio en una ciudad sobre la capacidad hotelera y se obtienen los siguientes resultados:

PLAZAS	Nº DE HOTELES
0-10	25
10-30	50
30-60	55
60-100	20

   a. Complete la tabla de frecuencias
   b. ¿Cuál es la proporción de hoteles que disponen de entre 15 y 50 plazas?

3. Se toma una muestra de 250 estudiantes. La siguiente tabla recoge la cantidad de tiempo empleado por cada uno de los miembros de dicha muestra en preparar un examen:

Tiempo de estudio (Horas)	[0 – 4]	[5 – 10]	[10 – 18]	[18 – 24]	[24 – 40]
Numero de estudiantes	9	21	24	15	6

Encuentre el número de estudiantes que le dedicaron entre 15 y 20 horas de estudio para prepararse?

4. Una fábrica de ropa vende en un mes 50 pantalones distribuidos en distintas tallas dados *a* continuación: 8 de la talla A, 7 de la talla B, 15 de la talla C, 10 de la talla D y el resto de la talla E. Realice la tabla de frecuencias y grafico

5. Se ha realizado un estudio entre 100 mujeres mayores de 15 años, observándose el número de hijos de las mismas. El resultado ha sido:

x_i: nº hijos	n_i: nº mujeres
0	13
1	20
2	25
3	20
4	11
5	7
6	4

   a. Complete la tabla de frecuencias
   b. Porcentaje de mujeres que tienen mas de 4 hijos?
   c. Numero de mujeres que tienen al menos tres hijos?
   d. Porcentaje de mujeres que tienen a los sumo 2 hijos

VI. ANALISIS UNIDIMENSIONAL DE DATOS

Se utiliza para encontrar un valor que me permita tomar decisiones con el mínimo error posible

4.1. Indicadores de tendencia central

Sirven para encontrar el punto medio de una serie de datos

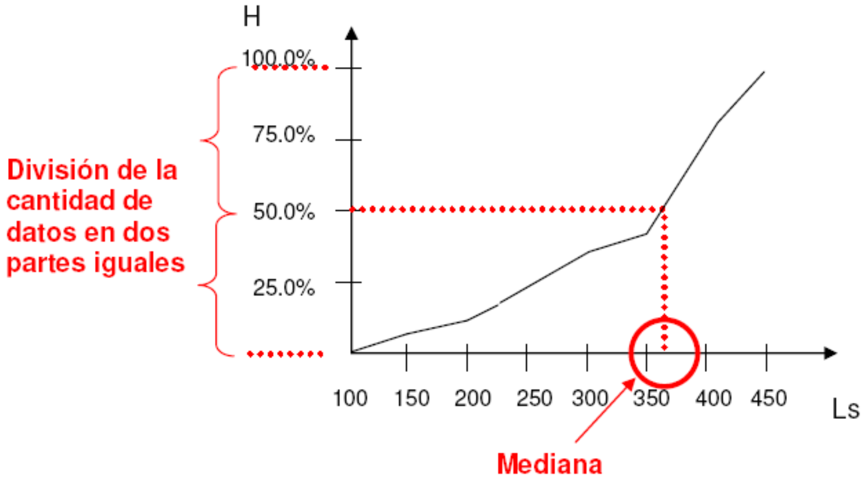
4.1.1. Media aritmética

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr><td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado
2 2 4 5 5 5 7 8 9</td><td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera
<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>[50 – 100]</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>[100 – 200]</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>[200 – 350]</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>[350 – 500]</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>[500 – 900]</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>Formula:
$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N} = 5.22$</td><td>Formula:
$\bar{X} = \sum_{i=1}^{m} \frac{\grave{X}_i n_i}{n} = 265$</td></tr>
<tr><td>Interpretación: El numero promedio de libros que leen los profesionales es de aproximadamente cinco por año</td><td>Interpretación: El ingreso promedio semanal de los profesionales recién graduados es de aproximadamente 265.000 pesos</td></tr>
</table>

4.1.2. Mediana

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr><td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado
n = 9 Impar
2 2 4 5 5 5 7 8 9
n = 8 Par
2 2 3 4 5 5 5 7</td><td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera
<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>[50 – 100]</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>[100 – 200]</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>[200 – 350]</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>[350 – 500]</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>[500 – 900]</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>Formula:</td><td>Formula:</td></tr>
</table>

Si n es impar la mediana es el datos que ocupa la posición $M_e = 5 = X_{\frac{n+1}{2}}$ de una serie ordenada Si n es impar la mediana es el datos que ocupa la posición $M_e = \frac{4+5}{2} = 4.5 = \frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$ de una serie ordenada	Se ubica el intervalo donde se encuentra el 50% de los datos $M_e = L_{i-1} + \left[\frac{\frac{n}{2}-N_{i-1}}{n_i}\right]C_i$ $M_e = 200 + \left[\frac{\frac{50}{2}-18}{21}\right](150) = 250$
Interpretación: El 50% de los profesionales leen 5 libros o menos al año	Interpretación: El 50% de los profesionales tiene ingresos inferiores a 250.000 pesos semanales su primer año de trabajo



4.1.3. Moda

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr><td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado
2 2 4 5 5 5 7 8 9</td><td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera
<table>
<tr><th>$I.R$</th><th>$X_i^`$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>[50 – 100]</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>[100 – 200]</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>[200 – 350]</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>[350 – 500]</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>[500 – 900]</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>Formula:
La moda es el dato de mayor frecuencia
$M_o = 5$</td><td>Formula:
Se ubica el intervalos de mayor frecuencia
$\Delta_1 = \frac{h_i}{C_i} - \frac{h_{i-1}}{C_{i-1}} = \frac{0.42}{150} - \frac{0.24}{100} = 0.0052$
$\Delta_2 = \frac{h_i}{C_i} - \frac{h_{i+1}}{C_{i+1}} = \frac{0.42}{150} - \frac{0.18}{150} = 0.0016$</td></tr>
</table>

	$M_o = L_{i-1} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] C_i$ $M_o = 200 + \left[\frac{0.0052}{0.0052 + 0.0016}\right](150) = 314.7$
Interpretación: Lo mas frecuente de encontrar es que los profesionales lean 5 libros por año	Interpretación: El ingreso mas frecuente de un profesional recién egresado es de 314.700 pesos semanales

4.1.4. Media geométrica

La media geométrica (*MG*) de un conjunto de *n* números positivos se define como la enésima raíz del producto de *n* números.

Ventajas y desventajas:

- En su cálculo intervienen todos los valores de la distribución.
- Los valores extremos tienen menor influencia que en la media aritmética.
- Es única.
- Su cálculo es más complicado que el de la media aritmética.
- Solo se puede calcular si no hay observaciones negativas.

Datos no Agrupados: La fórmula para su cálculo es:

$$\boxed{MG = \sqrt[n]{(X_1)(X_2)(X_3)\ldots(X_n)}}$$

Donde MG es media geométrica, *n* es el número total de datos y *X* es el valor de cada observación de la variable de interés.

Datos Agrupados

$$\boxed{MG = \sqrt[\Sigma f_i]{(y_1^{f_1})(y_2^{f_2})\ldots(y_n^{f_n})}}$$

Donde MG es media geométrica, *yi* es marca de clase, *fi* la frecuencia de clase correspondiente, n el número total de datos utilizados.

¿Cuándo se debería utilizar este tipo de media? Lo veremos a través de un par de ilustraciones.

APLICACIONES DE LA MEDIA GEOMÉTRICA:

- Es útil para encontrar el promedio de porcentajes, razones, índices o tasas de crecimiento.

- Se usa cuando se trabaja con observaciones, donde cada una tiene una razón aproximadamente constante respecto a la anterior.
- Para mostrar los efectos multiplicativos en el tiempo de los cálculos del interés compuesto, la inflación y el crecimiento poblacional.
- En estadística para calcular el crecimiento o decrecimiento de las poblaciones, en donde los valores están dados en sucesión geométrica.
- Se sugiere usar la media geométrica siempre que se desee calcular el cambio porcentual promedio en el tiempo para algunas variables.
- En ciertas situaciones, las respuestas obtenidas con la media aritmética no difieren mucho de las correspondientes a la media geométrica, pero incluso diferencias pequeñas pueden generar malas decisiones.

Ilustración 4. Las tasas de interés vigentes de tres bonos son 5%, 7% y 4%. La media geométrica es por lo tanto:

$$MG = \sqrt[3]{(5)(7)(4)} = 5.192\%$$

Comparativamente, la media aritmética correspondiente sería de:

X = (5 + 7 + 4)/3 = 5.333%.

Como puede observase, la *MG* da una cifra de ganancia más conservadora porque no tiene una ponderación alta para la tasa de 7%.

OTRA APLICACIÓN DE LA MEDIA GEOMÉTRICA

Otra aplicación de la media geométrica es para determinar el porcentaje promedio del incremento en ventas, producción u otros negocios o series económicas de un periodo a otro. La fórmula para este tipo de problema es:

$$\boxed{MG = \sqrt[n]{(valor\ al\ final\ del\ período)/(valor\ al\ inicio\ del\ período)} - 1}$$

Donde *n* es el número de años comprendido entre el inicio del período y el final del período de interés.

Se ejemplifica la aplicación de la media geométrica utilizando la fórmula anterior.

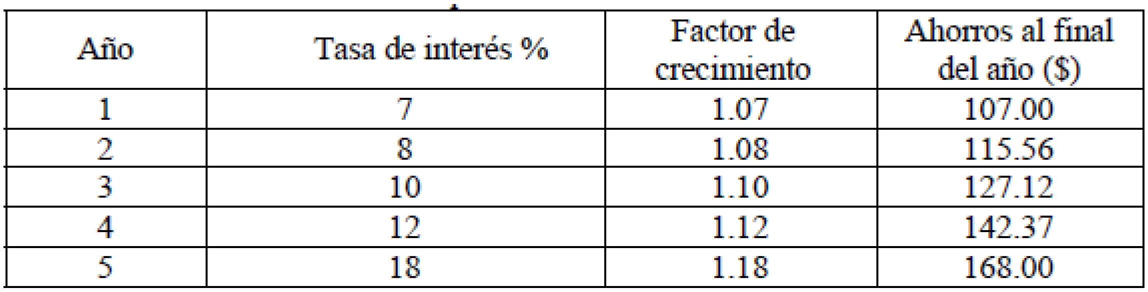
Ilustración 5. El número total de mujeres inscritas en las distintas universidades del país aumentó de 755,000 en 1996 a 835,000 en el año 2005.

Aquí n = 10 (años comprendidos entre 1996 y 2005), así (n - 1) = 9.

Es decir, la media geométrica de la tasa de crecimiento del número de mujeres inscritas en las distintas universidades del país es 1.27%.

Ilustración 6. A continuación se muestra el crecimiento de un depósito de ahorro de $100 durante cinco años, de acuerdo a las tasas de interés de 7, 8, 10, 12 y 18% para los años 1, 2, 3, 4 y 5 respectivamente.

Crecimiento de un Depósito de $100 en una Cuenta de Ahorro



Año	Tasa de interés %	Factor de crecimiento	Ahorros al final del año ($)
1	7	1.07	107.00
2	8	1.08	115.56
3	10	1.10	127.12
4	12	1.12	142.37
5	18	1.18	168.00

Si se usa la media aritmética simple sería:

X = (1.07 + 1.08 + 1.10 + 1.12 + 1.18)/5 = 1.11

Por lo que si se multiplica el promedio de la tasa de interés de los cinco años por la inversión inicial, se obtiene:

$100 x 1.11 x 1.11 x 1.11 x 1.11 x 1.11 = $168.51

Como puede verse en la tabla anterior, la cifra real ganada fue sólo de $168.00. Por lo tanto, el factor de crecimiento promedio correcto debe ser ligeramente menor a 1.11.

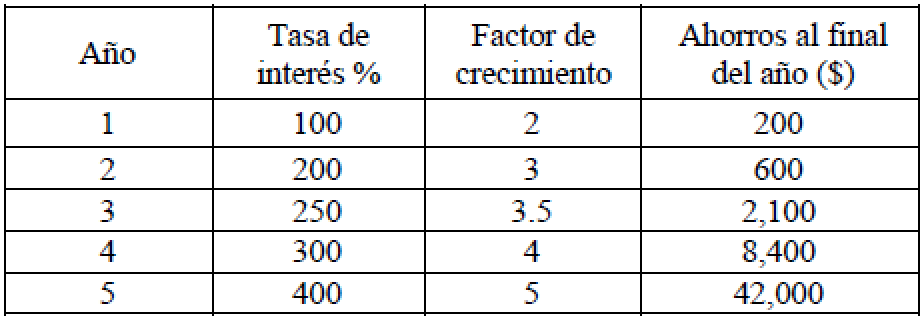
Para obtener el valor exacto, se debe de utilizar la media geométrica:

$$MG = \sqrt[5]{(1.07)(1.08)(1.10)(1.12)(1.18)} = 1.109328$$

Así, el factor de crecimiento es de 1.10

Ilustración 7. En las economías con un alto índice de inflación, los bancos deben pagar altas tasas de inflación y los bancos deben pagar altas tasas de interés para atraer clientes.

Suponga que en un período de cinco años, en una economía con inflación alta los bancos pagan tasas de interés anual de 100, 200, 250, 300 y 400%.



Año	Tasa de interés %	Factor de crecimiento	Ahorros al final del año ($)
1	100	2	200
2	200	3	600
3	250	3.5	2,100
4	300	4	8,400
5	400	5	42,000

Por lo tanto, la tasa media de crecimiento geométrico se calcula así:

$$MG = \sqrt[5]{(2)(3)(3.5)(4)(5)} = 3.347$$

Este factor de crecimiento corresponde a una tasa de:

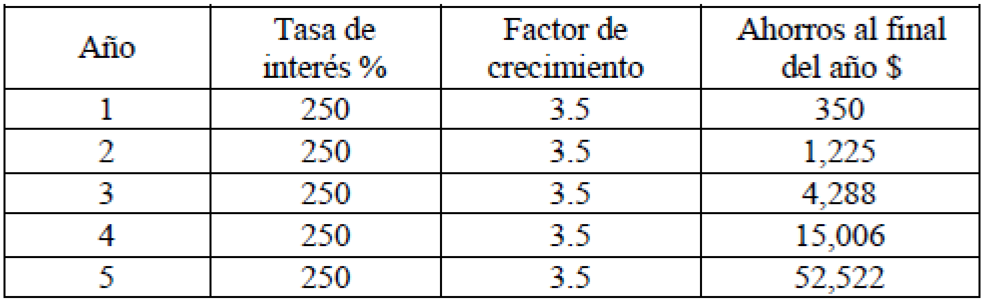
3.347 – 1 = 2.347 = 234.7%

El factor de crecimiento como media aritmética sería de:

X = (2+3+3.5+4+5)/5 = 3.5

que corresponde a una tasa de interés promedio anual del 250%.

Como se puede apreciar en la tabla a continuación:



Año	Tasa de interés %	Factor de crecimiento	Ahorros al final del año $
1	250	3.5	350
2	250	3.5	1,225
3	250	3.5	4,288
4	250	3.5	15,006
5	250	3.5	52,522

4.1.5. Media armónica

La media armónica (*MH*) se define como la recíproca de la media aritmética de los recíprocos de un conjunto de datos.

Datos no agrupados

La fórmula correspondiente para su cálculo es la siguiente:

$$\boxed{MH = \frac{n}{\sum (1/y_i)}}$$

Donde *MH* es la media armónica, *n* es el numero de datos, *yi* cada valor observado correspondiente a la variable de interés.

Obsérvese que la inversa de la media armónica es la media aritmética de los inversos de los valores de la variable. No es aconsejable en distribuciones de variables con valores pequeños.

Ventajas y desventajas:

- En su cálculo intervienen todos los valores de la distribución
- Su cálculo no tiene sentido cuando algún valor de la variable tomo valor cero
- Es única

Datos agrupados

La fórmula correspondiente para su cálculo es la siguiente:

$$MH = \frac{n}{\sum (f_i / y_i)}$$

Donde *MH* es la media armónica, *n* es el numero de datos, *fi* el valor de cada frecuencia, *yi* cada valor observado correspondiente a la variable de interés.

¿Cuándo se debería utilizar este tipo de media?

APLICACIONES DE LA MEDIA ARMÓNICA

Esta medida se emplea para promediar variaciones con respecto al tiempo tales como productividades, tiempos, rendimientos, cambios, etc., tal como se describe a continuación.

Precio promedio

Si se compran varios tipos de productos con distintas cantidades de unidades de cada tipo, pero gastando en ellos igual cantidad de dinero, el precio promedio por unidad es igual a la media armónica de los precios por unidad de cada tipo de producto.

Rendimiento promedio de producción

En un grupo puede haber operarios con distinta velocidad para producir un artículo. Si cada una de estas personas tiene que elaborar igual cantidad de artículos, el promedio de velocidad de rendimientos de tal grupo, es igual al promedio armónico de las velocidades de rendimiento de cada una de los operarios que lo integran.

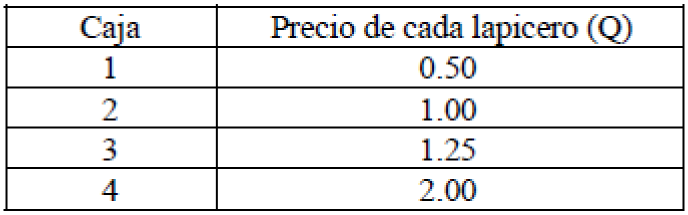
Rendimiento Promedio de la Producción

Si *v1, v2, …vn* son las velocidades de rendimiento de cada uno de las operarios, que aunque sea en distinta cantidad de tiempo, producen igual cantidad de productos, el promedio de velocidad de rendimiento del grupo es:

$MH = n / (1/v1 + 1/v2 + \dots 1/vn)$

Donde *n* es el número de operarios.

Ilustración 8. Se compran 4 cajas de bolígrafos. Las cuatro cajas costarán Q20.00 cada una. El precio de cada lapicero es:



Caja	Precio de cada lapicero (Q)
1	0.50
2	1.00
3	1.25
4	2.00

Este problema puede ser resuelto por dos métodos, los cuales se describen a continuación:

Primer Método

Precio promedio = cantidad total gastada / cantidad total de lapiceros comprada

Número de lapiceros = precio de la caja / precio de cada lapicero

Caja	Precio de cada lapicero (Q.)	Número de lapiceros
1	0.50	40
2	1.00	20
3	1.25	16
4	2.00	10

Total gastado = Q.20.00/caja * 4 cajas = Q80.00 en total

Total de lapiceros comprados = 40 + 20 + 16 + 10 = 86 lapiceros

Precio promedio = Q.80.00 / 86 lapiceros = Q.0.93 / lapicero

Segundo Método:

Como las 4 cajas cuestan 20 quetzales, el precio promedio de los lapiceros que contienen es igual al promedio armónico de los precios de los lapiceros de cada caja.

$MH = n$ / [(1/$p1$) + (1/$p2$) + (1/$p3$) + (1/$p4$)]
MH = 4 / [(1/0.50) + (1/1.00) + (1/1.25) + (1/2.00)]
MH = Q.0.93 / lapicero

Ilustración 9. Si un mensajero conduce 100 millas en una vía rápida a 60 millas/hora y las siguientes 10 millas después de la vía rápida las conduce a 30 millas/hora. ¿Cuál es la velocidad promedio?

Distancia recorrida = 20 millas
Tiempo recorrido:

- Vía rápida = 0.1667 horas
- Vía normal = 0333 horas
- Tiempo total = 0.5 horas

El promedio del tiempo es 20/0.5 = 40 millas / h
$MH = n / \Sigma (1/yi)$ = 2 /(1/60+1/30) = 40 millas/h

Por lo tanto, el promedio de la velocidad en que conduce el mensajero es de 40 millas por hora.

4.1.6. Media cuadrática

Una media cuadrática (MC) se define como la raíz cuadrada de la media aritmética de los cuadrados de los valores de la variable.

Datos No Agrupados

Para datos no agrupados su fórmula puede expresarse como:

$$MC = \sqrt{\frac{\sum y_i^2}{n}}$$

Donde *MC* es la media cuadrática, *yi* el valor correspondiente a cada dato observado de la variable de interés, *n* el número total de datos.

Datos Agrupados

Para datos agrupados se puede encontrar mediante la siguiente fórmula:

$$MC = \sqrt{\frac{\sum \left(y_i^2\right)\left(f_i\right)}{n}}$$

Donde *MC* es la media cuadrática, *yi* el valor correspondiente a cada dato observado de la variable de interés, *fi* la frecuencia correspondiente a cada valor observado, *n* el número total de datos.

¿Cuándo se debería utilizar este tipo de media? Este tipo de media se utiliza mucho en cálculos científicos.

APLICACIONES DE LA MEDIA CUADRÁTICA

Ilustración 10. A continuación se muestra una serie de datos (agrupados en intervalos) de 56 mediciones de temperatura en grados centígrados. Se va a calcular la media geométrica, armónica y cuadrática para mostrar la diferencia que se obtiene con cada medida.

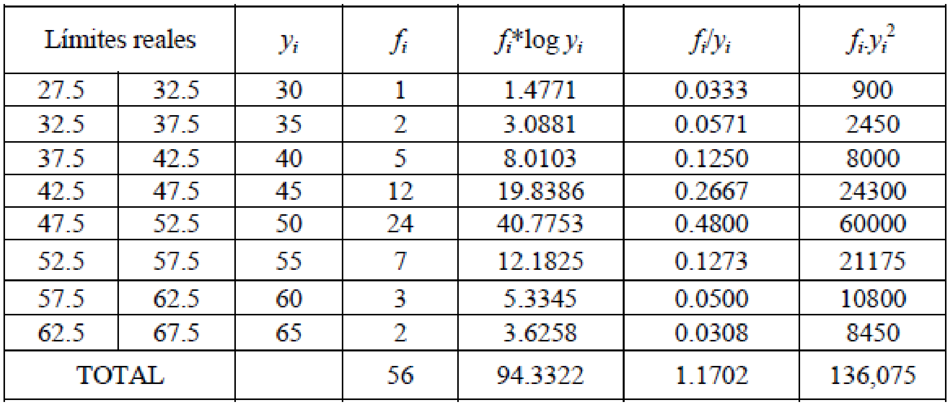
Por medio de la fórmula de la media geométrica se obtienen los siguientes resultados:

log *MG* = (Σ *fi* * log *yi*)/*n* = 94.3322/56 = 1.6845
MG = 48.36 °C

Si se aplica la media armónica se obtiene el siguiente resultado:
MH = *n* / Σ (*fi*/*yi*) = 56 / 1.170 = 47.86°C

Y con el uso de la media cuadrática se obtiene el siguiente resultado:

$$MC = \sqrt{\sum y_i^2 f_i / n} = \sqrt{136,076 / 56} = 49.29°C$$



Límites reales		y_i	f_i	f_i*log y_i	f_i/y_i	$f_i.y_i^2$
27.5	32.5	30	1	1.4771	0.0333	900
32.5	37.5	35	2	3.0881	0.0571	2450
37.5	42.5	40	5	8.0103	0.1250	8000
42.5	47.5	45	12	19.8386	0.2667	24300
47.5	52.5	50	24	40.7753	0.4800	60000
52.5	57.5	55	7	12.1825	0.1273	21175
57.5	62.5	60	3	5.3345	0.0500	10800
62.5	67.5	65	2	3.6258	0.0308	8450
TOTAL			56	94.3322	1.1702	136,075

RELACIÓN ENTRE LAS MEDIAS ARITMÉTICA, GEOMÉTRICA, ARMÓNICA Y CUADRÁTICA

La relación entre la media armónica, la media geométrica y la media cuadrática puede expresarse de la siguiente manera:

MH ≤ MG ≤ MC

Como puede observarse a través de la relación anterior, el máximo valor medio de una serie de datos se tiene al calcular la media cuadrática (*MC*) y el mínimo valor medio se obtiene al calcular la media armónica (*MH*)

En distribuciones simétricas los valores de las medias armónica, aritmética, geométrica y cuadrática, son iguales entre sí, es decir:

MH = MA = MG = MC

A continuación se describe un ejemplo en donde se aplica la media cuadrática y la armónica, para ilustrar estas relaciones.

Ilustración 11. Un ingeniero obtuvo los siguientes datos de concentración de mercurio en partes por millón (ppm) en ocho localidades a lo largo de un arroyo:

0.064 0.071 0.066 0.062 0.073 0.065 0.061 0.066

Desea determinar la concentración máxima y la concentración mínima de mercurio.

La concentración máxima y mínima de mercurio corresponde a la media cuadrática y a la media armónica, ya que estos datos dan los valores extremos de la serie de datos:

MC = √ 0.0642 + 0.0712 + ... + 0.066 2 = 0.066113 (máximo)
MH = 8 / (1/0.064 +1/0.071 + ... + 1/0.066) = 0.06577 (mínimo)

Concentración máxima de mercurio = 0.0661 ppm
Concentración mínima de mercurio = 0.06577 ppm

CONCLUSIONES

- La media cuadrática tiene aplicaciones científicas. El máximo valor medio de una serie de datos se tiene al calcular la media cuadrática (*MC*) y el mínimo valor medio se obtiene al calcular la media armónica (*MH*).
- En distribuciones simétricas los valores de las medias armónica, aritmética, geométrica y cuadrática, son iguales entre sí.
- Se suele utilizar para promediar variables tales como productividades, velocidades, tiempos, rendimientos, cambios, etc.
- La media geométrica se utiliza para determinar el porcentaje promedio del incremento en ventas, producción u otros negocios o series económicas de un periodo a otro.
- Para mostrar los efectos multiplicativos en el tiempo de los cálculos del interés compuesto, la inflación y el crecimiento poblacional.
- También se utiliza en estadística para calcular el crecimiento o decrecimiento de las poblaciones, en donde los valores están dados en sucesión geométrica.

4.2. Indicadores de dispersión

Sirven para encontrar el grado de homogeneidad de una serie de datos

4.2.1. Varianza

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr>
<td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9</td>
<td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>$[50 - 100]$</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>$[100 - 200]$</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>$[200 - 350]$</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>$[350 - 500]$</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>$[500 - 900]$</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table>
</td>
</tr>
<tr>
<td>Formula:
$S^2 = \sum_{i=1}^{N} \frac{[X_i - \bar{X}]^2}{N} = 5.28$</td>
<td>Formula:
$S^2 = \sum_{i=1}^{m} \frac{[\grave{X}_i - \bar{X}]^2 n_i}{n-1} = 20127.55$</td>
</tr>
</table>

4.2.2. Desviación estándar

DATOS NO AGRUPADOS

Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9

Formula:

$$S = \sqrt{S^2} = 2.298$$

DATOS AGRUPADOS

Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

$I.R$	$X_i^{`}$	n_i	h_i	N_i	H_i
$[50 - 100]$	75	6	0.12	6	0.12
$[100 - 200]$	150	12	0.24	18	0.36
$[200 - 350]$	275	21	0.42	39	0.78
$[350 - 500]$	425	9	0.18	48	0.96
$[500 - 900]$	700	2	0.04	50	1.00
		50			

Formula:

$$S = \sqrt{S^2} = 141.87$$

4.2.3. Coeficiente de variación

En algunas aplicaciones, más que la dispersión absoluta, interesa la dispersión relativa. En esos casos suele usarse el coeficiente de variación:

$$CV = \frac{S}{\overline{\overline{X}}} * 100$$

Interpretación del coeficiente de variación

$CV < 20\%$	Muestra homogénea Comportamientos regulares Aplican muestreos no probabilisticos
$20 < CV \leq 40\%$	Muestra relativamente homogénea Comportamientos esporádicos Indiferente el tipo de muestreo
$CV > 40\%$	Muestra heterogénea Comportamientos casuales Muestreos probabilisticos

DATOS NO AGRUPADOS

Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9

DATOS AGRUPADOS

Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

$I.R$	$X_i^{`}$	n_i	h_i	N_i	H_i
$[50 - 100]$	75	6	0.12	6	0.12
$[100 - 200]$	150	12	0.24	18	0.36
$[200 - 350]$	275	21	0.42	39	0.78
$[350 - 500]$	425	9	0.18	48	0.96
$[500 - 900]$	700	2	0.04	50	1.00
		50			

<table>
<tr>
<td>Formula:
$C.V = \frac{S}{\bar{X}} * 100 = \frac{2.298}{5.22} * 100 = 44\%$</td>
<td>Formula:
$C.V = \frac{S}{\bar{X}} * 100 = \frac{141.87}{265} * 100 = 53.5\%$</td>
</tr>
</table>

4.2.4. Desviación media

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr>
<td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9</td>
<td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

<table>
<tr><th>$I.R$</th><th>$X_i^{`}$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>$[50 - 100]$</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>$[100 - 200]$</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>$[200 - 350]$</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>$[350 - 500]$</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>$[500 - 900]$</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table>
</td>
</tr>
<tr>
<td>Formula:
$DM = \sum_{i=1}^{N} \frac{|X_i - \bar{X}|}{N} = 1.85$</td>
<td>Formula:
$DM = \sum_{i=1}^{m} \frac{|X_i^{`} - \bar{X}| n_i}{n - 1} = 100.8$</td>
</tr>
</table>

4.2.5. Teorema de Tchebyshev

Independiente de la forma de la distribución de frecuencias de un conjunto de datos, se puede estimar un valor de la media de la población de la siguiente forma:

$$h(\bar{x} - ks \leq \mu \leq \bar{x} + ks) \leq 1 - \frac{1}{k^2}$$

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr>
<td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9</td>
<td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

<table>
<tr><th>$I.R$</th><th>$X_i^{`}$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>$[50 - 100]$</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>$[100 - 200]$</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>$[200 - 350]$</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>$[350 - 500]$</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>$[500 - 900]$</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td>50</td><td></td><td></td><td></td><td></td></tr>
</table>
</td>
</tr>
<tr>
<td>Formula: 95% - K = 4.47

$h((-5.05) \leq \mu \leq (15.49)) \leq 0.95$</td>
<td>Formula:

$h((-369.16) \leq \mu \leq (899.16)) \leq 0.95$</td>
</tr>
</table>

4.3. Indicadores de asimetría

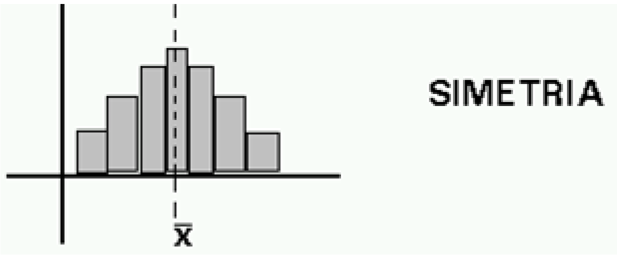
Sirven para encontrar como están distribuidos los datos con respecto a la media

4.3.1. Índice de asimetría de Pearson

$$Si\ \bar{X}\ > M_e \gg M_o \quad Entonces\ se\ tiene\ una\ distribucion\ con\ asimetria\ positiva$$

$$Si\ \bar{X} < M_e \leq M_o \quad Entonces\ se\ tiene\ una\ distribucion\ con\ asimetria\ negativa$$

$$Si\ \bar{X} = M_e = M_o \quad Entonces\ se\ tiene\ ditribucion\ simetrica$$



4.3.2. Coeficiente de asimetría de Fisher

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr><td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado
2 2 4 5 5 5 7 8 9</td><td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera
<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>[50 – 100]</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>[100 – 200]</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>[200 – 350]</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>[350 – 500]</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>[500 – 900]</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>Formula:
$$CA_F = \sum_{i=1}^{N} \frac{[X_i - \bar{X}]^3}{NS^3} =$$</td><td>Formula:
$$CA_F = \sum_{i=1}^{m} \frac{[\grave{X}_i - \bar{X}]^3 n_i}{(n-1)S^3} =$$</td></tr>
</table>

$$\boldsymbol{Si\ \ CA_F}\ > 0\ \ \boldsymbol{Entonces\ se\ tiene\ una\ distribucion\ con\ asimetria\ positiva}$$

$$\boldsymbol{Si\ \ CA_F}\ < 0\ \ \boldsymbol{Entonces\ se\ tiene\ una\ distribucion\ con\ asimetria\ negativa}$$

$$\boldsymbol{Si\ \ CA_F\ = 0\ \ Entonces\ se\ tiene\ ditribucion\ simetrica}$$

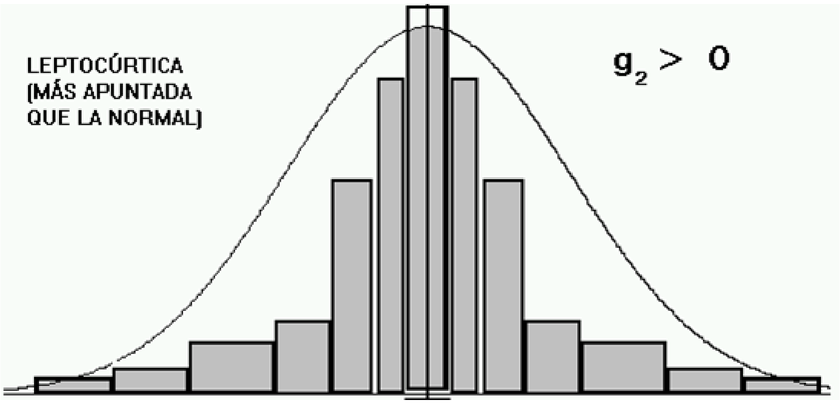
4.4. Indicadores de apuntamiento

Sirven para comoi están distribuidos los datos con respecto a la distribución ideal

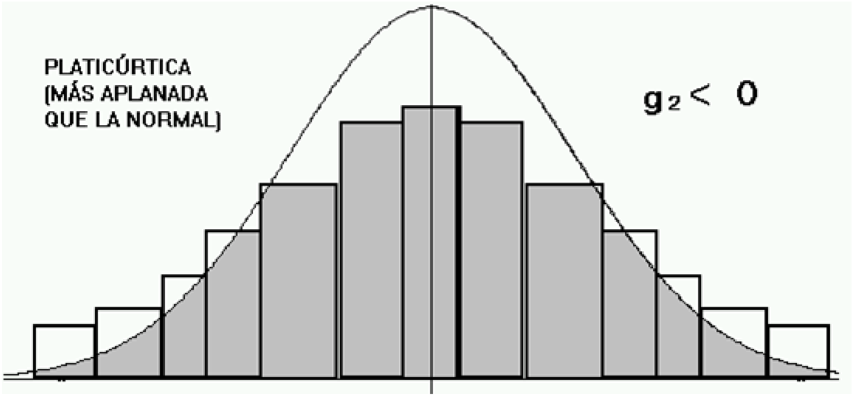
4.4.1. Coeficiente de curtosis

<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr><td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado
2 2 4 5 5 5 7 8 9</td><td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera
<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>[50 – 100]</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>[100 – 200]</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>[200 – 350]</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>[350 – 500]</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>[500 – 900]</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>Formula:
$$Curtosis = K = \sum_{i=1}^{N} \frac{[X_i - \bar{X}]^4}{NS^4} =$$</td><td>Formula:
$$Curtosis = K = \sum_{i=1}^{m} \frac{[\grave{X}_i - \bar{X}]^4 n_i}{(n-1)S^4} =$$</td></tr>
</table>

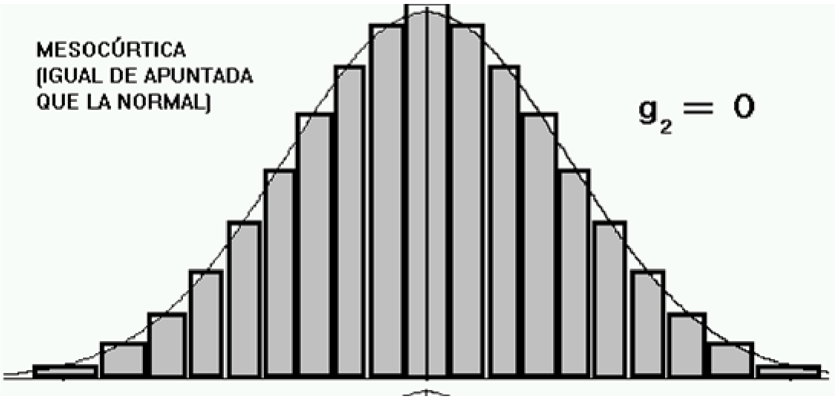
$Si\ \ K\ > 3\quad Entonces\ se\ tiene\ una\ distribucion\ Leptocurtica$



$Si\ \ K\ < 3\quad Entonces\ se\ tiene\ una\ distribucion\ Platocurtica$



$Si\ \ K = 3\quad Entonces\ se\ tiene\ ditribucion\ Mesocurtica$



4.5. Indicadores de normalidad

Sirven para tomar decisiones con niveles de confianza datos

4.5.1. Coeficiente Jarque Bera

Este indicador sirve para encontrar si los datos utilizados siguen una distribución normal con un nivel de confianza deseado (95%)

Calculo del estadístico Jarque Bera (JB)

$$JB = \frac{n}{6}\left[(CA_{Fisher})^2 + \frac{(K-3)^2}{4}\right]$$

Interpretación del coeficiente Jarque Bera con un 95% de confianza

Si "$JB < \chi^2_2 = 5.99$" Con un 95% de confianza se puede afirmar que la distribución de los datos tienen un comportamiento normal y se pueden garantizar las estimaciones con la media aritmética de la variable estudiada

Si "$JB \geq \chi^2_2 = 5.99$" se pude afirmar con un 95% de confianza se puede afirmar que la distribución de los datos no tiene un comportamiento normal y no se pueden garantizar las estimaciones con la media aritmética de la variable estudiada

Si la distribución no tiene un comportamiento normal; se pueden tomar las siguientes decisiones con respecto a los indicadores de tendencia central:

- Si la distribución no es normal y la distribución de los datos es asimétrica y platocurtica la medida de tendencia central que garantiza las estimaciones mas confiables es la mediana
- Si la distribución no es normal y la distribución de los datos es asimétrica y leptocurtica la medida de tendencia central que garantiza las mas confiables estimaciones es la moda
- Y en otros casos utilizar la media aritmética , pero con niveles de confianza bajos

4.6. Indicadores de concentración

Las medidas de concentración nos informan de la concentración de la distribución, entendida en un sentido distinto al de la antinomia "dispersión/ concentración": miden lo que podríamos llamar la concentración en sentido "económico":miden el mayor o menor "grado de igualdad en el reparto de la totalidad de los valores de la variable.

De esta manera si una pequeña parte de la población (unos pocos individuos) tiene una gran parte del total de la variable (renta, salario, capital, etc.), la variable estará muy concentrada (en pocas manos).Sin embargo, si se guardan las proporciones entre individuos y parte del total que se reparten la distribución será igualitaria, homogénea, poco o nada concentrada.

Las dos situaciones extremas serán:

IGUALDAD EN EL REPARTO:

Si suponemos la distribución agrupada por intervalos:

la proporción del monto total que se haya acumulado en el primer intervalo es igual a la proporción sobre el total de los individuos que se lo reparten.

la proporción del monto total que se haya acumulado en el segundo intervalo es igual a la proporción sobre el total de los individuos que se lo reparten.

y así sucesivamente hasta acumular todo el monto total y toda la población.

MÁXIMA CONCENTRACIÓN: Sólo un individuo acumula el total del monto y el resto de los individuos no tienen nada.

CURVA DE LORENZ

Una forma de representar la concentración de la distribución es a través de esta gráfica:

Suponiendo la distribución agrupada por intervalos y siendo xi la marca de clase de cada intervalo (el punto medio) el monto acumulado en el primer intervalo será: $u_1 = x_1 . n_1$

El monto acumulado hasta el segundo intervalo: $u_2 = x_1 . n_1 + x_2 . n_2$

El acumulado hasta el tercero: $u_3 = x_1 . n_1 + x_2 . n_2 + x_3 . n_3$

y así sucesivamente: ....

Hasta el monto acumulado hasta el último intervalo: $u_n = \square\ x_i . n_i$

Los porcentajes del monto total acumulados hasta cada intervalo (cada tramo) podrán obtenerse como:

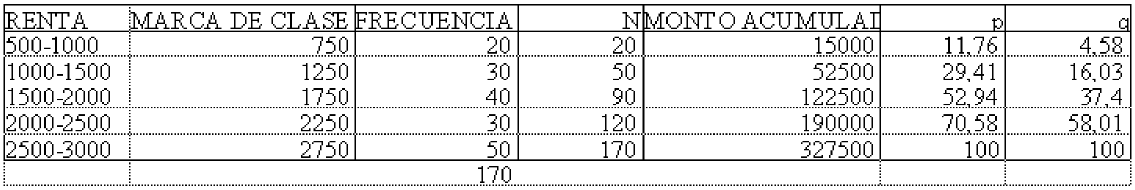
$$q_i = u_i / u_n .100$$

De la misma manera podemos obtener, para cada intervalo, el porcentaje (sobre el total de individuos) de los individuos que lo integran: $p_i = N_i / N .100$

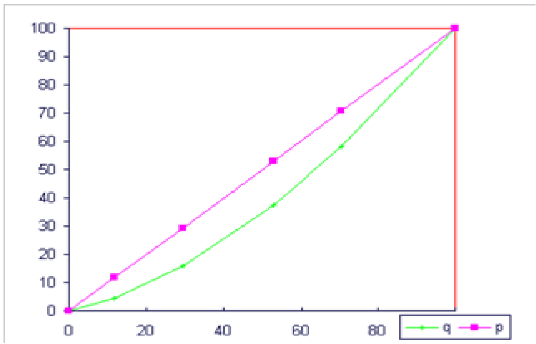
Considerando esto, es evidente que , en el caso de máxima igualdad en el reparto, q_i y p_i deben ir creciendo exactamente en la misma medida.

Una manera de representar la concentración de la distribución será a través de un gráfico de ejes coordenados en el que se representen los pares de puntos (q_i , p_i).La curva que una esos puntos será la Curva de Lorenz , que debería coincidir con la diagonal en el caso de máxima homogeneidad.

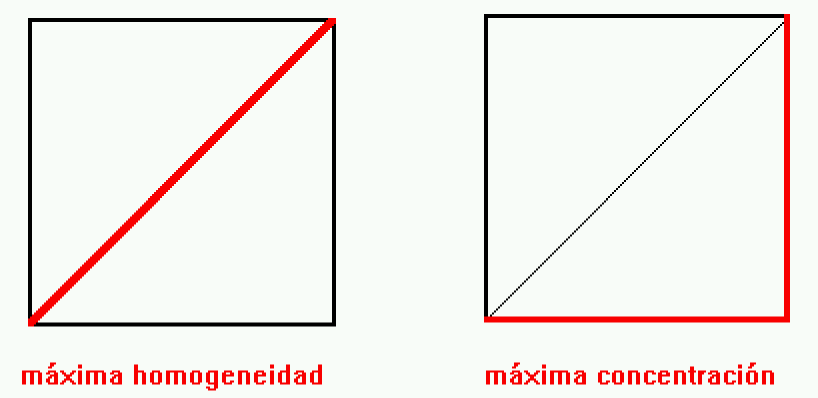
En un ejemplo de la distribución de la renta de 170 familias tendríamos:



RENTA	MARCA DE CLASE	FRECUENCIA	N	MONTO ACUMULAI	p	q
500-1000	750	20	20	15000	11,76	4,58
1000-1500	1250	30	50	52500	29,41	16,03
1500-2000	1750	40	90	122500	52,94	37,4
2000-2500	2250	30	120	190000	70,58	58,01
2500-3000	2750	50	170	327500	100	100
		170				



Así pues a través de la curva de Lorenz puede representarse gráficamente la concentración de la distribución. La curva obtenida puede compararse con la que se da en los dos casos extremos para darnos una idea de la concentración de la distribución estudiada:



Un indicador numérico de la concentración es el índice de concentración de Gini, que equivale al doble del área encerrada entre la curva de Lorenz y la diagonal. El índice de Gini tomará, entonces el valor 1 en el caso de máxima concentración y el valor 0 en el caso de máxima uniformidad. El índice de Gini admite la siguiente expresión de cálculo:

$$I_g = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

4.7. Indicadores de localización

Sirven para encontrar el valor sobre el cual se encuentra un porcentaje dado

4.7.1. Cuartiles

DATOS NO AGRUPADOS

Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9

DATOS AGRUPADOS

Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

$I.R$	$\grave{X}_i$	n_i	h_i	N_i	H_i
$[50-100]$	75	6	0.12	6	0.12
$[100-200]$	150	12	0.24	18	0.36
$[200-350]$	275	21	0.42	39	0.78
$[350-500]$	425	9	0.18	48	0.96
$[500-900]$	700	2	0.04	50	1.00
		50			

DATOS NO AGRUPADOS

Formula:

El lugar o posición donde se encuentran los cuartiles para ***n*** datos ordenados es:

$$Q_K = X_{\frac{K(n+1)}{4}}$$

DATOS AGRUPADOS

Formula:

Se encuentra el intervalo donde esta el cuartil buscado

$$Q_K = L_{i-1} + \left[\frac{\frac{Kn}{4} - N_{i-1}}{n_i}\right] C_i$$

Donde K = 0 , 1 , 2 y 3

4.7.2. Deciles

DATOS NO AGRUPADOS

Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9

DATOS AGRUPADOS

Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

$I.R$	$\grave{X}_i$	n_i	h_i	N_i	H_i
$[50-100]$	75	6	0.12	6	0.12
$[100-200]$	150	12	0.24	18	0.36
$[200-350]$	275	21	0.42	39	0.78
$[350-500]$	425	9	0.18	48	0.96
$[500-900]$	700	2	0.04	50	1.00
		50			

Formula: El lugar o posición donde se encuentran los deciles para ***n*** datos ordenados es: $D_K = X_{\frac{K(n+1)}{10}}$	Formula: Se encuentra el intervalo donde esta el decil buscado $D_K = L_{i-1} + \left[\dfrac{\frac{Kn}{10} - N_{i-1}}{n_i}\right] C_i$ Donde K = 0 , 1 , 2 …. 9

4.7.3. Percentiles

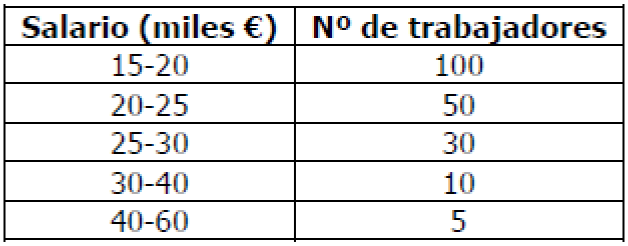
<table>
<tr><th>DATOS NO AGRUPADOS</th><th>DATOS AGRUPADOS</th></tr>
<tr>
<td>Ejemplo: Se tiene la información de nueve profesionales sobre el numero de libros que leyeron el año pasado

2 2 4 5 5 5 7 8 9</td>
<td>Se tiene la información de los ingresos semanales en miles de pesos de 50 profesionales el año próximo después de culminar su carrera

<table>
<tr><th>$I.R$</th><th>$\grave{X}_i$</th><th>n_i</th><th>h_i</th><th>N_i</th><th>H_i</th></tr>
<tr><td>$[50 - 100]$</td><td>75</td><td>6</td><td>0.12</td><td>6</td><td>0.12</td></tr>
<tr><td>$[100 - 200]$</td><td>150</td><td>12</td><td>0.24</td><td>18</td><td>0.36</td></tr>
<tr><td>$[200 - 350]$</td><td>275</td><td>21</td><td>0.42</td><td>39</td><td>0.78</td></tr>
<tr><td>$[350 - 500]$</td><td>425</td><td>9</td><td>0.18</td><td>48</td><td>0.96</td></tr>
<tr><td>$[500 - 900]$</td><td>700</td><td>2</td><td>0.04</td><td>50</td><td>1.00</td></tr>
<tr><td></td><td></td><td>50</td><td></td><td></td><td></td></tr>
</table>
</td>
</tr>
<tr>
<td>Formula:
El lugar o posición donde se encuentran los percentiles para ***n*** datos ordenados es:
$P_K = X_{\frac{K(n+1)}{100}}$</td>
<td>Formula:
Se encuentra el intervalo donde esta el percentil buscado
$P_K = L_{i-1} + \left[\dfrac{\frac{Kn}{100} - N_{i-1}}{n_i}\right] C_i$
Donde K = 0 , 1 , 2 …. 99</td>
</tr>
</table>

EJERCICIOS DE COMPLEMENTO

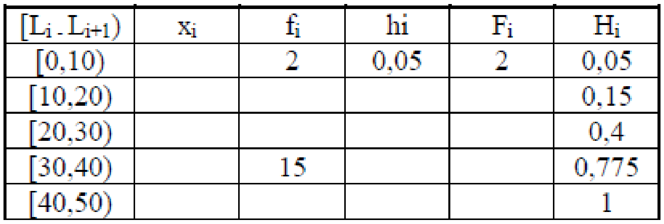
1. La distribución de los salarios percibidos por los trabajadores de una empresa es:



Salario (miles €)	Nº de trabajadores
15-20	100
20-25	50
25-30	30
30-40	10
40-60	5

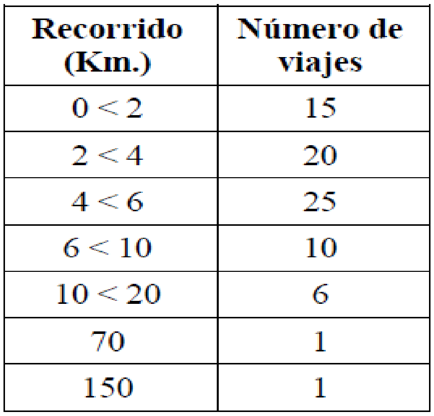
   a. Complete la tabla de frecuencias
   b. Calcule e interprete la media aritmética, Mediana y Moda? ¿Qué tipo de información ofrecen estos indicadores? ¿A qué son debidas las diferencias entre ambos?
   c. Calcule e interprete el coeficiente de variación?
   d. Calcule de interprete el coeficiente de asimetría?
   e. Calcule de interprete el coeficiente de curtosis
   f. Calcule e interprete el estadístico Jarque Bera?
   g. Calcule e interprete el percentil 48?

2. Completar la siguiente tabla:



$[L_i . L_{i+1})$	x_i	f_i	hi	F_i	H_i
[0,10)		2	0,05	2	0,05
[10,20)					0,15
[20,30)					0,4
[30,40)		15			0,775
[40,50)					1

3. Una empresa anota todos los meses los viajes efectuados y las distancias recorridas en cada viaje que realizan sus comerciales. Sistemáticamente agrupa las distancias de la misma forma y el último mes, la distribución es la que se muestra en la tabla adjunta:



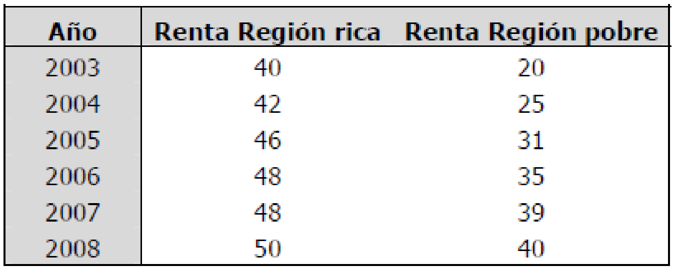
Recorrido (Km.)	Número de viajes
0 < 2	15
2 < 4	20
4 < 6	25
6 < 10	10
10 < 20	6
70	1
150	1

   a. Complete la tabla de frecuencias

b. ¿Cuántos km recorrieron los taxis ese día?, teniendo en cuenta únicamente los datos de la tabla. ¿Coincidirá obligatoriamente con la suma de km que registren los cuenta-kilómetros de cada taxi? ¿Por qué?.

c. Calcule e interprete el coeficiente de variación. Si al día siguiente se hicieron el mismo número de viajes, pero ninguno superior a 20 km, razone que variación cabría esperar en la desviación típica.

4. Se dispone de la renta per capita de la región más rica y más pobre de un determinado país.



Año	Renta Región rica	Renta Región pobre
2003	40	20
2004	42	25
2005	46	31
2006	48	35
2007	48	39
2008	50	40

Para la aplicación de una política territorial encaminada a reducir las diferencias existentes entre las distintas regiones del país, se pide calcular los siguientes resultados:

a. Calcule la renta media del periodo 2003-2008 en cada región. ¿En cuál de las dos regiones la renta media es más representativa?

b. A qué tasa media anual acumulativa ha crecido cada una de las dos regiones durante el periodo 2003-2008?

c. Suponiendo que a partir de 2008 se mantienen las mismas tasas de crecimiento medio anual ¿Cuántos años serán precisos para que la región pobre iguale el nivel de renta per capita de la región rica?

5. Observando el tipo de alquiler de en 390 viviendas da la capital federal se ha obtenido la siguiente distribución:

Tipos de alquiler	**F_i**
0-500	20
500-1000	140
1000-1500	180
1500-2000	40
2000-2500	10

a. Complete la tabla de frecuencias

b. Calcule e interprete la media aritmética, Mediana y Moda? ¿Qué tipo de información ofrecen estos indicadores? ¿A qué son debidas las diferencias entre ambos?

c. Calcule e interprete el coeficiente de variación?

d. Calcule de interprete el coeficiente de asimetría?

e. Calcule de interprete el coeficiente de curtosis

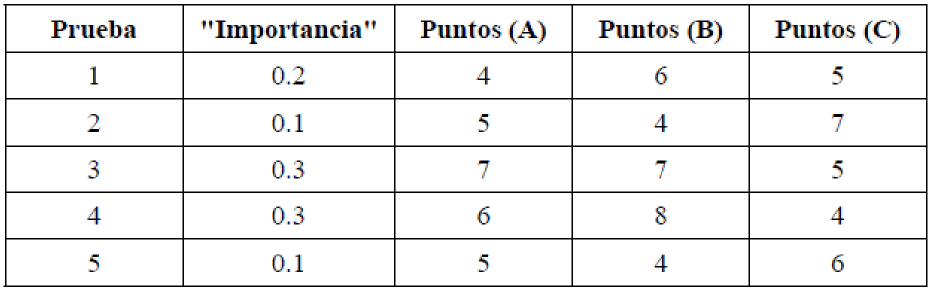
f. Calcule e interprete el estadístico Jarque Bera?

g. Calcule e interprete el percentil 48?

6. Dado el siguiente cuadro donde la variable es el "monto de venta" en pasos, calcular el promedio aritmético o monto mínimo por ventas:

X= monto de ventas	**Xi**	**Fi**	**Xi * Fi**
100-105	102,5	3	307,5
105-110	107,5	4	430
110-115	112,5	9	1012,5
115-120	117,5	6	705
120-125	122,5	2	245
125-130	127,5	1	127,5

7. La tabla muestra las puntuaciones obtenidas en una serie de 5 pruebas por tres candidatos (A, B, C) que se disputan una plaza.



Prueba	"Importancia"	Puntos (A)	Puntos (B)	Puntos (C)
1	0.2	4	6	5
2	0.1	5	4	7
3	0.3	7	7	5
4	0.3	6	8	4
5	0.1	5	4	6

Se pide:

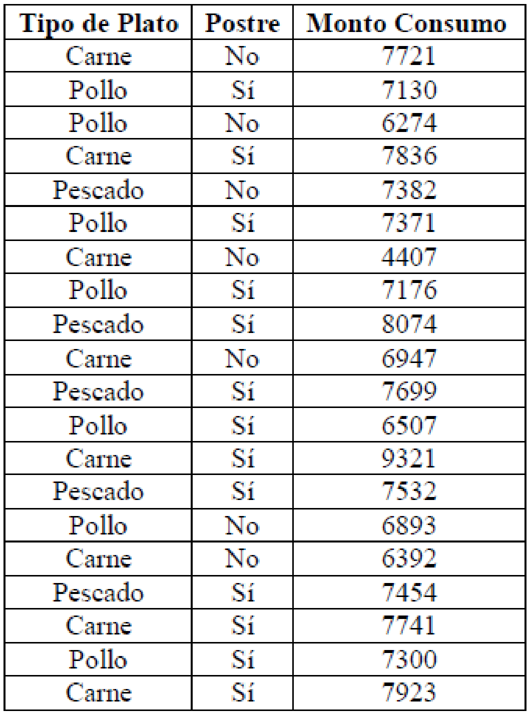
a. La media aritmética ponderada de cada candidato.
b. La media aritmética ponderada de los tres aspirantes.
c. Para este tipo de pruebas la media aritmética ponderada es, por lo general, de 7'4 puntos.
d. ¿Qué comentarios le merecen los tres candidatos que se han presentado?

8. Una Aerolínea recopila información en un año, sobre el número de vuelos y la edad de sus clientes. La siguiente tabla entrega los resultados en una muestra de 36 pasajeros.

Número de vuelos				
Años cumplidos	**1-3**	**3-5**	**5-7**	**Total**
15-25	1	1	2	**4**
25-45	2	8	10	**20**
45-75	1	6	5	**12**
Total	**4**	**15**	**17**	**36**

Con base en estadístico analice si el número de vuelos depende del tiempo de la aerolínea en el mercado

9. El dueño de un restaurante está interesado en estudiar los patrones de consumo de sus clientes. Con este fin tomó una muestra de 20 clientes para los que registró el tipo de plato que ordenó, si ordenó o no postre y el monto en bolívares del consuno realizado por cada cliente. Los datos se presentan en la siguiente tabla:



Tipo de Plato	Postre	Monto Consumo
Carne	No	7721
Pollo	Sí	7130
Pollo	No	6274
Carne	Sí	7836
Pescado	No	7382
Pollo	Sí	7371
Carne	No	4407
Pollo	Sí	7176
Pescado	Sí	8074
Carne	No	6947
Pescado	Sí	7699
Pollo	Sí	6507
Carne	Sí	9321
Pescado	Sí	7532
Pollo	No	6893
Carne	No	6392
Pescado	Sí	7454
Carne	Sí	7741
Pollo	Sí	7300
Carne	Sí	7923

a. El dueño está interesado en saber cómo está relacionado el monto del consumo con el tipo de plato, para saber a qué tipo de plato le daría prioridad en la preparación. Ayúdelo en su selección, tomando en cuenta todas las características de los datos y construyendo un gráfico que facilite la comprensión. Escriba sus conclusiones.
b. ¿Los datos apoyan la siguiente afirmación: “Los clientes que consumen carnes blancas en su mayoría piden postre”?
c. Si un cliente cualquiera desea ir al restaurante a comer pescado, ¿entre qué valores oscilaría el monto de su consumo?
d. ¿Cuánto sería el monto del consumo mínimo del 10% de los clientes que más gastan en el restaurante?

10. La siguiente tabla representa los resultados en la prueba de aptitud académica de un grupo de 1000 jóvenes que aspiran ingresar a cierta universidad:

Calificación	[300-350)	[350-400)	[400-450)	[450-500)	[500-550)	[550-600)
%Hi	6	28	45	63	95	100

a. Complete la tabla de frecuencias
b. Porcentaje de aspirantes cuya calificación es superior a 420 puntos pero inferior a 510
c. N° de estudiantes que obtuvieron 500 puntos o más
d. La mayor nota del 30% que obtuvo la nota más baja
e. Porcentaje que obtuvo más de 480 puntos
f. Coeficiente de asimetría de Pearson e interprete
g. La curtosis e intérprete.

11. Cierta cartera de valores ha pasado en 16 años de tener un valor de 1.000 dólares a valer 4.728 dólares. Halle el tipo de interés medio anual, o tasa media anual acumulativa, al que ha crecido esta inversión.

12. La rentabilidad media anual de un capital de 3.000 dólares invertido a un plazo de 10 años ha sido del 5,9%. En el transcurso de ese plazo el tipo de interés anual ha ido cambiando: los tres primeros años obtuvo un interés del 5,2% y los tres siguientes del 6,4%. Suponiendo que en los cuatro últimos años el tipo de interés anual no cambió, ¿cuál fue la rentabilidad anual en esos últimos cuatro años?

13. El volumen de ventas (millones de pesos) de una empresa de telefonía en el año 2009 se repartió de la siguiente manera:

    - En telefonía móvil las ventas de la empresa fueron de 7,51 millones de pesos. En el sector de telefonía móvil la media fue 6,61 y varianza de 86,5.
    - En telefonía fija la empresa alcanzó unas ventas de 8,41 millones de pesos. La media en el sector de telefonía fija fue de 7,2 y varianza de 117,79.

    a. ¿En qué unidades vendrá medida la varianza?
    b. ¿En cuál de los dos sectores está mejor situada la empresa en cuanto a su volumen de ventas? Razone la respuesta.

14. En una pequeña encuesta a 80 personas, se reveló que tales personas están dispuestas a no dar dinero, o dar hasta un máximo de US$ 16 para una campaña política. Del sondeo, 30 de ellas darían menos de US$ 8.

    a. Cuántas estarían dispuestas a dar más de US$ 11.
    b. Si el monto en el aporte de estas personas se incrementara en un 8%, ¿cuál sería el nuevo coeficiente de variación?

15. Para comparar la precisión de 2 instrumentos de medición, un técnico de laboratorio estudia mediciones hechas con ambos instrumentos. El primero se usó recientemente para medir el diámetro de un rodamiento y las mediciones tuvieron una media de 4,92 mm. con una desviación estándar de 0,018 mm. El segundo se empleó hace poco para medir la longitud sin extender de un resorte y las mediciones tuvieron una media de 2,54 pulgadas con una desviación estándar de 0,012 pulgadas. ¿Cuál de los 2 instrumentos es relativamente más exacto?. Sol: el primero

16. La población caleña se incremento de 875.000 habitantes a 1.700.000 habitantes de 1972 a 1993. Con base en esa información se pide calcular e interpretar:

    a. El incremento porcentual anual durante este periodo de tiempo
    b. Que población tenia Cali en 1984?
    c. Si la población caleña siguió a este ritmo de crecimiento; que población se espera que tenga Cali hoy

17. En un club de fútbol hay equipos que juegan en 3 categorías. Hay un 10% de jugadores que juegan en primera división, un 30% en segunda y un 60% en tercera división. Se sabe que en la temporada 2007-08 el sueldo medio para los jugadores de primera división fue de 500.000 euros al año, el de segunda división 300.000 € al año y para los de tercera 175.000 € al año.

    a. ¿Cuál fue el sueldo medio de los jugadores de todo el club?
    b. Para la temporada 2008-09 se ha mantenido la plantilla de los trabajadores, pero se han negociado incrementos salariales distintos para cada categoría. Se conocen sólo algunos aspectos de dicha negociación. El salario medio para el conjunto de la empresa será exactamente de 250.000 € anuales. El incremento previsto para los de primera división será del 10%, y para los de segunda un 8%. Tras conocer esta información los jugadores de tercera división deciden convocar una huelga indefinida en tanto en cuanto no se

revisen los incrementos salariales pactados ya que se consideran claramente desfavorecidos. Según la información de que dispone ¿Estaría de acuerdo con la actitud de dichos jugadores? Justifique su respuesta.

18. José Pérez es un directivo de una empresa de planificación financiera que asesora a quienes quieren establecer sus carteras de inversión personales. Hace poco José estaba interesado en las tasas de rendimiento que habían ofrecido dos fondos de inversión diferentes a lo largo de los 5 últimos años. FIVENEZ presentaba tasas de retorno a lo largo de ese período de 12, 10, 13,9 y 11%; mientras que Corporación Dinámica había producido 13, 12, 14, 10, y 6%. Un cliente se puso en contacto con el señor Pérez y expresó su interés por uno de estos fondos de inversión. ¿Cuál de ellos deberá elegir Pérez para su cliente? Sol: FIVENEZ.

19. La señorita Disbier Araque utiliza 2 máquinas diferentes para fabricar productos de salida de papel destinado a copiadoras Kodak. Los conductos de una muestra de la primera máquina medían 12,2; 11,9 ; 11,8 ; 12,1 ; 11,9 ; 12,4 ; 11,3 y 12,3 pulgadas. Los conductos hechos con la segunda máquina medían 12,2 ; 11,9 ; 11,5 ; 12,1 ; 12,2 ; 11,9 y 11,8 pulgadas. Disbier tiene que utilizar la máquina que produzca conductos de tamaños más uniformes. ¿Qué máquina deberá utilizar? Sol: la segunda máquina.

20. Cuatro fábricas A, B, C y D, producen un mismo objeto. La fábrica B produce el doble de C, la D 10% menos que la C y la A el 60% menos que la B. Los costos de producción (en dólares) por unidad de estas fábricas son respectivamente: 0.2, 0.3, 0,2, y 0.5. Calcular el precio medio de venta si se quiere ganar el 20% por unidad.

21. La señora Lorena de Ugueto es una agente de inversiones que encuentra 2 valores prometedores. El primero conduce a un rendimiento medio del 10% con una desviación típica del 1,2%; el segundo produce una tasa de rendimiento medio del 20% con una desviación típica del 5%. Con ayuda del CV como medida del riesgo, Lorena aconseja a su cliente más conservador que invierta en el primer valor. ¿Estaría usted de acuerdo?. Sol: sí estaría de acuerdo.

22. En un examen final de microeconomía, la puntuación media de 150 estudiantes fue 12,8 puntos y la desviación típica 2,3 puntos. En estadística el promedio fue 10,2 puntos y la desviación típica 1,6 puntos.

    a. En qué materia hay mayor dispersión relativa? Sol. Microeconomía
    b. ¿En qué materia destaca más un alumno que obtuvo 14 puntos en ambas? Sol. Estadística

23. En cierta evaluación para optar por una beca, Yulimar Prato obtuvo una calificación de 310 puntos en habilidad verbal y 218 puntos en habilidad numérica. Los parámetros de c/u son:

Habilidad verbal:	x = 245	S = 900
Habilidad numérica:	x =150	S = 24

    a. ¿En cuál de las dos pruebas obtuvo mejor calificación?
    b. ¿En cuál de las dos pruebas el grupo es más homogéneo?

24. Los salarios mensuales de 4 individuos son Bs. 150000, 160000, 165000 y 200000 Hallar el salario medio. Ahora entra a trabajar una nueva persona en la empresa, percibiendo un salario de Bs. 500000 mensuales. ¿Se verá afectado el salario medio tras esta incorporación?. ¿Crees que la media es una medida de centralización adecuada en los dos casos?. En caso de que no lo sea, propón y calcula otra medida de centralización más adecuada.

25. Tenemos dos variables X e Y con el mismo recorrido y media, siendo sus varianzas 4 y 9 respectivamente. ¿Para cual de las dos variables el valor de la media es más representativo?

26. Sea una variable con media 8 y desviación típica 0. ¿Qué se puede afirmar sobre el comportamiento de esta variable?.

27. Se les pidió a un grupo de estudiantes de la Universidad que hiciera una valoración global de la vida en una residencia universitaria. El dueño de la residencia quisiera alojar sólo a hombres porque piensa que éstos son mucho menos exigentes que las mujeres. Los resultados promedios obtenidos en una encuesta de valoración de la residencia entre 1(baja) y 10(alta), fueron los siguientes.

Hombres:	6,8 ;	9,1 ;	8,0	8,8 ;	8,8 ;	9,59 ;	7,90 ;	8,0 ;	8,0 ;	8,3 ;	10,0 ;	
Mujeres:	7,0 ;	7,1 ;	7,4 ;	7,3 ;	7,0 ;	7,5 ;	7,2 ;	7,15 ;	6,9 ;	6,8 ;	7,0	7,3

Realice un análisis descriptivo de los datos que incluya gráficos y las medidas descriptivas adecuadas para investigar si realmente el dueño tiene razón en lo planteado.

28. Complete las líneas en blanco que aparecen a continuación.

    a. Las medidas de tendencia central que siempre existen son _________ y _________.

    b. Cuando existen datos extremos no es adecuado el empleo de la ___________ como medida de tendencia central.

    c. La medida de variabilidad que es adecuada calcular cuando las medias de los grupos difieren es ______________.

    d. Cuando aproximadamente el 65% de los datos está en el intervalo [x-s; x+s] la distribución de los datos es:_________________.

    e. La clase ________es aquella dónde la frecuencia es mayor.

    f. El rango intercuantil se calcula como_______________.

    g. Los _______________ dividen a la distribución en cuatro partes iguales.

    h. La varianza indica la distancia promedio de cualquier observación del conjunto de datos con respecto a _______________.

    i. La diferencia entre el valor más alto de un conjunto de datos y el mínimo se conoce como ____________.

29. Complete las siguientes afirmaciones:

    a. Un _______________ es un gráfico conformado por una serie de rectángulos cuyas alturas dependen de la frecuencia absoluta o relativa de cada clase.

    b. Una _______________ es una colección de todos los elementos de un grupo. Un subconjunto de todos los elementos de un grupo es ____________.

30. Las notas de un profesor de Matemáticas tienen una media de 3.5 y una desviación media de 1.5. Otro profesor tiene una media de 3.0 y una desviación media de 0.45.

    a. Si deseas aprobar, ¿qué profesor elegirías?,
    b. ¿y si necesitas sacar nota?

c. ¿Siempre se aprueba con el último profesor?

31. Una empresa de servicios tiene dos sedes, una en Cali y otra en Medellín). Las ganancias mensuales netas en Cali son, por término medio, de 24.568 pesos con una desviación típica de 2.562 pesos. En Medellín las ganancias medias son de 8.700 pesos con una desviación típica de 2.750. Elige, en virtud de los respectivos coeficientes de variación, la respuesta adecuada:

   a. El mercado en Motril no es mucho más inestable que en Granada.
   b. El mercado en Motril unas dos veces más inestable que en Granada.
   c. El mercado en Motril unas tres veces más inestable que en Granada.

32. Los salarios mensuales de 4 individuos son: $150.000, $160.000, $165.000 y $200.000. Hallar el salario promedio. Ahora entra a trabajar una nueva persona en la empresa, percibiendo un salario de $500.000 mensuales.

   a. Se vera afectado el salario promedio tras esta incorporación?. Justifique
   b. Crees que la media es una medida de tendencia central adecuada en los dos casos? En caso de que no lo sea propón y calcula otra medida de tendencia central adecuada

33. En una encuesta sobre los ingresos anuales en miles de pesos de un grupo de familias se obtuvo la siguiente información:

$\left[L_i - L_s\right)$	f_i
100 – 300	20
300 – 500	
500 – 700	
700 - 900	20

Además, $\bar{x} = 540$ y $f_2 / f_3 = 1/5$, calcular el número de familias con ingreso no menos de 500 mil pesos.

34. Dada la siguiente distribución de frecuencias, calcular el valor de “n” sabiendo que la moda es 60 y pertenece al tercer intervalo.

$\left[L_i - L_s\right)$	f_i
16 – 32	6
32 – 48	n
48 – 64	8
64 – 80	3n
80 - 96	3

35. Cien estudiantes divididos en cuatro grupos A, B, C y D dan un examen y obtienen un promedio general de 72 (calificación centesimal). Los puntajes medios de los grupos A, B, C son 75, 62, 80, respectivamente. Los registros del grupo D se extraviaron; pero se sabe que en el grupo A están el 40% del total de alumnos, en el grupo B un cuarto del total, en el grupo C había 15 alumnos más que en el grupo D. Determinar el promedio del grupo D.

36. En una clase de 30 jóvenes la nota promedio de los aprobados fue de 3.0 y la de los que no aprobaron fue de 1.5; sabiendo que la nota media de la clase fue de 2.5 ¿Que % hubo de aprobados y no aprobados?

37. El sueldo promedio de 200 empleados de una empresa es $400. Se proponen dos alternativas de aumento: a) $75 a cada uno, b) 15% de su sueldo más 10 pesos a cada uno. Si la empresa dispone a lo más de $94,000 para pagar sueldos, ¿cuál alternativa es más conveniente?.

38. Bill Karl compró 20 acciones a $ 15 cada una, 50 acciones a $20 cada una, 100 acciones a $30 cada una y 75 acciones a $35 cada una. ¿Cuál es el precio promedio por acción?

39. Completa los datos que faltan en la siguiente tabla estadística:

X_i	n_i	N_i	h_i
1	4		0.08
2	4		
3		16	0.16
4	7		0.14
5	5	28	
6		38	
7	7	45	
8			

Calcula la media, mediana y moda.

40. Durante los últimos 10 días del mes de septiembre el tren " de las nubes " llego tarde a su destino de acuerdo con los siguientes retrasos en minutos (Los números negativos significan que llego antes)

Tiempo de retrazo (En minutos)

N	Válidos	10
Media		14,30
Mediana		4,00
Moda		3
Desv. típ.		38,77
Asimetría		3,093
Error típ. de asimetría		,687
Curtosis		9,679
Error típ. de curtosis		1,334
Mínimo		-4
Máximo		124
Percentiles	25	-1,50
		3,00
		7,00

a. Si el ferrocarril lo contratara a usted para demostrar que proporciona un buen servicio: Que medida utilizaría y porque?
b. Si lo contratara una estación de televisión para demostrar que el tren proporciona un mal servicio. Que medida utilizaría y porque?
c. Intente juzgar en forma objetiva el desempeño del ferrocarril.

41. Suponga que usted presenta un examen a la universidad del valle con la intención de matricularse a medicina, pero para la cual presentaron cinco exámenes a saber asignándole a cada uno el nivel de importancia de la siguiente forma: Ciencias el 50%, Matemáticas el 20%, actitud verbal el 20%, Cultura general el 5% y razonamiento abstracto el 5%. Usted sabe que para ser admitido necesita estar sobre un puntaje promedio de 80 puntos en una escala de O a 100. Si sus resultados fueron: 50 en ciencias, 70 en matemáticas, 95 en actitud verbal, 98 en cultura general y 100 en razonamiento abstracto.

a. Tiene esperanzas de ser admitido
b. Donde radica el problema de no ser admitido
c. Cual sugeriría usted para ser admitido y porque

42. Por el incremento del costo de la vida se plantean dos alternativas de aumento para le mes de septiembre. **La primera consiste** en un aumento del 30% de los salarios de agosto a los empleados que ganan menos 1.000.000 de pesos y del 5% a los empleados que ganan mas de 1.000.000 de pesos y un aumento de adicional de 100.000 pesos a todos los empleados. **La segunda propuesta** consiste en un aumento general de 250.000 pesos. Analice

    a. Para el empleado que gana menos de 1.000.000 de pesos. Que propuesta le convendría?
    b. Si se acepta la primera propuesta. Es la distribución de septiembre mas homogénea que la distribución de salarios del mes de agosto?

43. Por error un ingeniero borro la calificación que obtuvo uno de sus 10 operarios. Si los otros 9 consiguieron las siguientes calificaciones es: 4.3, 6.6, 4.7, 6.4, 5.0, 5.2, 7.0, 5.8 y 6.2 y si la calificación media de los diez operarios fue de 5.8. Que calificación borro el ingeniero?

44. De 2000 al 2005 el costo de los alimentos se incremento en un 53% en cierta ciudad, el costo que el trabajador destina al alquiler subió en un 40%, el transporte creció en un 34%, si el porcentaje de salario que el trabajador destina a la alimentación es de un 28%, al alquiler 35% y al transporte 14%. Cal es el incremento porcentual promedio del costo de estos bienes?

45. A partir de una muestra de 10 datos se obtuvieron los siguientes resultados: media aritmética = 4, mediana = 5. realizados los cálculos se descubre que la observación con valor mas pequeño estaba equivocada en lugar de 2 era 1.

    - Cual es el valor correcto de la media aritmética? Fundamente su respuesta
    - Cual es el valor correcto de la mediana? Fundamente su respuesta

46. Los porcentajes de atención en las cuatro ultimas elecciones generales de un determinado país fueron: 36.63%, 25.53%, 39.5% y 48.9%. Cual es el promedio electoral en estos últimos años?

47. Un automovilista participa en una competición en la cual obtiene, para los distintos recorridos, las siguientes velocidades medias

Recorrido	Distancia (Km)	Velocidad media (Km/h.)
A-B	400	50
B-C	600	60
C-A	1000	100

Calcule la velocidad media conseguida en la competición?

48. Se tiene una muestra de 130 miembros de una Cooperativa de la cual se conoce que la edad mínima era de 21 años y la edad máxima era 66 años. Con base en estos datos:

    a. Indique el rango
    b. Indique el numero de intervalos a realizar
    c. Calcule la amplitud del intervalo
    d. Calcule los dos primeros intervalos reales
    e. Calcule los dos primeros intervalos de clase
    f. Calcule la marca de clase para estos dos intervalos

VII. CONCEPTOS BASICOS DE PROBABILIDAD

1.1. Conjuntos

Conjunto o colección: La idea de conjunto es una idea intuitiva. Se representa generalmente por una letra mayúscula.

C = { átomo, d, mesa, perro}.

Elemento: Elemento es todo aquello que constituye a un conjunto dado. En el ejemplo anterior la letra d es un elemento del conjunto C.

Observación 1: En un conjunto dado ninguno de sus elementos debe aparecer repetido.

B = { a , a , b } , debe escribirse : B = { a , b }.

Observación 2: Un conjunto formado por un solo elemento es conceptualmente distinto a dicho elemento.

{ v } es distinto a: v.

Relación de pertenencia ($\in$) : Si un elemento está en un conjunto dado, se dice que pertenece a él y ésto se indica mediante el símbolo $\in$.

En el ejemplo 1, $d \in C$.

Para indicar lo contrario se usa $\notin$.

En el ejemplo 2, $d \notin B$.

Determinación de un conjunto

Por extensión: Indicando cada uno de los elementos que lo forman.

D = { Gabriela Mistral, Pablo Neruda}.

Por comprensión: Indicando alguna(s) propiedad(es) que cumplen todos sus elementos y solamente ellos.

D = { Poetas chilenos que han obtenido el Premio Nobel de Literatura }

Universo, espacio o conjunto referencial (U o *Ω*) : Se nombra así al conjunto formado por todos los elementos de un tema dado.

U = { a , e , i , o , u } (Tema : vocales minúsculas del abecedario castellano) .

Conjunto vacío ($\varnothing$) : Es el conjunto que no tiene elementos. También puede decirse que ningún elemento del universo cumple la condición dada en él.

$$\{ \text{Especies de insectos de 10 patas} \} = \{ \ \} = \varnothing$$

Relación de igualdad (=) : Dos conjuntos son iguales si y sólo si están formados por los mismos elementos.

$$A = B \iff \forall x (x \in A \iff x \in B)$$

$$\{ 1 , 2 , 3 \} = \{ 2 , 1 , 3 \}.$$

Relación de inclusión ($\subset$) : Sean A y B conjuntos, entonces A está incluido en B, o bien A es un subconjunto de B, si y sólo si cada elemento de A lo es también de B .

$$A \subset B \iff \forall x (x \in A \Rightarrow x \in B)$$

Si $A = \{ p , q \}$ y $B = \{ m , n , p , q , r \}$, entonces $A \subset B$.

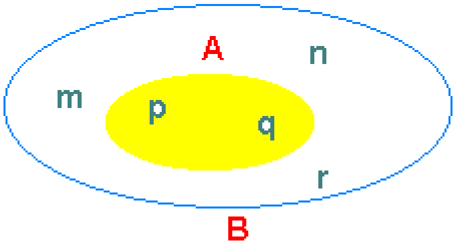


Diagrama de Venn - Euler **Diagrama lineal**

Teorema 1: El conjunto vacío está incluido en cada conjunto. Ejemplo: $\varnothing \subset A$.

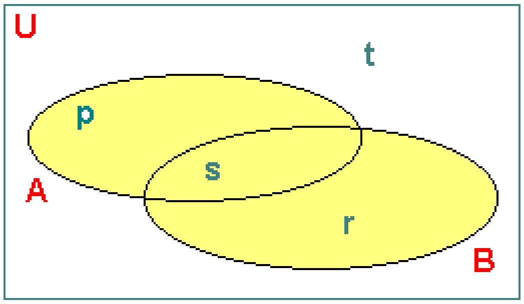
Teorema 2: Cada conjunto está incluido en su universo respectivo. Ejemplo: $A \subset U$.

Teorema 3: Cada conjunto está incluido en sí mismo. Ejemplo: $A \subset A$.

Operaciones con conjuntos

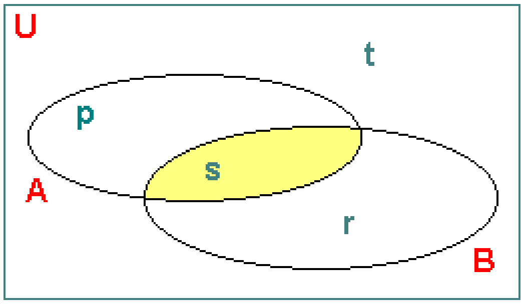
Para ejemplificar, sean $U = \{ p , r , s , t \}$, $A = \{ p , s \}$ y $B = \{ r , s \}$.

Unión o reunión ($\cup$) : El conjunto $A \cup B$ está formado solamente por todos los elementos que pertenecen a A, o a B, o a ambos.



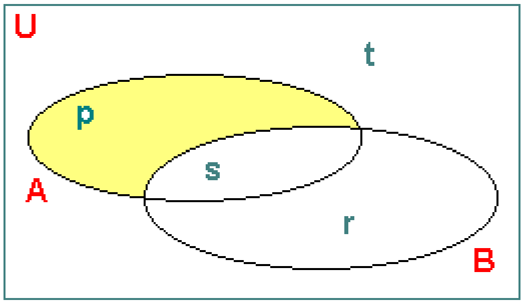
$$A \cup B \ = \ \{ x : \ x \in A \ \vee \ x \in B \} \ = \ \{ p , r , s \}$$

Intersección ($\cap$): El conjunto $A \cap B$ está formado solamente por todos los elementos que pertenecen a A y a B simultáneamente.



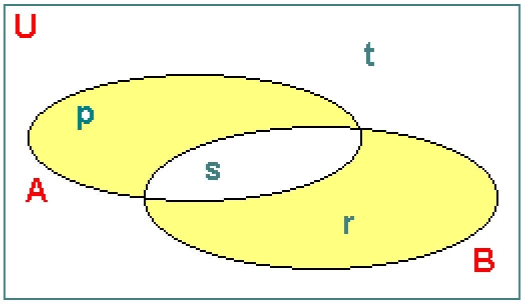
$$A \cap B \ = \ \{ x : \ x \in A \ \wedge \ x \in B \} \ = \ \{ s \}$$

Diferencia (–): El conjunto A – B está formado solamente por todos los elementos que pertenecen a A, pero que no pertenecen a B.



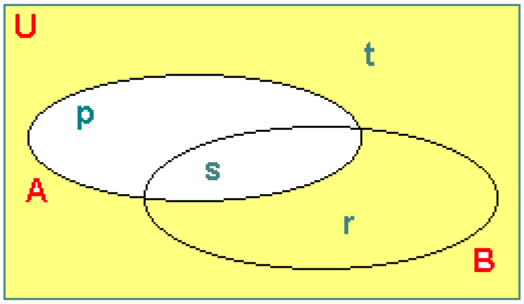
$$A - B \ = \ \{ x : \ x \in A \ \wedge \ x \notin B \} \ = \ \{ p \}$$

Diferencia simétrica (Δ): El conjunto $A \,\Delta\, B$ está formado solamente por todos los elementos que pertenecen a A o a B, pero no a ambos.



$$\mathbf{A \Delta B = \{x : x \in A \oplus x \in B\} = \{p, r\}}$$

Complemento (') (^{c}): El conjunto A' está formado solamente por todos los elementos del **U** que no pertenecen a A.



$$\mathbf{A' = \{x : x \notin A\} = \{r, t\}}$$

Observación 3: Dos conjuntos son disjuntos si y solo si no tienen elementos comunes, es decir, su intersección es vacía.

$A \cap A' = \varnothing \quad \therefore \quad A$ y A' son disjuntos.

Conjunto potencia: Sea A un conjunto dado, entonces su conjunto potencia, P(A), está formado por todos los subconjuntos de A y sólo por ellos.

$$P(A) = \{X : X \subset A\}$$

Además:

$$\# P(A) = 2^{\# A}$$

Ejemplo Sea $A = \{a, b, e\}$, entonces :

$$P(A) = \{\varnothing, \{a\}, \{b\}, \{e\}, \{a, b\}, \{a, e\}, \{b, e\}, A\}.$$

Par ordenado: Es una pareja de elementos en un orden dado. Ejemplo: (x, y) .

Teorema 4: $(x, y) = (z, w) \quad \Leftrightarrow \quad x = z \wedge y = w$.
Teorema 5: $(x, y) = (y, x) \quad \Leftrightarrow \quad x = y$.

Producto cartesiano: Dados dos conjuntos A y B su producto cartesiano (A × B) se define así:

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

Ejemplo 15: Sean $A = \{a, b, c\}$ y $B = \{d, e\}$, entonces su producto cartesiano es :

$$A \times B = \{(a, d), (a, e), (b, d), (b, e), (c, d), (c, e)\}.$$

Tipo de conjuntos

Conjuntos mutuamente excluyentes: se dice que dos o mas conjuntos son mutuamente excluyentes si y solo si no tienen ningún elemento en común

$$A \cap B \cap C \cap D = \varphi$$

Conjuntos colectivamente exhaustivos: se dice que dos o más conjuntos son colectivamente exhaustivos si la unión de ellos nos conduce la conjunto universal

$$A \cup B \cup C \cup D = \Omega$$

1.2. Conceptos básicos

Notación factorial

$$n! = x! = n(n-1)(n-2)(n-3)\ldots\ldots\ldots 1$$

Ejemplo: 4! = 4x3x2x1 = 24

Permutaciones

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

$$P(5,2) = \frac{5!}{(5-2)!} = \frac{5!}{3\cdot} = 20$$

Combinatorios

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

$$C(5,2) = \binom{5}{2} = \frac{5!}{(5-2)!\,2!} = \frac{5!}{3! * 2!} = 10$$

1.3. Técnicas de conteo

1.3.1. Regla de la multiplicación

$$Numero\ de\ muestras = \prod_{i=1}^{n} E_i = E_1E_2E_3\ldots.E_n = \mathbb{N}_1\mathbb{N}_2\mathbb{N}_3\ldots..\mathbb{N}_n$$

Ejemplo. Jorge propietario de un concesionario ofrece a sus clientes automóviles con ocho opciones de color, cuatro paquetes de interior y tres diseños diferentes de techo corredizo. Entre cuantos automóviles pueden escoger los clientes de Jorge?

$$Numero\ de\ escogencias = \prod_{i=1}^{3} E_i = E_1E_2E_3 = \mathbb{N}_1\mathbb{N}_2\mathbb{N}_3 = 8x4x3 = 96$$

1.3.2. Muestras ordenadas con sustitución

$$\boldsymbol{P`(n,r) = n^r}$$

Ejemplo: De cuantas formas puede un estudiante contestar un examen de 8 preguntas si sus respuestas son de tipo falso o verdadero

$$\boldsymbol{P`(2,8) = 2^8 = 256}$$

1.3.3. Muestras ordenadas sin sustitución

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

Ejemplo. Un una carrera compiten 10 caballos. En los boletos hay que indicar el nombre del primero, segundo y tercero. Cuantos debemos llenar para asegurarnos que ganaremos?

$$\boldsymbol{P(10,3) = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 720}$$

1.3.4. Muestras no ordenadas con sustitución

$$\boldsymbol{C`(n,r) = \binom{n+r-1}{r}}$$

Ejemplo. De cuantos formas nos pueden regalar un cono con tres bolillas; si la persona que lo regala dispone de 6 sabores diferentes?

$$\boldsymbol{C`(6,3) = \binom{6+3-1}{3} = \binom{8}{3} = 56}$$

1.3.5. Muestras no ordenadas sin sustitución

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

Ejemplo. Un club tiene 15 miembros. Desean elegir un presidente y alguien más como vicepresidente. ¿De cuántas formas pueden llenarse esos cargos?

$$\boldsymbol{C(15,2) = \binom{15}{2} = \frac{15!}{(13)!2!} = 105}$$

1.3.6. Particiones

$$\boldsymbol{P[n; r: r_1 \;\; r_2 \quad r_3 \ldots\ldots\ldots\ldots r_n] = \frac{n!}{r_1!\, r_2!\, r_3! \ldots \ldots . . r_k!}}$$

Ejemplo. Un grupo de jóvenes aventureros han decidido conocer diez países de Europa, de los cuales no tienen ninguna información. Luego de marcharse de cada país, de acuerdo a la vista en su corta visita califican como ^Muy agradable ^, ^Agradable ^ y ^No agradable ^. De cuantas maneras podrán calificar a cuatro países como muy agradables, tres como agradables y 3 como no agradables?

$$\boldsymbol{P[10; 3: 4 \;\; 3 \quad 3] = \frac{10!}{4!\, 3!\, 3!} = 4200}$$

EJERCICIOS DE COMPLEMENTO

1. Sean los sucesos A y B tales que; P(A) = 1/5 y P(B) = 1/3. Determine el valor de $P(A^c \cap B^c)$ para cada una de las siguientes condiciones:

    a. Los dos eventos son mutuamente excluyentes
    b. Los dos eventos son independientes
    c. A C B (Si A esta contenida en B
    d. $P(A \cap B) = 1/9$

2. Si A, B y C son eventos correspondientes a un cierto experimento aleatorio; determine gráficamente los siguientes sucesos:

    a. Ocurre solo A
    b. Ocurre A y B pero no C
    c. Ocurren los tres
    d. Ocurre al menos uno
    e. Ocurre al menos dos
    f. Ocurre uno y solo uno
    g. Ocurren dos y solo dos
    h. No ocurre ninguno
    i. No ocurren mas de dos

3. Una encuesta sobre 200 personas revelo los siguientes datos acerca del consuno de tres productos: A, B y C: 30 personas consumían A, 85 consumían B, 103 consumían C, 1 consumían A y C pero no B, 13 personas consumían A y C, 18 personas B y C, y 5 personas consumían A y B pero no C.

    a. Cuantas personas no consumieron ninguno de los tres productos? 18
    b. Cuantas personas consumieron los tres productos? 3
    c. Cuantas personas consumieron A pero no B ni C? 12
    d. Cuantas personas no consumieron A? 170
    e. Cuantas personas consumieron por lo menos uno de los tres productos? 181

4. Se llevó a cabo una investigación con 1000 personas, para determinar que medio utilizan para conocer las noticias del día. Se encontró que 400 personas escuchan las noticias en forma regular por TV, 300 personas escuchan las noticias por la Radio y 275 se enteran de las noticias por ambos medios.

    a. ¿Cuántas de las personas investigadas se enteran de las noticias solo por la TV?
    b. ¿Cuántas de las personas investigadas se enteran de las noticias solo por Radio?
    c. ¿Cuántas de las personas investigadas no escuchan ni ven las noticias?

5. En una encuesta aplicada a 1000 empleados de un centro comercial sobre el tipo de transporte que utilizan para ir de sus casas al trabajo se obtuvo la siguiente información:

    431 empleados utilizan metro.
    396 empleados utilizan autobús.
    101 empleados utilizan metro y trolebús pero no autobús.
    176 empleados no utilizan ninguno de los tres medios considerados.
    341 utilizan trolebús.
    634 utilizan metro o trolebús.
    201 utilizan sólo metro.

a. ¿Cuántos empleados utilizan metro o trolebús pero no autobús? R/ 428
b. ¿Cuántos empleados utilizan sólo uno de los tres medios de transporte mencionados? R/ 517
c. ¿Cuántos empleados utilizan sólo trolebús? R/126
d. ¿Cuántos empleados utilizan metro, trolebús y autobús? R/37

6. La revista semana clasifico 120 ciudades colombianas de acuerdo con la calidad de vida, con base en parte del porcentaje de empleados que tenían titulo universitario; donde A es menos del 15% con titulo universitario, B entre el 15 y 20% con titulo universitario y C mas del 20% con titulo universitario. Los resultados se presentan en la siguiente tabla:

	Calidad de vida			
% con titulo universitario	Pobre (P)	Bueno (B)	Excelente (E)	Total
A	10	20	10	40
B	10	30	20	60
C	0	10	10	20
Total	20	60	40	120

Responda las siguientes preguntas de acuerdo a la tabla:

a. P(A)
b. $P(P \cap B)$
c. $P(P \cap C)$
d. Dado un rango de excelente; cual de las tres categorías porcentuales es mas probable que ocurra
e. Si el 19% de los empleados de una ciudad tienen titulo universitario; cual es la probabilidad de que la calidad de vida se clasifique como pobre, bueno o excelente
f. Si mas del 20% de los empleados de una ciudad tienen titulo, Cual es la probabilidad de que la ciudad sea clasificada como excelente
g. Si una ciudad es clasificada como excelente , Cual es la probabilidad que mas del 20% de sus empleados tengan titulo universitario

7. Para armar la siguiente tabla se han tenido en cuenta las clasificaciones: N, A, S.

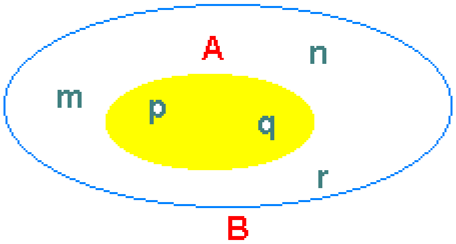
	sexo		
calificación	**Mujer**	**Varon**	TOTAL
N	7	9	16
A	10	8	18
S	2	4	6
TOTAL	19	21	40

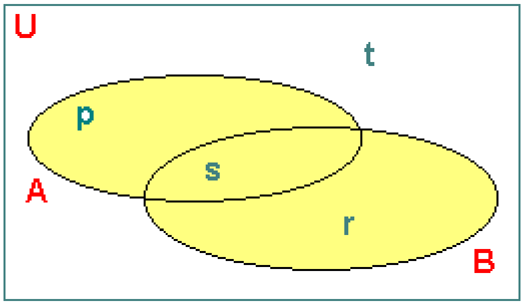
Si entre los 40 alumnos de dicho curso, se elige 1 al azahar, hallar la probabilidad de que:

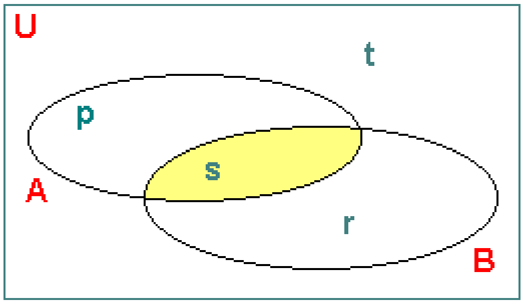
a. Haya obtenido A en la evaluación
b. Haya obtenido A sabiendo que el alumno elegido es varón.

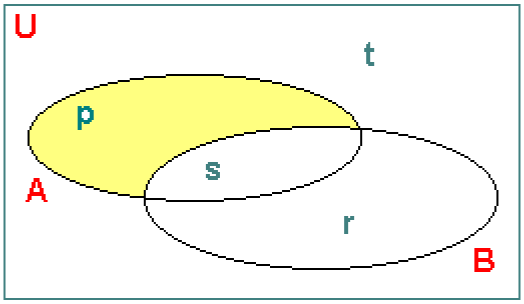
8. En una clase de 20 alumnos se van a conceder tres premios: uno al más destacado en matemáticas, otro al mejor en historia y otro al mejor deportista. De cuantas formas distintas podemos hacerlo?

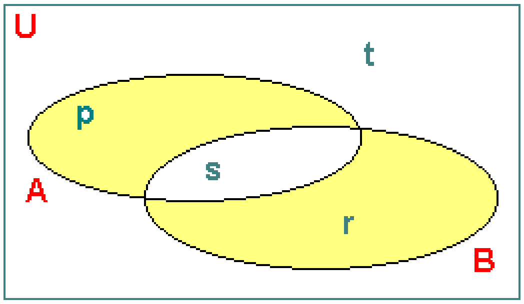
9. Identifique que conjunto describe la grafica presentada

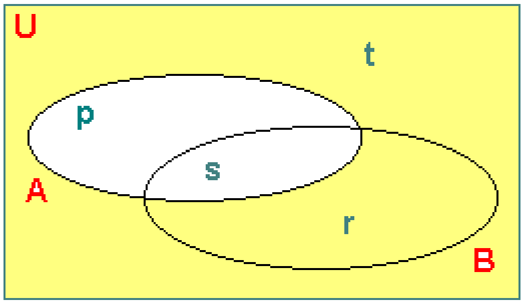












10. Identifica las regiones que comprende cada uno de los conjuntos siguientes en un diagrama de Venn adecuado:

a. $A^C \cup B$

b. $A^C \cap B^C$

c. $(A \cup C) - B$

d. $[(A \cup B) \cup (B \cap C) \cup (C \cap A)] - (A \cap B \cap C)$

e. $C^C - (A^C \cap B)$

f. $B \cap \left(A \cup C^{C}\right)$

11. La siguiente tabla muestra la distribución de 400 personas según hábito de fumar y presencia de bronquitis.

HÁBITO DE FUMAR	BRONQUITIS		TOTAL
	SI	NO	
FUMA	140	110	250
NO FUMA	50	100	150
TOTAL	190	210	400

a. Si se elige una persona al azar ¿Cuál es la probabilidad de que:
    - ✓ Fume y tenga bronquitis
    - ✓ No fume dado de que tiene bronquitis
    - ✓ No tenga bronquitis dado que fuma
    - ✓ No fume o tenga bronquitis.

b. Los sucesos "Fumar" y "Tener bronquitis" son independientes?

12. Un grupo de jóvenes fue entrevistado sobre sus preferencias por ciertos medios de transporte (bicicleta, moto y auto). Los datos de la encuesta fueron los siguientes: Moto solamente: 5 ; Moto: 38 , No gustan de auto: 9 ; Moto y bicicleta, pero no auto: 3 ; Moto y auto pero no bicicleta: 20 ; No gustan de bicicleta: 72 ; Ninguna de las tres cosas: 1 y No gustan de la moto: 61 ¿A cuántos les gustaba las tres cosas?

13. Ocho amigos van de viaje llevando para ello solo dos coches. Si deciden ir 4 en cada coche:

    a. De cuantas formas pueden ir si todos tienen licencia de conducción
    b. De cuantas formas pueden ir si solo tres tienen licencia de conducir?

14. Una señora tiene tres frutas: piña, guayaba y mango. Cuantos sabores diferentes de jugos puede preparar con estas frutas?

15. ¿Cuántas palabras de once letras pueden formarse con la palabra Mississippi?.

16. Un matrimonio quiere invitar a sus amigos a cenar. Debido a las dimensiones de la casa solo puede invitar a cinco cada vez. Si se quieren invitar a 10 amigos. De cuantas formas pude invitar a cinco de ellos

17. Se distribuyen tres regalos entre cinco chicos. De cuantas formas puede hacerlo si:

    a. Cada chico solo puede recibir un regalo R/60
    b. A cada chico le puede tocar mas de un regalo R/125
    c. Cada chico solo puede recibir un regalo, pero los tres son idénticos R/ 10

18. De cuantas formas pueden cubrir los puestos de presidente y secretario de una comunidad de vecinos, contando con 10 vecinos para ello?

19. Un hombre tiene tiempo para jugar a la ruleta a lo sumo cinco veces. En cada juego gana o pierde un dólar. El hombre empieza con 1 dólar y dejara de jugar si antes de la quinta vez pierde todo su dinero o si gana tres dólares.. Esto es si tiene cuatro dólares. Hallar el numero de casos en que la apuesta puede ocurrir?

20. Un restaurante ofrece cebolla, salsa, mostaza y picante como condimento para su agregado a una hamburguesa simple. Cuántas clases de hamburguesas puede preparar si los sabores se clasifican en: sin sabor, con uno, con dos, tres o cuatro condimentos a la vez?

21. Un entrenador dispone de 22 jugadores para formar un equipo de fútbol. Cuantas alineaciones de 11 jugadores puede hacer?. R/ 705432

22. Tres matrimonio se reúnen para celebrar el aniversario de uno de ellos. Desean que se les haga una fotografía de forma que estén todos los hombres y también las mujeres. De cuantas formas distintas pueden colocarse?

23. Un una carrera compiten 10 caballos. En los boletos hay que indicar el nombre del primero, segundo y tercero. Cuantos debemos llenar para asegurarnos que ganaremos?

24. En una clase de 20 alumnos se van a conceder tres premios: uno al mas destacado en matemáticas, otro al mejor en historia y otro al mejor deportista. De cuantas formas distintas podemos hacerlo?

25. Ocho amigos van de viaje llevando para ello solo dos coches. Si deciden ir 4 en cada coche:
    - De cuantas formas pueden ir si todos tienen licencia de conducción
    - De cuantas formas pueden ir si solo tres tienen licencia de conducir?

26. Cinco amigos que están en una piscina, después de haberse lanzado por el deslizadero gigante, observan que cada vez que llegan a la parte superior para el nuevo lanzamiento hacen cola en distinto orden. ¿De cuántas formas podrán hacer cola para arrojarse de nuevo?

27. En un campeonato suramericano de Fútbol llegan a un cuadrangular final los cuatro seleccionados de Brasil, Argentina, Colombia y Uruguay. Formar las diferentes clasificaciones para los cuatro primeros puestos del torneo. ¿Cuántas hay?

28. Un apostador tiene el presentimiento de que en la próxima jornada futbolística (en un torneo nacional con 28 equipos) ganarán 9 equipos en casa, empatarán 3 y ganarán en campo contrario (de visitantes) 2. ¿Cuántas apuestas deberá realizar para asegurarse un pleno de 14?

29. En una frutería ofrecen entre sus productos distintas mezclas con zumos de frutas. El cliente puede seleccionar entre 6 zumos de frutas diferentes y obtener algún sabor en particular de la mezcla de dos zumos en partes iguales. ¿Entre cuántos sabores distintos puede el cliente hacer su pedido?

30. En una clase de 20 alumnos se van a conceder tres premios: uno al más destacado en matemáticas, otro al mejor en historia y otro al mejor deportista. De cuantas formas distintas podemos hacerlo?

31. Como respuesta a un anuncio de trabajo se presentan 15 personas para cubrir tres cargos administrativos. ¿Cuántos grupos diferentes de tres personas se pueden formar?

32. Un experto ladrón de joyas intenta abrir un maletín que posee un sistema de seguridad con una clave de cuatro dígitos. El ladrón sabe que el primer digito puede ser cuatro o cinco y que el tercero es exactamente 8. Cual será el numero máximo de intentos que deberá realizar el ladrón para abrir el maletín?

33. Resolver $P'(n:2)+5P'(n-2:2)=70$

34. Resolver $P(n:3)-5P(n:2)=0$

35. Resolver $C(n:3)=5C(n-1:4)$

36. Una familia formada por los padres y tres hijos van al cine. Se sientan en cinco butacas consecutivas:

   a. De cuantas maneras distintas pueden sentarse R/120 = P[5,5]
   b. Y si los padres se sientan en los extremos R/ 12 = 2 P[3,3)
   c. Y si los padres no deciden sentarse en los extremos R36 = P[3, 2] P[3,3]

37. Hay 6 carreteras entre A y B y 4 carreteras entre B y C. a) ¿De cuántas maneras se puede viajar de A a C pasando por B? b) ¿De cuántas maneras se puede hacer el viaje de ida y vuelta de A a C pasando por B? c) ¿De cuántas maneras se puede hacer el viaje de ida y vuelta de A a C sin usar la misma carretera más de una vez B?

38. En Cali las placas de los autos constan de tres letras seguidas por tres números. Cuantas placas distintas pueden hacerse?

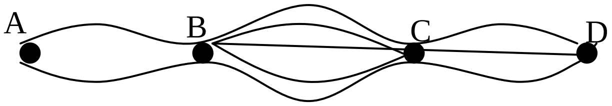
39. Un examen consta de 8 preguntas: las 5 primeras con dos respuestas posibles y las 3 últimas con cinco posibles respuestas cada una. De cuantas formas posibles puede responderse este examen?

40. Un estudiante tiene que elegir un idioma y cuatro asignaturas; entre 4 idiomas y 6 asignaturas. Hallar el numero de formas distintas en que puede hacerlo?

41. Un matrimonio quiere invitar a sus amigos a cenar. Debido a las dimensiones de la casa solo puede invitar a cinco cada vez. Si se quieren invitar a 10 amigos. De cuantas formas pude invitar a cinco de ellos

42. De cuantas formas pueden cubrir los puestos de presidente y secretario de una comunidad de vecinos, contando con 10 vecinos para ello?

43. Para ir de la ciudad A a la ciudad D hay que pasar por las ciudades B y C a través de las carreteras que se indican en la figura



El número de posibles recorridos distintos es

44. Sean los conjuntos $A=\{a,b,c,d\}$ $B=\{c,d,e,f,g\}$ y $C=\{b,d,e,g\}$ Determine:

   a. $A-B$
   b. $B-A$
   c. $C-B$
   d. $(A\cup C)-B$
   e. $A-(B\cap C)$
   f. $(A\cup B)-(A\cap C)$

45. Dados los conjuntos $A=\{1,2,3,4,5\}$, $B=\{1,2,4,6,8\}$ y $C=\{2,4,5,7\}$ Obtenga un conjunto *X* tal que $X \subset A$ y $A-X=B\cap C$

46. Clasifique en verdadero o falsa las siguiente sentencias (utilizando ejemplos numéricos):

a. $(A-B)\cup(B-A)=(A\cup B)-(A\cap B)$

b. $A\subset B \Rightarrow B^{C}\subset A^{C}$

c. $A-B\subset A^{C}$

d. $A-B\subset B^{C}$

47. Escriba por extensión los siguientes conjuntos descritos por comprensión:_

a. $A=\{x/x^{2}-5x-6=0\}$

b. $B=\{\text{x / x es la letra de la palabra excusa}\}$

c. $C=\{x/x^{2}-9=0 \text{ o } 2x-1=9\}$

48. Sea $E=\{a,\{a\}\}$. Diga cuales de las proposiciones de más abajo son verdaderas:

a. $a\in E$

b. $\{a\}\in E$

c. $a\subset E$

d. $\{a\}\subset E$

e. $\phi\in E$

f. $\phi\subset E$

49. Dado los conjuntos *A* y *B* tales que # *A* = 4, # *B* = 5 y # $A\cap B=3$, determine el número de subconjuntos de $A\cup B$

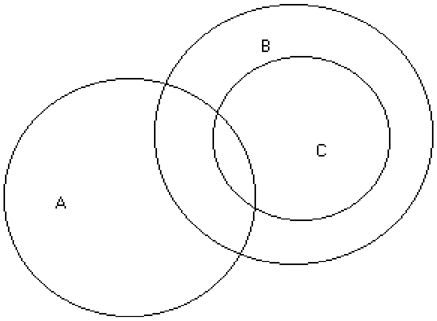
50. La tabla siguiente muestra la distribución de personas según **hábito de fumar**, **padecer bronquitis,** y **presión sistólica.**

	HABITO DE FUMAR			
	SI		**NO**	
Bronquitis	**Presión Sistólica**		**Presión Sistólica**	
	ALTA	**NORMAL**	**ALTA**	**NORMAL**
SI	**400**	**300**	**150**	**100**
NO	**200**	**50**	**40**	**30**

a. Determine el número de personas que fuman o tienen bronquitis
b. De las personas fumadoras; ¿cuántas tiene presión sistólica normal o tienen bronquitis?
c. De las personas con bronquitis; ¿cuántas tiene presión sistólica alta o son fumadoras?

51. En una escuela que tiene 415 alumnos, 221 estudian inglés, 163 estudian francés y 52 estudian ambas lenguas. ¿Cuántos alumnos estudian inglés o francés?, ¿Cuántos alumnos no estudian ninguna de las dos lenguas?.

52. Considere los conjuntos dibujados en el gráfico y además sabiendo que # $(A \cup B) = 24$ # $(A \cap B) = 4$, # $(B \cup C) = 16$,# $(A - C) = 11$, # $(B - C) = 10$



Se pide calcular:

a. $\#(A - B)$
b. $\#(A \cap B \cap C)$
c. $\#(B - (C \cup A))$
d. $\#((A \cap B) - C)$
e. $\#(B - (A \cap B))$

53. Una población consume tres tipo de jabón : A, B y C. Hecha una investigación de mercado , conociéndose los resultados la tabla siguiente,.

Marca	A	B	C	A y B	B y C	C y A	A, B y C	Ninguna de la tres
Nº de consumidores	109	203	162	25	41	28	5	115

Responda:

a. El número de personas consultadas
b. El número de personas que sólo consumen la marca A
c. El número de personas que no consumen las marcas A o C.
d. El número de personas que consumen al menos dos marcas.

54. De todos los empleados de una firma, 30% optaron por un plan de asistencia médica. La firma tiene la casa matriz en la capital y sólo dos filiales, una en Antofagasta y la otra en Calama. 45% de los empleados trabajan en la casa matriz y 20% de los empleados trabajan en la filial de Antofagasta. Sabiendo que el 20% de los empleados de la capital optaron por el plan de asistencia médica y que 35% de los empleados de la filial de Antofagasta lo hicieron ¿cuál es el porcentaje de los empleados de la filial de Calama que optaron por el plan?

55. En una cierta comunidad hay individuos de tres razas: blanca, negra, y amarilla. Sabiendo que 70 son blancos, 350 son negros y 50% son de raza amarilla, responda:

a. ¿Cuántos individuos tiene la comunidad?
b. ¿Cuántos individuos son de raza amarilla?

56. a) Si A y B son conjuntos cualesquiera. Desarrolle completamente cada una de las siguientes operaciones de conjuntos usando propiedades:

a. $(A-B)\cap B=$

b. $(A\cup B)\cup\left(A^{C}\cap B^{C}\right)=$

c. $\left[(A\cap B)-A\right]\cup\left[B-(A\cap B)\right]=$

d. $(A-B)\cup(B-A)=$

e. $\left(A\cap B^{C}\right)\cup(A\cap B)\cup\left(B\cap A^{C}\right)=$

57. Si A es el conjunto de los pacientes con "tifoidea" y B es el conjunto de pacientes con "ascaris" . Exprese las siguientes expresiones verbales como operaciones de los conjunto A y B.

i. El paciente tiene sólo una de las dos enfermedades.
ii. El paciente tiene al menos una de las dos enfermedades.
iii. El paciente no tiene las enfermedades descritas.
iv. El paciente tiene sólo tifoidea.

VIII. CALCULO DE PROBABILIDAD

Espacios de probabilidad ($(\boldsymbol{\Omega};\ \boldsymbol{\mathcal{T}}; \boldsymbol{P})$

2.1. Ejemplo

Un profesor asigna una semana antes del examen un conjunto de 10 problemas. El examen consistirá de 5 problemas elegidos al azar de entre los 10 asignados. Un estudiante solo pudo resolver 7 de esos problemas. Cuál es la probabilidad de que el estudiante conteste bien 3 de las 5 preguntas?

2.1.1. **Espacio muestral ($\boldsymbol{\Omega}$)** : Son todos los resultados posibles de un experimento aleatorio

$$\boldsymbol{Numero\ total\ examenes = \ \Omega = C(10,5) = \binom{10}{5} = 252}$$

2.1.2. **Familia de eventos ($\boldsymbol{\mathcal{T}}$):** Numero de casos favorables de un experimento aleatorio

$$A{\sim}Numero\ de\ examens\ que\ tienen\ e\ ejercicios\ de\ lo\ siete\ que\ pudo\ realizar\ el\ estudiante$$

$$\boldsymbol{A = C(7,3)C(3,2) = \binom{7}{3}\binom{3}{2} = 35 * 3 = 105}$$

2.2.3. **Probabilidad (P):** Probabilidad es un numero entre 0 y 1 que mide la posibilidad de que llegue a ocurrir un evento determinado

$$0 \leq P(\mathcal{T}) \leq 1$$

$$P(\mathcal{T}) = \frac{Numero\ de\ casos\ favorables}{Numero\ total\ de\ casos}$$

$$P(A) = \frac{Numero\ de\ casos\ favorables}{Numero\ total\ de\ casos} = \frac{A}{\Omega} = \frac{105}{252} = \frac{5}{12} = 0.42$$

EJERCICIOS DE COMPLEMENTO

1. Un profesor asigna una semana antes del examen un conjunto de 10 problemas. El examen consistirá de 5 problemas elegidos al azar de entre los 10 asignados. Un estudiante solo pudo resolver 7 de esos problemas. Cuál es la probabilidad de que el estudiante Tenga por lo menos 4 preguntas buenas?

2. Según la revista Semana la Coca Cola y la Pepsi ocuparon el primero y segundo lugar en la preferencia de las personas. Suponga que en un grupo de 10 personas; seis prefieren coca cola y cuatro prefieren Pepsi. Se selecciona una muestra aleatoria de tres miembros de ese grupo. Cual es la probabilidad de:

    i. Exactamente dos prefieran coca cola?
    ii. Al menos dos prefieran Pepsi Cola
    iii. Alo sumo dos prefieran Coca Cola

3. De una lata que contiene 18 galletitas de salvado y 10 de agua, se extraen 2 galletitas al azar, sucesivamente y sin repetición. Calcular la probabilidad de que la primera galletita extraída sea de salvado y la segunda de agua.

4. El juego de la LOTTO consiste en acertar 6 números entre el 1 y el 48. El primer premio se otorga a los que aciertan los 6 números, el segundo premio a los que aciertan 5 de los 6 y el tercer premio a los que aciertan 4 de los 6. Si una persona compra un boleto de la LOTTO. Cuál es la probabilidad de que se gane
    a. El primer premio?
    b. El segundo premio?
    c. El tercer premio?

IX. TEOREMAS DE PROBABILIDAD

3.1. Ejemplo

El 18% de las familias de un barrio tienen vehículo propio, el 20% tienen vivienda de su propiedad y el 12% tienen vivienda y vehículo propio.

Datos: $\boldsymbol{P(V) = 0.18}$ $\boldsymbol{P(C) = 0.20}$ $\boldsymbol{P(V \cap C) = 0.12}$

3.2. Probabilidad complemento

$$\boldsymbol{P\left(A^C\right) = P\left(A^`\right) = 1 - P(A)}$$

$$\boldsymbol{P\left(A^C/B\right) = 1 - P(A/B)}$$

Ejemplo. Si se escoge una familia al azar; cual es la probabilidad que no tenga vivienda propia

$$\boldsymbol{P\left(C^C\right) = P\left(C^`\right) = 1 - P(C) = 1 - 0.20 = 0.80}$$

3.3. Probabilidad unión

$$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Ejemplo. Si se escoge una familia al azar; Cual es la probabilidad de tener vivienda o vehículo propio

$$\boldsymbol{P(V \cup C) = P(V) + P(C) - P(V \cap C) = 0.18 + 0.20 - 0.12 = 0.26}$$

3.4. Probabilidad conjunto

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

3.4.1. Conjuntos mutuamente excluyentes

$$\boldsymbol{P(A \cap B) = 0}$$

Ejemplo. Son mutuamente excluyentes estos dos eventos

$\boldsymbol{P(V \cap C) = 0.12 \neq 0}$ **; No son mutuamente excluyentes**

3.4.2. Eventos independientes

$$\boldsymbol{P(A \cap B) = P(A)P(B)}$$

Ejemplo. Son independientes estos dos eventos

$$\boldsymbol{P(V\cap C) = P(V)P(C)}$$

$\boldsymbol{0.12 \neq (0.18)(0.20) = 0.036}$ **;** No son independientes

Ejemplo. Un avión de alto rendimiento contiene tres computadoras idénticas. Se utiliza únicamente una para operar el avión; las dos restantes son repuestos que pueden activarse en caso de que el sistema primario falle. Durante una hora de operación la probabilidad de que una falle en la computadora primaria (o de cualquiera de los sistemas de repuesto activados) es 0,0005. Suponiendo que cada hora representa un ensayo independiente, ¿Cuál es el tiempo promedio para que fallen las tres computadoras?

$$\boldsymbol{P(I\cap II\cap III) = P(I)P(II)P(III) = (0.0005)(0.0005)(0.0005) = (0.0005)^3}$$

3.5. Probabilidad condicional

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

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

3.5.1. **Ejemplo**. De las familias que tienen vivienda propia ; cual es la probabilidad de que tenga vehículo propio

$$\boldsymbol{P(V/C) = \frac{P(V\cap C)}{P(C)} = \frac{0.12}{0.20} = 0.60}$$

3.5.2. **Ejemplo**. De las familias que no tienen vivienda propia; cual es la probabilidad de que tenga vehículo propio

$$\boldsymbol{P\left(V/C^C\right) = \frac{P(V\cap C^c)}{P(C^c)} = \frac{P(C^c/V)P(V)}{1-P(C)} = \frac{(1-P(C/V))P(V)}{1-P(C)} = \frac{\left[1-\frac{0.12}{0.18}\right](0.18)}{1-0.20} = 0.075}$$

3.5.3. **Ejemplo.** De las familias que tienen vivienda propia; cual es la probabilidad de que no tenga vehículo propio

$$\boldsymbol{P(V^c/C) = 1 - P(V/C) = 1 - \frac{0.12}{0.20} = 0.40}$$

3.5.4. **Ejemplo.** De las familias que no tienen vivienda propia; cual es la probabilidad de que no tenga vehículo propio

$$\boldsymbol{P(V^c/C^c) = 1 - P(V/C^c) = 1 - 0.075 = 0.925}$$

3.6. Probabilidad marginal o total

$$\boldsymbol{P(D) = \sum_{i=1}^{m} P[D \cap A_i] = P[D \cap A_1] + P[D \cap A_2] + P[D \cap A_2] + \cdots \ldots + P[D \cap A_m]}$$

Ejemplo. Un tubo de vacío puede provenir de cualquiera de tres fabricantes con probabilidad: P1=0,20 P2=0,5 P3=0,30. Las probabilidades de que el tubo funcione correctamente durante un período de tiempo específico son: 0,1; 0,2; 0,4. Respectivamente para los 3 fabricantes. Calcular la probabilidad de que el tubo elegido al azar funcione correctamente.

$$\boldsymbol{P(I) = 0.20 \qquad P(II) = 0.50 \qquad P(III) = 0.30}$$

$$\boldsymbol{P[C/I] = 0.10 \qquad P[C/II] = 0.20 \qquad P[C/III] = 0.40}$$

$$\boldsymbol{P(C) = \sum_{i=1}^{m} P[C \cap A_i] = P[C \cap I] + P[C \cap II] + P[C \cap III]}$$

$$\boldsymbol{P(C) = P[C/I]P(I) + P[C/II]P(II) + P[C/III]P(III) = 0.10(0.20) + 0.20(0.50) + 0.40(0.30) = 0.24}$$

3.7. Teorema de bayes

$$\boldsymbol{P[A_i/D] = \frac{P(A_i \cap D)}{P(D)}}$$

Ejemplo. Un tubo de vacío puede provenir de cualquiera de tres fabricantes con probabilidad: P1=0,25 P2=0,5 P3=0,25. Las probabilidades de que el tubo funcione correctamente durante un período de tiempo específico son: 0,1; 0,2; 0,4. Respectivamente para los 3 fabricantes. Si un tubo funciona correctamente; cual es la probabilidad de que haya sido fabricado por P1?

$$\boldsymbol{P[I/C] = \frac{P(I \cap C)}{P(C)} = \frac{0.10(0.20)}{0.24} = 0.083}$$

EJERCICIOS DE COMPLEMENTO

1. Un tubo de vacío puede provenir de cualquiera de tres fabricantes con probabilidad: P1=0,25 P2=0,5 P3=0,25. Las probabilidades de que el tubo funcione correctamente durante un período de tiempo específico son: 0,1; 0,2; 0,4. Respectivamente para los 3 fabricantes. Calcular la probabilidad de que el tubo elegido al azar funcione correctamente.

2. Sean A y B dos sucesos asociados con un experimento que P(a) = 0,4 mientras que P(A u B) =0,7: Sea por comodidad P (A u B) = P

    a. ¿para que elección de P(b) son Ay B mutuamente excluyentes?
    b. ¿para que elección de P(b) son Ay B mutuamente independientes?

3. Tres caballos A,B,C, intervienen en una carrera. A tiene el doble de probabilidad de ganar que B, y B tiene el doble que C. ¿Cuales son las respectivas probabilidades de ganar de cada caballo?

4. Dos equipos A y B juegan un torneo. La probabilidad de que A gane un juego es de un 57% y la probabilidad de que B gane un juego es de un 43%. Si los equipos juegan tres partidos. Calcule la probabilidad de que:

    a. El equipo A gane todos los partidos
    b. El equipo A gane solamente un partido
    c. El equipo B gane solamente un juego
    d. El equipo B gane solamente dos partidos
    e. El equipo B gane los dos primeros juegos

5. Dados los sucesos A y B. Si P(A) = 0.50, P(B) = 0.40 y $P(A \cap B) = 0.20$. Calcula las probabilidades de los siguientes eventos:

    a. $P(A \cap B)$
    b. $P(A^c \cap B)$
    c. $P(A \cap B)^c$
    d. $P(A^c/B)$
    e. $P(A/A \cap B)$

6. Sea un dado cargado, tal que la posibilidad de salir un número cuando se lanza el dado es proporcional a dicho número. Por ejemplo el 6 tiene el doble de probabilidad que 3.

7. En un grupo de 120 estudiantes; 60 de ellos cursan por lo menos estadística, 50 de ellos por lo menos matemáticas y 20 cursan ambas asignaturas. Si se selecciona un estudiante al azar. Cuál es la probabilidad de que:

    a. No curse ninguna de las dos asignaturas
    b. Curse matemática, pero no estadística
    c. Curse estadística; sabiendo que no cursa matemáticas

8. Hallar la probabilidad de que el fuerte sea destruido, si sobre el se lanzan 4 bombas con probabilidades de impactos iguales a: 0,3- 0,4- 0,6- 0,7- respectivamente. ¿cuál es la probabilidad que el fuerte sea destruido con cada una de las bombas?

9. En un cierto país, el 99% de los detenidos y sometidos a juicio son culpables del delito que se les imputa. Los jueces, al emitir veredicto, aciertan en el 95% de los casos, tanto si el acusado es culpable como inocente. Según estos datos, calcúlese la probabilidad de que:

    a. un ciudadano inocente haya sido declarado culpable.
    b. sea culpable, si ha sido declarado inocente.

10. El 30% de las ventas de una tienda departamental son en efectivo; el 30% son pagadas por cheque en el momento de la compra y el 40% son a crédito. De las compras mayores de 500 mil pesos: el 20% son pagadas en efectivo, el 90% son pagadas con cheque y el 60% son pagadas a crédito.

    a. Qué porcentaje de las ventas de las tiendas departamentales son mayores de 500 mil **pesos**
    b. Qué porcentaje de las ventas de las tiendas departamentales son menores de 500 mil **pesos**
    c. En este momento se está realizando una compra de un millón de pesos; Cual es la probabilidad sea pagada en efectivo
    d. En este momento se está realizando una compra de un millón de pesos; Cual es la probabilidad sea pagada con cheque
    e. En este momento se está realizando una compra de un millón de pesos; Cual es la probabilidad sea pagada a crédito

11. De acuerdo a una investigación realizada en una determinada ciudad acerca d e mujeres mayores de 20 años se ha comprobado que entre otras cosas el 68% están casadas, de estas el 40 % trabaja fuera del hogar. De las que no están casadas, el 72 % trabajan fuera del hogar:

    a. Que porcentaje de mujeres mayores de 20 años trabaja fuera del hogar.
    b. Si se selecciona al azar una mujer mayor de 20 años, ¿cuál es la probabilidad de que no este casada ni trabaje fuera?

12. Un obrero atiende tres telares. Supongamos que la posibilidad que los telares no requieran de la atención del obrero en una hora sea para el primer telar de 0,9, para el segundo de 0,8 y para el tercero 0,85. Se desea saber cual es la probabilidad de que ninguno de los telares reclame la atención del obrero durante 1 hora.

13. En la fabricación de un cierto articulo se encuentra que se presenta un tipo de defecto con una probabilidad 0,1 y defecto de un segundo tipo con probabilidad de 0,05. Se supone la independencia entre ambos tipos de defecto. ¿cuál es la probabilidad de que?

    a. Un artículo no tenga ambas clases d e defecto.
    b. Un artículo sea defectuoso.

14. Un avión de alto rendimiento contienen tres computadoras idénticas. Se utiliza únicamente una para operar el avión; las dos restantes son repuestos que pueden activarse en caso de que el sistema primario falle. Durante una hora de operación la probabilidad de que una falle en la computadora primaria(o de cualquiera de los sistemas de repuesto activados) es 0,0005. Suponiendo que cada hora representa un ensayo independiente,

    a. ¿Cuál es el tiempo promedio para que fallen las tres computadoras?
    b. ¿Cuál es la probabilidad de que las tres computadoras fallen en un vuelo de 5 horas?

15. En una casa viven tres estudiantes: Juan Pedro, y Pablo, se escucha el sonido de un vaso roto. La vecina que es un poco fisgona se pregunta quien habrá sido el autor de dicho destrozo, asi que coincidiendo con uno de los estudiantes en el ascensor le pregunta quien ha sido el que rompió el vaso fregando. El estudiante que por aquel entonces estaba preparando el examen de estadística le dice, que ha sido aquel cuya probabilidad de haber roto el vaso es mas alta y

además le da la siguiente información: de los siete días de la semana Juan y pedro friegan dos días cada uno y el Pablo friega tres días. Y además le dice que Juan rompe un de cada 20 vasos que friega; Pedro uno de cada 50 vasos y Pablo cuatro de cada 100 vasos que friega. Con estos datos Pablo le dice que puede saber quien rompió el vaso.

   a. A que conclusión debería llegar la vecina?
   b. Se puede decir que el culpable es el mas torpe?

16. En la universidad hay estudiantes correspondientes al plan nuevo y al plan antiguo. Terminan la carrera el 75% del plan antiguo y el 55% del plan nuevo. Se sabe que del total de alumnos el 70% pertenecen al plan nuevo. Si se elige un estudiante al azar y se pide:

    a. La probabilidad de que sea del plan nuevo y haya terminado la carrera
    b. Nos dicen que ha terminado la carrera. Cual es la probabilidad que sea del plan antiguo?
    c. La probabilidad de que sea del plan antiguo y no haya terminado la carrera?
    d. La probabilidad de que no habiendo terminado la carrera: sea del plan antiguo?

17. Suponga que tiene las siguientes probabilidades: $P(A) = 0.4$ y $P(A\cup B) = 0.7$. Encuentre la $P(B)$ cuando:

    a. A y B con mutuamente excluyentes
    b. A y B son independientes

18. El 34% de los árboles de un bosque tienen más de 15 años. El 54% son de la variedad A. De los de la variedad A, el 7% tiene más de 15 años. Si se elige un árbol al azar:

    a. ¿Cuál es la probabilidad de que tenga más de 15 años y sea de la variedad A?
    b. ¿Cuál es la probabilidad de que teniendo menos de 15 años, sea de la variedad A?

19. Una fábrica tiene tres fábricas para producir bombillas. La maquina A produce el 35% del total de las bombillas, la maquina B produce el 50% y la maquina C produce el 15% de las bombillas. Sin embargo las maquinas no son perfectas; la maquina A daña el 10% de las bombillas que produce; la maquina B daña el 5% y la maquina C daña el 20%.

    a. La fábrica produce 10.000 bombillas sin defectos en un día. Cuantas de estas corresponden a la maquina A?. Cuantas daña al día?
    b. Si seleccionamos una bombilla de la maquina C. Cual es la probabilidad de que este defectuosa?
    c. Luego de fabricadas pero antes de ser probadas, las bombillas se colocan juntas en un salón. Si se selecciona una bombilla al azar. Cual es la probabilidad de que este defectuosa?
    d. Si se comprueba que una bombilla no esta defectuosa. Cual es la probabilidad que provenga de la maquina C?

X. VARIABLES ALEATORIAS

4. Variables aleatorias

4.1. Variables aleatorias discretas

4.1.1. Ejemplo

Suponga que alguien le propone un juego que consiste en lanzar al aire una moneda 3 veces, si consigue 3 caras ganará $200, si consigue 2 caras ganará $100, si consigue 1 cara perderá $120 y si son todos sellos perderá $200.

4.1.2. Función de probabilidad y Función de distribución

$$X \sim Variable\ aleatoria\ que\ cuenta\ el\ numero\ de\ caras\ en\ el\ lanzamiento$$

Eventos	X_i	$P(X = x_i)$	$F(x_i) = P(X \leq x_i)$
(c,c,c)	200	0.125	0.125
(ccs)(scc)(csc)	100	0.375	0.500
(ssc)(scs)(css)	-120	0.375	0.875
(sss)	-200	0.125	1.000

4.1.3. Calculo de probabilidad

Cual es la probabilidad de que el jugador se gane al menos 100 pesos?

$$P(X \geq 100) = 1 - P(X < 100) = 1 - P(-120) = 1 - 0.875 = 0.125$$

4.1.4. Valor esperado

Cual seria la ganancia esperada en el juego?

$$E(X) = \sum_{i=1}^{4} X_i\, P(X = X_i)$$

$E(X) = \sum_{i=1}^{4} X_i\, P(X = X_i)$= 200(0.125) + 100(0.375) – 120(0.375) – 200(0.125) = - 7.5

4.1.5. Varianza

$$V(X) = E(X^2) - [E(X)]^2$$

$E(X^2) = \sum_{i=1}^{4} X_i^2\, P(X = X_i)$= $200^2$0.125) + 100^2(0.375) + $(-120)^2$(0.375) + $(-200)^2$(0.125) = 19150

$$V(X) = 19150 - [-7.5)]^2 = 19093.75$$

$$CV = \frac{\sqrt{V(X)}}{E(X)} = \frac{\sqrt{19093.75}}{7.5} = 581.7\%$$

4.2. Variables aleatorias continuas

4.2.1. Ejemplo

El tiempo que tarda un usuario en atenderse en el Servicio de Tesorería, es una variable aleatoria, descrita mediante la siguiente función (expresada en minutos):

$$f(x)\begin{cases} 3x^2 & 0 < x < 1 \\ 0 & Para\ cualquier\ otro\ valor \end{cases}$$

4.2.2. Función de densidad

$$f(X) = 3X^2 \qquad 0 \leq X \leq 1$$

4.2.3. Función de distribución

$$F(k) = P(X \leq k) = \int_0^k f(x)dx = \int_0^k 3X^2\,dx = X^3 \qquad 0 \leq X \leq 1$$

$$F(k) = K^3 \qquad 0 \leq K \leq 1$$

4.2.4. Calculo de probabilidad

4.2.4.1. ¿Cuál es la probabilidad de que una persona se demore menos de 0.5 minutos en atenderse?

$$P(X \leq 0.5) = F(0.5) = (0.5)^3 = 0.125$$

4.2.4.2. ¿Cuál es la probabilidad de que una persona se demore entre 0.75 y 1 minutos?

$$P(0.75 \leq X \leq 1.0) = P(X \leq 1.0) - P(X \leq 0.75) = F(1.0) - F(0.75) = (1.0)^3 - (0.75)^2 = 0.578$$

4.2.4.3. ¿Cuál es la probabilidad de que una persona se demore mas de 0.8 minutos?

$$P(X > 0.8) = 1 - P(X \leq 0.8) = 1- F(0.8) = 1 - (0.8)^3 = 0.488$$

4.2.4.4. ¿Cuál es la probabilidad de que una persona se demore más de un minuto?

$$P(X > 1.0) = 1 - P(X \leq 1.0) = 1- F(1.0) = 1 - (1.0)^3 = 0$$

4.2.5. Valor esperado

$$E(X) = \int_0^1 Xf(x)dx = \int_0^1 X(3X^2)dx = \frac{3}{4} = 0.75$$

4.2.6. Varianza

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^1 X^2 f(x)dx = \int_0^1 X^2(3X^2)dx = \frac{3}{5} = 0.60$$

$$V(X) = 0.60 - [0.75]^2 = 0.0375$$

$$CV = \frac{\sqrt{V(X)}}{E(X)} = \frac{\sqrt{0.0375}}{0.75} = 25.8\%$$

EJERCICIOS DE COMPLEMENTO

1. Supóngase que la producción de un día de 850 piezas manufacturadas contiene 50 piezas que no cumplen con los requerimientos del cliente. Se seleccionan del lote dos piezas al azar y sin reemplazo. Sea la variable aleatoria X igual al número de piezas de la muestra que no cumplen. ¿Cuál es la función de distribución acumulada de X?

2. Usted dispone de $1.000.000 para invertir. Tiene dos proyectos alternativos; Proyecto A y Proyecto B

Rendimiento proyecto A	
$	Probabilidad
80	0.40
90	0.30
70	0.30

Rendimiento proyecto B	
$	Probabilidad
90	0.40
60	0.30
100	0.30

Que proyecto escogería usted y explique su respuesta?

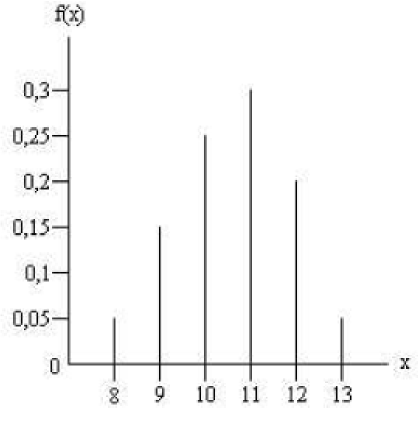
3. Una muestra aleatoria con reposición de tamaño n=2 se selecciona del conjunto {1,2,3} produciendo el espacio equiprobable de 9 elementos.

S={(1,1),(1,2),(1.3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)}

Sea X la suma de los dos números.

   a. Encuentre la distribución f de X.
   b. Encuentre el valor esperado E(X).

4. A partir de la gráfica siguiente, de una función de probabilidad:



   a. Construya la tabla de la función de probabilidad.
   b. Encuentre el valor esperado de la variable aleatoria. Sol: E(**X**) = 10,6

5. Suponga que alguien le propone un juego que consiste en lanzar al aire una moneda 3 veces, si consigue 3 caras ganará 20$, si consigue 2 caras ganará 10$, si consigue 1 cara perderá 12$ y si son todos sellos perderá 20$.

   a. Determine la función de probabilidad siendo la variable aleatoria el número de caras.

   b. Con base a las probabilidades obtenidas anteriormente y siendo ahora la variable aleatoria la cantidad de dinero que gana o pierde en cualquier lanzamiento. ¿Aceptaría usted jugar? Sol: no

6. Supóngase un juego con un dado, en este juego, el jugador gana 20$ si obtiene un 2, 40$ si obtiene un 4, pierde 30$ si obtiene un 6, en tanto que ni pierde ni gana si obtiene otro resultado. Hallar la suma esperada de dinero ganado. Sol: una ganancia de 5$.

7. Las cantidades (en decenas de miles de euros) ingresadas por declaraciones de la renta en una Delegación de Hacienda presentan la siguiente distribución de probabilidad: **f(x)=a+bx2** si **0<x≤1**.

   a. Calcular el valor de los parámetros "**a**" y "**b**" sabiendo que el ingreso esperado en dicha Delegación es de **7.500** euros.
   b. Obtener la distribución acumulativa de probabilidades.
   c. ¿Cuál es el ingreso más probable?
   d. El **20%** de los contribuyentes que más abonan ¿cuánto abonan como mínimo?

8. Suponga que en una determinada comarca, el nº de hijos por familia puede representarse mediante una v.a. con la siguiente función de probabilidad:

$$p(x) = \begin{cases} 0.5 \cdot k & x = 0,1 \\ 1.5 \cdot k & x = 2,3 \\ k & x = 4,5 \end{cases}$$

   a. Hallar el valor de *k* para que *p*(*x*) sea una función de probabilidad.
   b. Hallar la media y la varianza
   c. Si se escoge al azar una familia de esta comarca ¿cuál es la probabilidad de que tenga por lo menos un hijo, si se sabe que tiene menos de dos?
   d. Si mensualmente cada familia incurre en un gasto de 150 € por cada hijo y tiene un gasto fijo de 900 € mensuales, hallar el coste mensual esperado y su varianza.

9. Se realiza una investigación de mercado con el propósito de conocer la factibilidad de abrir una nueva tienda de autoservicio en cierta zona de la ciudad. Coma parte del estudio se considero la variable Z que representa el número de veces al mes que cada una de las amas de casa asiste al supermercado. La distribución de Z esta dada por:

Z	3	4	5	6	7
Numero de amas casa	30	50	15	5	10

   a. Calcule e interprete el valor esperado y la varianza
   b. El tiempo en minutos en el que las amas de casa realizan sus compras en el supermercado esta dado por T = 30Z +20. Calcule el valor esperado y la varianza de T .

10. Se sabe que el precio de un boleto de una rifa es igual al valor medio de la ganancia (Ganancia esperada), entonces la rifa es equitativa. Entonces si Juan participa en un rifa donde la probabilidad de ganar $5000 pesos es de 0.001 y la probabilidad de ganar $2000 pesos es de 0.004. Cuanto debe pagar por boleto para que la rifa sea equitativa?

11. El tiempo que vive un virus es una variable aleatoria que tiene la siguiente función de densidad:

   $f(x) = K e^{-x}$ si $0 \le X \le 1$

a. Hallar K para la función de densidad
b. Hallar la función de distribución
c. Hallar la probabilidad de que un virus elegido al azar viva mas de 5 horas

12. Un empresario va a participar en un negocio en el que la probabilidad de ganar $100.000 pesos es de 0.6 y la probabilidad de perder $20.000 pesos es de 0.4. Cual es la utilidad esperada?. Si usted fuera el empresario invertiría en el negocio? Justifique su respuesta.

13. Un fabricante de leche pasteurizada la vende a $1000 pesos el litro y le cuesta a $500 pesos el litro. Por cada litro que este en malas condiciones el le devuelve al comprador $200. Las frecuencias relativas del porcentaje de leche en malas condiciones son las siguientes:

% litros de leche en malas condiciones	Frecuencia relativa
5%	0.60
10%	0.20
25%	0.20

Cual es la ganancia esperada por litro?

14. La demanda mensual de un hogar de cierto producto (Miles de pesos) es una variable aleatoria y tiene una funcion de densidad:

$$f(x) = Ke-^{2x} \qquad X > 0$$

a. calcule el valor de K?
b. Halle la funcion de distribución
c. Calcule la probabilidad de que la demanda este comprendida entre 1.5 y 2.5 mil pesos
d. Que Stock debe disponer la empresa al iniciar el mes para garantizar que se atiende a la demanda mensual con una probabilidad del 90%?
e. Calcule e interprete el valor esperado y varianza

15. En el pronóstico de las ganancias de una compañía, 2 analistas llegaron a los siguientes escenarios con sus respectivas probabilidades.

	Analista 1	
	Ganancias	Prob
Escenario Optimista	180	0.15
Escenario Neutral	120	0.6
Escenario Pesimista	80	0.25

Analista 2	
Ganancias	Prob
160	0.45
120	0.15
90	0.4

Estime el valor esperado según cada analista. ¿Qué analista estima una mayor volatilidad en las ganancias esperadas?

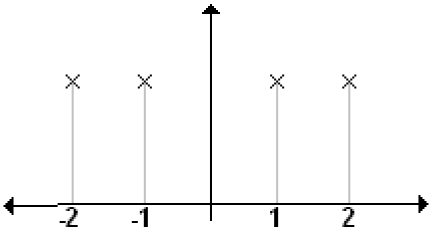
16. Sea $f(x,5)=\frac{1}{5}$ para $x=-2,-1,0,1,2$, calcular E(x), V(x), $P(x\leq 0)$ y $P(x<-2)$.

$f(-2)=f(-1)=f(0)=f(1)=f(2)=\frac{1}{5}$

$E(x)=0, \quad E(x^2)=V(x)=2$

$P(x\leq 0)=f(-2)+f(-1)+f(0)=\frac{3}{5}$

$P(x<-2)=0$



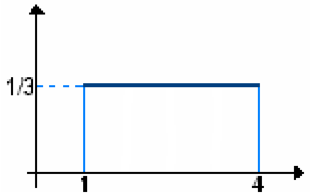
17. Sea $f(x,3)=\frac{1}{3}$ para $1\leq x\leq 4$. Calcular E(x), V(x) y $P(x\leq 4)$

E(x)=2.5

$$V(x)=E(x^2)-E^2(x)$$

$$=7-\frac{25}{4}$$

$$=\frac{3}{4} \quad y \quad P(x\leq 2)=\int_1^2 \frac{1}{3}\,dx=\frac{1}{3}$$



$P(x\leq 4)=f(1)+f(2)+f(3)+f(4)=F(4)=1$

18. Calcular E(x), E(x^2) y V(x) si x es una v.a. continua con $f(x,k)=\frac{1}{k}$, para $a\leq x\leq a+k$

$$E(x)=\int_a^{a+k} x f(x)dx=\frac{1}{k}\int_a^{a+k} x dx=\frac{1}{2k}x^2\Big|_a^{a+k}$$

$$\boxed{E(x)=a+\frac{k}{2}}$$

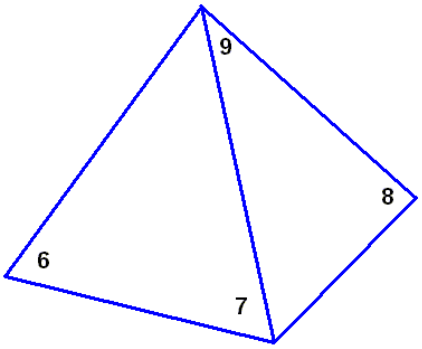
$$E(x^2)=\int_a^{a+k} x^2 f(x)dx=\frac{1}{k}\int_a^{a+k} x^2 dx=\frac{1}{3k}x^3\Big|_a^{a+k}$$

$$E(x^2)=a^2+ak+\frac{k^2}{3}$$

$$V(x)=E(x^2)-E^2(x)=\left(a^2+ak+\frac{k^2}{3}\right)-\left(a^2+ak+\frac{k^2}{4}\right)$$

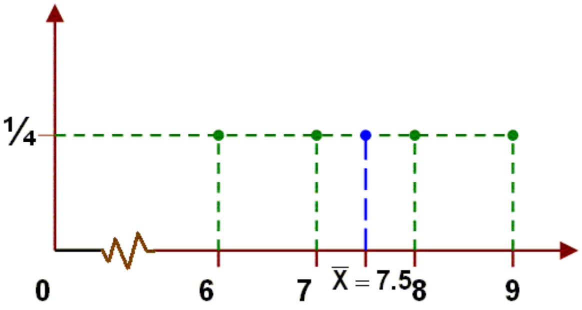
$$\boxed{V(x)=\frac{k}{12}}$$

19. Consideremos un dado regular en forma de tetraedro que tiene sus vértices marcados con los números 6, 7, 8 y 9. Sea X la v. a. que indica el número en el vértice superior cuando se lanza el tetraedro. Halle S, R_x; la fdp; $E(X)$, $V(X)$, $P(x\leq E(x))$, $P(7\leq x\leq 8)$ y $P(x=E(x))$.



$$S = Rx = \{6,7,8,9\}$$

$$f(x,4) = \frac{1}{4}$$



$$E(X) = \frac{1}{k}\sum_{1}^{4} x_i = \frac{1}{4}(6+7+8+9) = 7.5 = \overline{X}$$

$$V(X) = \frac{1}{k}\sum_{1}^{4} (X_i - \overline{X})^2 = \frac{1}{4}\left(2\left[0.5^2 + 1.5^2\right]\right) = 1.25.$$

$P(x \leq E(x)) = P(6 \leq x \leq 7.5) = P(x=6) + P(x=7) = \frac{2}{4}$

$P(7 \leq x \leq 8) = P(x=7) + P(x=8) = \frac{2}{4}$

$P(x= E(x)) = P(x=7.5) = 0$

XI. VARIABLES ALEATORIAS CONTINUAS

Ejemplo los gastos en alimentos de las familias colombianas se pueden representar mediante la siguiente función de densidad:

$$f(x) = \frac{1}{400} \qquad 250 < x < 650$$

a. Función de distribución
b.

$$F(x) = \int_{250}^{x} \frac{1}{450} dx = \frac{x-250}{400} \qquad 250 < x < 650$$

c. Calculo de probabilidades

$$F(a) = P(X \leq a) = P(X < a)$$

- Que porcentaje de las familias colombianas gastan en alimentos menos de 300 mil pesos

$$P(X < 300) = P(X \leq 300) = F(300) = \frac{300-250}{400} = 0.125 = 12.5\%$$

- Que porcentaje de las familias colombianas gastan en alimentos mas de 400 mil pesos

$$P(X > 400) = 1 - P(X \leq 400) = 1 - F(400) = 1 - \left(\frac{400-250}{400}\right) = 1 - 0.375 = 0.625 = 62.5\%$$

- Que porcentaje de las familias colombianas gastan en alimentos entre 300 y 400 mil pesos

$$P(300 < X < 400) = P(X \leq 400) - P(X < 300) = F(400) - F(300) = 0.375 - 0.125 = 0.25 = 25\%$$

d. Valor esperado

$$E(X) = \int_{250}^{650} \frac{1}{400} dx = \frac{650^2}{800} - \frac{250^2}{800} = 450$$

e. Varianza

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_{250}^{650} X^2 \frac{1}{400} dx = \frac{650^3}{1200} - \frac{250^3}{1200} = 215.833.3$$

$$V(X) = E(X^2) - [E(X)]^2 = 215.833.3 - 450^2 = 13.333.3$$

$$CV = \frac{\sqrt{V(X)}}{E(X)} x100 = \frac{\sqrt{13.333.3}}{450} x100 = 25.66\%$$

XII. DISTRIBUCIONES DE PROBABILIDAD DISCRETAS

1.1. Distribución Binomial

1.1.1. Ejemplo

El jefe de personal de una empresa está buscando empleados para un nuevo puesto de trabajo. Se sabe que sólo el 15% de las solicitudes recibidas reúnen las condiciones necesarias para dicho puesto. Si se revisan 10 solicitudes Cual es la probabilidad de:

1.1.2. Notación

X es la variable aleatoria que cuenta el número de solicitudes que cumplen con las condiciones necesarias para el puesto

$$X \sim B(n; p) = B(10; 0.15)$$

1.1.3. Distribución de probabilidad

$$P(X = k) = \binom{n}{k}(P^k)(1 - P)^{n-k} \quad \text{k = 0, 1, 2, …., n}$$

$$P(X = k) = \binom{10}{k}(0.15)^k(0.85)^{10-k} \quad \text{k = 0, 1, 2, …., 10}$$

1.1.4. Calculo de probabilidad

i	(X=k)	F(k) = P(X≤K)
	1968744	0,19687440
	34742540	0,54429980
	27589670	0,82019650
	12983370	0,95003020
	04009570	0,99012590
	00849090	0,99861680
	00124860	0,99986540
	00012590	0,99999130
	00000840	0,99999970
	00000030	1,00000000
0	00000000	1,00000000

1.1.5. Valor esperado

$$E(X) = \sum_{i=1}^{m} X_i \, P(X \le X_i) = nP = 1.5$$

1.1.6. Varianza

$$V(X) = E(X^2) - [E(X)]^2 = nP(1-P) = 10(0.15)(0.85) = 1.275$$

1.1.7. Un gran lote de tubos catódicos de televisión es verificado inspeccionando uno a uno los tubos hasta encontrar el primero defectuoso. El número de tubos inspeccionados determina que el lote se acepte o no. Sea k el número crítico, es decir, el lote es aceptado si k o más tubos son inspeccionados sin que aparezca ninguno defectuoso.

a. Calcular la probabilidad de que un lote que contiene una proporción p de defectuosos sea aceptado.
b. Si k=5 y p=0.1 calcular la probabilidad de aceptar el lote.
c. Determinar el valor de k si se quiere rechazar por los menos con probabilidad del 90% un lote que contenga el 25% de defectuosos.

1.1.8. De la producción diaria de un determinado tipo de artículos se extrae una muestra de 20 artículos y si en ella hay más de dos defectuosos se inspecciona toda la producción. Suponiendo independencia en la fabricación de los distintos artículos, y que la frecuencia observada en un gran lote de artículos es del 5% de defectuosos, calcular la probabilidad de los siguientes sucesos:

a. Inspeccionar toda la producción un determinado día.
b. Que en 15 días consecutivos no sea necesario realizar una inspección completa.
Que en 5 días sea inspeccionada la producción completa en dos ocasiones

1.2. Distribución Poisson

1.2.1. Ejemplo

Un empleado del departamento de atención al público de una gran empresa recibe, en promedio 20 llamadas por hora. La probabilidad de que una de estas llamadas sea una reclamación de un cliente es de 0,1. Se pide:

a. Probabilidad de tener 15 llamadas en media hora.
b. Si se han recibido 50 llamadas, ¿cuál es la probabilidad de que haya habido al menos 8 reclamaciones?.
c. El empleado decide que no se irá almorzar hasta que no haya atendido la tercera reclamación ¿cuál es el número medio de llamadas que debe atender?.

1.2.2. Notación

X es la variable aleatoria que cuenta el número de llamadas realizadas en una hora

$$X \sim P(\lambda) = P(20)$$

1.2.3. Distribución de probabilidad

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, 2, …., n$$

$$P(X = k) = e^{-20}\frac{20^k}{k!} \quad k = 0, 1, 2, …., n$$

1.2.4. Calculo de probabilidad

X	P(X=k)	P(X≤K)
0	0,00000000	0,00000000
1	0,00000000	0,00000000
2	0,00000050	0,00000050
3	0,00000270	0,00000320
4	0,00001370	0,00001690
5	0,00005500	0,00007190
6	0,00018320	0,00025510
7	0,00052350	0,00077860
8	0,00130870	0,00208730
9	0,00290810	0,00499540
10	0,00581630	0,01081170
11	0,01057510	0,02138680
12	0,01762520	0,03901200
13	0,02711560	0,06612760
14	0,03873670	0,10486430
15	0,05164880	0,15651310
16	0,06456110	0,22107420
17	0,07595420	0,29702840
18	0,08439350	0,38142190
19	0,08883540	0,47025730
20	0,08883530	0,55909260

1.2.5. Valor esperado

$$E(X) = \lambda = 20$$

1.2.6. Varianza

$$V(X) = \lambda = 20$$

1.2.7. Ejercicio 1. En un cruce de carreteras se producen accidentes a razón de 2 por semana (en media), siguiendo una distribución de Poisson. Reconociendo que la frecuencia anterior es intolerable para la Dirección General de Tráfico se ha decidido instalar un semáforo en dicho cruce. La siguiente semana de la instalación sólo ocurre un accidente.

a. ¿ Se puede afirmar que el semáforo es efectivo ?.
b. ¿ Cuál sería la conclusión si se hubiera producido un accidente en dos semanas ?.
   ¿ Y si se hubiera producido un accidente en 4 semanas

1.2.8. Ejercicio 2. El número de hombres que llegan a un comercio sigue una distribución de Poisson, a razón media de uno por minuto. El número de mujeres que llegan al mismo comercio sigue también una distribución de Poisson, a razón media de dos por minuto. Suponiendo la independencia de los dos sucesos, calcular:

a. La probabilidad de que lleguen menos de tres clientes en un minuto.
b. La probabilidad de que hayan llegado cinco hombres en media hora, si en esa media hora han llegado diez clientes en total.

1.3. Aproximación Binomial ~ Poisson

Si n>60 y P < 0.05; si np > 5 y np(1-p) entonces se considera que estas dos expresiones son aproximadamente iguales:

$$P(X=k)=\binom{n}{k}(P^k)(1-P)^{n-k}=e^{-np}\frac{np^k}{k!}$$

Ejemplo.

De una población de cierto país se conoce que la probabilidad de adquirir una determinada enfermedad es de 5 por cada 10000. Si se toma una ciudad de este país que tiene 350.000 habitantes cual es la probabilidad de encontrar menos de 3 personas contagiadas por esta enfermedad?

Calculo de probabilidad

X	Binomial P(X=k)	Poisson P(X=k)
0	0,00000000	0,00000000
1	0,00000050	0,00000050
2	0,00000380	0,00000380
3	0,00002240	0,00002240
4	0,00009810	0,00009820
5	0,00034340	0,00034340
6	0,00100150	0,00100170
7	0,00250390	0,00250430
8	0,00547750	0,00547810
9	0,01065090	0,01065190
10	0,01863960	0,01864080

Ejercicio 1. La probabilidad de que una persona de 50 a 60 años muera de cierta enfermedad durante un período de un año es 0.00001. Si una compañía de seguros tiene 100.000 pólizas en este grupo. ¿Cuál es la probabilidad aproximada de que deba pagar más de 4 reclamaciones debido a muertes por esta enfermedad en un año?.

1.4. Distribución hypergeometrica

1.4.1. Ejemplo

Para comprobar la experiencia de sus auditores, una compañía les manda examinar 25 transacciones, de las cuales 8 tienen algún tipo de error. Si uno de los auditores selecciona 5 al azar ¿cuál es la probabilidad de seleccionar correctamente las 5 transacciones erróneas? ¿Cuál es la probabilidad de que sólo seleccione 2 de las erróneas?.

1.4.2. Notación

X es la variable aleatoria que cuenta el número de transacciones que tiene algún tipo de error

$$X \sim HYPER(N, r, n) = HYPER(25; 8: 5)$$

1.4.3. Distribución de probabilidad

$$P(X = k) = \frac{\binom{N-r}{n-K}\binom{r}{k}}{\binom{N}{n}} = \frac{\binom{25-8}{5-k}\binom{8}{k}}{\binom{25}{5}} \quad \text{K= 0, 1 2…..8}$$

1.4.4. Calculo de probabilidad

	(X=k)	k) = P(X<K)
	11646900	11646900
	35836630	47483530
	35836630	83320160
	14334650	97654810
	02239790	99894600
	00105400	00000000
	00000000	00000000
	00000000	00000000
	00000000	00000000

1.4.5. Valor esperado

$$E(X) = n\frac{r}{N} = 5\frac{8}{25} = 1.6$$

1.4.6. Varianza

$$V(X) = \frac{n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)(N-n)}{N-1} = 0.906$$

XIII. DISTRIBUCIONES DE PROBABILIDAD CONTINUAS

1.5. Distribución uniforme

1.5.1. Ejemplo

Se sabe que las ventas diarias de una empresa se distribuyen uniformemente en el intervalo (100, 200). Se pide:

a. Probabilidad de que las ventas realizadas estén en el intervalo (120,250)
b. Cuál sería la cifra media de ventas? ¿Y su desviación típica?
c. Si se supone que el precio (Y) se puede tomar como función de las ventas (X) por medio de la relación Y□= 220 - X. Calcular el precio medio.

1.5.2. Notación

X es la variable aleatoria que mide las ventas diarias en miles de pesos por la empresa

$$X\sim U(Minimo; Maximo)\sim U(120; 250)$$

1.5.3. Función de densidad

$$f(X) = \frac{1}{b-a} = \frac{1}{130} \qquad 120 < X < 250$$

1.5.4. Función de distribución

$$F(X) = \frac{x-a}{b-a} = \frac{x-120}{250-120} \qquad 120 < X < 250$$

1.5.5. Calculo de probabilidad

- Que porcentaje de las ventas de la empresa fueron superiores de 150 mil pesos diarios

$$P(X > 150) = 1 - P(X < 150) = 1 - F(150) = 1 - \left(\frac{150-120}{130}\right) = 0.77$$

1.5.6. Valor esperado

$$E(X) = \frac{a+b}{2} = \frac{120+250}{2} = 185$$

1.5.7. Varianza

$$V(X) = \frac{(b-a)^2}{12} = \frac{(250-120)^2}{12} = 1408.3$$

$$CV = \frac{\sqrt{V(x)}}{E(x)} x100 = \frac{\sqrt{1408.3}}{185} x100 = 20.28\%$$

1.5.8. Ejercicio

De una estación parte un tren cada 20 minutos. Un viajero llega de improviso. Hallar:

a. Función de distribución de la variable aleatoria "tiempo de espera".
b. Probabilidad de que espere al tren menos de 7 minutos.

c. Esperanza y Varianza de la variable aleatoria “tiempo de espera”.
d. Probabilidad de que espere exactamente 12 minutos.

1.6. Distribución exponencial

1.6.1. Ejemplo

Si el tiempo en minutos de atención de un médico a un paciente en una sala de urgencias de un hospital es una exponencial de parámetro 1/9 calcular las probabilidades de:

a. Que el médico tarde menos de 20 minutos en atender a 2 pacientes.
b. Que tarde menos de 20 minutos en atender a 3 pacientes.

1.6.2. Notación

X es la variable que mide el tiempo en minutos que un medico demora con un paciente

$$X \sim Exp(\lambda = \frac{1}{9})$$

1.6.3. Función de densidad

$$f(x) = \frac{1}{9} e^{-\frac{x}{9}} \qquad x > 0$$

1.6.4. Función de distribución

$$F(x) = 1 - e^{-\frac{x}{9}} \qquad x > 0$$

1.6.5. Calculo de probabilidad

$$P(X < 20) = F(20) = 1 - e^{-\frac{20}{9}} = 0.89$$

1.6.6. Valor esperado

$$E(X) = \frac{1}{\lambda} = 9$$

1.6.7. Varianza

$$V(X) = \frac{1}{\lambda^2} = 81$$

1.6.8. Ejercicios

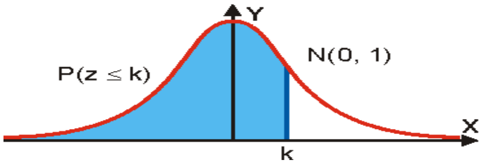
El tiempo de servicio en una ventanilla es una variable aleatoria con distribución exponencial. Un cliente se considera satisfecho si su tiempo de servicio es inferior a 2. Se sabe que el 95% de los clientes están satisfechos.

a. Calcular el tiempo medio de servicio.
b. Si se escogen al azar 60 clientes, calcular la probabilidad de encontrar menos de 4 clientes satisfechos.

XIV. DISTRIBUCION NORMAL

a. Tabla de distribución normal estándar

TABLA DE LA DISTRIBUCION NORMAL



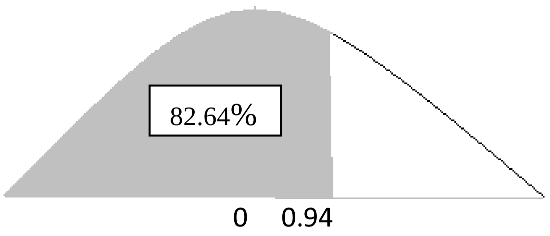
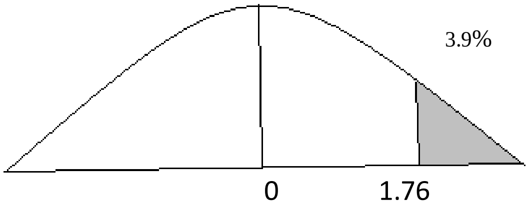
z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7704	0,7793	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8364	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8930
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9235	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9934	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9975	0,9975	0,9976	0,9977	0,9978	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0.9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

b. Manejo de tabla

Realice la grafica y encuentre el valor de las siguientes probabilidades

a. $P[Z \le 0.94]$	b. $P[0.87 \le Z \le 1.28]$	c. $P[Z \ge 1.76]$	d. $P[Z \ge -0.65]$
e. $P[\lvert Z\rvert \ge 1.65]$	f. $P[Z \le k] = 0.281$ k=?	g. $P[Z \le -0.85]$	h. $P[-0.34 \le Z \le 0.62]$
i. $P[\lvert Z\rvert \le 1.28]$	j. $P[-2.98 \le Z \le -1.32]$	k. $P[\lvert Z - 0.2\rvert \le 0.9]$	l. $P[k \le Z \le 1.25] = 0.48$ k=?

$F_z(- k) = 1 - F_z(k)$

$P[Z \le 0.94] = F_Z(0.94) = 0.8264$  	$P[Z \ge 1.76] = 1 - F_z(1.76) = 1 - 0.9608 = 0.0392$ 

c. Ejemplo de aplicación

El contenido de grasa (medido en %) de un cierto artículo de consumo es una variable aleatoria con distribución N(10,0.5). Se considera que el artículo es defectuoso si su contenido en grasa es superior a 11. Calcular la probabilidad de dicho evento. Si se examina un lote de 100 artículos, calcular la probabilidad de encontrar al menos dos.

i. Función de densidad

X es la variable aleatoria que mide el contenido de grasa que tiene el artículo

$$X \sim N(\mu;\ \sigma) = N(10; 0.5)$$

$$f(x) = \frac{1}{\sqrt{2\Pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma}} \qquad x > 0$$

ii. Función de distribución

$$X \sim N(\mu;\ \sigma) = N(0; 1)$$

$$F(z) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{(Z)^2}{2}} \qquad z > 0$$

iii. Calculo de probabilidad

$$P(X < k) = P\left(Z < \frac{x - \mu}{\sigma}\right) = F_z\left(\frac{x - \mu}{\sigma}\right)$$

$$P(X > 11) = 1 - P(X < 11) = 1 - P\left(Z < \frac{11-10}{0.5}\right) = 1 - F_z(2.0) = 1 - 0.9713 = 0.0287$$

iv. Valor esperado

$$E(X) = \mu = 10$$

v. Varianza

$$V(x) = \sigma^2 = 0.25$$

d. Aproximación Normal ~ Binomial o Poisson

El Banco Atlántico recibe el 42% de sus solicitudes de trabajo de personas solteras. ¿Cuál es la probabilidad de que en 150 solicitudes 66 o más personas sean solteras?

$$P(X \leq k) = P\left[Z \leq \frac{k - np + 0.5}{\sqrt{np(1-p)}}\right]$$

$$P(X < k) = P\left[Z \leq \frac{k - np - 0.5}{\sqrt{np(1-p)}}\right]$$

$$\mathrm{P}(\mathrm{X} \geq 66) = 1 - \mathrm{P}(\mathrm{X} < 66) = 1 - P\left[Z \leq \frac{66-150(0.42)-0.5}{\sqrt{150(0.42)(0.58)}}\right] = 1 - P[Z \leq 0.41] = 1 - F_z(0.41) = 1 - 0.6591 = 0.3409$$

XV. APLICACIONES DE LA DISTRIBUCION NORMAL

a. La vida de los tubos de imagen para televisores producidos por la TENSOR CORPORATION están normalmente distribuidos. El 92.5% de los tubos tienen una vida mayor de 2.160 horas y el 3.92% mayor de 17.040 horas. ¿ Qué media y desviación típica tiene la distribución ?.

$$P(X > 2.160) = 0.925 \qquad P\left(Z < \frac{2160-\mu}{\sigma}\right) = 0.075$$

$$P(X > 17.040) = 0.0392 \quad P\left(Z < \frac{17.040-\mu}{\sigma}\right) = 0.9608$$

$$\frac{2160-\mu}{\sigma} = -1.44 \qquad\qquad \frac{17040-\mu}{\sigma} = 1.76$$

$$\mu - 1.44\sigma = 2160$$
$$\mu + 1.76\sigma = 17040$$

$\sigma = 4.650$ $\mu = 8.856$

b. Ejercicio 1. Teniendo en cuenta que el diámetro de las naranjas de exportación sigue una distribución normal, un determinado inspector conoce por su dilatada experiencia que el 30% de las naranjas que examina tienen un diámetro inferior a 60 mm y el 20% tienen el diámetro superior a 100 mm. El país A exige que el diámetro esté comprendido entre 75 y 90 mm.

   a. Calcula la media y la desviación estándar
   b. Calcular la probabilidad de que esto ocurra en una determinada partida.
   c. El país B exige que el diámetro no baje de los 50 mm. La inspección se realiza midiendo el diámetro de 10 naranjas y rechazando una determinada partida si se encuentran más de dos naranjas con un diámetro inferior a 50 mm. Calcular la probabilidad de que una partida sea aceptada.

XVI. REGRESION LINEAL SIMPLE

MODELOS DE REGRESION LINEAL

Se tiene la siguiente información de un grupo de 10 familias: X - representara el ingreso y ; Y - el consumo

Consumo	Ingreso
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

CORRELACIÓN

Es una medida cuantitativa del grado de asociación entre las variables o sea el grado de bondad de la manera como una ecuación, describe o expresa la relación entre ellas. Cuando todos loa valores se pueden expresar perfectamente en términos de una fracción matemática

Clases de correlación

- Cuando el problema es bivariado la correlación es simple
- Cuando el problema multivariable se dice que la correlación es múltiple
- Si se escogen dos variables de un grupo de mas de tres se dice que la correlación es parcial
- Si al aumentar una variable también se aumenta la otra y también la correlación es directamente positiva en cero o lo contrario se dice que es inversamente proporcional

Correlación parcial: mide el grado de afinidad entre cualquier par de variables, cuando se controla el efecto de otras

Correlación múltiple: mide el grado de variación total de la variable dependiente causada por las variables independientes, como si actuaran todas juntas.

SUMATORIAS INDISPENSABLES EN LOS MODELOS DE REGRESION LINEAL

A continuación se presentan las sumatorias necesarias a ser utilizadas para realizar la estimación de los parámetros y el coeficiente de determinación para cada uno de los cuatro modelos de regresión que se presentan a continuación:

y	x	Ln(y)	Ln(x)	xy	xLn(y)	yLn(x)	Ln(x)Ln(y)
70	80	4.25	4.38	5600.00	339.88	306.74	18.62
65	100	4.17	4.61	6500.00	417.44	299.34	19.22
90	120	4.50	4.79	10800.00	539.98	430.87	21.54
95	140	4.55	4.94	13300.00	637.54	469.46	22.50
110	160	4.70	5.08	17600.00	752.08	558.27	23.86
115	180	4.74	5.19	20700.00	854.09	597.19	24.64
120	200	4.79	5.30	24000.00	957.50	635.80	25.37
140	220	4.94	5.39	30800.00	1087.16	755.11	26.65
155	240	5.04	5.48	37200.00	1210.42	849.50	27.64
150	260	5.01	5.56	39000.00	1302.77	834.10	27.86
1110	1700	46.71	50.72	205500.00	8098.85	5736.37	237.91
132100	322000	218.95	258.60	5534870000.00	7543936.34	3649836.83	5757.07

n = 10	$\sum X = 1.110$	$\sum X^2 = 132.100$
$\sum Y^2 = 322.000$	$\sum XY = 205.500$	$\sum Ln(x) = 50.72$
$\sum Y = 1.700$	$\sum(Ln(x))^2 = 258.60$	$\sum y\, Ln(x) = 5736.37$
$\sum Ln(y) = 46.71$	$\sum(Ln(y))^2 = 218.95$	$\sum x\, Ln(y) = 8098.8$
	$\sum Ln(x)\ Ln(y) = 237.91$	

MODELO DE REGRESIÓN LINEAL SIMPLE

Estimación de los parámetros de un modelo de regresión lineal de la forma:

$$Y = A + \beta X$$

$B = [\, n \sum XY - (\sum X)(\sum Y)] / [\, n \sum X^2 - (\sum X)^2\,] = 0.509$

$A = [\sum Y - B \sum X\,] / n = 24.454$

$r_{xy} = [n \sum XY - (\sum X)(\sum Y)] / [\, \sqrt{} [n \sum X^2 - (\sum X)^2] \, [n \sum Y^2 - (\sum Y)^2\,] = 0.98$

$R^2 = (r_{xy})^2 = 0.96$

El modelo estimado seria: $Y = 24.454 + 0.509\, X$

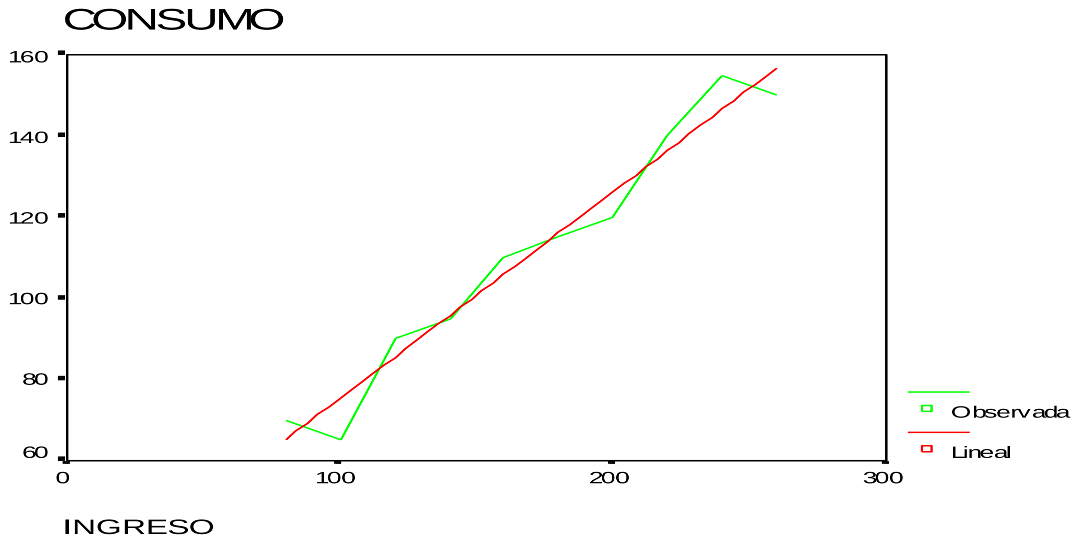
La interpretación de estos parámetros seria la siguiente:

- **R^2 = 0.96** ; indica que el ingreso esta explicando un 96% de la variación el consumo de las familias

- **B = 0.509** ; indica que ante un incremento del ingreso en $1.000 el consumo de las familias se incrementa en $ 509 pesos.
- **A = 24.45** ; indica que el consumo esperado de las familias cuando esta no percibe ingresos es de $24.450
- Xo = 100 ; indica que el consumo esperado de las familias cuando su ingreso es de $100.000 es de aproximadamente de $75.350.
- Xo = 300 ; indica que el consumo esperado de las familias cuando su ingreso es de $300.000 es de aproximadamente de $177.150.

El modelo estimado bajo el programa SPSS es el siguiente :

```
Dependent variable.. CONSUMO           Method.. Lineal

Listwise Deletion of Missing Data

Multiple R           ,98085
R Square             ,96206
Adjusted R Square    ,95732
Standard Error      6,49300

            Analysis of Variance:

              DF   Sum of Squares      Mean Square

Regression     1        8552,7273        8552,7273
Residuals      8         337,2727          42,1591

F =     202,86792       Signif F =  ,0000

-------------------- Variables in the Ecuación --------------------

Variable                  B        SE B       Beta         T  Sig T

INGRESO             ,509091     ,035743    ,980847    14,243  ,0000
(Constant)        24,454545    6,413817                3,813  ,0051
```



MODELO LOGARÍTMICO

Estimación de los parámetros de un modelo de regresión lineal logarítmico de la forma:

$$e^{Y} = AX^{\beta} \qquad \Leftrightarrow \qquad Y = Ln(A) + \beta Ln(X)$$

B = [n ∑ Ln X *Y - (∑ Ln X)(∑ Y)] / [n ∑(Ln X)^2 - (∑ Ln X)^2] = 78.06

A = [∑ Y - B ∑ Ln X] / n = - 284.88

r_{xy} = [n ∑Ln X * Y - (∑ Ln X)(∑ Y)] / [√ [n ∑(Ln X)^2 - (∑ Ln X)^2] [n ∑ Y^2 - (∑ Y)^2] = 0.96

R^2 = (r_{xy})^2 = 0.92

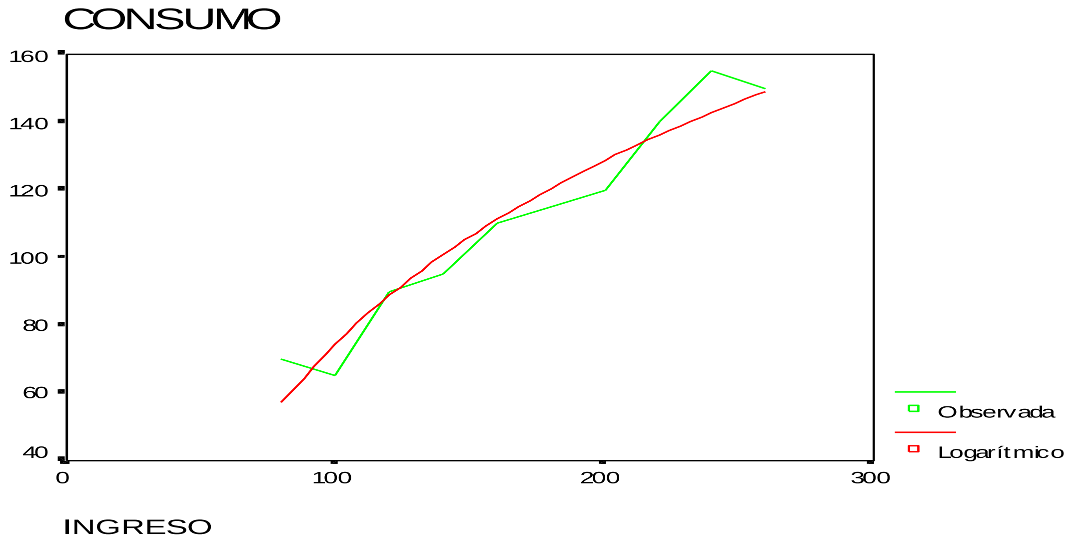
El modelo estimado seria: $Y = -2.4.88 + 78.057\, Ln(X)$

La interpretación de estos parámetros seria la siguiente:

- R^2 = 0.92; indica que el ingreso esta explicando un 92% de la variación el consumo de las familias.
- B = 78.05; indica que ante un incremento del ingreso en un 1% el consumo de las familias se incrementa en $ 78.057 pesos.
- Xo = 100 ; indica que el consumo esperado de las familias cuando su ingreso es de $100.000 es de aproximadamente de $75.580
- Xo = 300 ; indica que el consumo esperado de las familias cuando su ingreso de $300.000 es de aproximadamente de $160.340.

El modelo estimado bajo el programa SPSS es el siguiente :

```
Dependent variable.. CONSUMO           Method.. Logaritmico

Listwise Deletion of Missing Data

Multiple R           ,96795
R Square             ,93692
Adjusted R Square    ,92904
Standard Error      8,37211

            Analysis of Variance:

              DF    Sum of Squares       Mean Square

Regression     1         8329,2621         8329,2621
Residuals      8          560,7379           70,0922

F =     118,83288        Signif F =  ,0000

-------------------- Variables in the Ecuación --------------------

Variable                    B          SE B        Beta          T  Sig T

INGRESO             78,057406    7,160540    ,967949     10,901  ,0000
(Constant)        -284,889420   36,413007                -7,824  ,0001
```



MODELO POTENCIAL

Estimación de los parámetros de un modelo de regresión lineal potencial de la forma:

$$Y = AX^{\beta} \qquad \Leftrightarrow \qquad Ln\,(Y) = Ln\,(A) + \boldsymbol{\beta}\, \boldsymbol{Ln}\,(\boldsymbol{X})$$

$B = [\, n\ \sum Ln\ X * Ln\ Y - (\sum Ln\ X)(\sum Ln\ Y)] / [\, n \sum(Ln\ X)^2 - (\sum Ln\ X)^2\,] = 0.752$

$A = [\sum Ln\ Y - B \sum Ln\ X\,] / n = 2.355$

$r_{xy} = [n \sum Ln\ X\ \ LnY - (\sum Ln\ X)(\sum Ln\ Y)] / [\, \sqrt{}\, [n \sum(Ln\ X)^2 - (\sum Ln\ X)^2\,]\, [n \sum Ln\ Y^2 - (\sum Ln\ Y)^2$

$= 0.97$

$R^2 = (r_{xy})^2 = 0.94$

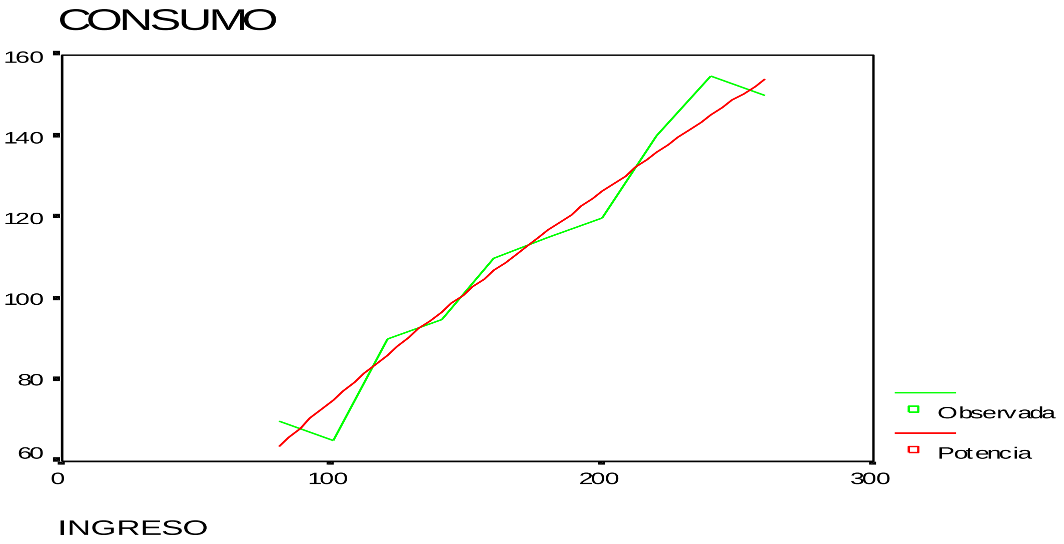
El modelo estimado seria: $Ln\,(Y) = Ln\,(2.355) + 0.752\, \boldsymbol{Ln}\,(\boldsymbol{X})$

La interpretación de estos parámetros seria la siguiente:

- $R^2 =$ 0.94 ; indica que el ingreso esta explicando un 94% de la variación el consumo de las familias
- B = 0.752 ; indica que ante un incremento del ingreso en un 1% ; el consumo de las familias se incrementa en un 75.2%
- Xo = 100 ; indica que el consumo esperado de las familias cuando su ingreso es de $100.000 es de aproximadamente de $75.160
- Xo = 300; indica que el consumo esperado de las familias cuando su ingreso de $300.000 es de aproximadamente de $171.706.

El modelo estimado bajo el programa SPSS es el siguiente:

```
Dependent variable.. CONSUMO          Method.. Potencial

Listwise Deletion of Missing Data

Multiple R           ,97364
R Square             ,94798
Adjusted R Square    ,94148
Standard Error       ,07282

            Analysis of Variance:

              DF   Sum of Squares      Mean Square

Regression     1        ,77294497        ,77294497
Residuals      8        ,04241643        ,00530205

F =     145,78219       Signif F =  ,0000

-------------------- Variables in the Ecuación --------------------

Variable                    B         SE B       Beta          T  Sig T

INGRESO             ,751943     ,062278    ,973642     12,074  ,0000
(Constant)         2,355685     ,746038                 3,158  ,0134
```



MODELO EXPONENCIAL

Estimación de los parámetros de un modelo de regresión lineal exponencial de la forma:

$$Y = Ae^{X\beta} \qquad \Leftrightarrow \qquad Ln\,(Y) = Ln\,(A) + \boldsymbol{\beta}\,\boldsymbol{X}$$

B = [n $\sum$ X * Ln Y - ($\sum$ X)($\sum$ Ln Y)] / [n $\sum$(X)^2 - ($\sum$ X)^2] = 0.0048

A = [$\sum$ Ln Y - B $\sum$ X] / n = 3.659

r_{xy} = [n $\sum$ X * Ln Y - ($\sum$ X)($\sum$Ln Y)] / [$\sqrt{}$ [n $\sum$(X)^2 - ($\sum$ X)^2] [n $\sum$ Ln Y^2 - ($\sum$ Ln Y)^2] = 0.96

R^2 = (r_{xy})^2 = 0.93

El modelo estimado seria: $Ln\,(Y) = Ln\,(47.067) + \mathbf{0.0048}\,\boldsymbol{X}$

La interpretación de estos parámetros seria la siguiente:

- R^2 = 0.93; indica que el ingreso esta explicando un 93% de la variación el consumo de las familias.
- B = 0.0048; indica que ante un incremento del ingreso de $1.000; el consumo de las familias se incrementa en un 0.48%.
- Xo = 100 ; indica que el consumo esperado de las familias cuando su ingreso es de $100.000 es de aproximadamente de $4.331

- Xo = 300; indica que el consumo esperado de las familias cuando su ingreso de $300.000 es de aproximadamente de $5.921.

El modelo estimado bajo el programa SPSS es el siguiente:

```
Dependent variable.. CONSUMO          Method.. Exponencial

Listwise Deletion of Missing Data

Multiple R           ,96913
R Square             ,93922
Adjusted R Square    ,93162
Standard Error       ,07871

            Analysis of Variance:

              DF   Sum of Squares      Mean Square

Regression     1        ,76580045        ,76580045
Residuals      8        ,04956095        ,00619512

F =     123,61352       Signif F =  ,0000

-------------------- Variables in the Ecuación --------------------

Variable                    B         SE B       Beta         T  Sig T

INGRESO              ,004817      ,000433    ,969132    11,118  ,0000
(Constant)         47,067465     3,659457               12,862  ,0000
```

XVII. EJERCICIOS DE COMPLEMENTO

1. El 20% de las ventas de automóviles nuevos en USA corresponde a los automóviles importados. Suponga que se seleccionan al azar 4 personas que han comprado un automóvil nuevo durante la semana pasada.

    a. Encuentre la probabilidad de que las 4 personas hayan comprado un automóvil importado.
    b. Encuentre la probabilidad de que sólo una de ellas haya comprado un automóvil importado.
    c. Encuentre la probabilidad de que ninguna de ellas hayan comprado un automóvil importado
    Sol: 0,0016 ; 0,4096 ; 0,4096 respectivamente

2. Los registros de mantenimiento revelan que solamente 1 de cada 100 máquinas de escribir de cierta marca requieren de una reparación mayor durante el 1er año de uso. El gerente de una oficina ordenó la compra de 10 máquinas de esta marca.

    a. Encuentre la probabilidad de que ninguna de las máquinas requieran una reparación mayor durante el 1er año de uso. Sol: 0,90438
    b. Encuentre la probabilidad de que 2 máquinas requieran una reparación mayor durante el 1er año de uso.

3. El centro de políticas energéticas de la agencia para la protección ambiental reporta que el 75% de las viviendas de Nueva Inglaterra utilizan quemadores de petróleo para calefacción. Si se sabe que una comunidad de Nueva Inglaterra tiene 2500 viviendas, encuentre el número esperado de las que usan quemadores de petróleo para calefacción y la desviación estándar.
    Sol: E(X) = 1875 viviendas, σ = 21,651 viviendas

4. El reporte anual es uno de los documentos más importantes producidos por las empresas de propiedad pública y su producción representa un gasto de importancia considerable. Sin embargo, un estudio reciente revela que 40% de los accionistas dedican 5 minutos o menos a la lectura del reporte anual de su compañía. Suponga que se eligen al azar 100 accionistas de empresas de propiedad pública.

    a. Encuentre el valor esperado del número de accionistas que dedican 5 minutos o menos a la lectura del reporte anual de su compañía. Sol: 40 accionistas
    b. Determine la desviación estándar. Sol: 4,898 accionistas
    c. Si se observa que 25 de los 100 accionistas seleccionados dedican no más de 5 minutos a la lectura del reporte anual. ¿Sería razonable pensar que la proporción de accionistas que dedican 5 minutos o menos a la lectura del reporte anual es 40%? Sol: No.

5. Suponga que una compañía de seguros vendió pólizas de seguros de vida a 5000 hombres de 42 años de edad. Si los estudios actuariales indican que la probabilidad de que un hombre de 42 años muera en un determinado año es 0,001. ¿Cuál es la probabilidad de que la compañía pague 4 indemnizaciones en un determinado año?. Sol: 0,1755200 (Por binomial) 0,1745 (Por aproximación de Poisson)

6. El gerente local de una compañía de renta de automóviles compra neumáticos en lotes de 500 para aprovechar los descuentos por compras al mayor. El gerente sabe, por experiencias anteriores, que el 1% de los neumáticos nuevos adquiridos en un determinado almacén salen defectuosos y se deben reemplazar durante la 1ra semana de uso. Encuentre la probabilidad de que en un envío de 500 neumáticos haya solamente uno defectuoso; no más de 3 defectuosos y ninguno defectuoso. Sol: 0,03368 ; 0,2650 ; 0,006737 respectivamente.

7. La central telefónica de un edificio de consultorios médicos puede manejar un máximo de 5 llamadas por minuto. Si la experiencia indica que se recibe un promedio de 120 llamadas por hora, encuentre la probabilidad de que en un determinado minuto la central esté sobrecargada. Sol: 0,0168

8. Un determinado antibiótico se envía a las farmacias en cajas de 24 botellas, el farmacéutico sospecha que la cantidad de antibiótico en algunos frascos es deficiente y decide analizar el contenido de 5 frascos. Suponga que 10 de las 24 botellas tienen cantidad deficiente de antibiótico.

   a. Encuentre la probabilidad de que ninguno de los frascos analizados tenga una cantidad deficiente de antibiótico. Sol: 0,0471
   b. Encuentre la probabilidad de que uno de los frascos analizados tenga una cantidad deficiente de antibiótico. Sol: 0,2355

9. Un embarque de 10 artículos contiene 2 unidades defectuosas y 8 no defectuosas. Al revisarlo, se tomará una muestra y las unidades se inspeccionarán. Si se encuentra una unidad defectuosa, se rechazará todo el embarque.

   a. Si se selecciona una muestra de 3 artículos. ¿Cuál es la probabilidad de rechazar el embarque?.
   b. Si se selecciona una muestra de 4 artículos. ¿Cuál es la probabilidad de rechazar el embarque?
   c. Si la gerencia estuviera de acuerdo en que hubiese una probabilidad de 0,47 de rechazar un embarque con 2 unidades defectuosas y 8 no defectuosas. ¿De qué tamaño se debe seleccionar la muestra?
      Sol: 0,5333 ; 0,6666 ; 7

10. Los pasajeros de las aerolíneas llegan al azar e independientemente a la sección de documentación en un gran aeropuerto internacional. La frecuencia promedio de llegadas es de 10 pasajeros por minuto.

    a. ¿Cuál es la probabilidad de que no lleguen pasajeros en un minuto?
    b. ¿Cuál es la probabilidad de que lleguen 3 pasajeros en un minuto?
    c. ¿Cuál es la probabilidad de no llegadas en período de 15 segundos?
    d. ¿Cuál es la probabilidad de al menos una llegada en período de 15 segundos?
       Sol: 0,000045 ; 0,0103 ; 0,0820 ; 0,918

11. Supongamos que la probabilidad de tener una unidad defectuosa en una línea de ensamblaje es de 0.05. Si el conjunto de unidades terminadas constituye un conjunto de ensayos independientes:

    a. ¿cuál es la probabilidad de que entre diez unidades dos se encuentren defectuosas?
    b. ¿y de que a lo sumo dos se encuentren defectuosas?
    c. ¿cual es la probabilidad de que por lo menos una se encuentre defectuosa?

12. El gerente de un restaurante que sólo da servicio mediante reservas sabe, por experiencia, que el 20% de las personas que reservan una mesa no asistirán. Si el restaurante acepta 25 reservas pero sólo dispone de 20 mesas, ¿cuál es la probabilidad de que a todas las personas que asistan al restaurante se les asigne una mesa? R/ 0.579

13. Una empresa electrónica observa que el número de componentes que fallan antes de cumplir 100 horas de funcionamiento es una variable aleatoria de Poisson. Si el número promedio de estos fallos es ocho,

    a. ¿cuál es la probabilidad de que falle un componente en 25 horas?
    b. ¿y de que fallen no más de dos componentes en 50 horas?
    c. ¿cuál es la probabilidad de que fallen por lo menos diez en 125 horas?

14. Supóngase que la producción de un día de 850 piezas manufacturadas contiene 50 piezas que no cumplen con los requerimientos del cliente. Se seleccionan del lote dos piezas al azar y sin reemplazo. Sea la variable aleatoria X igual al número de piezas de la muestra que no cumplen. ¿Cuál es la función de distribución acumulada de X?
15. Un avión de alto rendimiento contienen tres computadoras idénticas. Se utiliza únicamente una para operar el avión; las dos restantes son repuestos que pueden activarse en caso de que el sistema primario falle. Durante una hora de operación la probabilidad de que una falle en la computadora primaria(o de cualquiera de los sistemas de repuesto activados) es 0,0005. Suponiendo que cada hora representa un ensayo independiente,

    a. ¿Cuál es el tiempo promedio para que fallen las tres computadoras?
    b. ¿Cuál es la probabilidad de que las tres computadoras fallen en un vuelo de 5 horas?

16. Un lote contiene 100 piezas de un proveedor de tubería local y 200 unidades de un proveedor de tubería del estado vecino. Si se seleccionan cuatro piezas al azar y sin reemplazo,

    a. ¿cuál es la probabilidad de que todas sean del proveedor local?
    b. ¿Cuál es la probabilidad de que dos o más piezas de la muestra sean del proveedor local?
    c. ¿Cuál es la probabilidad de que al menos una pieza de la muestra sea del proveedor local?

17. Supongamos que el número de imperfecciones en un alambre delgado de cobre sigue una distribución Poisson con una media de 2.3 imperfecciones por milímetro.

    a. Determine la probabilidad de 2 imperfecciones en un milímetro de alambre.
    b. Determine la probabilidad de 10 imperfecciones en 5 milímetros de alambre.
    c. Determine la probabilidad de al menos una imperfección en 2mm de alambre

18. La contaminación constituye un problema en la fabricación de discos de almacenamiento óptico. El número de partículas de contaminación que ocurre en un disco óptico tiene una distribución de Poisson y el número promedio de partículas por centímetro cuadrado de superficie del disco es 0.1. El área de un disco bajo estudio es 100 centímetros cuadrados.

    a. Encuentre la probabilidad de que ocurran 12 partículas en el área del disco bajo estudio.
    b. La probabilidad de que ocurran cero partículas en el área del disco bajo estudio
    c. Determine la probabilidad de que 12 o menos partículas ocurran en el área del disco bajo estudio

19. Una caja contiene 8 bombillos, de los cuales están 3 están defectuosos. Se selecciona un bombillo de la caja y se prueba. Si este sale defectuoso se selecciona y se prueba otro bombillo, hasta que se escoja un bombillo no defectuoso. Encuentre el número esperado E de bombillos seleccionados.

20. Una prueba consta de 200 preguntas de verdadero o falso, para un sujeto que respondiese al azar ¿Cual sería la probabilidad de que acertase:

    a. 50 preguntas o menos.
    b. Más de 50 y menos de 100.
    c. Más de 120 preguntas.

21. Una gran tienda de artículos eléctricos descubre que el número *x* de tostadores vendidos por semana obedece a una ley de Poisson de media 10. La ganancia de cada tostador vendido es de 500 ptas. Sin embargo, un lunes se encuentran con que solo les quedan 10 tostadores, y que a lo largo de esa semana no van a poder traer más del almacén. Determinar la distribución de las ganancias totales (en ptas.) en concepto de tostadores de pan a lo largo de esa semana.

22. El tiempo de reparación de unas máquinas de escribir tiene una distribución aproximadamente exponencial, con media 22 minutos.

    a. Hallar la probabilidad de que el tiempo de reparación sea menor que diez minutos.
    b. El costo de reparación es de 2000 pts. por cada media hora o fracción. ¿Cuál es la probabilidad de que una reparación cueste 4000 pts.?
    b. Para efectuar una programación, ¿cuanto tiempo se debe asignar a cada reparación para que la probabilidad de que cualquier tiempo de reparación mayor que el tiempo asignado sea solo de 0.1?

23. Una alumna trae cada día a la Universidad una tableta de chocolate de 16 cm., y de cuando en cuando le da un mordisco y se come la mitad de lo que le queda. Asumiendo que esta golosa apetencia aparece en la mañana siguiendo una distribución de Poisson de media un mordisco por hora:

    a. Calcular la distribución del tiempo que transcurre hasta que aparece la primera mordida.
    b. ¿Cuantos centímetros de chocolate esperas que le quede tras las cinco horas de clases?
    c. ¿Qué probabilidad hay de que soporte una hora de clase sin morder su tableta?
    b. Si un día, entre las 9:00 y las 14:00 horas, la ha mordido en cuatro ocasiones, ¿Qué probabilidad hay de que lo haya hecho durante las tres primeras horas de clase?

24. Los pesos de 2 000 soldados presentan una distribución normal de media 65 kg y desviación típica 8 kg. Calcula la probabilidad de que un soldado elegidoal azar pese:

    a. Más de 61 kg.
    b. Entre 63 y 69 kg.
    c. Menos de 70 kg.
    d. Más de 75 kg

25. En un proceso de fabricación de tornillos se sabe que el 2% son defectuosos. Los empaquetamos en cajas de 50 tornillos. Calcula la probabilidad de que en una caja haya este número de tornillos defectuosos:

    a. Ninguno.
    b. Uno.
    c. Más de dos.
    d. ¿Cuántos tornillos defectuosos habrá, por término medio, en cada caja?

26. Calcula las probabilidades de las siguientes distribuciones binomiales mediante aproximación a la normal correspondiente (en todas ellas, ten en cuenta el ajuste de media unidad que hay que hacer al pasar de una variable discreta a una continua).

    a. x es B (100; 0,1). Calcula $P[x = 10]$, $P[x < 2]$ y $P[5 < x < 15]$
    b. x es B (1 000; 0,02). Calcula $P[x > 30]$ y $P[x < 80]$
    c. x es B (50; 0,9). Calcula $P[x > 45]$ y $P[x \leq 30]$

27. El departamento de control de calidad de una empresa que fabrica pañuelos sabe que el 5% de su producción tiene algún tipo de defecto .Los pañuelos se empaquetan en cajas con 15 elementos. Calcular la probabilidad de que una caja contenga:

    a. 2 elementos defectuosos .
    b. Menos de 3 elementos defectuosos
    c. Entre 3 y 5 elementos defectuosos(ambos incluidos)

28. Una prueba de inteligencia consta de diez cuestiones cada una de ellas con cinco respuestas de las cuales una sola es verdadera .UN alumno responde al azar ¿Cuál es la probabilidad de que responda al menos a dos cuestiones correctamente?¿Cuál la de que responda bien a seis?¿Cuál la de que responda bien como máximo a dos cuestiones?

29. Si se contesta sin pensar un test de 10 preguntas en las que hay que contestar si es cierto o falso. ¿Cuál es la probabilidad de acertar el 70 % o mas de las preguntas?, ¿y exactamente 7 de las 10 respuestas?

30. Los mensajes que llegan a una computadora utilizada como servidor lo hacen de acuerdo con una distribución de Poisson con una tasa promedio de 0.1 mensajes por minuto.

    a. ¿Cuál es la probabilidad de que lleguen como mucho 2 mensajes en una hora?
    b. Determinar el intervalo de tiempo necesario para que la probabilidad de que no llegue ningún mensaje durante ese lapso de tiempo sea 0.8.

31. El número de pinchazos en los neumáticos de cierto vehículo industrial tiene una distribución de Poisson con media 0.3 por cada 50000 kilómetros .Si el vehiculo recorre 100000 km, se pide:

    a. probabilidad de que no tenga pinchazos
    b. Probabilidad de que tenga menos de tres pinchazos
    c. Número de km recorridos para que la probabilidad de que no tenga ningún pinchazo sea 0.4066

32. La probabilidad de que el Banco “Riu Sec” reciba un cheque sin fondos es 1%.

    a. Si en una hora reciben 20 cheques, ¿cuál es la probabilidad de que se tenga algún cheque sin fondos?
    b. El banco dispone de 12 sucursales en la ciudad, ¿cuál es la probabilidad de que al menos 4 de las sucursales reciban algún cheque sin fondos?
    c. Si la media del valor de los cheques sin fondos es de $580 y el banco trabaja 6 horas diarias, ¿qué cantidad total de pesos no se espera pagar?
    d. Si se computaran los primeros 500 cheques, ¿cuál es la probabilidad de recibir entre 3 y 6 (inclusive) cheques sin fondos?

33. La duración de un láser semiconductor a potencia constante tiene una distribución normal con media 7.000 horas y desviación típica de 600 horas.

    a. ¿Cuál es la probabilidad de que el láser falle antes de 5.000 horas?
    b. ¿Cuál es la duración en horas excedida por el 95 % de los láseres?
    c. Si se hace uso de tres láseres en un producto y se supone que fallan de manera independiente.¿ Cuál es la probabilidad de que tres sigan funcionando después de 7.000 horas?

34. Una máquina fabrica tornillos cuyas longitudes se distribuyen normalmente con media 20 mm y varianza 0.25 mm. Un tornillo se considera defectuoso si su longitud difiere de la media más de 1 mm. Los tornillos se fabrican de forma independiente. ¿Cuál es la probabilidad de fabricar un tornillo defectuoso? Si los envasamos en envases de 15 tornillos, probabilidad de que en un envase no tenga más de 2 defectuosos.

35. Se supone que la probabilidad de que un pasajero opte por una compañía aérea dada para hacer un viaje es 0,5. Tomando un grupo de 400 pasajeros potenciales, esta compañía vende billetes a cualquiera que se lo solicita y la capacidad de su avión es de 230 pasajeros. Se pide:

    a. La probabilidad de que la compañía tenga un overbooking, es decir, que un pasajero no tenga asiento.

  b. Si existen 10 compañías aéreas que realizan el mismo viaje y cuyas condiciones son similares a la anterior, ¿cuál será la probabilidad de que al menos dos de ellas tenga un overbooking?

36. En un quiosco de periódicos se supone que el número de ventas diarias se distribuye normalmente con media 30 y varianza 2. Determinar:

  a. Probabilidad de que en un día se vendan entre 13 y 31 periódicos
  b. Determinar el máximo número de periódicos que se venden en el 90% de las ocasiones
  c. Supongamos que en una ciudad hay 10 quioscos independientes del mismo tipo y con las mismas características. Determinar la probabilidad de que más de dos quioscos vendan entre 13 y 31 periódicos

37. Un operador elige al azar entre “n” chips de una caja. La probabilidad de que sea defectuoso es 0,2.

  a. Si n = 7, ¿cuál es la probabilidad de que al menos 3 chips sean defectuosos?
  b. Si n = 50, ¿cuál es la probabilidad de tener entre 9 y 12 chips defectuosos?
  c. ¿Cuántos chips hay en la caja si la varianza es 32?

38. El volumen que una máquina de llenado automático deposita en latas de una bebida gaseosa tiene una distribución normal con media 34 cl. Y una desviación típica 1,5 cl.

  a. Si se despachan aquellas latas que tienen menos de 33 cl., ¿cuál es la proporción de latas desechadas?
  b. La máquina de llenado puede ser ajustada para cambiar el volumen medio ö para que únicamente el 1% de las latas tuviera menos de 33 cl.?
  c. Si tomamos 10 latas llenadas con la máquina tal y como figura originalmente, ¿cuál es la probabilidad de que ninguna sea desechada?
  d. Si ahora tomamos 500 latas llenadas con la máquina tal y como figura originalmente, ¿cuál es la probabilidad de que al menos 100 sean desechadas?

39. La confianza de un fusible eléctrico corresponde a la probabilidad de que un fusible. Escogido al azar de una línea de producción, funcione adecuadamente bajo condiciones de diseño. Calcule la probabilidad de obtener 27 ó mas fusibles defectuosos en una muestra de 1000 fusibles, sabiendo que la probabilidad de que un fusible elegido al azar no sea defectuoso es de 0,98.

40. Un banco recibe en promedio 6 cheques falsos al día, suponiendo que el número de cheques falsos sigue una distribución de Poisson, hallar:

  a. Probabilidad de que se reciban cuatro cheques falsos en un día.
  b. Probabilidad de que se reciban más de 30 cheques falsos en una semana

41. El número de fallos de un instrumento de prueba debidos a las partículas de un producto es una variable de Poisson con media 0,2 fallos por hora.

  a. ¿Cuál es la probabilidad de que el instrumento no falle en una jornada de 8 horas?
  b. ¿Cuál es la probabilidad de que haya entre 20 y 40 fallos (ambos incluidos) en un periodo de una semana (funcionando los 7 días, 24 horas diarias)?

42. Los agricultores de una región están preocupados po la calidad de sus cosechas, ya que se ha detectado en ciertas áreas la existencia de sustancias contaminantes en el suelo. Para analizarla, se segmenta la tierra en parcelas de 100 m2, y se concluye que hay una probabilidad de 0.6 de encontrar estos contaminantes en una determinada parcela. Se pide:

  a. Si un agricultor posee 15 de estas parcelas. ¿Qué probabilidad hay de que tenga alguna contaminada?

b. Una cooperativa posee 200 parcelas del tipo anterior. ¿Qué probabilidad hay de que tenga entre 100 y 150 parcelas contaminadas?
c. Si por cada parcela contaminada la cooperativa sufre unas perdidas de 1000 €, ¿cuál es la perdida que la cooperativa espera tener?.

43. En un estudio estadístico sobre la altura de los Colombianos y de los ingleses. Se han obtenido los siguientes datos:

Nacionalidad	Colombiano	Ingles
Media	170.2	175.4
Desviación típica	6.4	5.9

a. ¿Quién es más alto en su país, un español que mide 177 cm o un inglés que mide 181 cm?
b. ¿Cuál es la probabilidad de que un español mida más de 180 cm? ¿cuál es la probabilidad de que un ingles mida entre 160 y 170 cm?
c. ¿Cuál es la probabilidad de que un español sea más alto que un inglés?

44. Supóngase que la concentración que cierto contaminante se encuentra distribuida de manera uniforme en el intervalo de 0 a 20 pares de millón. Si se considera tóxica una concentración de 8 o más. ¿Cuál es la probabilidad de que al tomarse una muestra la concentración de esta sea tóxica?. Concentración media y varianza. Probabilidad de que la concentración sea exactamente 10.

45. El personal de la compañía Onda S.L. usa una Terminal para realizar sus pedidos internacionales. Si el tiempo que cada comercial gasta en una sesión en la Terminal tiene una distribución exponencial con media 36 minutos, encontrar:

a. Probabilidad de que un comercial utilice la Terminal 30 minutos o menos.
b. Si un comercial a estado 30 minutos en la Terminal, ¿Cuál es la probabilidad de que pase al menos una hora más en la Terminal?.
c. El 90% de las sesiones terminan en menos de R minutos. ¿Cuánto vale R?

46. El peso de las naranjas de un determinado calibre, fluctúa normalmente con media 150 gr. Y desviación típica 30 gr. Si una bolsa se llena con 15 naranjas seleccionadas al azar. ¿Cuál es la probabilidad de que el peso total de la bolsa sea inferior a 2 kilos?

47. Una fábrica produce en cada turno 100000 bolas para rodamientos de forma que la probabilidad de defectuosa es 0.04. En el control de las bolas se revisan todas depositando las defectuosas (que se detectan todas) en un recipiente que se vacía al final de cada turno. ¿Cuántas bolas ha de contener el recipiente para que la probabilidad de que su capacidad no se vea rebasada sea 0.95?

48. En un experimento de laboratorio se utilizan 10 gramos de 210*Po* 84 . Sabiendo que la duración media de un átomo de esta materia es de 140 días, ¿cuántos idas transcurrirán hasta que haya desaparecido el 90% de este material?

49. Se ha comprobado que el tiempo de vida de cierto tipo de marcapasos sigue una distribución exponencial con media de 16 años. ¿Cuál es la probabilidad de que a una persona a la que se le ha implantado este marcapasos se le deba reimplantar otro antes de 20 años? Si el marcapasos lleva funcionando correctamente 5 años en un paciente, ¿cuál es la probabilidad de que haya que cambiarlo antes de 25% años?

50. Supongamos que el cuerpo de bomberos de una gran ciudad es capaz de atender asta un máximo de 300 servicios por día. Si el numero medio de servicios diarios es de 250 con distribución de Poisson, ¿Cuál es la probabilidad de que un día determinado no se puedan atender todos los servicios requeridos? ¿Que probabilidad existe de que en 31 días haya al menos un día en el que no se puedan atender todos los servicios requeridos? 58. Supongamos que el cuerpo de bomberos de una gran

ciudad es capaz de atender asta un máximo de 300 servicios por día. Si el numero medio de servicios diarios es de 250 con distribución de Poisson, ¿Cuál es la probabilidad de que un día determinado no se puedan atender todos los servicios requeridos? ¿Que probabilidad existe de que en 31 días haya al menos un día en el que no se puedan atender todos los servicios requeridos?

51. Supongamos que el consumo familiar de un cierto producto se distribuye como una variable aleatoria de distribución uniforme, con esperanza igual a 10 y varianza unidad. Determina la probabilidad de que dicho consumo este comprendido entre 8 y 12 unidades.

52. En un juego solo caben dos posibilidades, ganar o perder. ¿Cuántas veces es necesario jugar para que la probabilidad, de que la frecuencia relativa de victorias difiera, de la verdadera probabilidad de ganar, en valor absoluto, en al menos 0,05, sea inferior o igual al 5%?

53. La probabilidad de que un deportista gane una competición es 0,4. ¿Cuantas veces habremos de verle competir para que haya una probabilidad de al menos 0,95, de que la frecuencia relativa de triunfos difiera de 0,4, en valor absoluto, como máximo en 0,02?

54. Un fabricante de faros para coches informa que en un envío de 4000 faros a un distribuidor, 500 tenían un ligero defecto. Si se compran al distribuidor 20 faros elegidos al azar. ¿Cuál es la probabilidad de que haya exactamente dos con defecto?

55. La probabilidad que un día cualquiera la agencia de viajes venda pasajes a Italia” es de 0.20 y la agencia trabaja 3 días. Si la atención de cada día es independiente a la atención de los días siguientes:

   a. Cual es la probabilidad que durante los próximos 3 días de trabajo, la agencia ‘venda pasajes a Italia” en exactamente 3 días?
   b. Cual es la probabilidad que en los próximos 3 días de trabajo, la agencia ‘venda pasajes a Italia” en 2 ó mas días?.

56. Para el caso de un experimento relacionado con la atención de clientes a través de un cajero automático que falla en 2 clientes por día.

   a. Cual es la probabilidad que durante el dia falle exactamente 3 veces?.
   b. Cual es la probabilidad que durante el dia no falle ninguna vez?.
   c. Cual es la probabilidad que durante el dia falle en más de un cliente?.
   d. Cual es la probabilidad que en medio dia falle 3 veces?.

57. En la recepción del hotel Astor hay 2O clientes de los cuales 15 estdn insatisfechos por la atención recibida. Se elige una muestra sin reposición de 4 clientes y se le pregunta su opinión sobre el servicio.

   a. Que 3 clientes de la muestra manifiesten estar insatisfechos
   b. Por lo menos 3 clientes de la muestra manifiesten estar insatisfechos.

58. Una compañía de seguros sabe que la probabilidad de que roben un auto en un año es 0.15 (nº robos/ nº autos) y que hay solicitudes para 1.500 seguros contra robo. El precio promedio de un auto es U$ 3.000. ¿Cuánto debe cobrar por seguro para que en un año no tenga pérdidas?

59. La probabilidad de que una persona tenga gripe es de *0.3 = p.*

   a. La probabilidad de que 5 (*x*) personas de 20 (*n*) tengan gripe es:

b. Encuentre la probabilidad de que cuando mucho 2 tengan gripe. Se calcula la probabilidad de 0, de 1 y de 2 y se suman.
c. La probabilidad de que un equipo funcione es de 0.9. Si se prueban 16 equipos, la probabilidad de que al menos14 funcionen es:

60. El equipo funciona o no funciona, esto indica un experimento de tipo binomial. Al menos 14 significa que deben funcionar, 14, 15 y 16, por lo que la probabilidad es:

61. Encuentre la probabilidad de que se atiendan a 12 personas en una hora si el valor esperado es de 10 personas atendidas por hora.

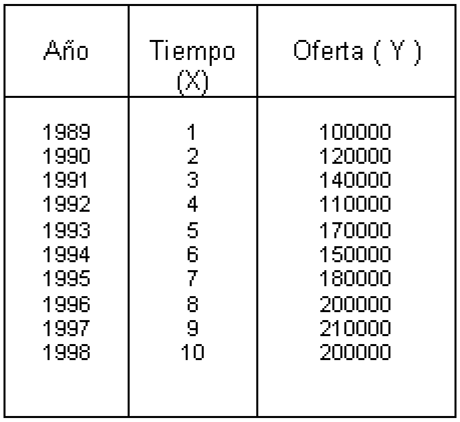
62. Se tienen 20 resistencias de 100 ohms, 15 resistencias de 200 ohms y 10 resistencias de 300 ohms. Si se toman aleatoriamente 6 resistencias, determine las siguientes probabilidades, escoger:

a. Dos de cada de cada tipo
b. 3 de 100 ohms , dos de 200 y una de 300
c. Las 6 de 200 ohms.
d. Las 6 de un mismo tipo

63. Una lata se debe llenar con un litro de producto. Un estudio en 36 latas proporciona una media de 0.98 litros con una desviación estándar de 0.1 litros.

a. La probabilidad de que una lata se llene con 1.1 litros es:
b. La probabilidad de que la lata se llene con menos de 0.96 litros es:
c. La probabilidad de que se llene con más 1.1 litros es:

64. Proyectar la Oferta de un cierto producto tomando en cuenta los datos obtenidos en el estudio de mercado, ver cual de los métodos o curvas de proyección se ajusta mejor a la nube de puntos y determinar la Oferta para los próximos diez años.



Año	Tiempo (X)	Oferta (Y)
1989	1	100000
1990	2	120000
1991	3	140000
1992	4	110000
1993	5	170000
1994	6	150000
1995	7	180000
1996	8	200000
1997	9	210000
1998	10	200000

65. Para graficar la curva de demanda hay que suponer que se mantienen constantes todas las variables salvo el precio.

Precio	Cantidad demandada
0	12
0.5	10
1	8
1.5	6
2	4
2.5	2

3 0

66. Suponga que un país se ve afectado por un fenómeno climatológico que provoca una menor cosecha de maíz, afectando negativamente la producción de raciones para la cría de aves. Analice los impactos de este fenómeno en los siguientes mercados: a) de raciones, b) de carne de aves, c) de carne vacuna.

67. Considere el mercado de pan de las panaderías de barrio de Montevideo. Determine qué ocurre con el precio y cantidad de equilibrio del pan ante cada una de las siguientes situaciones:

   a. Cae el precio del pan envasado de producción fabril.
   b. Aumenta el precio de la harina.
   c. Aumenta el número de panaderías.
   d. Se corre el rumor de que una toxina ha contaminado la producción de trigo, produciendo efectos negativos sobre la salud para quienes comen productos fabricados con harina.
   e. Se realiza una campana a la población para que disminuya el consumo de conservantes que afecta negativamente la demanda de consumo de pan envasado de producción fabril.

68. *Explique cada una de las afirmaciones siguientes, utilizando gráficos de oferta y demanda:*

   a. Cuando una ola de frío sacude Salto, el precio del jugo de naranja sube en los supermercados de Colombia.
   b. Todos los veranos, cuando comienza a hacer calor en Punta del Este y Mar del Plata, caen los precios de las habitaciones de hotel de los centros turísticos del Caribe.
   c. Cuando estalla una guerra en Medio Oriente, sube el precio del combustible mientras que baja el de los autos de gran cilindrada.

69. Mostrar con gráficos de oferta y demanda cómo afectan los siguientes acontecimientos al mercado de buzos de algodón:

   a. Desastres naturales provocan la destrucción de la cosecha de algodón.
   b. Baja el precio de las camperas de cuero.
   c. Todos los liceos obligan a hacer ejercicio físico con el atuendo adecuado.
   d. Se inventan nuevas máquinas de tejer.

70. Suponga que en el ano 2002 el número de nacimientos es temporalmente alto. Cómo afecta esta explosión de natalidad al precio de los servicios de cuidar niños en el 2007 y en el 2017? Pista: Los niños de 5 anos necesitan cuidadores mientras que los de 15 pueden ser cuidadores.

71. El mercado de chivitos tiene las siguientes tablas de demanda y oferta:

Precio	cantidad demandada	cantidad ofrecida
4	135	26
5	104	53
6	81	81
7	68	98
8	53	110
9	39	121

   a. Represente gráficamente las curvas de demanda y oferta e indique cuál es el precio y la cantidad de equilibrio.
   b. Si en este mercado el precio efectivo fuera superior al de equilibrio, qué llevaría al mercado al equilibrio?
   c. Si fuera inferior al de equilibrio, qué llevaría al mercado al equilibrio?

72. Un avance tecnológico reduce el costo de producir chips para computadoras. Muestre con gráficos de oferta y demanda la influencia de este avance en el precio y cantidad de equilibrio en los siguientes mercados:

   a. El mercado de computadoras.
   b. El mercado de programas informáticos.

73. Analice qué ocurre en los siguientes mercados, utilizando gráficos de oferta y demanda:

   a. Hay una mala cosecha de naranjas.
   b. Se hace publicidad sobre la bondad de una dosis de vitamina C diaria.

74. Analice qué ocurre en el mercado de servicios de peluquería en los siguientes casos:

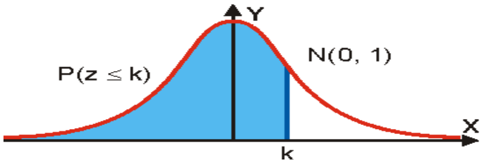
   a. Cae el ingreso del país.
   b. Se pone de moda el pelo lacio.
   c. Aumenta el precio de los alquileres.
   d. Aumenta el número de peluquerías.

XVIII. REPASO DE LA DISTRIBUCION NORMAL

1.1. MANEJO DE TABLA DE DISTRIBUCION NORMAL

Realice la grafica y encuentre el valor de las siguientes probabilidades

$$X \sim N(0,1) \qquad P(X \le k) = F_Z(k)$$



a. $P[Z \le 0.94]$	b. $P[0.87 \le Z \le 1.28]$	c. $P[Z \ge 1.76]$	d. $P[Z \ge -0.65]$
e. $P[\lvert Z \rvert \ge 1.65]$	f. $P[Z \le k] = 0.281$ k=?	g. $P[Z \le -0.85]$	h. $P[-0.34 \le Z \le 0.62]$
i. $P[\lvert Z \rvert \le 1.28]$	j. $P[-2.98 \le Z \le -1.32]$	k. $P[\lvert Z - 0.2 \rvert \le 0.9]$	l. $P[k \le Z \le 1.25] = 0.48$ k=?

1.2. CALCULO DE PROBABILIDAD CON UNA DISTRIBUCION ESPECIFICA

$Si\ X \sim N(\mu;\ \sigma)$ Entonces $P(X \le K) = P\left[\frac{\bar{X}-\mu}{\sigma} \le \frac{k-\mu}{\sigma}\right] = P\left[Z \le \frac{k-\mu}{\sigma}\right] = F_Z\left[\frac{k-\mu}{\sigma}\right]$

Si X es una variable con distribución normal con media de 8.47 y desviación estándar de 1.15. Hallar:

a. $P[X \le 9.12]$	b. $P[X \le 6.42]$	c. $P[X \ge 12.62]$
d. $P[X \ge 10.53]$	e. $P[6.12 \le X \le 11.92]$	f. $P[5.06 \le X \le 6.84]$

1.3. APLICACIONES DE LA DISTRIBUCION NORMAL

1.3.1. Se han utilizado dos tipos de pruebas A y B para medir los conocimientos sobre cierta materia en una misma población. Los resultados en ambos tienen una distribución normal. La prueba A tiene una media de 78.3 con una desviación estándar de 4.2. Y la prueba B tiene una media de 85.1 con una desviación estándar de 3.2. Una persona ha obtenido 83.1 en la prueba A Y 87.5 en la prueba B. Cual de os dos se encuentra en mejor posición? Porque?

1.3.2. Las ventas semanales de un quiosco de prensa no es fija y el propietario del quiosco sabe por experiencia acumulada que las ventas medias son de 200 revistas con una desviación estándar de 100.

a. Que volumen de revistas debe pedir semanalmente, si quiere satisfacer la demanda en al menos un 90%?

b. Si para otro quisco con el mismo numero de ventas medias resulta que su varianza es mayor. El volumen de revistas que debería pedir semanalmente si quiere satisfacer la demanda en al menos un 90% será mayor, menor o igual que el quisco del punto anterior? Explica tu respuesta

1.3.3. El departamento de recursos humanos de una gran compañía financiera ha realizado un examen para contratar empleados para cubrir los puestos de trabajo durante las vacaciones de verano. Se sabe que las calificaciones se comportan como una normal con media 8 y desviación estándar 2. Además ha realizado una prueba de manejo de programas informáticos que también se comporta siguiendo una normal con media 5 y desviación estándar 1.2 (independientes)

a. Si la prueba de manejo informático se aprueba con mas de 6 puntos. Que porcentaje de individuos la aprobaran?
b. Una persona será contratada si la calificación total obtenida como suma de ambas pruebas supera los 16 puntos. Que probabilidad hay que una persona sea contratada?
c. En una sucursal se han presentado 10 personas, suponiendo que la probabilidad de aprobar es la misma para todos y es de 0.10. Que probabilidad hay que sean al menos sean contratadas al menos 2 de las 10 presentadas?
d. En una convocatoria general para todo Colombia se han presentados 700 personas. Suponiendo también que la probabilidad de ser contratado sigue siendo la misma para todos los candidatos. Que probabilidad hay que sean contratados mas del 10% de los presentados. Cual es el numero esperado de contratados

1.3.4. Una explotación agraria cultiva pepinos. Por cosechas anteriores se sabe que la longitud media de los pepinos es de 10 centímetros con una desviación estándar de 0.5. la cosecha se pasa por una maquina que rechaza aquellos pepinos cuya longitud difiera de la media en valor absoluto en mas de un centímetro. Los pepinos rechazados se venden a 50 pesos. Cuanto cree que habrá cobrado por los pepinos rechazados un agricultor que ha producido 10.000 pepinos en la presente cosecha?

1.3.5. En una ciudad una de cada tres familias poseen teléfono. Si se eligen al azar 90 familias. Cual es la probabilidad:

a. Halla por lo menos 30 familias con teléfono
b. Halla menos de 60 familias con teléfono
c. Halla entre 30 y 50 familias con teléfono

1.3.6. Un bus de transporte MIO llega a una estación con un determinado número de viajeros. Se sabe que el numero de viajeros que suben a este tipo de vehículos sigue una distribución normal con media de 40 y varianza 81, mientras que el numeró de pasajeros que se bajan sigue una distribución normal con media 30 y desviación estándar de 12. Suponiendo independientes el numero de pasajeros que suben y bajan en este. Calcular la probabilidad de que cuando salga el vehiculo lo haga con mas pasajeros que cuando llego a la estación.

1.3.7. Una empresa dedicada a la fabricación de vestidos de señora los fabrica con longitudes comprendidas entre 110 y 170 cm y con tallas que se diferencian entre sí en 10 cm y una talla extra de 200 cm. La empresa sabe que las alturas de las mujeres potenciales clientes medidas desde el hombro hasta los pies se distribuyen normalmente con media 135 y con desviación típica 15 y que la clienta cuya altura no coincide exactamente con una de las tallas se comprará la inmediatamente mayor. Si la empresa proyecta fabricar para la próxima temporada 50.000 vestidos ¿cuántas debería razonablemente hacer de cada una de las siete tallas previstas?

1.3.8. El responsable de la sección de frutería de una gran superficie sabe que el número medio de kilos de patatas vendidos diariamente es de 250 kg con una desviación estándar de 25 kg. Además se ha comprobado que las ventas diarias de patatas son independientes entre si. Se pide:

   a. Cuantos kilos de patatas deberá tener disponibles cada día si quiere satisfacer la demanda de al menos un 90%?
   b. Esta gran superficie abre 320 días al año. Cual es la probabilidad de vender entre 80.000 y 81.000 kg de patatas en un año?

1.3.9. El gasto mensual de la familia Robles sigue una distribución normal de media de 3.000 pesos y varianza 500. Supongamos que el gasto de cada mes es independiente del de los otros meses. Si el ingreso anual es de 37.000 pesos, ¿cuál es la probabilidad de que no gasten más de lo que ganan? ¿Cuánto deberían ganar para tener una seguridad del 99% de que no gastarán más de lo que han ganado?

1.3.10. Un servicio de reparto de pizzas a domicilio reparte en una residencia de estudiantes. Los tiempos de entrega siguen una distribución normal con media de 20 minutos y desviación típica de 4 minutos.

   a. Cual es la probabilidad de que se tarde entre 15 y 25 minutos la entrega de una pizza
   b. La pizza no tiene cargo si no es entregada en menos de 30 minutos. Cual es probabilidad de comerse una pizza gratis, si se hace un único pedido.
   c. Durante la semana de exámenes finales. un estudiante planea pedir una pizza cinco noches consecutivas .Supongamos que los tiempos de entrega de las pizzas son independientes entre si. Cual es probabilidad de que este estudiante consiga al menos una pizza gratis.
   d. Encontrar el rango mas corto en el cual este contenido el 40% de los tiempos de entrega

1.3.11. Los empleados de una empresa trabajan un promedio de 55.8 horas semanales, con una desviación estándar se 9.8 horas semanales. Los asensos son mas probables para los empleados que están dentro del 10% de los que pasan mas tiempo trabajando. Cuanto debe trabajar usted para mejorar sus oportunidades de ascenso?

1.3.12. Los costos de producción mensual en una imprenta de Cali son de $1.025.000 pesos en promedio con una desviación estándar de $217.500 pesos. El gerente promete al propietario de la tienda mantener los costos por debajo de los $750.000 pesos mensuales. Si los costos están distribuidos normalmente. El propietario puede creerle al gerente?

1.3.13. Una empresa de contabilidad descubre que el tiempo que se toma en realizar un proceso de auditoria esta distribuido normalmente con un tiempo promedio de 17.2 días y una desviación estándar de 3.7 días. El dueño de la empresa promete iniciar un trabajo de auditoria para su firma dentro de 20 días , pero debe terminar una ya comenzada. Que tan probable es que cumpla su promesa?

1.3.14. Las ventas promedio de Ingrid son de $1.250.000 pesos con una desviación estándar de $38.000. Gana una comisión de $250.000 solo si sus ventas exceden de $1.325.000 pesos en promedio. Cual es la comisión esperada por cada 25 ventas?

1.3.14. Los estudiantes inscritos a una prueba de aptitud gerencial para graduados obtienen 812 en promedio con una desviación estándar se 145. Solo quienes están entre el 20% de los mejores pueden aplicar a una beca especifica. El genio Miguel recibió un puntaje de 900 puntos en la prueba. Puede aplicar?

XIX. DISTRIBUCIONES DE MUESTRALES

2. **Distribución muestral:** es aquella que incluye todos los valores posibles que puede tomar un estadístico como una media muestral para un tamaño de muestra dado

2.1. Una población se compone de los cinco números: 2, 3, 6, 8, 11.

- Hallar la media de la población, μ
- Hallar la varianza de la población, σ^2
- Calcule el número de muestras posibles de tamaño dos que se pueden extraer de esta población (sin sustitución).
- Indique cuáles son todas esas muestras posibles de tamaño dos que se pueden extraer de esta población y calcule sus respectivas medias, X_i
- Grafique la Distribución Muestral de Medias.
- Calcule la media de la distribución muestral de medias, $\mu_{\overline{X}}$. ¿Se verifica que $\mu_{\overline{X}}$ = μ?
- Usando las medias de las muestras (X_i), calcule la varianza de la distribución muestras de medias, $\sigma^2_{\overline{X}}$
- Usando el valor de σ^2 y la expresión correspondiente, calcule la varianza de la distribución muestral de medias, $\sigma^2_{\overline{X}}$,.. Verifique que este resultado es el mismo que el del inciso anterior.
- Calcule el error muestral de cada una de las muestras del inciso

2.2. Una población se compone de los cinco números: 2, 3, 6, 8, 11.

- Hallar la media de la población, μ
- Hallar la varianza de la población, σ^2
- Calcule el número de muestras posibles de tamaño dos que se pueden extraer de esta población (con sustitución).
- Indique cuáles son todas esas muestras posibles de tamaño dos que se pueden extraer de esta población y calcule sus respectivas medias, X_i
- Grafique la Distribución Muestral de Medias.
- Calcule la media de la distribución muestral de medias, $\mu_{\overline{X}}$. ¿Se verifica que $\mu_{\overline{X}}$ = μ?
- Usando las medias de las muestras (X_i), calcule la varianza de la distribución muestras de medias, $\sigma^2_{\overline{X}}$
- Usando el valor de σ^2 y la expresión correspondiente, calcule la varianza de la distribución muestral de medias, $\sigma^2_{\overline{X}}$,.. Verifique que este resultado es el mismo que el del inciso anterior.
- Calcule el error muestral de cada una de las muestras del inciso

2.3. El departamento de control de calidad tiene como empleados a cinco técnicos en el turno matutino. A continuación aparece el número de veces que cada técnico indico al supervisor de producción que interrumpiera el proceso durante la última semana.

TÉCNICO	INTERRUPCIONES	PROFESIONALES
Luís	4	Si
Pedro	3	No
Jaime	5	No
Juan	3	Si
Rodrigo	2	Si

- Cuantas muestras de dos técnicos se pueden formar con esa población
- Enumere todas las muestras de dos observaciones que pueden formar y calcule la media de cada muestra
- Compare las medias de la muestra con la media de la población
- Compare la desviación estándar de las muestras y la desviación estándar de la población. Razone

2.4. El departamento de control de calidad tiene como empleados a cinco técnicos en el turno matutino. A continuación aparece cuales son profesionales.

TÉCNICO	INTERRUPCIONES	PROFESIONALES
Luís	4	Si
Pedro	3	No
Jaime	5	No
Juan	3	Si
Rodrigo	2	Si

- Calcule la media y desviación estándar de la proporción poblacional
- Cual es la distribución de muestras de una proporción para una muestra de tamaño 2
- Cual es la media y la desviación de la distribución de muestras
- Que relación tiene la desviación estándar de la población y la desviación estándar de la muestra ; justifique
- Realice y analice una grafica de las proporciones muéstrales

2.5. **Teorema central del límite:** establece que si se selecciona una muestra aleatoria suficientemente grande de n observaciones de una población, la distribución muestral de las medias de las muestras se aproximara a una distribución normal. Cuando mas grande sea el tamaño de la muestra, n mejor será la aproximación normal a la distribución muestral de las medias de las muestras. Se pueden establecer las siguientes propiedades:

- La media de la distribución muestral es igual a la media de la población μ = X
- La desviación estandar de la distribución muestral de las medias (error muestral) es igual a la desviación estandar de la población dividida entre la raiz cuadrada del tamaño de las muestras $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ para n>33 o $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ para n $\leq$ 33
- La distribución muestral de las medias es aproximadamente normal para los tamaños de muestra menores de 33

2.6. Notación a utilizar en las distribuciones muéstrales

DISTRIBUCION NORMAL	***DISTRIBUCION t STUDENT***

a. Distribución

$$Z = \frac{\overline{X}-\mu}{\sigma} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$$

b. Función de densidad

$$f(Z) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{Z^2}{2}} \qquad - \infty \le Z \le \infty$$

c. Función de distribución

$$F(Z) = P(Z \le z) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{Z} e^{-\frac{Z^2}{2}} dt$$

d. Valor esperado y varianza

$$E(Z) = 0 \qquad V(Z) = 1$$

e. Notación

$$P[Z \le k] = \alpha \quad \leftrightarrow \quad Z_\alpha = k$$

a. Distribución

Si Y ~ N(0,1) y U ~ χ^2_k. Suponiendo que Y y U son independientes

$$T = \frac{Y}{\sqrt{\frac{U}{k}}} \sim t_k$$

b. Función de densidad

$$f(X) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{X^2}{n}\right)^{-\frac{n+1}{2}} \qquad - \infty \le X \le \infty$$

c. Valor esperado y varianza

$$E(T) = 0 \qquad V(Z) = \frac{k}{k+2}$$

d. Notación

$$P[t_n \ge k] = \alpha \quad \leftrightarrow \quad t_{\alpha;n} = k$$

DISTRIBUCION CHI CUADRADO

a. Distribución

Si Z ~ N(0,1) y $\sum_{i=1}^{n} Z_i^2 = \frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n-1}$.

b. Función de densidad

$$f(X) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} X^{\frac{n-2}{2}} e^{-\frac{nx}{2}} \qquad X > 0$$

c. Valor esperado y varianza

$$E(T) = n \qquad V(Z) = 2n$$

DISTRIBUCION F SNEDECOR

a. Distribución

Si $Z_1 \to \chi^2_{n_1}$ y $Z_2 \to \chi^2_{n_2}$. Si Z_1 y Z_2 son independientes

$$F = \frac{\frac{Z_1}{n_1}}{\frac{Z_2}{n_2}} \sim F\left(\frac{n_1}{n_2}\right).$$

Donde n_1 numero de grados de libertad en el numerador y n_2 numero de grados de libertad en el denominador

b. Valor esperado y varianza

d. Notación $P\left[\chi^2_n \geq k\right] = \alpha \quad \leftrightarrow \quad \chi^2_{n;\alpha} = k$	$E(F) = \frac{n_2}{n_2 - 2}$ $V(F) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$ c. Notación $P\left[F_{n_1;n_2} \geq k\right] = \alpha \quad \leftrightarrow \quad F_{n_1;n_2,\alpha} = k$

2.7. Notación, manejo y ejercicios de aplicación de la distribución normal

2.7.1. Manejo de tabla

DISTRIBUCION NORMAL	
a. $P\left[Z \leq -1.96\right]$	d. $Z_{0.117}$
b. $P\left[Z \leq K\right] = 0.975$ K=?	e. $Z_{\alpha} = 2.63$ ∝ = ?
c. $P\left[Z \geq 0.95\right]$	f. $Z_{0.937}$

2.7.2. Un empresario afirma que la duración de las bombillas de 100 W que fabrica su empresa sigue una distribución normal con una desviación típica de 120 horas de duración. Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. Cual es la probabilidad que rechace el lote por no cumplir la garantía?

2.7.3. El control de calidad una fábrica de pilas y baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Hasta ahora el tiempo de duración en conversación seguía una distribución normal con media 300 minutos y desviación típica 30 minutos. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que de una muestra de 60 baterías el tiempo medio de duración en conversación fue de 290 minutos. Suponiendo que ese tiempo sigue siendo Normal con la misma desviación típica: ¿Se puede concluir que las sospechas del control de calidad son ciertas?

2.7.4. Un concesionario vende dos tipos de vehículos, unos de gama alta y otros de gama media. Las ventas de coches de gama alta suponen el 30% del total de coches vendidos. ¿Cuál es la probabilidad de que entre los últimos vehículos vendidos, se elijan 100 al azar y resulte que más del 35% sean de gama alta?

2.7.5. La empresa Grano Sol vende galletas ecológicas en paquetes de 60 unidades. Los dueños saben que el peso de cada galleta es una variable aleatoria que tienen una media de 71 gr. y una dispersión, medida a través de la desviación típica, de 10 gr. a) ¿Cuál es la probabilidad de que en un paquete de 60 galletas escogidas aleatoriamente, el peso medio de las galletas sea superior a 70 gramos? b) ¿Cuál sería el resultado si la varianza poblacional fuera desconocida? (Suponga que la desviación típica muestral es de 5 kg, y una cuasidesviación típica de 5,04).

2.7.6. Se desea analizar las diferencias de calificaciones entre dos grupos de alumnos. Unos proceden del grupo 22 y otros del grupo 23. Para estudiar la distribución en el muestreo de la diferencia de medias se toman m.a.s. independientes de ambas poblaciones obteniéndose la siguiente tabla:

DATOS	**GRUPO 22**	**GRUPO 23**
Tamaño de la población	200	150
Tamaño de la muestra	100	75
Media de la población	4,10	5,18
Media de la muestra	4,2153	5,3247
Desviación típica de la población	1,55	1,95
Desviación típica de la muestra	1,5635	1,8238

¿Cuál es la probabilidad de que la diferencia de medias muestrales sea mayor que uno?

2.7.7. Según los resultados de un estudio exhaustivo de la población un 80% de las mujeres entrevistadas afirman utilizar algún producto cosmético todos los días, mientras que en el caso de los hombres este porcentaje en la actualidad asciende 55%. Una pequeña firma de cosmética se plantea sacar al mercado una crema hidratante de uso específico para hombres, pero antes de crear esa nueva línea de negocio, decide realizar su propia encuesta sobre una pequeña muestra aleatoria: selecciona a 50 mujeres y a 60 hombres y les pregunta sobre sus hábitos cosméticos. Calcule la probabilidad de que la diferencia entre la proporción de mujeres que utiliza cosméticos respecto a la proporción de hombres que los utiliza sea inferior al 20%.

2.7.8. La duración de las bombillas de 100 W que fabrica una empresa sigue una distribución normal con una desviación típica de 120 horas de duración. Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. , ¿habría que rechazar el lote por no cumplir la garantía?

2.7.9. Un fabricante de lámparas eléctricas está ensayando un nuevo método de producción que se considerará aceptable si las lámparas obtenidas por este método dan lugar a una población normal de duración medias 2400 horas, con una desviación típica igual a 300. Se toma una muestra de 100 lámparas producidas por este método y esta muestra tiene una duración media de 2320 horas. ¿Se puede aceptar la hipótesis de validez del nuevo proceso de fabricación con un riesgo igual o menor al 5%?

2.7.10. El control de calidad una fábrica de pilas y baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Hasta ahora el tiempo de duración en conversación seguía una distribución normal con media 300 minutos y desviación típica 30 minutos. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que de una muestra de 60 baterías el tiempo medio de duración en conversación fue de 290 minutos. Suponiendo que ese tiempo sigue siendo Normal con la misma desviación típica: ¿Se puede concluir que las sospechas del control de calidad son ciertas a un nivel de significación del 2%?

2.8. Notación, manejo y ejercicios de aplicación de la distribución t de student

2.8.1. Manejo de tabla

DISTRIBUCION t STUDENT	
a. $P[t_{12} \geq 1.782]$	e. $t_{0.10;25}$
b. $P[t_n \geq 1.10] = 0.85$ n=?	f. $t_{\alpha;10} = 0.397$ ∝ = ?
c. $P[t_{23} \leq 2.50]$	g. $t_{0.25;n} = 0.684$ n=?
d. $P[0.85 \leq t_{18} \leq 1.50]$	h. $t_{\alpha;17} = 1.50$ ∝ = ?

2.8.2. Una empresa indica en un paquete se arroz que el peso medio del paquete es de 900 gramos. En un inspección hemos analizado el peso en gramos de 10 paquetes de arroz y hemos obtenido los siguientes datos:

890 901 893 893 896 895
895 904 899 894

Cual es la probabilidad de que la distancia entre la media poblacional y la media muestral sea mayor de 3 gramos?

2.8.3. Hemos hecho una encuesta entre los hombres de una población determinada y, a partir de los resultados, deducimos que el peso de los hombres de esta población sigue una distribución normal de media 72 kg. Para saber si los datos que hemos obtenido son fiables, pesamos a cuatro de los encuestados y obtenemos una media de 77,57 kg, con una desviación típica de 3,5 kg. ¿Tenemos suficientes motivos para pensar que los encuestados han mentido cuando nos han dicho su peso?

2.9. Notación, manejo y ejercicios de aplicación de la distribución Chi cuadrado

2.9.1. Manejo de tabla

DISTRIBUCION CHI CUADRADO	
a. $P[\chi^2_{12} \geq 11.34]$	e. $\chi^2_{0.20;15}$
b. $P[\chi^2_n \geq 12.34] = 0.50$ n=?	f. $\chi^2_{\alpha;18} = 36$ ∝=?
c. $P[\chi^2_{20} \geq k] = 0.10$ k=?	g. $\chi^2_{0.25;n} = 21.92$ n=?
d. $P[8.17 \leq \chi^2_{12} \leq 11.34]$	h. $\chi^2_{0.10;12}$

2.9.2. Se sabe por los datos censales que la variabilidad de la altura de alumnos de una clase medida a través de la varianza es de 15,3. No obstante, para estudiar la variabilidad en el muestreo de la varianza muestral se decide tomar una m.a.s. de 15 alumnos. ¿Cuál es la probabilidad de que la varianza muestral sea mayor que 15?. Nota: Suponga que la estatura es una variable aleatoria normalmente distribuida.

2.10. Notación, manejo y ejercicios de aplicación de la distribución F Snedecor

2.10.1. Manejo de tabla

DISTRIBUCION F SNEDECOR	
$P\left[F_{10:8} \geq 3.07\right]$	$F_{0.05:30,60}$
$P\left[F_{10:20} \leq 2.77\right]$	$F_{0.95:60,30}$
$P\left[F_{8:10} \geq 3.35\right]$	$F_{\alpha:60,\infty} = 1.39$ ∝=?

2.10.2. Se desea analizar las diferencias de calificaciones entre dos grupos de alumnos. Unos proceden del grupo 22 y otros del grupo 23. Para estudiar la distribución en el muestreo de la diferencia de medias se toman muestreo aleatorio simple. independientes de ambas poblaciones obteniéndose la siguiente tabla:

DATOS	GRUPO 22	GRUPO 23
Tamaño de la población	200	150
Tamaño de la muestra	100	75
Media de la población	4,10	5,18
Media de la muestra	4,2153	5,3247
Desviación típica de la población	1,55	1,95
Desviación típica de la muestra	1,5635	1,8238

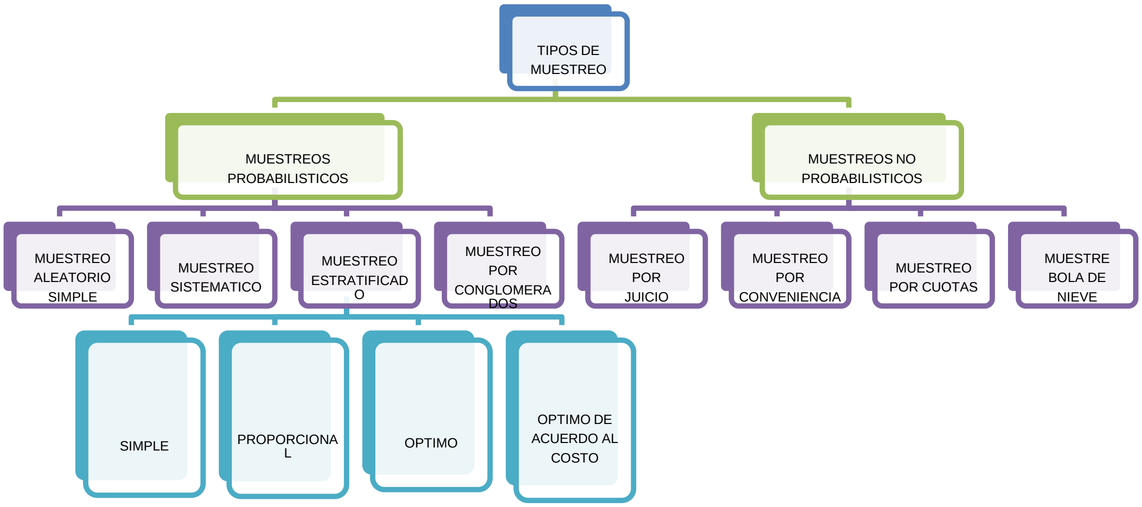
¿Cuál debe ser cociente de sus varianzas para afirmar en un 95% que el grupo 23 tiene mayor variabilidad que el grupo 24?

2.11. Muestreo

2.11.1. Cálculos de tamaños de muestra

PARA MEDIAS	PARA PROPORCIONES
a. Media poblacional = μ b. Desviación est. poblacional = σ c. Nivel de confianza = 1 - ∝ d. $Z_{\propto/2}$ o $t_{\propto/2;\ n-1}$ e. Error muestral = $\overline{X}$ - μ = β $\overline{X}$	a. Proporción poblacional = P b. Nivel de confianza = 1 - ∝ c. $Z_{\propto/2}$ o $t_{\propto/2;\ n-1}$ d. Error relativo = β
Tamaño de muestra inicial $n' = \frac{Z_{\alpha/2}^{2} S^{2}}{e^{2}} = \frac{Z_{\alpha/2}^{2} CV^{2}}{\beta^{2}}$	**Tamaño de muestra inicial** $n' = \frac{Z_{\alpha/2}^{2} P(1-P)}{\beta^{2}}$
Si se conoce el tamaño de la población $n = \frac{n'}{1 + \frac{n'}{N}}$	

2.11.2. Tipos de muestreo



2.12. EJERCICIOS DE COMPLEMENTO

2.12.1. En un lote de frascos de medicina con una población de 8000 unidades. Se desea estimar la media de la capacidad en centímetros cúbicos de los mismos. A través de una prueba piloto se ha estimado que la desviación estándar es de 2 centímetros cúbicos. Si queremos obtener una precisión de 0.25 centímetros cúbicos, con un nivel de significancia del 5%. De que tamaño se debe tomar la muestra?

2.12.2. Un estudiante de ingeniería desea determinar el ingreso medio de los miembros del consejo urbano que cuenta con 900 personas. Desea un error de estimación inferior a 100 dólares, con un nivel de confianza del 90%. El estudiante encontró un informe presentado por el departamento de trabajo que estima la desviación estándar en 1000 dólares. Cual debe ser el tamaño de muestra requerido para este estudio?

2.12.3. Un fabricante de pilas alcalinas sabe que el tiempo de duración en horas de las pilas que fabrica sigue una distribución normal con media desconocida y varianza 3600. Con una muestra de producción elegida al azar y un nivel de confianza del 95% ha obtenido para la media el siguiente intervalo de confianza de 372.6 a 392.2. a. Calcule el valor que obtuvo para la media de la muestra y el tamaño muestral utilizado? b. Cual sería el error de su estimación; si hubiese utilizado un tamaño de muestra de 225 y un nivel de confianza del 90&

2.12.4. Queremos ajustar una máquina de refrescos de modo que el promedio del líquido dispensado quede dentro de cierto rango. La cantidad de líquido vertido por la máquina sigue una distribución normal con desviación estándar 0.25 decilitros. Deseamos que el valor estimado que se vaya a obtener comparado con el verdadero no sea superior a 0.02 decilitros con una confianza del 95%.¿De qué tamaño debemos escoger la muestra?

2.12.5. Es necesario estimar entre 20.000 establos, el número de vacas lecheras por establo con un error de estimación de 6 y un nivel de confianza del 95%. Sabemos que la varianza es 1.400. ¿Cuántos establos deben visitarse para satisfacer estos requerimientos?

2.12.6. La siguiente es una muestra piloto de 17 familias de un total de 1200. Se está pesquisando: el tamaño de la familia, el ingreso mensual y si ven o no un determinado programa de TV.

CODIGO	**TAMAÑO**	**INGRESO MENSUAL**	**PREFERENCIA TV.**
1	2	3.250.000	SI
2	3	1.800.000	NO
3	3	1.100.000	NO
4	5	8.500.000	NO
5	4	950.000	SI
6	7	3.000.000	SI
7	2	2.800.000	NO
8	4	6.000.000	NO
9	2	2.500.000	SI
10	5	2.500.000	NO
11	3	1.900.000	NO
12	6	2.600.000	NO
13	4	7.000.000	SI
14	4	1.100.000	NO
15	2	2.000.000	NO
16	5	1.800.000	NO
17	3	1.200.000	SI

Determine un tamaño de muestra para cada uno de los items, si se desea un nivel de confianza del 95,5 % y un error del 5 %.

2.12.7. Suponga que dispone de la siguiente información en cuatro estratos:

Estrato 1	**Estrato 2**	**Estrato 3**	**Estrato 4**
6	50	22	2
4	16	47	13
33	37	7	14
8	34	21	19
25	39	49	27
23	11	29	40
35	44	17	43
12	3	38	
36	45	10	
24	32	30	
9	1	18	
42	41		
26	15		
46	20		
5	28		
31			
48			

Con base en los datos de la tabla; saque una muestra aleatoria de tamaño 12: teniendo en cuenta el método de muestreo:

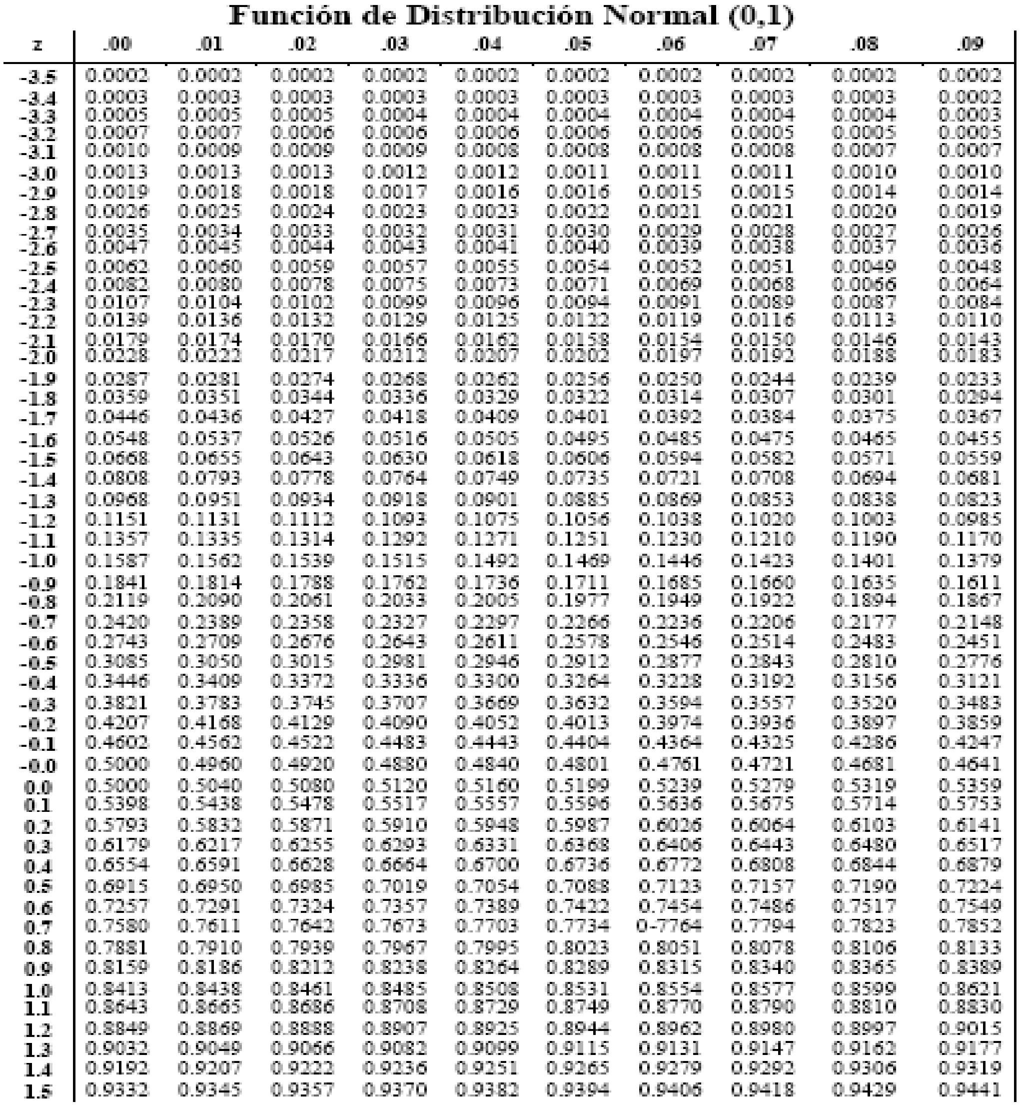
a. Aleatorio simple
b. Sistemático
c. Estratificado
    - Simple
    - Estratificado
    - Optimo
d. Por conglomerados

2.12.8. Explique y describa un caso de aplicación real de un muestreo no probabilistico:

a. Por conveniencia
b. Por juicio
c. Por cuotas
d. Bola de nieve

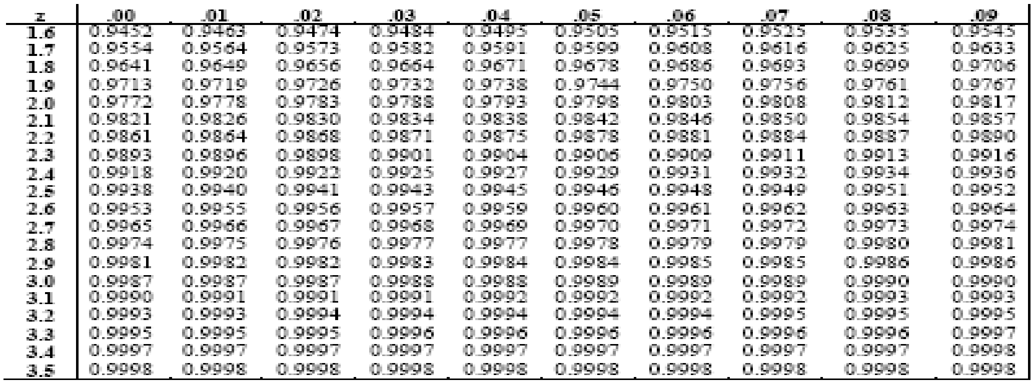
2.13. Tablas estadísticas a utilizar

2.13.1. Distribución normal



Función de Distribución Normal (0,1)

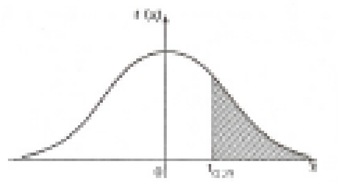
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0-7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

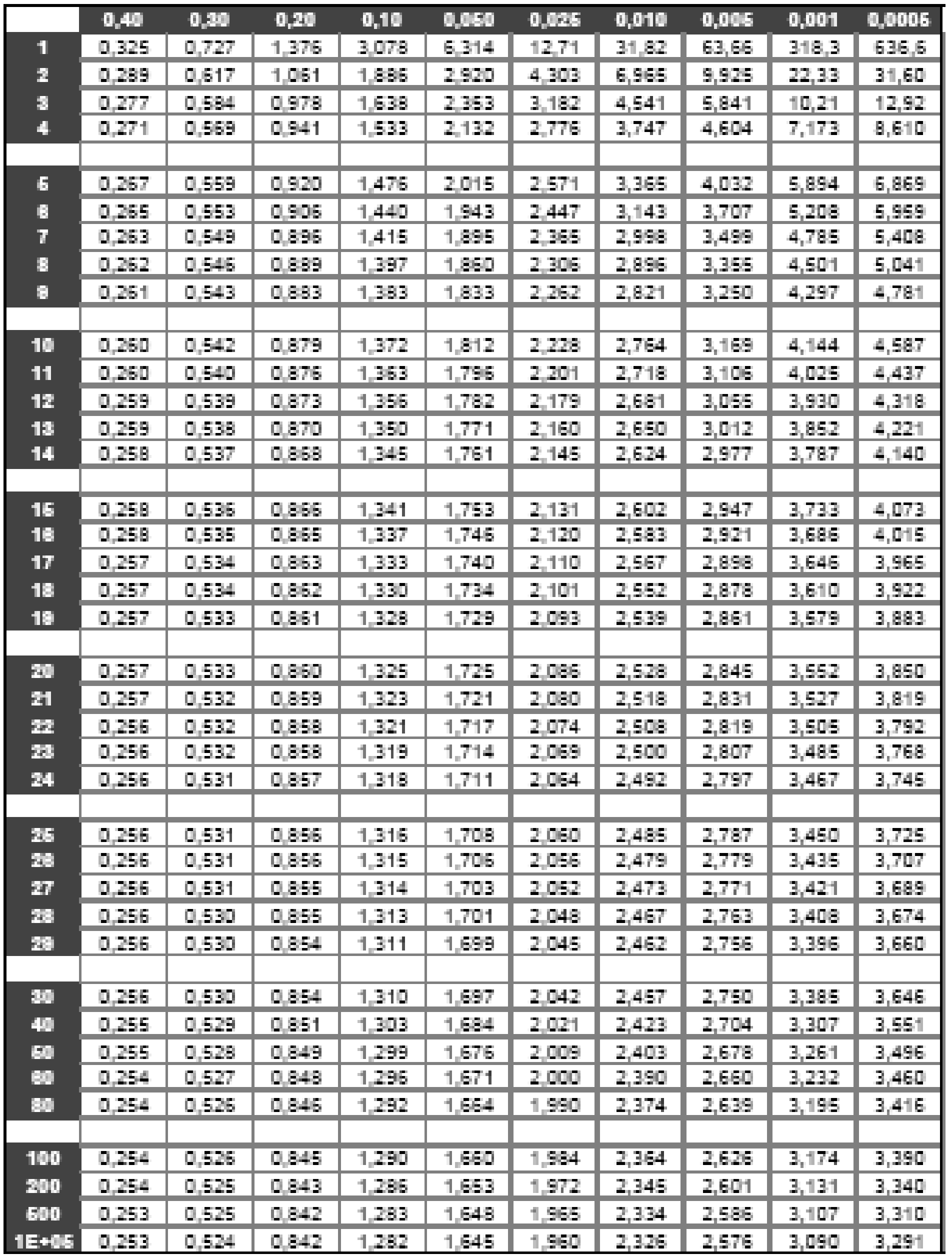


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

2.13.2. Distribución T de Student

Distribución t de Student

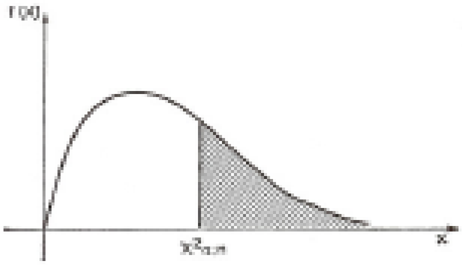


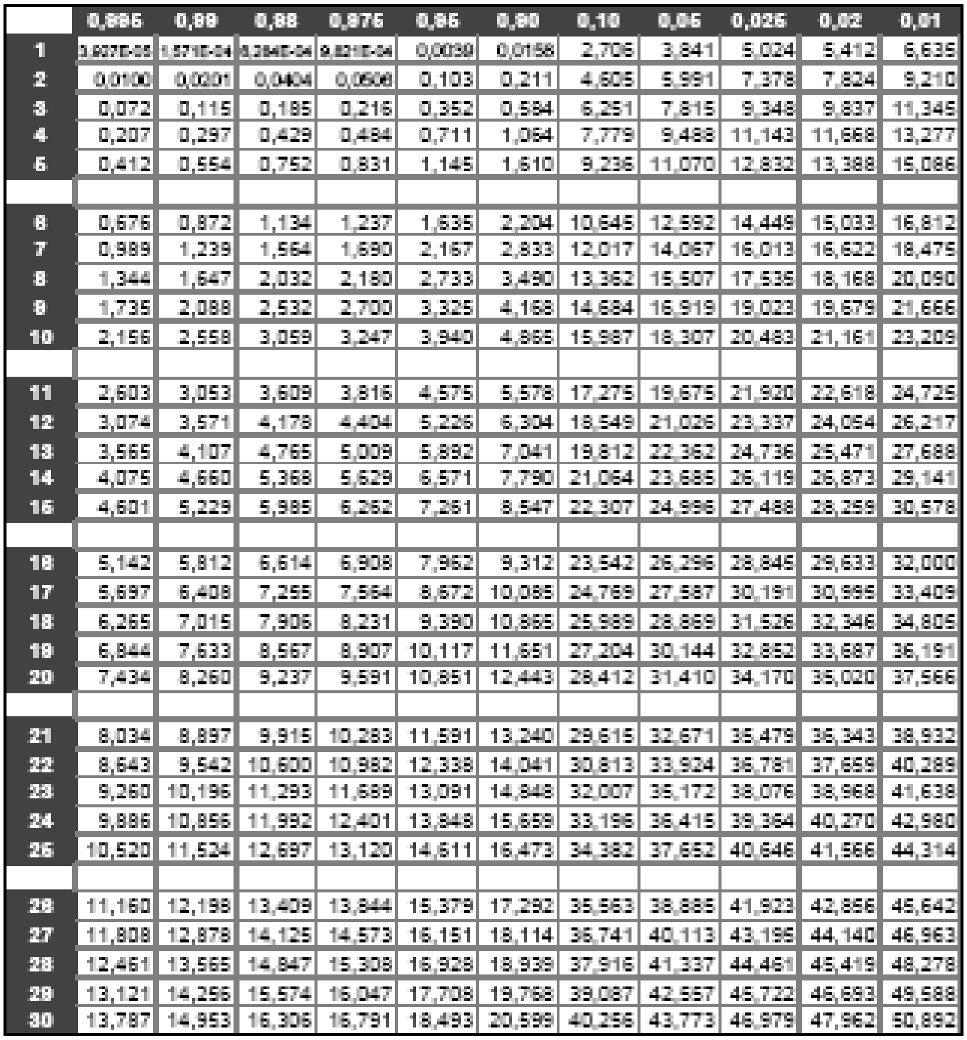


	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,21	12,92
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,660
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,299	1,676	2,009	2,403	2,678	3,261	3,496
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,526	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,254	0,526	0,845	1,290	1,660	1,984	2,364	2,626	3,174	3,390
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,340
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,107	3,310
1E+06	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

2.13.3. Distribución Chi cuadrado

Distribución χ^2

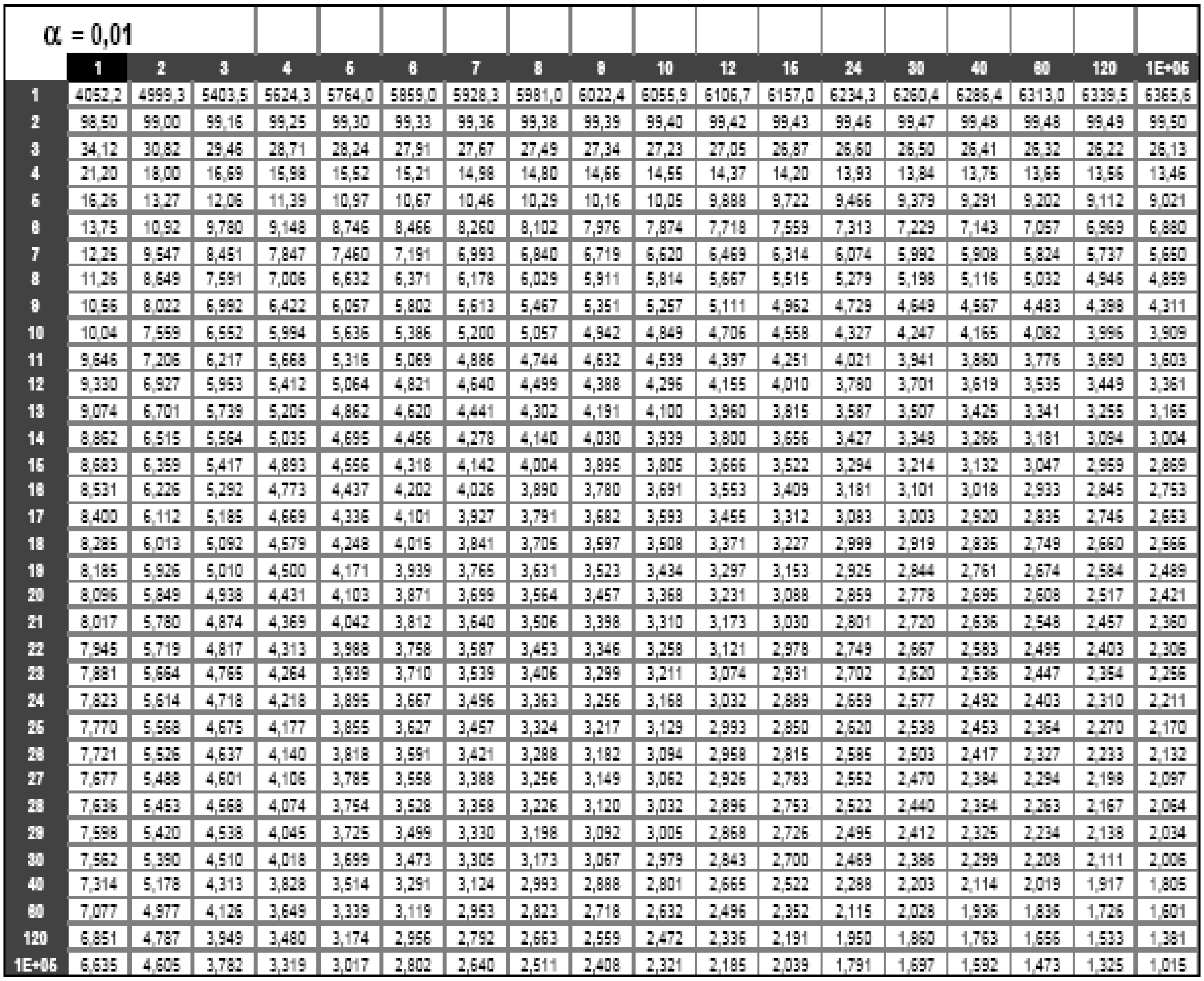




	0,995	0,99	0,98	0,975	0,95	0,90	0,10	0,05	0,025	0,02	0,01
1	3,927E-05	1,571E-04	6,284E-04	9,821E-04	0,0039	0,0158	2,706	3,841	5,024	5,412	6,635
2	0,0100	0,0201	0,0404	0,0506	0,103	0,211	4,605	5,991	7,378	7,824	9,210
3	0,072	0,115	0,185	0,216	0,352	0,584	6,251	7,815	9,348	9,837	11,345
4	0,207	0,297	0,429	0,484	0,711	1,064	7,779	9,488	11,143	11,668	13,277
5	0,412	0,554	0,752	0,831	1,145	1,610	9,236	11,070	12,832	13,388	15,086
6	0,676	0,872	1,134	1,237	1,635	2,204	10,645	12,592	14,449	15,033	16,812
7	0,989	1,239	1,564	1,690	2,167	2,833	12,017	14,067	16,013	16,622	18,475
8	1,344	1,647	2,032	2,180	2,733	3,490	13,362	15,507	17,535	18,168	20,090
9	1,735	2,088	2,532	2,700	3,325	4,168	14,684	16,919	19,023	19,679	21,666
10	2,156	2,558	3,059	3,247	3,940	4,865	15,987	18,307	20,483	21,161	23,209
11	2,603	3,053	3,609	3,816	4,575	5,578	17,275	19,675	21,920	22,618	24,725
12	3,074	3,571	4,178	4,404	5,226	6,304	18,549	21,026	23,337	24,054	26,217
13	3,565	4,107	4,765	5,009	5,892	7,041	19,812	22,362	24,736	25,471	27,688
14	4,075	4,660	5,368	5,629	6,571	7,790	21,064	23,685	26,119	26,873	29,141
15	4,601	5,229	5,985	6,262	7,261	8,547	22,307	24,996	27,488	28,259	30,578
16	5,142	5,812	6,614	6,908	7,962	9,312	23,542	26,296	28,845	29,633	32,000
17	5,697	6,408	7,255	7,564	8,672	10,085	24,769	27,587	30,191	30,995	33,409
18	6,265	7,015	7,906	8,231	9,390	10,865	25,989	28,869	31,526	32,346	34,805
19	6,844	7,633	8,567	8,907	10,117	11,651	27,204	30,144	32,852	33,687	36,191
20	7,434	8,260	9,237	9,591	10,851	12,443	28,412	31,410	34,170	35,020	37,566
21	8,034	8,897	9,915	10,283	11,591	13,240	29,615	32,671	35,479	36,343	38,932
22	8,643	9,542	10,600	10,982	12,338	14,041	30,813	33,924	36,781	37,659	40,289
23	9,260	10,196	11,293	11,689	13,091	14,848	32,007	35,172	38,076	38,968	41,638
24	9,886	10,856	11,992	12,401	13,848	15,659	33,196	36,415	39,364	40,270	42,980
25	10,520	11,524	12,697	13,120	14,611	16,473	34,382	37,652	40,646	41,566	44,314
26	11,160	12,198	13,409	13,844	15,379	17,292	35,563	38,885	41,923	42,856	45,642
27	11,808	12,878	14,125	14,573	16,151	18,114	36,741	40,113	43,195	44,140	46,963
28	12,461	13,565	14,847	15,308	16,928	18,939	37,916	41,337	44,461	45,419	48,278
29	13,121	14,256	15,574	16,047	17,708	19,768	39,087	42,557	45,722	46,693	49,588
30	13,787	14,953	16,306	16,791	18,493	20,599	40,256	43,773	46,979	47,962	50,892

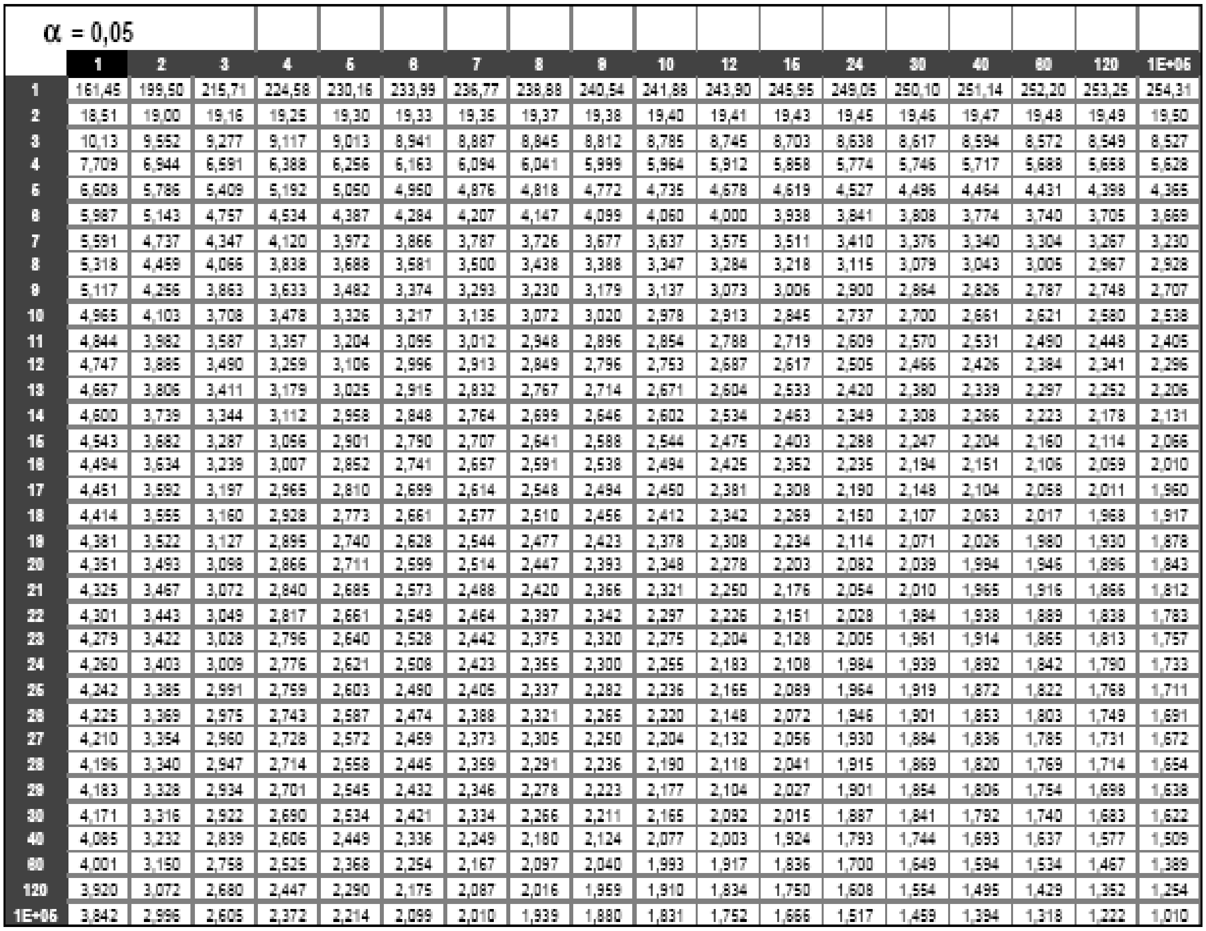
2.13.4. Distribución F de Snedecor

Distribución *F* de Fisher-Snedecor



α = 0,01	1	2	3	4	5	6	7	8	9	10	12	15	24	30	40	60	120	1E+05
1	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5928,3	5981,0	6022,4	6055,9	6106,7	6157,0	6234,3	6260,4	6286,4	6313,0	6339,5	6365,6
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,42	99,43	99,46	99,47	99,48	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,87	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,888	9,722	9,466	9,379	9,291	9,202	9,112	9,021
6	13,75	10,92	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,718	7,559	7,313	7,229	7,143	7,057	6,969	6,880
7	12,25	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,469	6,314	6,074	5,992	5,908	5,824	5,737	5,650
8	11,26	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,667	5,515	5,279	5,198	5,116	5,032	4,946	4,859
9	10,56	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257	5,111	4,962	4,729	4,649	4,567	4,483	4,398	4,311
10	10,04	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,706	4,558	4,327	4,247	4,165	4,082	3,996	3,909
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,632	4,539	4,397	4,251	4,021	3,941	3,860	3,776	3,690	3,603
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388	4,296	4,155	4,010	3,780	3,701	3,619	3,535	3,449	3,361
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	3,960	3,815	3,587	3,507	3,425	3,341	3,255	3,165
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,800	3,656	3,427	3,348	3,266	3,181	3,094	3,004
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805	3,666	3,522	3,294	3,214	3,132	3,047	2,959	2,869
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780	3,691	3,553	3,409	3,181	3,101	3,018	2,933	2,845	2,753
17	8,400	6,112	5,185	4,669	4,336	4,101	3,927	3,791	3,682	3,593	3,455	3,312	3,083	3,003	2,920	2,835	2,746	2,653
18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597	3,508	3,371	3,227	2,999	2,919	2,835	2,749	2,660	2,566
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523	3,434	3,297	3,153	2,925	2,844	2,761	2,674	2,584	2,489
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,231	3,088	2,859	2,778	2,695	2,608	2,517	2,421
21	8,017	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398	3,310	3,173	3,030	2,801	2,720	2,636	2,548	2,457	2,360
22	7,945	5,719	4,817	4,313	3,988	3,758	3,587	3,453	3,346	3,258	3,121	2,978	2,749	2,667	2,583	2,495	2,403	2,306
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299	3,211	3,074	2,931	2,702	2,620	2,536	2,447	2,354	2,256
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256	3,168	3,032	2,889	2,659	2,577	2,492	2,403	2,310	2,211
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129	2,993	2,850	2,620	2,538	2,453	2,364	2,270	2,170
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182	3,094	2,958	2,815	2,585	2,503	2,417	2,327	2,233	2,132
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149	3,062	2,926	2,783	2,552	2,470	2,384	2,294	2,198	2,097
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120	3,032	2,896	2,753	2,522	2,440	2,354	2,263	2,167	2,064
29	7,598	5,420	4,538	4,045	3,725	3,499	3,330	3,198	3,092	3,005	2,868	2,726	2,495	2,412	2,325	2,234	2,138	2,034
30	7,562	5,390	4,510	4,018	3,699	3,473	3,305	3,173	3,067	2,979	2,843	2,700	2,469	2,386	2,299	2,208	2,111	2,006
40	7,314	5,178	4,313	3,828	3,514	3,291	3,124	2,993	2,888	2,801	2,665	2,522	2,288	2,203	2,114	2,019	1,917	1,805
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,496	2,352	2,115	2,028	1,936	1,836	1,726	1,601
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559	2,472	2,336	2,191	1,950	1,860	1,763	1,656	1,533	1,381
1E+05	6,635	4,605	3,782	3,319	3,017	2,802	2,640	2,511	2,408	2,321	2,185	2,039	1,791	1,697	1,592	1,473	1,325	1,015

Distribución *F* de Fisher-Snedecor



α = 0,05

	1	2	3	4	5	6	7	8	9	10	12	15	24	30	40	60	120	1E+06
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,90	245,95	249,05	250,10	251,14	252,20	253,25	254,31
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,745	8,703	8,638	8,617	8,594	8,572	8,549	8,527
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,858	5,774	5,746	5,717	5,688	5,658	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,619	4,527	4,496	4,464	4,431	4,398	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000	3,938	3,841	3,808	3,774	3,740	3,705	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575	3,511	3,410	3,376	3,340	3,304	3,267	3,230
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	3,284	3,218	3,115	3,079	3,043	3,005	2,967	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,006	2,900	2,864	2,826	2,787	2,748	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,845	2,737	2,700	2,661	2,621	2,580	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,719	2,609	2,570	2,531	2,490	2,448	2,405
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,617	2,505	2,466	2,426	2,384	2,341	2,296
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,533	2,420	2,380	2,339	2,297	2,252	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,463	2,349	2,308	2,266	2,223	2,178	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,403	2,288	2,247	2,204	2,160	2,114	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,352	2,235	2,194	2,151	2,106	2,059	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,308	2,190	2,148	2,104	2,058	2,011	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,269	2,150	2,107	2,063	2,017	1,968	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,234	2,114	2,071	2,026	1,980	1,930	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,203	2,082	2,039	1,994	1,946	1,896	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321	2,250	2,176	2,054	2,010	1,965	1,916	1,866	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,226	2,151	2,028	1,984	1,938	1,889	1,838	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,204	2,128	2,005	1,961	1,914	1,865	1,813	1,757
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,183	2,108	1,984	1,939	1,892	1,842	1,790	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,165	2,089	1,964	1,919	1,872	1,822	1,768	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,148	2,072	1,946	1,901	1,853	1,803	1,749	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,132	2,056	1,930	1,884	1,836	1,785	1,731	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,118	2,041	1,915	1,869	1,820	1,769	1,714	1,654
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,104	2,027	1,901	1,854	1,806	1,754	1,698	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,092	2,015	1,887	1,841	1,792	1,740	1,683	1,622
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,003	1,924	1,793	1,744	1,693	1,637	1,577	1,509
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,917	1,836	1,700	1,649	1,594	1,534	1,467	1,389
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910	1,834	1,750	1,608	1,554	1,495	1,429	1,352	1,254
1E+06	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,939	1,880	1,831	1,752	1,666	1,517	1,459	1,394	1,318	1,222	1,010

XX. ESTIMACION

1.1. PROPIEDADES DE LOS ESTIMADORES

- **a.** **Estimador insesgado:** Si tenemos un gran número de muestras de tamaño n y obtenemos el valor del estimador en cada una de ellas, sería deseable que la media de todas estas estimaciones coincidiera con el valor de μ. Se dice que un estimador es insesgado si su esperanza matemática coincide con el valor del parámetro a estimar.
- **b.** **Estimador eficiente:** Se dice que los estimadores son eficientes cuando generan una distribución muestral con el mínimo error estándar, es decir, entre dos estimadores insesgados de un parámetro dado es más eficiente el de menor varianza.
- **c.** **Estimador consistente:** Un estimador se dice consistente cuando su valor tiende hacia el verdadero valor del parámetro a medida que aumenta el tamaño de la muestra. Es decir, la probabilidad de que la estimación sea el verdadero valor del parámetro tiende a 1.
- d. **Estimador suficiente:** Se dice de un estimador que es suficiente cuando es capaz de extraer de los datos toda la información importante sobre el parámetro.

1.2. La variable aleatoria “renta de las familias” del municipio de Cali se distribuye siguiendo un modelo N (μ,σ). Se extraen muestras aleatorias simples de tamaño 4. Como estimadores del parámetro μ, se proponen los siguientes:

$$\mu_1^* = \frac{x_1 + 2x_2 + 3x_3}{6} \qquad \mu_2^* = \frac{x_1 - 4x_2}{-3} \qquad \mu_3^* = \overline{x}$$

Se pide:

- a. Comprobar si los estimadores son insesgados
- b. ¿Cuál es el más eficiente?
- c. Si tuviera que escoger entre uno de ellos, ¿cuál escogería? Razone su respuesta a partir del Error Cuadrático Medio.

1.3. ESTIMACIÓN PUNTUAL

Una estimación puntual es un solo valor que se mide a partir de una muestra y se usa como un estimacion del parámetro poblacional correspondiente

1.3.1. Los siguientes datos corresponden a los pesos (en kilogramos) de 15 hombres escogidos al azar y que trabajan en una empresa: 72, 68, 63, 75, 84, 91, 66, 75, 86, 90, 62, 87, 77, 70, 69. Estime el peso promedio, la desviación estándar y el error estandar del peso promedio.

1.3.2. Entre los miembros de una comunidad se escogieron 150 personas al azar y se les preguntó si estaban de acuerdo con los programas que el gobierno estaba desarrollando para prevenir el consumo de drogas; la encuesta dio como resultado que 130 sí estaban de acuerdo. Estime la proporción de los que estaban de acuerdo y el error estándar.

1.3.3. De las 50 aulas que tiene un edificio de la facultad de matemáticas se escogieron al azar 5 y se determinó el número de alumnos que había en cada una de ellas en la primera hora de clases. Estime el número de alumnos que hay en el edificio si todas las aulas se encuentran ocupadas a esa hora, y si el número de alumnos en cada una de las aulas inspeccionadas fue: 24, 35, 16, 30, 28. El error del numero total de estudiantes

1.3.4. Un puesto de hamburguesas a $2500 cada una. Las ventas tienen una distribución normal con una media de 530 y una desviación estándar de 69

a. Hallar la media de los ingresos diarios por venta de hamburguesas
b. Hallar la desviación estándar de los ingresos diarios por la venta de hamburguesas
c. Los costos en pesos vienen dados por: C = 60 + 0.58 X; X es el numero de hamburguesas vendidas. Cual es la ganancia esperada?
d. Con un 95% de confianza. Que cifra superara diariamente las ganancias?

1.3.5. De una población se escogieron al azar 10 personas y se les tomo la estatura. Los resultados en cm fueron: 160, 170, 170, 150, 160, 180, 160, 170, 130, 150. Estime la media y la varianza.

1.3.6. En una universidad se desea conocer la opinión de los estudiantes acerca de ciertas medidas que han tomado las directivas. De 120 estudiantes consultados, 90 estuvieron a favor. Estime la proporción de estudiantes que están a favor de las medidas.(Sol. 75%).

1.3.7. Un conjunto residencial está formado por 200 apartamentos. Se seleccionaron 18 apartamentos y se observó que, en promedio, viven 4´5 personas por apartamento. Estime el total de personas que viven en el conjunto residencial. (Sol. 900 personas).

1.3.8. De un lote de 1.000 licuadoras se escogen aleatoriamente 30 y se encontró que 2 de ellas estaban estropeadas;¿cuántas licuadoras se estima que estén estropeadas? (Sol. 67 licuadoras).

1.3.9. Una agencia de encuesta selecciona 900 familias y calcula la proporción de éstas que utilizan cierto tipo de detergente. Si la proporción estimada es 0´35 ¿Cuál es el error estándar estimado? (Sol. 0´016).

1.3.10. En el estudio de cierta característica X de una población se sabe que la desviación estándar es 3. Se va a escoger una muestra de tamaño 100, halle el error estándar de la media muestral. (Sol. 0´3).

1.3.11. Se escogió al azar una muestra de 10 clientes de un banco y se les preguntó el número de veces que habían utilizado el banco para llevar a cabo alguna transacción comercial. Los resultados fueron los siguientes: 0, 4, 2, 3, 2, 0, 3, 4, 1, 1. Estime el error estándar del número de transacciones promedio. (Sol. 0´47).

1.3.12. Pedro candidato líder en la carrera por la gobernación desea desarrollar un intervalo de confianza con un ancho de 3 puntos porcentuales y un nivel de confianza del 99% para hallar la diferencia entre la proporción de hombres y mujeres que están a favor de su candidatura. Que tan grandes deberán tomarse las muestras, debido a que una prueba pilote revelo que el 40% de los hombres y el 30% de las mujeres están a favor de su candidatura. Si el ancho del intervalo dseado es de un 3%(0.03/2).

1.4. ESTIMACIÓN POR INTERVALOS

La estimación por intervalos establece un intervalo dentro del cual es muy probable que se encuentre el parámetro poblacional. **El coeficiente de confianza** se usa para indicar la probabilidad de que una estimación por intervalo contenga el parámetro poblacional. **El nivel de confianza** es el coeficiente de confianza expresado como porcentaje

1.4.1. Notación utilizada

1.4.1.1. Estimación para la media, con una muestra

DESCRIPCIÓN	INTERVALO DE CONFIANZA

Estimación de μ con sigma conocida, muestra grande n>30	$P\left[\overline{X}-\lvert Z_{\alpha/2}\rvert\left(\frac{\sigma}{\sqrt{n}}\right)\le\mu\le\overline{X}+\lvert Z_{\alpha/2}\rvert\left(\frac{\sigma}{\sqrt{n}}\right)\right]=1-\alpha$

Ejercicio 1. Un proceso produce bolsas de azúcar refinada. El peso del contenido de estas bolsas tiene una distribución normal con una desviación típica de 15 gramos. Los contenidos de una muestra aleatoria de 36 bolsos tienen un peso medio de 100 gramos. Calcule e interprete un intervalo de confianza del 95% para el verdadero peso medio de todas las bolsas de azúcar producidas por el proceso?

Estimación de μ con sigma desconocida, muestra grande n>30, se toma la desviación estándar. de la muestra S	$P\left[\overline{X}-\lvert Z_{\alpha/2}\rvert\left(\frac{S}{\sqrt{n}}\right)\le\mu\le\overline{X}+\lvert Z_{\alpha/2}\rvert\left(\frac{S}{\sqrt{n}}\right)\right]=1-\alpha$

Ejercicio 2. Para estimar el gasto promedio de los clientes en el Mc Donald local, los estudiantes de una clase de estadística tienen una muestra de 200 clientes y encuentran un gasto promedio de 5.67 dólares, con una desviación estándar de 1.10 dólares. Calcule e interprete un intervalo de confianza al 99% para estimar los gastos promedios de todos los clientes de Mc Donald?

Estimación de μ con muestras pequeñas, n < 30 y sigma desconocida	$P\left[\overline{X}-\lvert t_{\alpha/2;n-1}\rvert\left(\frac{S}{\sqrt{n}}\right)\le\mu\le\overline{X}+\lvert t_{\alpha/2;n-1}\rvert\left(\frac{S}{\sqrt{n}}\right)\right]=1-\alpha$

1.4.1.2. Estimación para la varianza , con una muestra

Ejercicio 3. Se desea investigar el consumo medio de gasolina de los autos de cierto modelo. Para ello se toma una muestra aleatoria de 6 coches de un determinado modelo y se analiza su consumo en kilómetros por litro. Los resultados son los siguientes:

18.6	18.4	19.2	20.8	19.4	20.5

Calcule e interprete un intervalo de confianza del 95% para estimar el consumo medio de todos los vehículos de este tipo?

Estimación de la σ	$P\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}\le\sigma^2\le\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}\right]=1-\alpha$

1.4.1.3. Estimación para proporciones , con una muestra

Ejercicio 4. Un trabajador tiene como cometido cubrir con una capa de plástico una superficie metálica. Se extrae una muestra aleatoria de 9 observaciones correspondientes al grosor de dicha capa plástica de valores obtenidos en milímetros. La cual arrojo los siguientes resultados:

19.8	21.2	18.6	20.4	21.6	19.8	19.9	30.3	20.8

Suponiendo que la población es normal. Halle un intervalo de confianza del 90%. Para la varianza poblacional?

Estimación de la proporción π	$P\left[\overline{P}-\left\|Z_{\alpha/2}\right\|\left(\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}\right)\le P\le \overline{P}+\left\|Z_{\alpha/2}\right\|\left(\sqrt{\frac{\overline{P}(1-\overline{P})}{n}}\right)\right]=1-\alpha$ $P\left[\overline{P}-\left\|t_{\alpha/2;n-1}\right\|\left(\sqrt{\frac{P(1-P)}{n}}\right)\le P\le \overline{P}+\left\|t_{\alpha/2;n-1}\right\|\left(\sqrt{\frac{P(1-P)}{n}}\right)\right]=1-\alpha$

Ejercicio 5. De una muestra aleatoria de 95 empresas pequeñas fabricantes. 29 de estas señalaron que lo mejor es la calidad como lo más importante para incrementar la competitividad de sus productos. Calcule e interprete un intervalo de confianza del 99% para la verdadera proporción de empresas que indican que la calidad es la que incrementa la competitividad?

Ejercicio 6. CNN informo que el 68% de todos los estudiantes de secundaria tienen computadores en sus casas. Si una muestra de 1020 estudiantes revela que 613 de estos tienen computadores en sus casas. Con un intervalo de confianza del 99% , que puede decir de la afirmación de CNN?

Ejercicio 7. CNN informo que el 68% de todos los estudiantes de secundaria tienen computadores en sus casas. Si una muestra de 30 estudiantes revela que 18 de estos tienen computadores en sus casas. Con un intervalo de confianza del 99% , que puede decir de la afirmación de CNN?

1.4.1.4. Estimación de diferencia de medias para dos muestras relacionadas

Diferencia de medias relacionadas	$P\left[\overline{d}-\left\|t_{\alpha/2;n-1}\right\|\left(\frac{S_d}{\sqrt{n}}\right)\le \mu_A-\mu_D\le \overline{d}+\left\|t_{\alpha/2;n-1}\right\|\left(\left(\frac{S_d}{\sqrt{n}}\right)\right)\right]=1-\alpha$

Ejercicio 8. Se extrajo una muestra aleatoria de 12 directivos intermedio de una compañía que asistió a un curso de capacitación en técnicas de modelos de dirección. Las puntuaciones que obtuvieron en evaluaciones por sus inmediatos superiores el año anterior y posterior a la realización de los cursos son las siguientes:

Antes	69	54	82	67	60	73	73	50	83	78	56	74
Después	75	78	64	72	70	63	74	87	69	72	77	74

Calcule e interprete un intervalo de confianza al 95% para estimar las verdaderas diferencias?

1.4.1.5. Estimación de diferencia de proporciones para dos muestras relacionadas

<table>
<tr>
<td>Diferencia de proporciones relacionadas</td>
<td>
<table>
<tr><td></td><td></td><td colspan="2">DESPUES</td><td></td></tr>
<tr><td></td><td></td><td>PRO</td><td>ANTI</td><td></td></tr>
<tr><td rowspan="2">ANTES</td><td>PRO</td><td>A</td><td>B</td><td>A+B</td></tr>
<tr><td>ANTI</td><td>D</td><td>C</td><td>D+C</td></tr>
<tr><td></td><td></td><td>A+D</td><td>B+C</td><td>1</td></tr>
</table>

$$P\left[(B-D)-\left|Z_{\alpha/2}\right|\left(\sqrt{\frac{B+D}{n^2}}\right)\le \mu_A-\mu_D\le (B-D)+\left|Z_{\alpha/2}\right|\left(\sqrt{\frac{B+D}{n^2}}\right)\right]=1-\alpha$$
</td>
</tr>
</table>

Ejercicio 9. Se tiene la siguiente información sobre la posibilidad de vender sus casas los habitantes de una pequeña ciudad después de un movimiento telúrico? Y para ello se cuneta con la siguiente información:

		DESPUES vender		
		Si	No	
ANTES Vender	Si	21%	13%	34%
	No	52%	14%	66%
		73%	27%	

Calcule e interprete un intervalo de confianza del 95% para las verdaderas diferencias?

1.4.1.6. Estimación de diferencia de medias para dos muestras independientes

<table>
<tr>
<td>Diferencia de medias independientes</td>
<td>

$$P\left[(\overline{X}_1-\overline{X}_2)-\left|Z_{\alpha/2;}\right|\left(S_{X_1-X_2}\right)\le \mu_1-\mu_2\le (\overline{X}_1-\overline{X}_2)+\left|Z_{\alpha/2;}\right|\left(S_{X_1-X_2}\right)\right]=1-\alpha$$

Para muestras grandes: $S_{X_1-X_2}=\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

Para muestras pequeñas: $S_{X_1-X_2}=\sqrt{\left(\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
</td>
</tr>
</table>

Ejercicio 10. Para una muestra de 96 fumadores de una empresa se encontró que el promedio de ausentismo laboral al mes fue de 2.15 con una desviación estándar de 2.09 horas al mes. Para una muestra de 206 trabajadores que nuca han fumado el numero promedio fue de 1.69, con una desviación estándar de 1.91 horas al mes. Calcule e interprete un intervalo de confianza del 99% para estimar la diferencia entre las dos poblaciones?

Ejercicio 11. En un estudio sobre los efectos de planificación en el rendimiento financiero de los bancos se extrajo una muestra aleatoria de seis instituciones financieras que contribuyen con el sistema de planificación formal y se encontró que el porcentaje medio de sus ingresos era del 9.97 y una desviación estándar de 7.47. Y la media de dicho crecimiento en otra muestra aleatoria que no recurre a la planificación fue de 2.08, con una desviación estándar de 10.83. Calcule e interprete un intervalo de confianza del 90% para estimar la verdadera diferencia?

1.4.1.7. Estimación de diferencia de proporciones para dos muestras independientes

Diferencia de proporciones independientes	$P\left[(\overline{P_1}-\overline{P_2})-\lvert Z_{\alpha/2;}\rvert\left(S_{P_1-P_2}\right)\le P_1-P_2\le(\overline{P_1}-\overline{P_2})+\lvert Z_{\alpha/2;}\rvert\left(S_{P_1-P_2}\right)\right]=1-\alpha$ Para muestras grandes: $S_{P_1-P_2}=\sqrt{\frac{P_1(1-P_1)}{n_1}+\frac{P_2(1-P_2)}{n_2}}$ Para muestras pequeñas: $S_{P_1-P_2}=\sqrt{\left(\frac{n_1P_1+n_2P_2}{n_1+n_2}\right)\left(1-\frac{n_1P_1+n_2P_2}{n_1+n_2}\right)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Ejercicio 12. De una muestra de 569 estudiantes de género masculino de un curso de contabilidad; 90 abandonaron a la mitad del curso y de una muestra de 567 mujeres de dicho cursos, 85 fueron las que lo dejaron. Calcule e interprete un intervalo de confianza del 95% para la verdadera diferencia poblacional.

Ejercicio 13. Una muestra de loa estudiantes de estadística de la universidad para dos grupos de estadística de distinto profesor, mostró los siguientes resultados de acuerdo a los que ganaron la asignatura. De una muestra de 12 estudiantes del profesor a, 8 ganaron la asignatura y de una muestra de 18 estudiantes del profesor B; 9ganaron la asignatura. Calcule e interprete un intervalo de confianza del 95% para la verdadera diferencia de la población?

Ejercicio 14. Unos grandes almacenes envían propaganda de sus colecciones de moda regularmente. Quieren comprobar si merece la pena enviarla en color, frente al estándar en blanco y negro que utilizaban. Se hace una prueba enviando 1000 sobres con folletos en blanco y negro y 1000 en color. Se reciben pedidos por correo de un 10% y un 12% de los clientes, respectivamente. Construya un intervalo de confianza para la diferencia de proporciones poblacionales. para un nivel de confianza del 95%.

1.4.1.8. Estimación de diferencia de varianzas para dos muestras indepedientes

Diferencia de varianzas	$P\left[\frac{S_1^2}{S_2^2}F_{\alpha;n_1,n_2}\le\frac{\sigma_2^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}F_{1-\alpha;n_1,n_2}\right]=1-\alpha$

Ejercicio 15. Para una muestra de 96 fumadores de una empresa se encontró que el promedio de ausentismo laboral al mes fue de 2.15 con una desviación estándar de 2.09 horas al mes. Para una muestra de 206 trabajadores que nuca han fumado el numero promedio fue de 1.69, con una desviación estándar de 1.91 horas al mes. Calcule e interprete un intervalo de confianza del 99% para estimar la diferencia entre las dos poblaciones?

1.5. Ejercicios de complemento

1.5.1. El peso (en gramos) de las cajas de cereales de una determinada marca sigue una distribución N (μ; 5). Se han tomado los pesos de 16 cajas seleccionadas aleatoriamente, y éstos han sido: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509 y 496.

a. Obtenga los intervalos de confianza del 90%, 95% y 99% para la media poblacional.
b. Determine cuál sería el tamaño muestral necesario para conseguir, con un 95% de confianza, un intervalo de longitud igual a 2 gramos.
c. Suponiendo ahora que la desviación típica poblacional es desconocida, calcule los intervalos de confianza para la media al 90%, 95% y 99%.

1.5.2. Se selecciona una muestra aleatoria de 600 familias, a las que se pregunta si tienen o no computador en casa. Contestaron afirmativamente 240 familias. Obtenga un intervalo de confianza al nivel del 95% para la proporción real de familias que poseen computador en casa.

1.5.3. El gasto diario en llamadas telefónicas de dos departamentos X e Y de una misma empresa sigue una distribución normal, con un gasto medio desconocido en ambos departamentos. Sin embargo, se conocen las desviaciones típicas, que son 100 y 110 mil pesos, respectivamente. La dirección ha observado que una muestra aleatoria simple de 20 días, el gasto medio diario en llamadas realizadas por el departamento X ha sido de 1.100 mil pesos, y de 1.400 en el departamento Y. Obtenga un intervalo de confianza al 90% para la diferencia de gastos medios entre ambos departamentos.

1.5.4. Según los dirigentes del partido A, la intención de voto del partido rival B, en Cali, es la misma que la que tiene en Palmira. Se realiza una encuesta a 100 personas en Cali de los que 25 mostraron su apoyo al partido B, y a otras 100 personas en Palmira de las que 30 se declaran simpatizantes del partido B.

a. Construya un intervalo de confianza al 90% para la proporción de personas que votarían al partido B en Cali.
b. ¿A cuántas personas habría que encuestar para tener un margen de error del ±2%, al nivel de confianza anterior?
c. Construya un intervalo de confianza del 90% para la diferencia de proporciones en la intención de voto del partido B en las dos comunidades. ¿Se puede afirmar que los dirigentes del partido A tienen razón?

1.5.5. Una muestra aleatoria de 36 cigarrillos de una marca determinada dio un contenido promedio de nicotina de 3 miligramos. Suponga que el contenido de nicotina de estos cigarrillos sigue una distribución normal con una desviación estándar de 1 miligramo.

a. Obtenga e interprete un intervalo de confianza del 95% para el verdadero contenido promedio de nicotina en estos cigarrillos.
b. El fabricante garantiza que el contenido promedio de nicotina es de 2,9 miligramos, ¿qué puede decirse de acuerdo con el intervalo hallado?

1.5.6. El tiempo (en minutos) que tardaron 15 operarios para familiarizarse con el manejo de una máquina moderna adquirida por la empresa fue: 3,4, 2,8, 4,4, 2,5, 3,3, 4, 4,8, 2,9, 5,6, 5,2, 3,7, 3, 3,6, 2,8, 4,8. Suponga que los tiempos se distribuyen normalmente.

a. Determine e interprete un intervalo del 95% de confianza para el verdadero tiempo promedio
b. el instructor considera que el tiempo promedio requerido por la población de trabajadores que recibe instrucción sobre esta m quina es superior a 5 minutos, ¿qué se puede decir de acuerdo con el intervalo hallado?

1.5.7. Se desea medir la diferencia entre dos categorías de empleados en la actividad de seguros. Una est formada por personas con título superior y la otra por personas que sólo tienen estudios secundarios. Tomamos una muestra de 45 empleados entre los primeros y la media de ventas resulta ser 32. Tomamos 60 empleados del segundo grupo y la media es 25. Suponga que las ventas de los dos grupos se distribuyen normalmente con varianzas de 48 para los titulados superiores y 56 para los de estudios secundarios.

a. Calcule e interprete un intervalo del 90% de confianza para la verdadera diferencia de las medias.
b. De acuerdo con el intervalo hallado, ¿hay evidencia de que las medias sean iguales?

1.5.8. Se registraron los siguientes datos, en minutos, que tardan algunos hombres y mujeres en realizar cierta actividad en una empresa, los cuales fueron seleccionados aleatoriamente.

HOMBRES	MUJERES
n_1=14	n_2=25
Media=17	Media=19
Varianza=1,5	Varianza=1,8

Suponga que los tiempos para los dos grupos se distribuyen normalmente y que las varianzas son iguales, aunque desconocidas.

a. Calcule e interprete un intervalo de confianza del 99% para la verdadera diferencia de medias.
b. De acuerdo con el intervalo hallado, ¿hay evidencia de que los dos tiempos promedio son iguales?

1.5.9. Una fábrica desea saber la proporción de amas de casa que preferirían una aspiradora de su marca. Se toma al azar una muestra de 100 amas de casa y 20 dicen que les gustaría la máquina. Calcule e interprete un intervalo del 95% de confianza para la verdadera proporción de amas de casa que preferirían dicha aspiradora.

1.5.10. Se está considerando cambiar el procedimiento de manufactura de partes. Se toman muestras del procedimiento actual así como del nuevo para determinar si este último resulta mejor. Si 75 de 1.000 artículos del procedimiento actual presentaron defectos y lo mismo sucedió con 80 de 2.500 partes del nuevo, determine un intervalo de confianza del 90% para la verdadera diferencia de proporciones de partes defectuosas.

1.5.11. Un fabricante de baterías para automóvil asegura que las baterías que produce duran en promedio 2 años con una desviación estándar de 0,5 años. Si cinco de estas baterías tienen duración 1,5, 2,5, 2,9, 3,2, 4 años, determine un intervalo del 95% para la varianza e indique si es cierta la afirmación del fabricante. Determine un intervalo del 90% de confianza para el cociente de varianzas

1.5.12. Una muestra aleatoria de 36 cigarrillos de una marca determinada dio un contenido promedio de nicotina de 3 miligramos. Suponga que el contenido de nicotina de estos cigarrillos sigue una distribución normal con una desviación estándar de 1 miligramo.

a. Obtenga e interprete un intervalo de confianza del 95% para el verdadero contenido promedio de nicotina en estos cigarrillos.
b. El fabricante garantiza que el contenido promedio de nicotina es de 2,9 miligramos, ¿qué puede decirse de acuerdo con el intervalo hallado?

1.5.13. El tiempo (en minutos) que tardaron 15 operarios para familiarizarse con el manejo de una máquina moderna adquirida por la empresa fue: 3,4, 2,8, 4,4, 2,5, 3,3, 4, 4,8, 2,9, 5,6, 5,2, 3,7, 3, 3,6, 2,8,4,8. Suponga que los tiempos se distribuyen normalmente.

a. Determine e interprete un intervalo del 95% de confianza para el verdadero tiempo promedio
b. el instructor considera que el tiempo promedio requerido por la población de trabajadores que recibe instrucción sobre esta m quina es superior a 5 minutos, ¿qué se puede decir de acuerdo con el intervalo hallado?

1.5.14. Se desea medir la diferencia entre dos categorías de empleados en la actividad de seguros. Una est formada por personas con título superior y la otra por personas que sólo tienen estudios secundarios. Tomamos una muestra de 45 empleados entre los primeros y la media de ventas resulta ser 32. Tomamos 60 empleados del segundo grupo y la media es 25. Suponga que las ventas de los dos grupos se distribuyen normalmente con varianzas de 48 para los titulados superiores y 56 para los de estudios secundarios.

a. Calcule e interprete un intervalo del 90% de confianza para la verdadera diferencia de las medias.
b. De acuerdo con el intervalo hallado, ¿hay evidencia de que las medias sean iguales?

1.5.15. Se registraron los siguientes datos, en minutos, que tardan algunos hombres y mujeres en realizar cierta actividad en una empresa, los cuales fueron seleccionados aleatoriamente.

HOMBRES	MUJERES
n_1=14	n_2=25
Media=17	Media=19
Varianza=1,5	Varianza=1,8

1.5.16. Suponga que los tiempos para los dos grupos se distribuyen normalmente y que las varianzas son iguales, aunque desconocidas.

a. Calcule e interprete un intervalo de confianza del 99% para la verdadera diferencia de medias.
b. De acuerdo con el intervalo hallado, ¿hay evidencia de que los dos tiempos promedio son iguales?

1.5.17. Una fábrica desea saber la proporción de amas de casa que preferirían una aspiradora de su marca. Se toma al azar una muestra de 100 amas de casa y 20 dicen que les gustaría la máquina. Calcule e interprete un intervalo del 95% de confianza para la verdadera proporción de amas de casa que preferirían dicha aspiradora.

1.5.18. Se está considerando cambiar el procedimiento de manufactura de partes. Se toman muestras del procedimiento actual así como del nuevo para determinar si este último resulta mejor. Si 75 de 1.000 artículos del procedimiento actual presentaron defectos y lo mismo sucedió con 80 de 2.500 partes del nuevo, determine un intervalo de confianza del 90% para la verdadera diferencia de proporciones de partes defectuosas.

1.5.19. Un fabricante de baterías para automóvil asegura que las baterías que produce duran en promedio 2 años con una desviación estándar de 0,5 años. Si cinco de estas baterías tienen duración 1,5, 2,5, 2,9, 3,2, 4 años, determine un intervalo del 95% para la varianza e indique si es cierta la afirmación del fabricante. , determine un intervalo del 90% de confianza para el cociente de varianzas

1.5.20. Una muestra aleatoria de 36 cigarrillos de una marca determinada dio un contenido promedio de nicotina de 3 miligramos. Suponga que el contenido de nicotina de estos cigarrillos sigue una distribución normal con una desviación estándar de 1 miligramo.

a. Obtenga e interprete un intervalo de confianza del 95% para el verdadero contenido promedio de nicotina en estos cigarrillos.
b. El fabricante garantiza que el contenido promedio de nicotina es de 2,9 miligramos, ¿qué puede decirse de acuerdo con el intervalo hallado?

1.5.21. Se ha medido la talla de 100 personas elegidas al azar, mediante muestra aleatoria s, de entre los estudiantes varones de bachillerato de una gran ciudad, obteniéndose una talla media de 1.75 m. Se sabe que la desviación típica de la población es 0,2 m. ¿Con qué nivel de confianza se ha construido el intervalo (1.73, 1.77) para la media poblacional?

1.5.22. La longitud de la ballena azul se distribuye según una ley Normal con desviación típica 7,5 m. En un estudio estadístico realizado a 25 ejemplares se ha obtenido el intervalo de confianza (21.06, 26.94) para la longitud media.

a. Calcule la longitud media de los 25 ejemplares de la muestra.
b. Calcule el nivel de confianza con el que se ha construido dicho intervalo.

1.5.23. Una muestra aleatoria de 100 alumnos que se presentan a Selectividad revela que la media de edad es de 18.1 años. Halla un intervalo de confianza del 95% para la media de la edad de todos los estudiantes que se presentan a Selectividad sabiendo que la desviación típica de la población es 0.4.

1.5.24. Un fabricante de bombillas sabe que la desviación típica de la duración de las bombillas es de 100 horas. Calcula el tamaño de la muestra que se ha de someter a prueba para tener una confianza del 95% de que el error de la duración media que se calcule sea menor que 10 horas.

1.5.25. En una muestra de 145 personas mayores de 65 años se ha encontrado que el nivel medio de colesterol es de x = 240 mg / 100 ml, con desviación típica s_x = 45 mg / 100 ml . Con un nivel de confianza del 95% , ¿ podemos admitir que la media de la población general es de 226 mg / 100 ml ?.

1.5.26. A 150 alumnos seleccionados aleatoriamente en determinada región se les preguntó si utilizaban la biblioteca de su instituto para la preparación de sus exámenes. El número de respuestas afirmativas fue de 60. A partir de dicha información : a) Estimar el porcentaje de alumnos de esa región que utilizan para la preparación de sus exámenes la biblioteca de su instituto. b) Obtener el error máximo cometido con dicha estimación para un nivel de confianza del 99%.

1.5.27. A una muestra de 169 deportistas seleccionados aleatoriamente en cierta población se les preguntó cuánto tiempo dedicaban diariamente a su entrenamiento. Como resumen de la información recogida, se obtuvo un tiempo medio de 4’3 horas y una desviación típica de 1’5 horas. Para un nivel de significación del 1% (α = 0’01), ¿ podríamos rechazar la hipótesis de que el tiempo medio al día que dedica un deportista de dicha población a su entrenamiento es de 4 horas ?.

XXI. PRUEBAS DE HIPOTESIS

La estimación supone el uso de la evidencia muestral para estimar las características desconocidas de una población**. Las pruebas de hipótesis** incluyen el uso de la evidencia muestral para evaluar la probabilidad de que una suposición sobre alguna característica de una población sea cierta.

27.1. Pasos de una prueba de hipótesis

a. Planteamiento de la hipótesis

Hipótesis nula:	Hipótesis alterna:
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$
$H_0 : \mu \leq \mu_0$	$H_1 : \mu > \mu_0$
$H_0 : \mu \geq \mu_0$	$H_1 : \mu < \mu_0$

b. **Población** : Normal

c. **Muestra**: n observaciones independientes

d. **Estatístico de contraste**

Si n > 33 observaciones

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

Si n ≤ 33 observaciones

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

e. Distribucion muestral

Z ~ N(0, 1) X ~ $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

f. Critério de decision

Si $\left|Z\right| > Z_{1-\alpha/2}$; se rechaza la hipótesis nula

Si $Z > Z_{1-\alpha}$; se rechaza la hipótesis nula

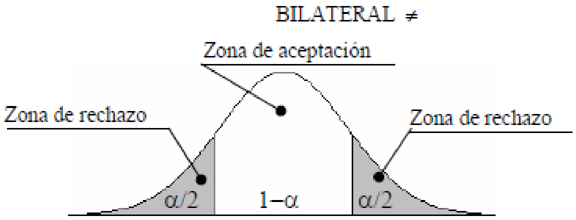
Si $Z < Z_{\alpha}$; se rechaza la hipótesis nula

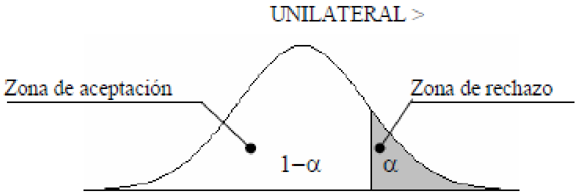
Si $\left|T\right| > T_{1-\alpha/2;n-1}$; se rechaza la hipótesis nula

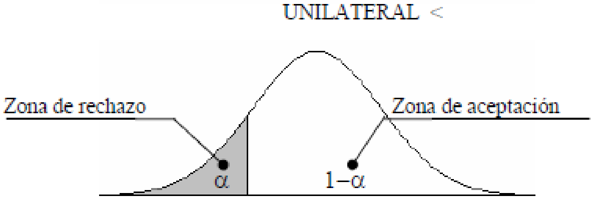
Si $T > T_{1-\alpha;n-1}$; se rechaza la hipótesis nula

Si $T < T_{\alpha;n-1}$; se rechaza la hipótesis nula

g. Region critica







h. Intervalos de confianza

$$P\left[\overline{X} - |Z_{\alpha/2}|\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \overline{X} - |Z_{\alpha/2}|\left(\frac{S}{\sqrt{n}}\right)\right] = 1 - \alpha$$

$$P[A \leq \mu \leq B] = 1 - \alpha$$

i. Interpretación del intervalo de confianza (Computador)

- Si A y B son positivos ; se dice que $\mu > \mu_0$
- Si A y B son negativos ; se dice que $\mu < \mu_0$
- Si A es negativo y B es positivo ; se dice que $\mu = \mu_0$

27.2. Tipos de error

El problema de decisión: rechazo/no rechazo, vendría expresado en las siguientes opciones en forma de tabla:

En el proceso de emplear una muestra para formar una decisión poblacional en una prueba de hipótesis, podemos cometer dos equivocaciones, al rechazar una hipótesis verdadera o al aceptar una hipótesis falsa; estas equivocaciones se conocen como:

a. Error tipo I. Se comete cuando se rechaza una hipótesis que por ser verdadera debería ser aceptada.
b. Error tipo II. Se comete cuando se acepta una hipótesis que por ser falsa debería ser rechazada.

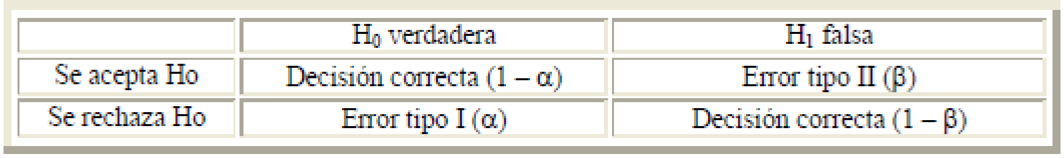
Hipótesis/Acción	No Rechazamos	Rechazamos
Es cierta	Correcto	Error Tipo I
Es falsa	Error Tipo II	Correcto

- Si la hipótesis nula (H_0) es cierta y nuestra decisión es no rechazarla, la decisión ha sido correcta.
- Si la hipótesis nula (H_0) es cierta y nuestra decisión es rechazarla, la decisión provoca un error. Dicho error se denomina error tipo I.
- Si la hipótesis nula (H_0) es falsa y nuestra decisión es no rechazarla, la decisión provoca un error. Dicho error se denomina error tipo II.
- Si la hipótesis nula (H_0) es falsa y nuestra decisión es rechazarla, la decisión ha sido correcta.

2.4. Nivel de significancia y nivel de confianza.

El nivel de significancia se refiere a la probabilidad $\propto$ de cometer error tipo I, es decir, rechazar una hipótesis verdadera.

El nivel de confianza se refiere a la probabilidad 1- $\propto$ de aceptar una hipótesis verdadera.



	H_0 verdadera	H_1 falsa
Se acepta Ho	Decisión correcta $(1 - \alpha)$	Error tipo II (β)
Se rechaza Ho	Error tipo I (α)	Decisión correcta $(1 - \beta)$

2.5. Ejercicios de pruebas de hipótesis

2.5.1. Dada la proposición de un problema, escriba la hipótesis nula y alterna correcta. (Suponiendo siempre que ya se ha decidido que el método a usar es un parametrico)

a. El Señor y la Señora Martínez suponen que el peso medio de una berenjena que crece en su jardín excede en promedio de 750 gramos que aparecen en el paquete de las semillas
b. En una presentación a posibles anunciadores , el canal de televisión dice que de la audiencia total de un sábado por la noche; mas del 75% estará viendo el programa de dicho canal
c. Alguien desea probar la afirmación de un productor de cigarrillos que afirma que la variabilidad de nicotina que contienen los cigarrillos de su marca es de al menos 1.3 miligramos
d. Un estudio arrojo que al comparar el tiempo que tardan hombres y mujeres en armar un producto determinado es superior en las mujeres en mas de 1.5 minutos
e. Una empresa de publicidad afirma que la efectividad de la campaña de publicidad incremento sus ventas en mas de 12 millones de pesos después de su campaña
f. Los automóviles estacionados en el estacionamiento de periodo prolongado del aeropuerto internacional de Baltimore permanecen un promedio de 2.5 días.
g. Una nueva marca de llantas radiales dura en promedio más de 48,000 millas.
h. El balance promedio de una cuenta de cheques en el First State Bank es de al menos $150.

i. Se reclama que al menos el 60% de las compras realizadas en cierta tienda por departamentos son artículos de especiales.
j. Se reclama que el 20% de los graduados de cierto colegio privado solicitan admisión a escuelas de medicina.

2.5.2. En los ejercicios determine si la región de rechazo para la hipótesis nula está en la cola izquierda, en la cola derecha, o ambas colas. Para el nivel de significancia $\propto$ dibuje la región de rechazo.

a. $H_0 : \mu \leq 11$; $H_1 : \mu > 11$
b. $H_0 : p = 0.4$; $H_1 : p > 0.4$
c. $H_0 : p \geq 0.3$; $H_1 : p < 0.3$
d. $H_0 : \mu \geq 5.8$; $H_1 : \mu < 5.8$
e. $H_0 : \mu = 110$; $H_1 : \mu \neq 110$
f. $H_0 : p \geq 0.8$; $H_1 : p < 0.8$

2.5.3. En los ejercicios use el método de la región de rechazo para probar la hipótesis.

a. $H_0 : 0.6$ $H_1 : p \neq 0.6$, $= 0.65$, $n = 100$, y $\propto = 0.01$
b. $H_0 : p = 0.29$ $H_1 : p \neq 0.29$, $= 0.26$, $n = 90$, y $\propto = 0.01$
c. $H_0 : p = 0.36$ $H_1 : p < 0.36$, $= 0.34$, $n = 630$, y $\propto = 0.05$

2.5.4. En los ejercicios use el método del valor-*p* para pruebas de hipótesis.

a. $H_0 : p = 0.2$ $H_1 : p > 0.2$, $= 0.245$, $n = 400$, y $\propto = 0.01$
b. $H_0 : p = 0.55$ $H_1 : p < 0.55$, $x = 175$, $n = 300$, y $\propto = 0.05$
c. $H_0 : p = 0.2$ $H_1 : p \neq 0.2$, $x = 235$, $n = 1000$, y $\propto = 0.02$

2.5.5. En los ejercicios establezca las hipótesis nula y alterna.

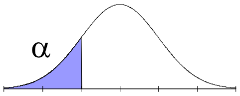
a. Los automóviles estacionados en el estacionamiento de periodo prolongado del aeropuerto internacional de Baltimore permanecen un promedio de 2.5 días.
b. Una nueva marca de llantas radiales dura en promedio más de 48,000 millas.
c. El balance promedio de una cuenta de cheques en el First State Bank es de al menos $150.
d. Se reclama que al menos el 60% de las compras realizadas en cierta tienda por departamentos son artículos de especiales.
e. Se reclama que el 20% de los graduados de cierto colegio privado solicitan admisión a escuelas de medicina.
f. Un dentista reclama que el 5% de sus pacientes sufren enfermedades en las encías.

2.5.6. Se acusa a una empresa de discriminación en sus prácticas de contratación.

a. ¿Que hipótesis se está probando si un jurado comete un error tipo 1 al encontrar que la compañía no es culpable?
b. ¿Que hipótesis Se está probando si un jurado comete un error tipo II al encontrar culpable a la empresa?

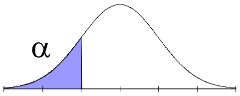
2.5.7. En los ejercicios complete la región de rechazo (encuentre el valor de z y t).

a.



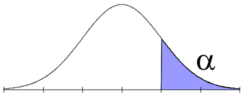
a) z, si $\propto = .05$ b) *t*, si $\propto = .025$ y n = 9

b.



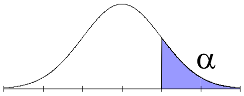
a) z, si $\propto = .01$ b) *t*, si $\propto = .05$ y n = 13

c.



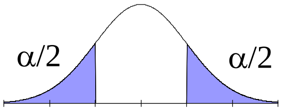
a) *z*, si $\propto = .02$ b) *t*, si $\propto = .01$ y n = 5

d.



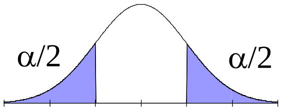
a) *z*, si $\propto = .025$ b) *t*, si $\propto = .01$ y n = 9

e.



a) *z*, si $\propto = .05$ b) *t*, si $\propto$=.05 y n = 10

f.



a) *z*, si $\propto = .01$ b) *t*, si $\propto$=0.1 y n = 7

2.5.8. Un directivo de uno de los grandes operadores de Internet está considerando la posibilidad de ofrecer tarifa plana a sus clientes. Según sus conocimientos sobre el tema, sabe que está trabajando con una variable aleatoria que se distribuye como una normal. Mantiene la hipótesis de que los hogares que tienen Internet se conectan una media de 5 horas semanales, y sabe por otros estudios que la dispersión de 7,24 horas. No obstante, existen otros estudios que sostienen que el tiempo de conexión es más alto. Para evaluar, a un 10% de significación, dicha hipótesis, el directivo decide encuestar aleatoriamente a 300 hogares, obteniendo una media de 5,34 horas de conexión. Que decisión debe tomar?

2.5.9. Un fabricante de pastas alimenticias asegura en su campaña publicitaria que el peso medio de los paquetes es de 250 gramos. Otro fabricante de la competencia pretende denunciarlo por engaño publicitario, ya que cree que es menor. Para contrastarlo selecciona una muestra aleatoria simple de 20 paquetes al azar siendo los pesos (en gramos) resultantes, 240, 225, 240, 220, 240, 250, 200, 215, 230, 140, 200, 216, 240, 250, 225, 240, 245, 220, 240, 240. Formula las hipótesis nula y alternativa y realiza un contaste a un nivel de significación del 5%? Suponga que el peso sigue una distribución normal.

2.5.10. Se sabe que un medicamento es efectivo en un 72% de los casos en los que se utiliza para tratar una determinada infección. Se ha desarrollado un nuevo medicamento y se ha comprobado que ha sido efectivo en 42 de los 50 casos tratados. ¿Estos datos proporcionan suficiente evidencia para demostrar que el nuevo medicamento tiene una efectividad distinta a la del antiguo? Utilice un 5% de nivel de significación.

2.5.11. Un fabricante de detergente afirma que el contenido de los paquetes que vende, pesa por termino medio al menos 200 gramos; si se sabe que la distribución de los pesos es normal con un desviación típica de 4 gramos. Una muestra aleatoria de 16 paquetes da un peso medio muestral de 198.4 gramos. Tiene razón el fabricante?

2.5.12. Una compañía asegura que las bombas de hule que fabrica tiene una vida media de servicio de 800 horas. Una muestra aleatoria simple de 36 bombas tomadas de un cargamento grande revela una vida media de 780 horas y una desviación típica de 90 horas. Cual es la conclusión adecuada de la prueba?

2.5.13. Antes de publicar un nuevo libro de cocina. Un editorial desea probar la hipótesis con un nivel de significancia del 2% de que el precio promedio de tales libros es de $101.500. Esta afirmación se sustenta si una muestra de 50 libros de cocina tiene una media de $95.613 y una desviación estándar de $13.323 pesos? Es cierta esta afirmación?

2.5.14. Un fabricante afirma que mediante el uso de un aditivo en la gasolina, los automóviles podrían recorrer por término medio, tres kilómetros más por litro. Se usa una muestra aleatoria de 100 automóviles para evaluar este producto. El incremento medio muestral alcanzado fue 2,4 kilómetros por litro, con una desviación típica de 1,8 kilómetros por litro.

2.5.15. Como gerente de compras para una gran empresa de seguros usted debe decidir si actualizar o no los computadores de la oficina. A usted se le ha dicho que el costo promedio de los computadores es de $3.150.000. Una muestra de 64 minoristas revela un precio promedio de $3.376.500, con una desviación estándar de $1.218.000. A un nivel de significancia del 5% parece que su información es correcta?

2.5.16. Los costos de producción mensual de una imprenta en Cali son en promedio de $410 mil pesos con una desviación estándar de $87 mil pesos. El gerente promete al propietario de la empresa promete al propietario de la imprenta mantener los costos por debajo de $300 mil pesos mensuales. Si los costos están distribuidos normalmente, el propietario puede creerle al gerente?

2.5.17. Una empresa de contabilidad descubre que el tiempo que se toma en realizar una auditoria esta distribuida normalmente con un tiempo promedio de 17.2 días y una desviación estándar se 3.7 días. El propietario promete iniciar una trabajo de auditoria para su firma dentro de 20 días, pero debe terminar uno ya comenzado. Que tan probable es que cumpla su promesa?

2.5.18. Como gerente de compras para una gran empresa de seguros usted debe decidir si actualizar o no los computadores de la oficina. A usted se le ha dicho que el costo promedio de los computadores es de 2.100 dólares. Una muestra de 64 minoristas revela que el precio promedio es de 2.251 dólares con una desviación estándar se 810 dólares. A un nivel de significancia del 5% parece que su información es correcta?

2.5.19. Un supermercado local gasto remodelando millones de pesos durante muchas semanas. Aunque la interrupción espanto a los clientes, el gerente espera que los clientes vuelvan a disfrutar de las nuevas comodidades. Antes de remodelar las ventas en la tienda promediaban $32.533.000 pesos por semana. Ahora que se ha terminado la remodelación el gerente toma una muestra de 36 semanas para ver si la construcción afecto de alguna manera el negocio. Se reporto un promedio de ventas semanales de $34.166.000 pesos con una desviación estándar de $12.955 pesos. Que puede decir el gerente a un nivel de significancia del 1%?

2.5.20. Antes de publicar un nuevo libro de cocina. Un editorial desea probar la hipótesis con un nivel de significancia del 2% de que el precio promedio de tales libros es de $101.500. Esta afirmación se sustenta si una muestra de 50 libros de cocina tiene una media de $95.613 y una desviación estándar de $13.323 pesos? Es cierta esta afirmación?

2.5.21. Deseamos conocer la postura de un gran colectivo frente al divorcio. Para ello interrogamos a 100 personas elegidas aleatoriamente e independientemente sobre este tema obligándolos a manifestarse en pro y en contra. El resultado es que 56 de los manifestantes esta a favor y 44 en contra. Es comprensible este resultado para determinar que existe la misma proporción en favor y en contra.

2.5.22. Usted ha estado trabajando para una empresa de publicidad durante cinco años. Ahora usted esta planeando iniciar su propia compañía, pero teme perder muchos clientes. Usted decide irse solo si por lo menos el 30% de las cuentas que usted maneja se irán con usted y le seguirán a su nuevo negocio. Como prueba usted descubre que 14 de las 54 cuentas que tomo como muestra expresan su deseo de irse con usted si usted deja la compañía. A un nivel de significancia del 7%. Debería usted comenzar su propia empresa?

2.5.23. De una muestra aleatoria de 802 clientes de supermercados, 378 fueron capaces de decir el precio del producto que acaban de comprar. Contrastar al 10% la hipótesis nula de que al menos la mitad de los clientes conocen los precios frente a la alternativa de que esa proporción es menor a la mitad.

2.5.24. Una empresa de mensajería anuncia que el 95% de los paquetes depositados en su oficina ¡¡asta la a AM, para repartir en la misma ciudad se entregaran hasta las 12 AM. Para contratar la empresa mensajera usted decide hacer una prueba con 100 paquetes que entrega antes de las 9 AM. Determine un procedimiento que le permita confirmar lo que anuncia la campaña de mensajería, con un nivel de confianza del 99%?

2.5.25. Un fabricante afirma que al menos 95% del equipo que ha surtido para cierta fábrica cumple con las especificaciones. Se examina una muestra de 700 piezas de equipo y se encuentra que 53 de ellas son defectuosas. ¿Puede decirse que los datos proporcionan suficiente evidencia para rechazar la afirmación del fabricante con un coeficiente de confianza de 0,95?.

2.5.26. Una empresa planea comercializar un nuevo producto solo si por lo menos al 40% del público le gusta. El departamento de investigación selecciona a 500 personas y encuentra que 225 lo prefieren al de la compañía, más cercana. A un nivel de significancia del 2%, la empresa debería comercializar el producto?

2.5.27. Un plan de reducción de peso estipula que el 75% de las personas que participan en el plan deberían perder entre el 5% y el 12% de su peso corporal en las primeras seis semanas. Si más del 75% pierden la cantidad estipulada, la dieta es demasiado severa y si menos del 75% de los participantes pierden la cantidad de peso estipulada es demasiado suave. De las 450 personas encuestadas 347 perdieron la cantidad de peso dentro del rango tolerable. A un nivel de significancia del 5%, que nos indican estas cifras sobre la dieta?

2.5.28. Usted ha estado trabajando para una empresa de publicidad durante cinco años. Ahora usted esta planeando iniciar su propia compañía, pero teme perder muchos clientes. Usted decide irse solo si por lo menos el 30% de las cuentas que usted maneja se irán con usted y le seguirán a su nuevo negocio. Como prueba usted descubre que 14 de las 54 cuentas que tomo como muestra expresan su deseo de irse con usted si usted deja la compañía. A un nivel de significancia del 7%. Debería usted comenzar su propia empresa?

2.5.29. Los diámetros de 31 tornillos elegidos al azar entre los producidos en cierta planta industrial han mostrado una varianza de 0.0625 mm^2 , Es compatible con este resultado experimental la hipótesis nula de 0.0484 mm^2 ,como supuesta varianza de los diámetros de la producción total de tornillos producidos en dicha planta industrial?

2.5.30. Elegimos 50 alumnos de ingeniería industrial de la CUAO y 120 de la UNIVALLE. La pertenecía de una a otra universidad depende de causas meramente fortuitas. Suponga que cada una de los dos universidades sigue un método distinto de enseñanza en cierta asignatura. Sea 3.7 el promedio obtenido en dicha asignatura por ellos estudiantes de la CUAO y 3.9 el promedio por los estudiantes de la UNIVALLE en dicha asignatura; sabiendo que la desviación típica de la CUAO es 0.6 y la UNIVALLE es de 0.9. Se desea saber si existe diferencia en los métodos de enseñanza.

2.5.31. Dos despliegues publicitarios se utilizan para ayudar a las ventas de un producto. La primera determina en ventas diarias: 110, 117, 82, 95, 123, 79, 92, 102, 108 y 113. Y el segundo despliegue produjo ventas de: 111, 85, 97, 117, 111, 89, 118, 121 y 109. Con un nivel de significancia del 5% se pude afirmar que los dos despliegues publicitarios produjeron los mismos dividendos?

2.5.32. Una planta manufacturera tiene dos máquinas ensambladoras que realizan operaciones idénticas en distintas líneas de ensamblado. Las interrupciones ocurren frecuentemente como resultado del constante uso. Se registraron los tiempos entre 10 descomposturas sucesivas para cada máquina, suponga que el tiempo entre descomposturas para cada máquina sigue una distribución normal con varianza común o^2. En la tabla se muestran las medias y las varianzas muéstrales de los tiempos entre descomposturas para cada máquina. ¿Presentan estos datos evidencia suficiente para indicar una diferencia en los tiempos medios entre descomposturas a un nivel de significación del 5 %?

Máquina 1	**Máquina 2**
$\overline{X} = 60{,}40$	$\overline{Y} = 65{,}3$
$\hat{S}_x^2 = 31{,}40$	$\hat{S}_y^2 = 44{,}82$

2.5.33. Se ha realizado una entrevista a cinco subdirectores y a cuatro analistas de mercado de una gran empresa. Se les preguntó a cada una cuál consideraba ser el porcentaje óptimo de cobertura de mercado para su compañía. Se obtuvieron las siguientes respuestas: **Subdirectores**: 22.5; 25; 30; 27.5; 20 - **Analistas de mercado**: 22.5; 17.5; 17; 20. ¿Sugieren estos datos que los subdirectores y los analistas de mercado están en desacuerdo cuando estiman la cobertura óptima de mercado para la empresa al 5 % de significación.

2.5.34. Una planta de producción tiene dos sistemas de fabricación extremadamente complejos, uno de ellos dos veces más viejo que el otro. A ambos sistemas se les aplica un tratamiento de mantenimiento cada dos semanas. Durante 30 días laborales se registra el número de productos terminados fabricados diariamente por cada uno de los sistemas y se obtienen los siguientes resultados:

Sistema nuevo	**Sistema viejo**
$\overline{X} = 246$	$\overline{Y} = 240$
$\hat{S}_x^2 = 15{,}16$	$\hat{S}_y^2 = 28.2$

¿Presentan los datos evidencia suficiente para concluir que la variabilidad en la producción diaria justifica un mantenimiento más intensivo para el sistema viejo con un nivel de confianza del 95 %?.

2.5.35. En un articulo sobre viajes comerciales la revista semana afirmo que el costo promedio de una cadena hotelera a nivel nacional era de $112.800 por noche y el de una segunda cadena era de $106.550 por noche. Se asume que estos estadísticos se basan en muestras de 82 y 97 respectivamente y que las varianzas poblacionales de cada cadena se sabe que son de $23.700 y $20.725 pesos al cuadrado respectivamente. Usted debe determinar cual cadena de hoteles utilizara su compañía. A un nivel se significancia del 1%, hace alguna diferencia en cual cadena utilizar?

2.5.36. Como director de producción de una empresa manufacturera usted debe decidir cual de las dos plantas debe responsabilizarse de producir los corchos utilizados para el vino en la vinería ABC Esta decisión se fundamentara en los niveles de productividad. Una muestra de 67 días en la planta de Cali produjo una media de 92.2 miles de corchos por día con una desviación estándar se 12.2. La planta de Palmira produjo un promedio de 89.2 miles de corchos por día con una desviación estándar de 15.4 miles en 54 días. A cual de las plantas responsabilizaría usted con un nivel de confianza del 98%?

2.5.37. Se desea conocer si los hombres estudian mas que las mujeres antes de un examen parcial en la universidad. Para verificar esta afirmación se toma una muestra aleatoria de 55 hombres y 45 mujeres antes de presentar un examen parcial y los resultados fueron los siguientes:

	Tamaño de la muestra	Media aritmética	Desviación estándar
Hombres	55	3.0	2.0
Mujeres	45	2.2	2.0

Con un nivel de confianza del 95% a que conclusión llego usted?

2.5.38. Para comparar los ingresos medios por ventas diarias, un minorista selecciona un tamaño de muestra de 12 semanas en una tienda del oriente de la ciudad y encuentra ingresos promedio de $1.250.400 pesos con una desviación estándar se $345.000 y una muestra de 15 semanas en otra tienda del norte de la ciudad tuvo una media de ingresos de $1.170.200 pesos con una desviación estándar se $210.500 pesos. Con un nivel de significancia del 5% donde le aconsejaría que distribuya sus productos?

2.5.39. Un estudio como objeto era valorar la relación entre la actividad cerebral de un sujeto mientras veía un anuncio en la televisión y la capacidad posterior de este sujeto de recordar el contenido del anuncio. En la tabla se presenta los datos para una serie de 10 productos. Se mostró a los sujetos anuncios de dos marcas diferentes. Para cada anuncio se midió la capacidad de recordarlo 24 horas después y cada anuncio de una pareja de anuncios se etiqueto como “bien recordado” y “mal recordado”.

Producto	Bien recordado	Mal recordado	d_i	d_i^2
1	137	53	84	7056
2	135	114	21	441
3	83	81	2	4
4	125	86	39	1521
5	47	34	13	169
6	46	66	- 20	400
7	114	89	25	625
8	157	113	44	1936
9	57	88	- 31	961
10	144	111	33	1089
Suma			210	14202

Que puede decir de la hipótesis de que no existe diferencia entre los niveles medios de actividad cerebral

2.5.40. Dos propagandas de publicidad para computadores las califican 15 clientes potenciales para determinar si existe alguna preferencia. los resultados se presentan aquí a un nivel de confianza del 90%. Cual es su conclusión?

Cliente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Publicidad 1	8	9	5	7	9	4	3	8	9	5	7	8	8	7	9
Publicidad 2	7	3	2	8	5	5	7	2	1	3	7	2	2	3	8

2.5.41. Se supone que usted esta trabajando como analista de mercados y desea medir la efectividad de un juego promocional del producto de su empresa. Antes del juego promocional usted selecciona 12 tiendas minoristas y registra las ventas en millones de pesos. Durante el segundo mes el juego promocional se complementa y se registran de nuevo las ventas. Que se presentan a continuación:

Tienda	1	2	3	4	5	6	7	8	9	10	11	12
Antes	42	57	38	49	63	36	48	58	47	51	83	27
Durante	40	60	38	47	65	39	49	50	47	52	72	33

La promoción incremento las ventas, con un nivel de significancia del 5%?

2.5.42. De una muestra de 300 mujeres se sabe que el 48% fuma y de una muestra de 400 hombres se sabe que el 54% fuma. Suponiendo que las 700 observaciones son independientes. Es compatible asegurar que no hay diferencias significativas entre las dos poblaciones.

2.5.43. En la prevención de cierta plaga que suele azotar los frutales de cierta región. Sus frutales han venido siendo fumigados unos con cierto producto A y otros con cierto producto B. El año pasado se ha observado que la diferencia entre la proporción de frutales inmunizados frente a la plaga tras la fumigación con A y la proporción de inmunizados tras su fumigación con B es de 0.10.

Este año las casas fabricantes en su campaña publicitaria afirman que sus productos han sido mejorados. La casa del producto A sostiene que la diferencia del año pasado será mayor este año. La casa del producto B sostiene que la diferencia será mucho menor. Para comprobar estas hipótesis observamos aleatoriamente 100 frutales con fumigación con producto A y 80 con producto B y vemos que entre los primeros 5 han sido afectados de algún modo por la enfermedad y entre los segundos 16 han sido afectados por la enfermedad. Es compatible la hipótesis.

2.5.44. Consideremos el posible cambio de actitud frente al aborto como consecuencia de un film sobre el mismo. Elegida una muestra al azar de 200 personas de entre los que han asistido a la proyección fílmica; constatamos la actitud antes y después de la proyección cinematográfica. Los resultados aparecen en la siguiente tabla:

		DESPUES			
		PRO	ANTI		
ANTES	PRO	80	32	112	0.56
	ANTI	70	18	88	0.44
		150	50	200	
		0.75	0.25		1.0

Se puede decir que la actitud de las personas cambio después de la proyección cinematográfica.

2.5.45. Un distribuidor que vende dos tipos de aparatos de calefacción (digamos A y B) a base de petróleo, decide comparar su eficacia. Para esto, se realiza un análisis de la eficacia de 8 calentadores del tipo A y 6 del tipo B. La eficacia de los 14 calentadores en porcentaje, se muestra en la tabla siguiente.

Tipo A	Tipo B
72	78
78	76
73	81
69	74
75	82
74	75
69	
75	

Encuentre un intervalo de confianza al 90% para la diferencia entre las eficacias medias de los dos tipos de calentadores.

2.5.46. En un estudio se comparan los efectos de dos promociones sobre las ventas. A continuación presentamos las ventas unitarias en miles de pesos durante la semana de una tienda que utilizo las dos promociones en semanas diferentes:

Muestras gratis	78	87	75	81	89	85
Descuentos	73	78	80	69	83	76

A que conclusión debe llegar con respecto a las dos promociones con un nivel de confianza del 95%?

2.5.47. Se desea averiguar si el período de vacaciones tiene algún efecto favorable sobre la productividad de los trabajadores de una empresa, entendiendo que, si existen diferencias, éstas pueden ser tanto positivas como negativas. Para ello se dispone de la siguiente muestra:

Producción semanal

Trabajador	Antes de vacaciones	Después de vacaciones
1	83	79
2	85	87
3	75	70
4	91	93
5	80	85
6	75	75
7	90	80
8	65	71
9	78	80
10	85	88
11	83	82
12	75	71
13	78	75
14	80	85
15	82	86
16	88	85
17	85	82
18	80	87
19	78	78
20	81	84
21	70	85
22	80	81

a. Se trata de un test para comparar medias de muestras dependientes o independientes?

b. Contraste si la producción semanal de antes y después de vacaciones (por separado) proviene de poblaciones normales (use α=0.05) . Escriba los valores en los cuales basó su decisión. ¿Cuál es el valor crítico p de su contraste? ¿Cambiaría su conclusión si α=0.10?

c. Según lo que haya concluido en el punto anterior, contraste con los métodos adecuados, si la media de la producción semanal es la misma antes y después de las vacaciones. Escriba la hipótesis nula y la alternativa. Escriba el valor del valor crítico p. Qué concluye?. Escriba también los valores de los estadísticos involucrados en la decisión.

2.5.48. Una tienda de muebles ha solicitado a una agencia de publicidad para que desarrolle una nueva campaña publicitaria. la gerente de comercialización de la tienda decide evaluar la efectividad de la campaña. Esta reúne los datos sobre las ventas mensuales antes y durante la campaña para 8 tiendas regionales y los resultados son los siguientes:

Tienda	1	2	3	4	5	6	7	8
Ventas antes	83	48	51	75	60	55	45	57
Ventas después	65	52	50	79	71	55	48	65

A que conclusión debe llegar con respecto a la efectividad de la campaña publicitaria con un nivel de confianza del 95%?

2.5.49. En un proceso de llenado de recipientes, la tolerancia en el peso es de 8 gramos. Para cumplir con este requisito, la máquina está calibrada para σ = 0.21 grs/recipiente. Se toman al azar 50 muestras y el resultado es una varianza de 0.04 grs/recipiente. Efectuar la dócima correspondiente y concluir al respecto. Use un N.S = 0.01.

2.5.50. Un médico compara la presión arterial de cuatro pacientes, antes y después del tratamiento con un médicamente. Las presiones fueron las siguientes

	Presión arterial	
Paciente	Antes del medicamento	Después del medicamento
1	158	135
2	169	146
3	154	143
4	155	141

Estime la disminución media de la presión arterial usando un intervalo de confianza del 90%.

2.5.51. Una disquería desea saber si un artista popular vende 3 veces más que un artista clásico, para ello efectuó un muestreo en 12 sucursales al azar, obteniendo los siguientes resultados: el artista popular vende en promedio 1500 copias de sus discos a la semana, con una desviación estándar de 5 discos por sucursal, en cambio el artista clásico vende 469 copias, con una varianza de 12 discos por sucursal.

a. Efectuar la dócima correspondiente usando un N.S = 0.05 y 0.01, y concluir respecto al tema.
b. Calcular la probabilidad de rechazar H_0, si es cierta.
c. Calcular el tamaño de muestra cuando la probabilidad de rechazo de Ho es de un 95%, si esta es falsa.

2.5.52. Es de interés conocer cómo actúa el ruido de una motosierra en el rendimiento de los trabajadores. Para controlarlo Forestal “CELCO” ha seleccionado a 48 personas al azar de entre el personal de sus contratistas para llevar acabo el estudio. De las 48, a 24 se les entregó motosierras que trabajaban a 200 decibeles y al resto motosierras que trabajan a 220 decibeles, obteniéndose los siguientes resultados, registrándose los resultados en número de arboles cortados en promedio por 1 semana. Supuestamente los trabajadores sometidos a 200 decibeles cortarían en promedio 2 árboles más que los otros.

200 decibeles: 53 54 52 51 50 49 56 48 52 55 53 52 51 54 51 52 51 51 52 56 55 51 51 220 decibeles: 48 47 46 55 51 52 53 56 54 51 57 49 45 47 47 49 50 51 52 56 55 51 52 50

Pruebe la hipótesis correspondiente y concluya al respecto. N.S. = 0.05

2.5.53. Una organización de salud se interesa en actualizar su información respecto a la proporción de hombres que fuman. Con base en estudios previos, se cree que esta es de un 40%. Para comprobarlo la organización lleva a cabo una encuesta en la que se seleccionan en forma aleatoria 1200 hombres a los cuales se les pregunta sus hábitos de fumador, de éstos resulto que 463 son fumadores. Emplee un método aproximado para determinar si esta evidencia apoya la noción de que la proporción de hombres que fuman es del 40%. Use un N.S. = 0.01 y 0.05.

2.5.54. El responsable de la campaña política del candidato A piensa que dado el ambiente de las últimas semanas previas a las elecciones, su candidato se encuentra en igual posición que su oponente B. Sin embargo han ocurrido algunas situaciones incómodas que hacen peligrar su elección (el factor Lenguisky), de tal forma una organización lleva a cabo una encuesta entre 1500 ciudadanos. Suponiendo que 720 personas indican una preferencia por el candidato A. a. ¿Existe alguna razón para indicar que el candidato A se encuentra en desventaja con respecto a B?, α=0.05 b. ¿Con qué n se podría empezar a tener la certeza de que no se pierda en las elecciones?, α=0.05

2.5.55. Un distribuidor de insumos forestales tiene 2 proveedores principales, A y B. Debido a una mejor estructura de precios, el distribuidor hace negocio únicamente con B, si es que la proporción de artículos defectuosos para A y B es la misma. De dos grandes lotes de producto, el distribuidor selecciona al azar 125 unidades de A y 100 unidades de B, al inspeccionar tales unidades se encuentra con que hay 7 artículos defectuosos por proveedor. Bajo las suposiciones adecuadas y con base en esta información, ¿existe alguna razón para no comprar en forma única a B?. α = 0.01.

2.5.56. Se cree que el promedio verbal para el número de respuestas correctas para la PAA es mayor para las mujeres que para los hombres, por más de 10 puntos. Las muestras aleatorias para ambos sexos arrojaron los siguientes resultados:

Hombres	Mujeres
n_1 = 125	n_2=100
$_{x1}$ = 480	$_{x2}$=460
$_{s1}$=60	$_{s2}$=52

a) Si se muestrearon dos poblaciones independientes normales, ¿se encuntra la creencia apoyada por la evidencia muestral con α = 0.05?. ¿Cuál es el valor de p?.
b) Supóngase una verdadera diferencia de 15 puntos. ¿Cuál es la potencia de la prueba anterior?.

2.5.57. El gerente de una planta de partes y piezas de muebles sospecha que el número de piezas que un trabajador arma varía día a día con una valor más allá del normal esperado. Para comprobarlo encarga a su jefe de producción que observe a un trabajador tipo en particular y que controle su desempeño durante 10 días. Los resultados fueron los siguientes 15, 12, 8, 13, 12, 15, 16, 9, 8 y 14. Si se sabe que la desviación estándar para los trabajadores en general es de 2 unidades, y si el número de éstas que se produce diariamente se encuentra de forma adecuada por una distribución normal, a un nivel α = 0.05. ¿Tiene apoyo la sospecha del gerente?. ¿Cuál es el valor de p?.

2.5.58. En un proceso de llenado, la tolerancia para el peso de los recipientes es de 8 gramos. Para reunir este requisito, la desviación estándar en el peso debe ser de dos gramos. Los pesos de 25 recipientes seleccionados al azar dieron como resultado una d.e. de 2.8 gramos.

a. Si los pesos se encuentran normalmente distribuidos, determinar si la varianza de estos es diferente del valor necesario, $\alpha = 0.01$.
b. ¿Para qué valores de la varianza muestral , no puede rechazarse la hipótesis de a)?.

2.5.59. Un contratista ordena un gran número de vigas de acero con longitud promedio de 5 m. Se sabe que la longitud de una viga se encuentra normalmente distribuida con d.e.= 0.02m. Después de recibir el embarque, se seleccionan 16 vigas al azar y se miden. Si el rpomedio es menor que el esperado, se enviará de vuelta el embarque al fabricante. Si la probabilidad de rechazar un embarque bueno es de 0.04. ¿Cuál debe ser el valor de la media muestral?. Y si el promedio real es de 4.98 m, ¿cuál es la potencia de la prueba anterior?

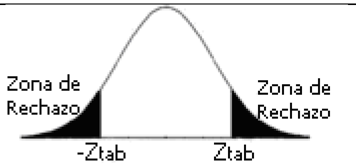
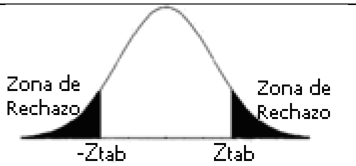
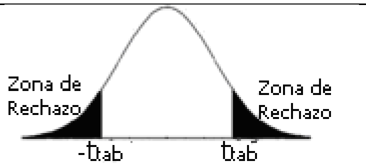
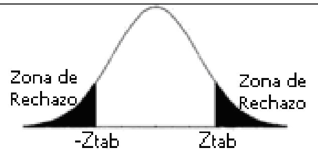
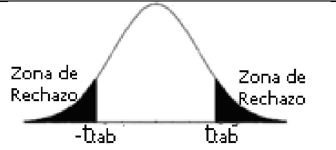
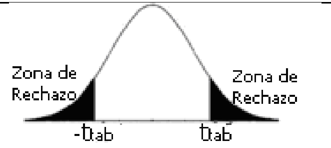
2.5.60. Se desea saber si la constante σ de una máquina lijadora está bien calibrada, para ello se compara con una máquina calibrada digitalmente, los datos que a continuación se presentan corresponden a los promedios de aserrín en grs por hora dejados como desecho, siendo la máquina B el testigo:

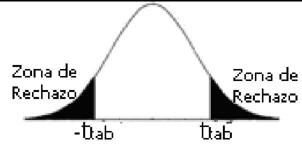
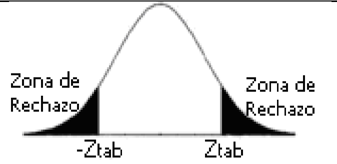
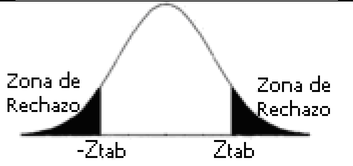
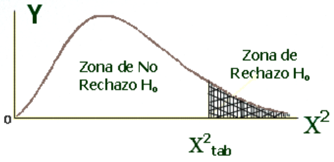
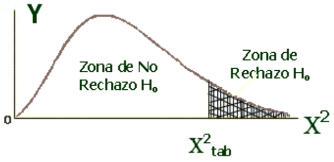
A: 254.4 254.1 259.1 263 298 351 237 267 249 254 269 254 215 236 214
B: 255 256 241 253 257 259 254 251 250 249.8 256 263 254 236 246 251 258

Realice la hipótesis correspondiente y concluya al respecto.

2.6. Tipos de pruebas de hipótesis

PRUEBAS DE HIPÓTESIS A TRAVÉS DEL CÁLCULO DEL ESTADÍGRAFO

Prueba de Hipótesis		**Hipótesis estadística**	**Nivel de significación**	**Estadígrafo**	**Región de rechazo**	**Regla de decisión**
Media para una muestra con σ conocida		H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	α = ?	$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$		Rechazar H_0 si $Z<-Z_{tab}$ o $Z>Z_{tab}$ de lo contrario no rechazo H_0
Media para una muestra con σ desconocida n≥30		H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	α = ?	$$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$		Rechazar H_0 si $Z<-Z_{tab}$ o $Z>Z_{tab}$ de lo contrario no rechazo H_0
Media para una muestra con σ desconocida n<30		H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	α = ?	$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$		Rechazar H_0 si $t<-t_{tab}$ o $t>-t_{tab\ t}$ de lo contrario no rechazo H_0
Diferencia de medias con σ₁ y σ₂ conocidas		H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	α = ?	$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}$$		Rechazar H_0 si $Z<-Z_{tab}$ o $Z>Z_{tab}$ de lo contrario no rechazo H_0
Diferencia de medias con σ₁ y σ₂ desconocidas para muestras Independientes	Realizar prueba de comparación de varianzas Si $\sigma_1 = \sigma_2$	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	α = ?	$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$	 Gl: n_1+n_2-2	Rechazar H_0 si $t<-t_{tab}$ o $t>-t_{tab\ t}$ de lo contrario no rechazo H_0
	Realizar prueba de comparación de varianzas Si $\sigma_1 \neq \sigma_2$	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	α = ?	$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1}{n_1} + \frac{S_2}{n_2}}}$$ $$Gl: \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 + 1}}$$		Rechazar H_0 si $t<-t_{tab}$ o $t>-t_{tab\ t}$ de lo contrario no rechazo H_0

Diferencia de medias con σ_1 y σ_2 desconocidas para muestras relacionadas		H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\alpha = ?$	$t = \frac{\bar{x}_d - \bar{\mu}_d}{\frac{s_d}{\sqrt{n}}}$ Nota: $\bar{x}_d$: Es la media de las diferencias de las observaciones antes y después, s_d : es la desviación estándar de las diferencias de las observaciones antes y después	 Gl: n_1+n_2-2	Rechazar H_0 si $t<-t_{tab}$ o $t>-t_{tab\ t}$ de lo contrario no rechazo H_0
Proporción para una muestra		H_0: $P = P_0$ H_1: $P \neq P_0$	$\alpha = ?$	$Z = \frac{x - nP_0}{\sqrt{nP_0Q_0}}$ ó $Z = \frac{p - P_0}{\sqrt{\frac{P_0Q_0}{n}}}$		Rechazar H_0 si $Z<-Z_{tab}$ o $Z>Z_{tab}$ de lo contrario no rechazo H_0
Diferencia de Proporciones		H_0: $P_1 = P_2$ H_1: $P_1 \neq P_2$	$\alpha = ?$	$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$		Rechazar H_0 si $Z<-Z_{tab}$ o $Z>Z_{tab}$ de lo contrario no rechazo H_0
Chi - cuadrado	Indepen-dencia	H_0: Existe Independen-cia …… H_1: No existe Independen-cia …..	$\alpha = ?$	$\chi^2 = \frac{\sum_{i=1}^{f}\sum_{j=1}^{c}\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$		Rechazar H_0 si $\chi^2>\chi^2_{tab}$ de lo contrario no rechazo H_0
	Homoge-neidad	H_0: Existe homogeneidad ….. H_1: No existe homoge-neidad …….	$\alpha = ?$	$\chi^2 = \frac{\sum_{i=1}^{f}\sum_{j=1}^{c}\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$		Rechazar H_0 si $\chi^2>\chi^2_{tab}$ de lo contrario no rechazo H_0

XXII. PRUEBAS NO PARAMETRICAS UNA MUESTRA

1.1. Prueba de Kolmogorov Smirnov

1.1.1. Ejemplo

El nuevo jefe de personal de una empresa desea saber si el número de piezas defectuosas producidas por turno están normalmente distribuidas y para ello cuenta con la siguiente información de diez turnos de trabajo:

6	3	2	5	0	5	5	7	6	5

1.1.2. Hipótesis

$H_0: Los\ tiempos\ estan\ distribuidos\ normalmente$
$H_1; Los\ tiempos\ no\ estan\ distribuidos\ normalmente$

1.1.3. Estadístico

	X	Z_1	n (X)	Y	Z_2	n (x)	$Max\|F_n(x) - S_n(x)\|$
1	0	,08	0189	,10	,4863	0686	0,0497
2	2	,13	1287	,20	,1560	1238	0,0048
3	3	,66	2544	,30	,8257	2045	0,0499
4	5	,28	6115	,40	,4954	3101	0,3113
5	5	,28	6115	,50	,1651	4344	0,1771
6	5	,28	6115	,60	1651	5656	0,0459
7	5	,28	6115	,70	4954	6899	0,0784
8	6	,76	7749	,80	8257	7955	0,0206
9	6	,76	7749	,90	1560	8762	0,1012
10	7	,23	8901	,00	4863	9314	0,0413
Media	,4			,55			
esv. Std.	,12			,30			

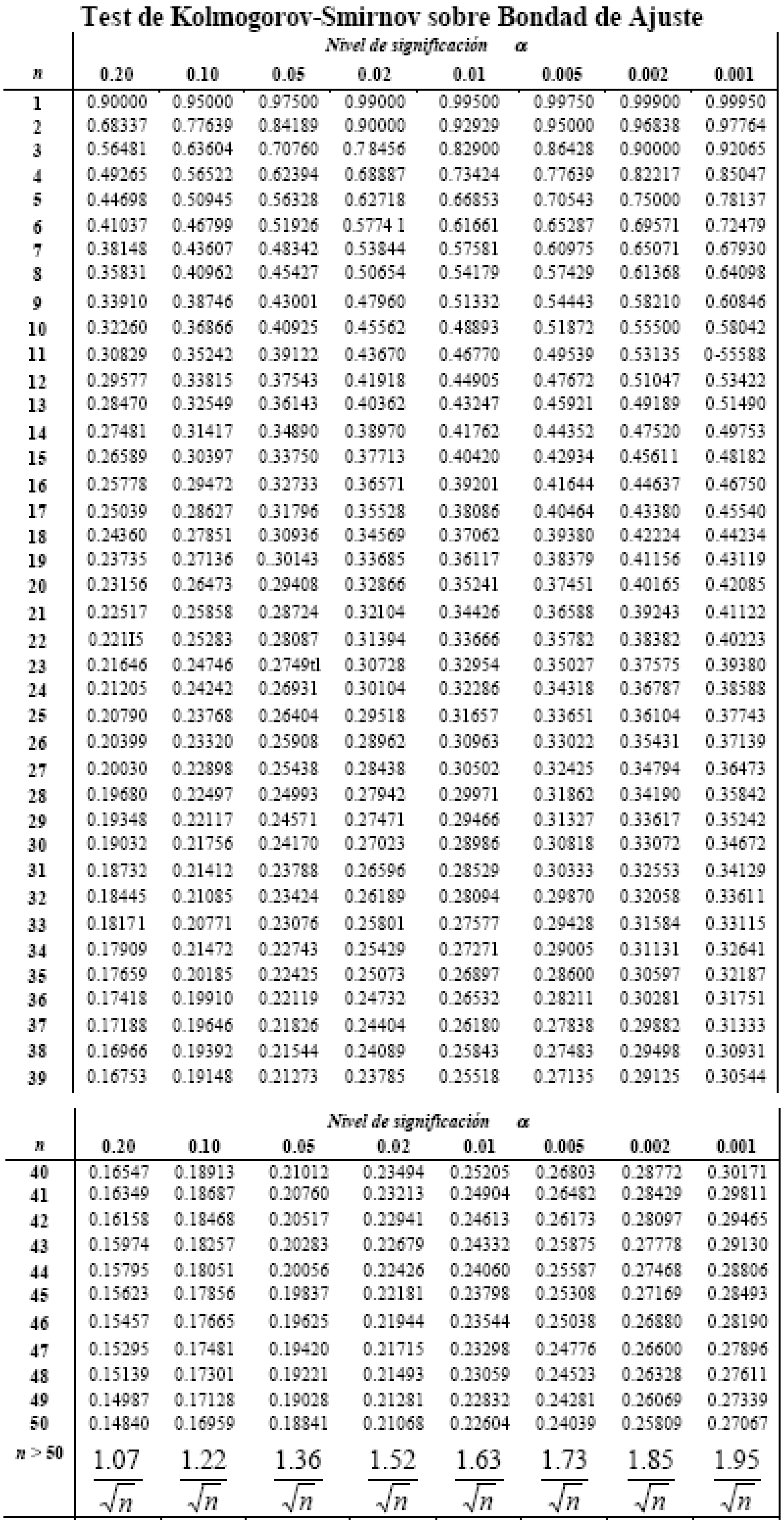
$$K - S = Max|F_n(x) - S_n(x)| = 0.3113$$

1.1.4. Criterio de decisión

$$Si\ K - S > K - S_{\propto,n} \quad \text{se rechaza } H_0$$

$$Si\ 0.3013 < 0.409 \quad \text{se acepta } H_0$$

1.1.5. Tabla de Kolmogorov Smirnov



Test de Kolmogorov-Smirnov sobre Bondad de Ajuste

	Nivel de significación α							
n	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.90000	0.95000	0.97500	0.99000	0.99500	0.99750	0.99900	0.99950
2	0.68337	0.77639	0.84189	0.90000	0.92929	0.95000	0.96838	0.97764
3	0.56481	0.63604	0.70760	0.7 8456	0.82900	0.86428	0.90000	0.92065
4	0.49265	0.56522	0.62394	0.68887	0.73424	0.77639	0.82217	0.85047
5	0.44698	0.50945	0.56328	0.62718	0.66853	0.70543	0.75000	0.78137
6	0.41037	0.46799	0.51926	0.5774 1	0.61661	0.65287	0.69571	0.72479
7	0.38148	0.43607	0.48342	0.53844	0.57581	0.60975	0.65071	0.67930
8	0.35831	0.40962	0.45427	0.50654	0.54179	0.57429	0.61368	0.64098
9	0.33910	0.38746	0.43001	0.47960	0.51332	0.54443	0.58210	0.60846
10	0.32260	0.36866	0.40925	0.45562	0.48893	0.51872	0.55500	0.58042
11	0.30829	0.35242	0.39122	0.43670	0.46770	0.49539	0.53135	0-55588
12	0.29577	0.33815	0.37543	0.41918	0.44905	0.47672	0.51047	0.53422
13	0.28470	0.32549	0.36143	0.40362	0.43247	0.45921	0.49189	0.51490
14	0.27481	0.31417	0.34890	0.38970	0.41762	0.44352	0.47520	0.49753
15	0.26589	0.30397	0.33750	0.37713	0.40420	0.42934	0.45611	0.48182
16	0.25778	0.29472	0.32733	0.36571	0.39201	0.41644	0.44637	0.46750
17	0.25039	0.28627	0.31796	0.35528	0.38086	0.40464	0.43380	0.45540
18	0.24360	0.27851	0.30936	0.34569	0.37062	0.39380	0.42224	0.44234
19	0.23735	0.27136	0..30143	0.33685	0.36117	0.38379	0.41156	0.43119
20	0.23156	0.26473	0.29408	0.32866	0.35241	0.37451	0.40165	0.42085
21	0.22517	0.25858	0.28724	0.32104	0.34426	0.36588	0.39243	0.41122
22	0.221I5	0.25283	0.28087	0.31394	0.33666	0.35782	0.38382	0.40223
23	0.21646	0.24746	0.2749tl	0.30728	0.32954	0.35027	0.37575	0.39380
24	0.21205	0.24242	0.26931	0.30104	0.32286	0.34318	0.36787	0.38588
25	0.20790	0.23768	0.26404	0.29518	0.31657	0.33651	0.36104	0.37743
26	0.20399	0.23320	0.25908	0.28962	0.30963	0.33022	0.35431	0.37139
27	0.20030	0.22898	0.25438	0.28438	0.30502	0.32425	0.34794	0.36473
28	0.19680	0.22497	0.24993	0.27942	0.29971	0.31862	0.34190	0.35842
29	0.19348	0.22117	0.24571	0.27471	0.29466	0.31327	0.33617	0.35242
30	0.19032	0.21756	0.24170	0.27023	0.28986	0.30818	0.33072	0.34672
31	0.18732	0.21412	0.23788	0.26596	0.28529	0.30333	0.32553	0.34129
32	0.18445	0.21085	0.23424	0.26189	0.28094	0.29870	0.32058	0.33611
33	0.18171	0.20771	0.23076	0.25801	0.27577	0.29428	0.31584	0.33115
34	0.17909	0.21472	0.22743	0.25429	0.27271	0.29005	0.31131	0.32641
35	0.17659	0.20185	0.22425	0.25073	0.26897	0.28600	0.30597	0.32187
36	0.17418	0.19910	0.22119	0.24732	0.26532	0.28211	0.30281	0.31751
37	0.17188	0.19646	0.21826	0.24404	0.26180	0.27838	0.29882	0.31333
38	0.16966	0.19392	0.21544	0.24089	0.25843	0.27483	0.29498	0.30931
39	0.16753	0.19148	0.21273	0.23785	0.25518	0.27135	0.29125	0.30544
40	0.16547	0.18913	0.21012	0.23494	0.25205	0.26803	0.28772	0.30171
41	0.16349	0.18687	0.20760	0.23213	0.24904	0.26482	0.28429	0.29811
42	0.16158	0.18468	0.20517	0.22941	0.24613	0.26173	0.28097	0.29465
43	0.15974	0.18257	0.20283	0.22679	0.24332	0.25875	0.27778	0.29130
44	0.15795	0.18051	0.20056	0.22426	0.24060	0.25587	0.27468	0.28806
45	0.15623	0.17856	0.19837	0.22181	0.23798	0.25308	0.27169	0.28493
46	0.15457	0.17665	0.19625	0.21944	0.23544	0.25038	0.26880	0.28190
47	0.15295	0.17481	0.19420	0.21715	0.23298	0.24776	0.26600	0.27896
48	0.15139	0.17301	0.19221	0.21493	0.23059	0.24523	0.26328	0.27611
49	0.14987	0.17128	0.19028	0.21281	0.22832	0.24281	0.26069	0.27339
50	0.14840	0.16959	0.18841	0.21068	0.22604	0.24039	0.25809	0.27067
$n > 50$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$	$\frac{1.73}{\sqrt{n}}$	$\frac{1.85}{\sqrt{n}}$	$\frac{1.95}{\sqrt{n}}$

1.1.6. Datos del computador

Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Numero de pieza defectuosas	10	4,40	2,12	0	7

Prueba de Kolmogorov-Smirnov para una muestra

		Numero de pieza defectuosas
N		10
Parámetros normales[a,b]	Media	4,40
	Desviación típica	2,12
Diferencias más extremas	Absoluta	,311
	Positiva	,125
	Negativa	-,311
Z de Kolmogorov-Smirnov		,985
Sig. asintót. (bilateral)		,286

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula (Los datos no tienen un comportamiento normal)

$Si\ \ Sig. > 0.05$ Se acepta la hipótesis nula (Los datos tienen un comportamiento normal)

1.2. Prueba Chi cuadrado

1.2.1. Ejemplo

De un grupo de estudiantes universitarios se desea saber si existen diferencias entre sus preferencias musicales y para ello se cuenta con la siguiente información

po de música	alor observado (O_i)	alor esperado (E_i)
alsa	18	33
etálica	7	33
achata	34	33
erengue	63	33
aladas	31	33
tra	45	33

1.2.2. Hipótesis

$$H_0: No\ hay\ preferencias \qquad\qquad H_1: Hay\ preferencias$$

1.2.3. Estadístico

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i} = 59.09$$

po de música	alor observado (O_i)	alor esperado (E_i)	**Estadistico** $\frac{(O_i - E_i)^2}{E_i}$
alsa	18	33	6,82
etálica	7	33	20,48
achata	34	33	0,03
erengue	63	33	27,27
aladas	31	33	0,12
tra	45	33	4,36
		χ^2	59,09

1.2.4. Criterio de decisión

$$Si\ \chi^2 > \chi^2_{\alpha,n-1} \quad \text{Se rechaza H}_0$$

$$Si\ 59 > \chi^2_{0.05;5-1} = 11.05 \quad \text{Se rechaza H}_0$$

1.2.5. Datos del computador

Tipos de musica

	N observado	N esperado	Residual
Salsa	18	33,0	-15,0
Metalica	7	33,0	-26,0
Bachata	34	33,0	1,0
Merengue	63	33,0	30,0
Baladas	31	33,0	-2,0
Otra	45	33,0	12,0
Total	198		

Estadísticos de contraste

	Tipos de musica
Chi-cuadrado[a]	59,091
gl	5
Sig. asintót.	,000

[a] 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 33,0.

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula (Hay preferencias)

$Si\ \ Sig. > 0.05$ Se acepta la hipótesis nula (No hay preferencias)

1.3. Prueba Binomial

1.3.1. Ejemplo

Un fabricante de televisores afirma en su póliza de garantía en el pasado que no más del 10% en sus aparatos necesitan reparación durante los primeros 2 años de operación. Una agencia de pruebas del gobierno selecciona una muestra de 100 televisores de este y encuentra que 14 de ellos necesitaron una reparación dentro los primeros dos años. Que puede usted decir de esta afirmación?

1.3.2. Hipótesis

$H_0 : P \geq 0.10$ $H_1 : P < 0.10$

1.3.3. Estadístico

$$Z = \frac{P_x - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.14-0.10}{\sqrt{\frac{0.10(1-0.10)}{100}}} = 1.33$$

1.3.4. Criterio de decisión

$Si\ Z < Z_{\propto}$ Se rechaza H_0

$Como\ \ 1.33 > -1.64$ Se acepta H_0

1.3.5. Datos del computador

Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
Se necesito reparacion	Grupo 1	Si	14	,1	,1	,122[a]
	Grupo 2	No	86	,9		
	Total		100	1,0		

[a] Basado en la aproximación Z.

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula

$$Si \ \ Sig. > 0.05$$ Se acepta la hipótesis nula

1.4. Prueba de rachas

1.4.1. Ejemplo

Juan voló a las Vegas y perdió 1.500 dólares jugando a la ruleta. Este piensa que ha sido estafado ya que había diseñado un sistema en casa de la ocurrencia del rojo y el negro en la ruleta. Debido a esto toma una muestra de la ocurrencia del rojo y negro en esta antes de regresarse a casa. La cual se presenta a continuación:

NNNRRNRNRRRNNRNRNRRRNNNNRRNNNRNRRRRRRNRNRRRNNNN

1.4.2. Hipótesis

$$H_0: Aleatoriedad \ (Pedio \ por \ mala \ suerte)$$
$$H_1: No \ Aleatoriedad \ (La \ ruleta \ estaba \ arreglada)$$

1.4.3. Estadístico

R= 24 Numero de rachas del juego
n_1=25 Numero de ocurrencias del negro
n_2=25 Numero de ocurrencias del rojo

$$E(R) = \frac{2n_1n_2}{n_1 + n_2} - 1 = 22.4$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} = 11.4$$

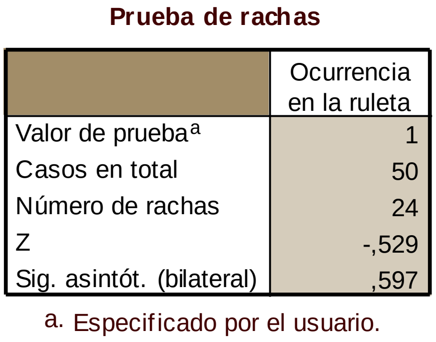
$$Z = \frac{R - E(R)}{\sqrt{V(R)}} = \frac{24 - 22.4}{\sqrt{11.4}} = 0.01$$

1.4.4. Criterio de decisión

$$Si \ |Z| > Z_{1-\frac{\propto}{2}}$$ Se rechaza H_0

$$Como \ \ |0.01| < 1.96$$ Se acepta H_0

1.4.5. Datos del computador



Prueba de rachas

	Ocurrencia en la ruleta
Valor de prueba[a]	1
Casos en total	50
Número de rachas	24
Z	-,529
Sig. asintót. (bilateral)	,597

a. Especificado por el usuario.

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula

$Si\ \ Sig. > 0.05$ Se acepta la hipótesis nula

XXIII. PRUEBAS NO PARAMETRICAS DOS MUESTRAS

1.5. Muestras independientes

1.5.1. Ejemplo – Prueba U de Mann Whiney

Dos personas A y B trabajan en el departamento de niños de una tienda. El gerente de la tienda piensa ampliar su negocio a otros locales, desde que leyó un artículo de marketing sobre la creciente popularidad de las tiendas de niños. Compara las ventas de los dos empleados por lo que toma la siguiente muestra:

VENDEDOR A			VENDEDOR B	
VENTAS	RANGO		VENTAS	RANGO
197	1		190	2
188	3		175	5
182	4		164	7,5
169	6		160	9
164	7,5		155	10
154	11		146	12
142	14		143	13
137	15		135	16
			133	17
			122	18
			118	19
			98	2
TOTAL	61,5		TOTAL	130,5

1.5.2. Hipótesis

$$H_0: No\ hay\ diferencias\ significativas\ en\ el\ rendimiento\ de\ los\ dos\ dependientes$$

$$H_1: Hay\ diferencias\ significativas\ en\ el\ rendimiento\ de\ los\ dos\ dependientes$$

1.5.3. Estadístico

$n_1 = 8$ Numero de registros del vendedor A

$n_2 = 12$ Numero de registros del vendedor B

$R_1 = 61.5$ Suma de rangos del vendedor A

$R_2 = 130.5$ Suma de rangos del vendedor B

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 130.5$$

$$E(U) = \frac{n_1 n_2}{2} = 48$$

$$V(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = 168$$

$$Z = \frac{U - E(U)}{\sqrt{V(U)}} = \frac{70.5 - 48}{\sqrt{168}} = 1.73$$

1.5.4. Criterio de decisión

$Si\ |Z| > Z_{1-\frac{\propto}{2}}$ Se rechaza H_0

$Como\ \ |1.73| < 1.96$ Se acepta H_0

1.5.5. Datos del computador

Rangos

	Vendedor	N	Rango promedio	Suma de rangos
Ventas (Millones de pesos)	Vendedor A	8	13,31	106,50
	Vendedor B	12	8,63	103,50
	Total	20		

Estadísticos de contraste[b]

	Ventas (Millones de pesos)
U de Mann-Whitney	25,500
W de Wilcoxon	103,500
Z	-1,737
Sig. asintót. (bilateral)	,082
Sig. exacta [2*(Sig. unilateral)]	,082[a]

a. No corregidos para los empates.

b. Variable de agrupación: Vendedor

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula

$Si\ \ Sig. > 0.05$ Se acepta la hipótesis nula

1.5.5.1. Reacciones extremas de Moses

Frecuencias

	Vendedor	N
Ventas (Millones de pesos)	Vendedor A (Control)	8
	Vendedor B (Experimental)	12
	Total	20

Estadísticos de contraste[a,b]

		Ventas (Millones de pesos)
Amplitud observada del grupo control		15
	Sig. (unilateral)	,187
Amplitud recortada del grupo control		12
	Sig. (unilateral)	,465
Valores atípicos recortados de cada extremo		1

a. Prueba de Moses

b. Variable de agrupación: Vendedor

1.5.5.2. Prueba de Kolmogorov Smirnov

Frecuencias

	Vendedor	N
Ventas (Millones de pesos)	Vendedor A	8
	Vendedor B	12
	Total	20

Estadísticos de contraste[a]

		Ventas (Millones de pesos)
Diferencias más extremas	Absoluta	,417
	Positiva	,417
	Negativa	,000
Z de Kolmogorov-Smirnov		,913
Sig. asintót. (bilateral)		,375

a. Variable de agrupación: Vendedor

2.2.1. Ejemplo – Prueba W de Wald Wolfowitz

Los alfareros queman doce piezas utilizando dos métodos I y II. Y se toma el número en minutos necesarios para que la pieza se enfriara. Y los resultados son los siguientes:

Método	1	2	3	4	5	6	7	8	9	10	11	12
I	27	31	28	29	39	40	35	33	32	36	37	43
II	34	24	38	28	30	24	37	42	41	44	37	40
Signo	-	+	-	+	+	+	-	-	-	-		+

2.2.2. Hipótesis

$$H_0: Las\ dos\ muestras\ provienen\ de\ la\ misma\ poblacion$$

$$H_1: Las\ dos\ muestras\ no\ provienen\ de\ la\ misma\ poblacion$$

2.2.3. Estadístico

$n_1 = 5$ Numero de registros positivos

$n_2 = 6$ Numero de registros negativos

R = 6 Numero de rachas

$$E(U) = \frac{n_1 n_2}{n} + 1 = 3.5$$

$$V(U) = \frac{2n_1 n_2(2n_1 n_2 - n)}{n^2(n-1)} = 0.163$$

$$Z = \frac{R - E(U)}{\sqrt{V(U)}} = \frac{6 - 3.5}{\sqrt{0.163}} = 6.19$$

2.2.4. Criterio de decisión

$$Si\ |Z| > Z_{1-\frac{\propto}{2}} \quad \text{Se rechaza } H_0$$

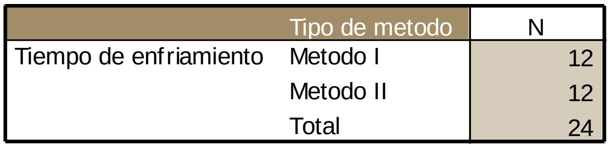
$$Como\quad |6.19| > 1.96 \quad \text{Se rechaza } H_0$$

2.2.5. Datos del computador

Estadísticos descriptivos

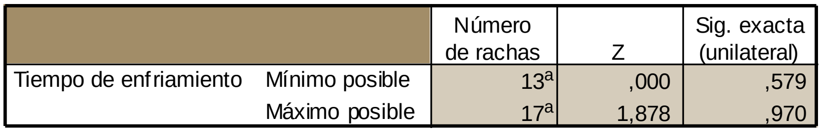
	N	Media	Desviación típica	Mínimo	Máximo
Tiempo de enfriamiento	24	34,54	5,93	24	44
Tipo de metodo	24	1,50	,51	1	2

Frecuencias



	Tipo de metodo	N
Tiempo de enfriamiento	Metodo I	12
	Metodo II	12
	Total	24

Estadísticos de contraste[b,c]



		Número de rachas	Z	Sig. exacta (unilateral)
Tiempo de enfriamiento	Mínimo posible	13[a]	,000	,579
	Máximo posible	17[a]	1,878	,970

a. Hay 3 empates inter-grupos que implican 7 casos.

b. Prueba de Wald-Wolfowitz

c. Variable de agrupación: Tipo de metodo

$Si\ \ Sig. < 0.05$ Se rechaza la hipótesis nula

$Si\ \ Sig. > 0.05$ Se acepta la hipótesis nula

2.3. Muestras relacionadas

2.3.1. Ejemplo (Prueba de Signos)

El gerente de un almacén desea conocer cual es la publicidad mas adecuada en la temporada escolar y para ello cuenta con la información de as ventas en millones de pasos cuando se realizaron dos tipos de publicidad:

PUBLICIDAD	1	2	3	4	5	6	7
I	8	9	5	7	9	7	4
II	7	10	2	8	5	7	3
Signo	+	-	+	-	+		+

2.3.2. Hipótesis

$$H_0: No\ hay\ diferencias\ significativas\ en\ las\ ventas\ con\ los\ dos\ tipos\ de\ publicidad$$

$$H_1: Hay\ diferencias\ significativas\ en\ las\ ventas\ con\ los\ dos\ tipos\ de\ publicidad$$

2.3.3. Estadístico

$$Z = \frac{(K\mp 0.5) - 0.5n}{0.5\sqrt{n}} = \frac{4 - 0.5 - 0.5(7)}{0.5\sqrt{7}} = -1.06$$

Si k $>\frac{n}{2}$ entonces se coloca k – 0.5; donde k es el signo de interes

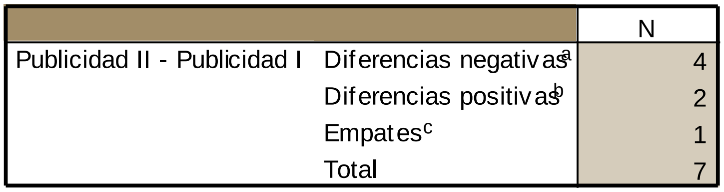
2.3.4. Criterio de decisión

$Si\ |Z| > Z_{1-\frac{\propto}{2}}$ Se rechaza H_0

$Como \quad |-1.06| < 1.96$ Se acepta H_0

2.3.5. Datos del computador

Frecuencias



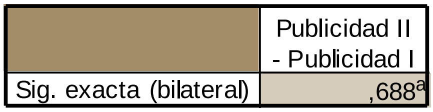
		N
Publicidad II - Publicidad I	Diferencias negativas[a]	4
	Diferencias positivas[b]	2
	Empates[c]	1
	Total	7

a. Publicidad II < Publicidad I

b. Publicidad II > Publicidad I

c. Publicidad I = Publicidad II

Estadísticos de contraste[b]



	Publicidad II - Publicidad I
Sig. exacta (bilateral)	,688[a]

a. Se ha usado la distribución binomial.

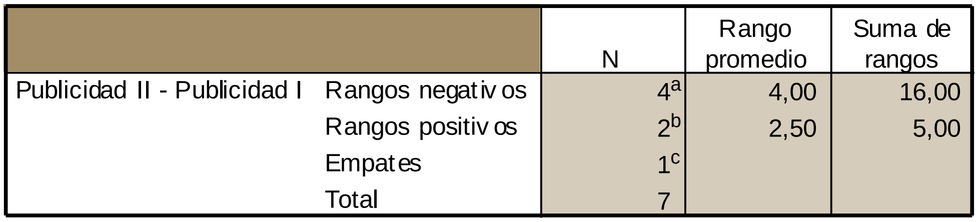
b. Prueba de los signos

$Si \;\; Sig. < 0.05$ Se rechaza la hipótesis nula

$Si \;\; Sig. > 0.05$ Se acepta la hipótesis nula

2.3.6. Prueba de rangos T de Wilcoxon

Rangos



		N	Rango promedio	Suma de rangos
Publicidad II - Publicidad I	Rangos negativos	4[a]	4,00	16,00
	Rangos positivos	2[b]	2,50	5,00
	Empates	1[c]		
	Total	7		

a. Publicidad II < Publicidad I

b. Publicidad II > Publicidad I

c. Publicidad I = Publicidad II

Estadísticos de contraste[b]

	Publicidad II - Publicidad I
Z	-1,186[a]
Sig. asintót. (bilateral)	,236

a. Basado en los rangos positivos.

b. Prueba de los rangos con signo de Wilcoxon

$$Si\ \ Sig. < 0.05$$ Se rechaza la hipótesis nula

$$Si\ \ Sig. > 0.05$$ Se acepta la hipótesis nula

2.3.7. Ejemplo (Prueba de Mc Nemar)

Un investigador de mercados esta interesado en estudiar la preferencia por los refrescos Coca Cola y Pepsi. Antes de llevar a cabo una prueba se sabor y después de esta. Se selecciono una muestra de 200 familias y los resultados se presentan a continuación:

		ESPUES		
ANTES		SI	NO	
	SI	.04	6	10
	NO	14	76	00
		.18	82	00

2.4.1. Hipótesis

$$H_0: No\ hay\ preferencias$$

$$H_1: Hay\ preferencias$$

2.4.2. Estadístico

		ESPUES		
ANTES		SI	NO	
	SI	A	B	+B
	NO	C	D	+D
		+C	3+D	n

$$Z = \frac{B-c}{\sqrt{b+c}} = \frac{6-14}{\sqrt{6+14}} = -1.79$$

2.4.3. Criterio de decisión

$$Si\ |Z| > Z_{1-\frac{\propto}{2}}$$ Se rechaza H_0

$Como \quad |-1.79| < 1.96$ Se acepta H_0

2.4.4. Datos del computador

Antes y Despues

	Despues	
Antes	0	1
0	76	14
1	6	104

Estadísticos de contraste[b]

	Antes y Despues
N	200
Sig. exacta (bilateral)	,115[a]

[a] Se ha usado la distribución binomial.

[b] Prueba de McNemar

$Si \ \ Sig. < 0.05$ Se rechaza la hipótesis nula

$Si \ \ Sig. > 0.05$ Se acepta la hipótesis nula

XXIV. PRUEBAS NO PARAMETRICAS VARIAS MUESTRAS

a. Varias muestras independientes

i. Ejemplo (Prueba de Kruskal Wallis)

Determinar si existen diferencias significativas entre el rendimiento de tres maquinas respecto al tiempo promedio en segundos que lleva a los trabajadores en efectuar completamente un proceso de llenado con ellas y los resultados se presentan a continuación:

MAQUINA I		MAQUINA II		MAQUINA III	
TIEMPO	RANGO	TIEMPO	RANGO	TIEMPO	RANGO
25,40	14	23,40	9	20.00	2
26,31	15	21,80	6	22,20	7
24,10	12	23,50	10	19,75	1
23,74	11	22,75	8	20,60	4
25,10	10	21,60	5	20,40	3
	62		38		17

ii. Hipótesis

$$H_0: Las\ mustras\ provienen\ de\ la\ misma\ poblacion$$

$$H_1: Las\ muestras\ no\ provienen\ de\ la\ misma\ poblacion$$

iii. Estadístico

$$H = \frac{12}{n(n+1)}\left[\sum \frac{R_i^2}{n_i}\right] - 3(n+1)$$

$$H = \frac{12}{15(15+1)}\left[\frac{62^2}{5} + \frac{38^2}{5} + \frac{17^2}{5}\right] - 3(15+1) = 11.58$$

iv. Criterio de decisión

$Si\ H > \chi^2_{\propto;k-1}$ Se rechaza H_0
$Si\ 11.58 > 5.99$ Se rechaza H_0

v. Comparaciones múltiples

$$Alcance\ critico = Z_{\propto}\sqrt{\frac{n(n+1)}{12}}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 6.76$$

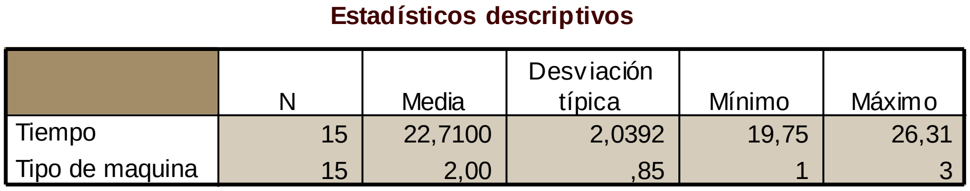
$[R_1 - R_2] = 5.4 < 6.76$ No hay diferencias entre la maquina I y II

$[R_1 - R_3] = 9.6 > 6.76$ Hay diferencias entre la maquina I y III

$[R_2 - R_3] = 4.2 < 6.76$ No hay diferencias entre la maquina II y III

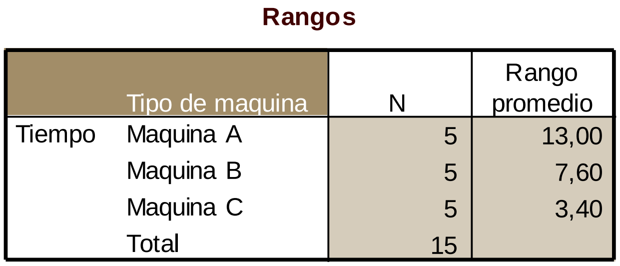
De acuerdo a estos los rendimientos se pueden organizar de la siguiente forma: I > II ≥ III

vi. Datos del computador



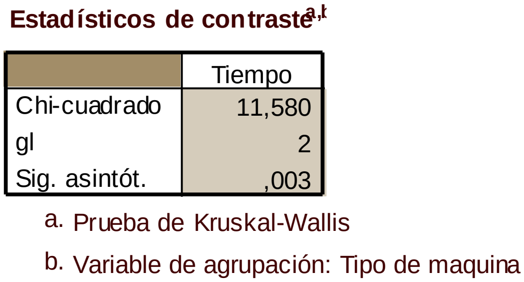
Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Tiempo	15	22,7100	2,0392	19,75	26,31
Tipo de maquina	15	2,00	,85	1	3



Rangos

	Tipo de maquina	N	Rango promedio
Tiempo	Maquina A	5	13,00
	Maquina B	5	7,60
	Maquina C	5	3,40
	Total	15	



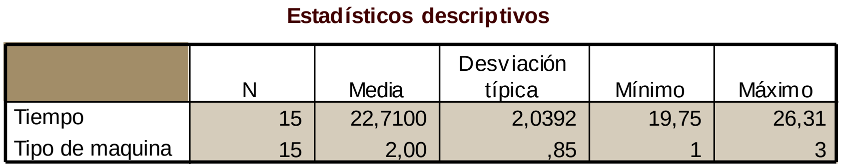
Estadísticos de contraste[a,t]

	Tiempo
Chi-cuadrado	11,580
gl	2
Sig. asintót.	,003

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Tipo de maquina

3.4.7 Prueba de la mediana

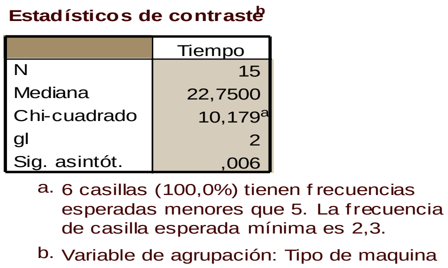


Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Tiempo	15	22,7100	2,0392	19,75	26,31
Tipo de maquina	15	2,00	,85	1	3

Frecuencias

		Tipo de maquina		
		Maquina A	Maquina B	Maquina C
Tiempo	> Mediana	5	2	0
	<= Mediana	0	3	5



Estadísticos de contraste[b]

	Tiempo
N	15
Mediana	22,7500
Chi-cuadrado	10,179[a]
gl	2
Sig. asintót.	,006

a. 6 casillas (100,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 2,3.

b. Variable de agrupación: Tipo de maquina

b. Varias muestras relacionadas

i. Prueba de Friedman

1. Ejemplo

Un empresario desea saber si las medianas de los rendimientos establecidos por publicidad son iguales en la evaluación del servicio de un restaurante por seis evaluadores

Sucursal A	R_1	Sucursal B	R_2	Sucursal C	R_3	Sucursal D	R_4
70	2	61	1	82	4	74	3
77	3	75	1	88	4	76	2
76	2	67	1	90	4	80	3
80	3	63	1	96	4	76	2
84	2.5	66	1	92	4	84	2.5
78	2.0	68	1	98	4	86	3
	14.5		6		24		15.5

2. Hipótesis

$$H_0\text{: } No\ hay\ diferencias\ en\ el\ servicio\ de\ los\ restaurantes$$

$$H_1\text{: } Hay\ diferencias\ en\ el\ servicio\ de\ los\ restaurantes$$

3. Estadístico

$$Q = \frac{12}{fc(c-1)}\left[\sum R_i^2\right] - 3f(c+1)$$

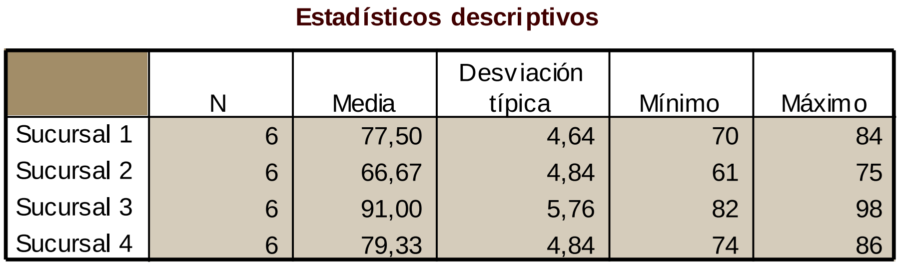
$$Q = \frac{12}{6(4)(6-1)}[14.5^2 + 6^2 + 15.5^2] - 3(6)(6+1) = 16.5$$

4. Criterio de decisión

$Si\ \ Q > \chi^2_{\propto;k-1}$ Se rechaza H_0

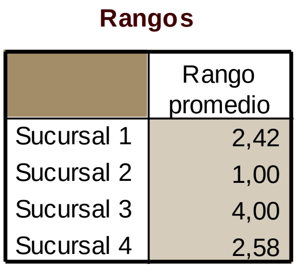
$Si\ \ 16.5 > 7.81$ Se rechaza H_0

5. Datos del computador



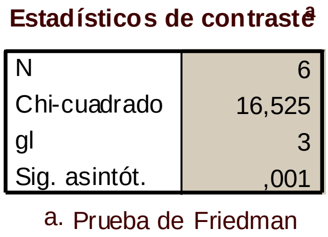
Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Sucursal 1	6	77,50	4,64	70	84
Sucursal 2	6	66,67	4,84	61	75
Sucursal 3	6	91,00	5,76	82	98
Sucursal 4	6	79,33	4,84	74	86



Rangos

	Rango promedio
Sucursal 1	2,42
Sucursal 2	1,00
Sucursal 3	4,00
Sucursal 4	2,58

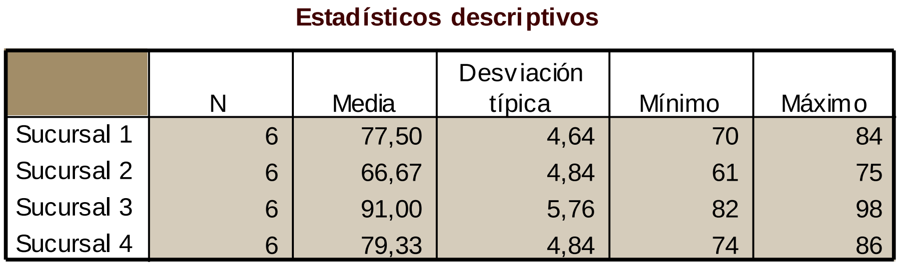


Estadísticos de contraste[a]

N	6
Chi-cuadrado	16,525
gl	3
Sig. asintót.	,001

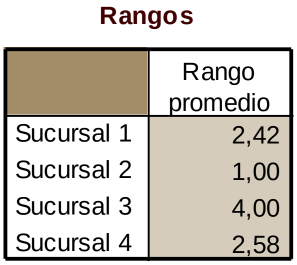
a. Prueba de Friedman

ii. Prueba W de Kendall



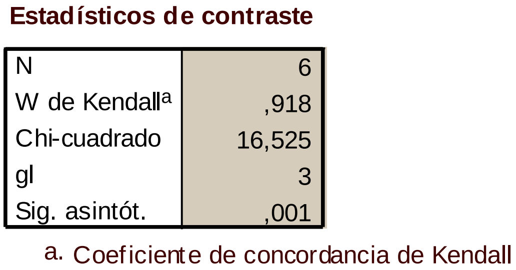
Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Sucursal 1	6	77,50	4,64	70	84
Sucursal 2	6	66,67	4,84	61	75
Sucursal 3	6	91,00	5,76	82	98
Sucursal 4	6	79,33	4,84	74	86



Rangos

	Rango promedio
Sucursal 1	2,42
Sucursal 2	1,00
Sucursal 3	4,00
Sucursal 4	2,58



Estadísticos de contraste

N	6
W de Kendall[a]	,918
Chi-cuadrado	16,525
gl	3
Sig. asintót.	,001

a. Coeficiente de concordancia de Kendall

iii. Prueba Q de Cochran

1. Ejemplo

Diferencias en los tratamientos

ID	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L_i^2
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
$\sum G_i$	[illegible]	[illegible]	[illegible]	[illegible]	9	47

2. Hipótesis

$$H_0: P_1 = P_2 = \cdots \ldots \ldots . . = P_k$$

$$H_1: Alguno\ es\ diferente$$

3. Estadístico

$$Q = \frac{k(k-1)\sum\left(G_j^2 - \bar{G}^2\right)}{k\sum L_i - \sum L_i^2}$$

$$Q = \frac{4(4-1)\sum\left(99 - \overline{475}^2\right)}{4(19) - 47} = 31.63$$

4. Criterio de decisión

$$Si\ \ Q > \chi^2_{\propto;k-1}$$ Se rechaza H_0

$$Si\ \ 31.63 > 7.81$$ Se rechaza H_0

5. Datos del computador

Frecuencias

	Valor	
	0	1
Tratamiento 1	6	4
Tratamiento 2	4	6
Tratamiento 3	2	8
Tratamiento 4	6	4

Rangos

	Rango promedio
Tratamiento 1	2,20
Tratamiento 2	2,60
Tratamiento 3	3,00
Tratamiento 4	2,20

Estadísticos de contraste

N	10
Q de Cochran	4,125[a]
gl	3
Sig. asintót.	,248

a. 0 se trata como un éxito.

XXV. ANALISIS DE VARIANZA

4.1. Planteamiento de hipótesis

Se forman k grupos con diferentes sujetos en cada uno

$H_0\colon \mu_1 = \mu_2 = \cdots \ldots \ldots = \ \mu_k$ (Todos los tratamientos tienen igual rendimiento)

$H_0\colon \mu_i \neq \mu_j \quad i \neq j$ (Algunos tienen diferente rendimientos)

4.2. Supuestos

- Independencia
- Normalidad
- Homocedasticidad

4.3. Tabla ANOVA

Factor	Observaciones	Totales T_j
A_1	Y_{11} Y_{12} .... Y_{n1}	T_1
A_2	Y_{12} Y_{22} .... Y_{n2}	T_2
.	.	
.	.	
.	.	
A_n	Y_{n1} Y_{n2} Y_{nm}	T_j
		T

Fuente de variación	Grados de libertad	Suma de cuadrados	uadrados medio	F
Regresión	K - 1	SCR	MR = (SCR/K – 1)	MR/CME
Error	N - k	SCE	ME = (SCE / N – K)	
Total	N - 1	SCT		

$$SCR = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

$$SCE = \sum Y_{ij}^2 - \sum \frac{T_j^2}{n_j}$$

$$SCT = \sum Y_{ij}^2 - \frac{T^2}{N}$$

4.4. Criterio de decisión

$Si \quad F > F_{\propto,k-1;N-k}$ Se rechaza H_0

4.5. Comparaciones múltiples

Diferencias mínimas significativas (DMS)

a. Modelos balanceados

$$\boldsymbol{DMS} = \sqrt{\frac{\boldsymbol{2(CME)F_{\propto.1;N-k}}}{\boldsymbol{r}}}$$ Donde r es el número de repeticiones

b. Modelos no balanceados

$$\boldsymbol{DMS_{i,k}} = \sqrt{\left[\frac{\boldsymbol{1}}{\boldsymbol{r_i}}+\frac{\boldsymbol{1}}{\boldsymbol{r_k}}\right]\boldsymbol{(CME)F_{\propto,c-k;N-k}}}$$ Donde r es el número de repeticiones

c. Evaluación

$\boldsymbol{Si}\ \left|\boldsymbol{\bar{X}_i - \bar{X}_j}\right| < DMS$ No hay diferencias significativas entre los dos tratamiento

4.6. Ejemplo de un modelos Balanceado

Se forman tres grupos de seis alumnos y a cada uno de estos se le aplica un método de enseñanza. Los resultados se presentan a continuación

Alumno	Presencial	Internet	Autodidacta	
1	2.4	2.5	0.8	
2	3.6	3.1	3.2	
3	2.7	2.7	2.0	
4	3.4	1.8	2.7	
5	4.3	2.1	1.2	
6	3.1	1.2	1.6	
$\sum X_i$	19.5	13.4	11.50	T = 44.4
$\sum X_i^2$	65.67	32.24	26.17	$T^2 = 1971.36$
$\bar{X}$	3.25	2.23	1.92	

a. Hipótesis

$$H_0\colon \mu_P = \mu_I = \mu_A$$ (Todos los tratamientos tienen igual rendimiento)

$$H_0\colon \mu_i \neq \mu_j \quad i \neq j$$ (Algunos tienen diferente rendimientos)

b. Supuestos

- Independencia
- Normalidad
- Homocedasticidad

c. Tabla ANOVA

Fuente de variación	Grados de libertad	Suma de cuadrados	uadrados medio	
Regresión	2	5.82	2.91	
Error	15	8.74	0.58	99
Total	17	14.56		

d. Criterio de decisión

$$Si \quad 4.99 > F_{0.05,2;15} = 3.68$$ Se rechaza H_0

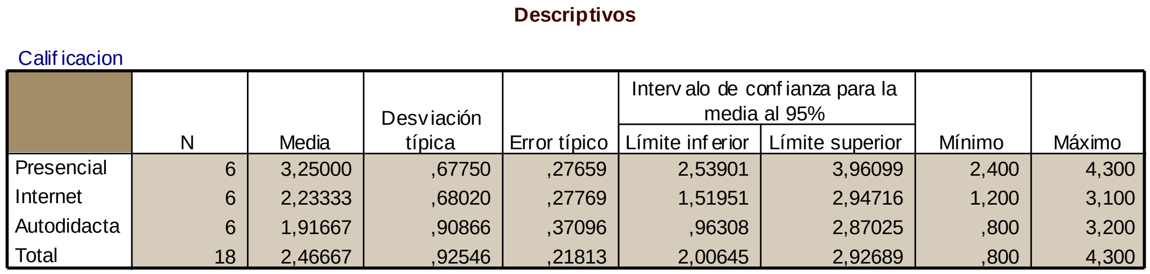
e. Comparaciones múltiples

$$\boldsymbol{DMS} = \sqrt{\frac{\boldsymbol{2(CME)F_{\propto,1;N-k}}}{\boldsymbol{r}}} = \sqrt{\frac{\boldsymbol{2(0.58)4.75}}{\boldsymbol{6}}}$$ **=0.958** Donde r es el número de repeticiones

- $\boldsymbol{Si}\ \left|\bar{X}_p - \bar{X}_I\right| = \boldsymbol{1.02} > DMS$ **H**ay diferencias significativas entre el método presencial e internet
- $\boldsymbol{Si}\ \left|\bar{X}_p - \bar{X}_A\right| = \boldsymbol{1.33} > DMS$ **H**ay diferencias significativas entre el método presencial e Autodidacta
- $\boldsymbol{Si}\ \left|\bar{X}_I - \bar{X}_A\right| < DMS$ No hay diferencias significativas entre el método de internet y autodidacta

Dela análisis de esto se destaca que el mejor método es el presencial, después el de mejor rendimiento es el de Internet y por ultimo se debe tomar el autodidacta

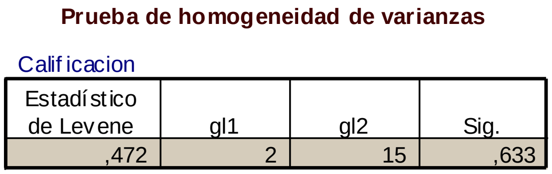
f. Datos del computador



Descriptivos

Calificacion

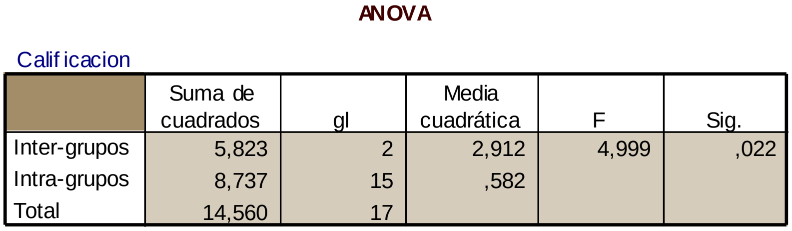
	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Presencial	6	3,25000	,67750	,27659	2,53901	3,96099	2,400	4,300
Internet	6	2,23333	,68020	,27769	1,51951	2,94716	1,200	3,100
Autodidacta	6	1,91667	,90866	,37096	,96308	2,87025	,800	3,200
Total	18	2,46667	,92546	,21813	2,00645	2,92689	,800	4,300



Prueba de homogeneidad de varianzas

Calificacion

Estadístico de Levene	gl1	gl2	Sig.
,472	2	15	,633



ANOVA

Calificacion

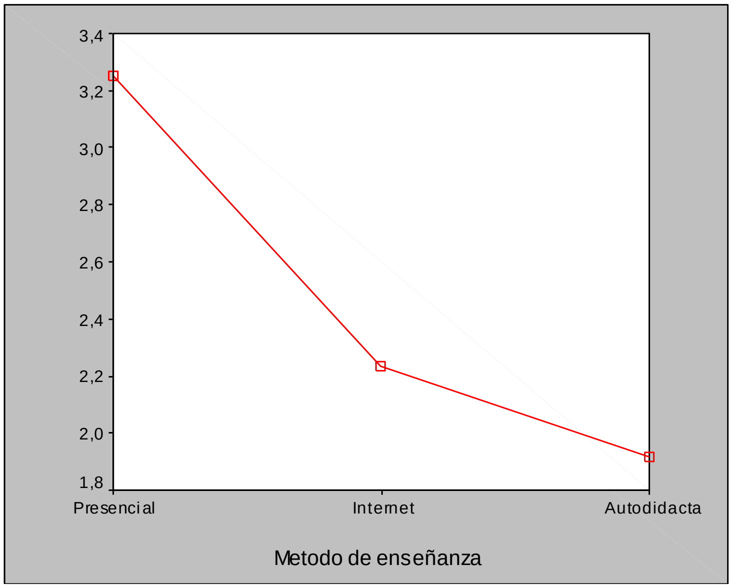
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	5,823	2	2,912	4,999	,022
Intra-grupos	8,737	15	,582		
Total	14,560	17			

Comparaciones múltiples

Variable dependiente: Calificacion
DMS

(I) Metodo de enseñanza	(J) Metodo de enseñanza	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Presencial	Presencial					
	Internet	1,01667*	,44062	,036	7,7502E-02	1,95583
	Autodidacta	1,33333*	,44062	,009	,39417	2,27250
Internet	Presencial	-1,01667*	,44062	,036	-1,95583	-7,75021E-02
	Internet					
	Autodidacta	,31667	,44062	,483	-,62250	1,25583
Autodidacta	Presencial	-1,33333*	,44062	,009	-2,27250	-,39417
	Internet	-,31667	,44062	,483	-1,25583	,62250
	Autodidacta					

*. La diferencia entre las medias es significativa al nivel .05.



4.7. Ejemplo de un modelo no balanceado

a. Ingreso a un parque de diversiones

Visitante	Acampar	Pescar	Pasear en bote	
1	38	30	19	
2	32	25	35	
3	35	31	20	
4	36	35	22	
5	38		25	
6	32			
$\sum X_i$	211	121	121	T = 453
$\sum X_i^2$	7457	3711	30.95	$T^2 = 205209$
$\bar{X}$	35.2	30.5	24.2	

b. Hipótesis

$$H_0: \mu_A = \mu_P = \mu_B$$

c.

$$H_0: \mu_i \neq \mu_j \quad i \neq j$$

d. Supuestos

- Independencia
- Normalidad
- Homocedasticidad

e. Estadístico de contraste

Fuente de variación	Grados de libertad	Suma de cuadrados	uadrados medio	=
Regresión	2	382.02	191.01	
Error	12	254.98	21.25	99
Total	14	637.4		

f. Criterio de decisión

$Si \quad F > F_{\propto,k-1;N-k}$ Se rechaza H_0

$Si \quad 8.99 > 3.89$ Se rechaza H_0

g. Comparaciones múltiples

$$\boldsymbol{DMS_{A,P}} = \sqrt{\left[\frac{1}{r_i}+\frac{1}{r_k}\right](\boldsymbol{CME})\boldsymbol{F_{\propto,c-k;N-k}}} = \sqrt{\left[\frac{1}{6}+\frac{1}{4}\right](\mathbf{21.25})(\mathbf{3.89})} = \mathbf{5.87}$$

$$\boldsymbol{DMS_{A,B}} = \sqrt{\left[\frac{1}{r_i}+\frac{1}{r_k}\right](\boldsymbol{CME})\boldsymbol{F_{\propto,c-k;N-k}}} = \sqrt{\left[\frac{1}{6}+\frac{1}{5}\right](\mathbf{21.25})(\mathbf{3.89})} = \mathbf{5.51}$$

$$\boldsymbol{DMS_{p,B}} = \sqrt{\left[\frac{1}{r_i}+\frac{1}{r_k}\right](\boldsymbol{CME})\boldsymbol{F_{\propto,c-k;N-k}}} = \sqrt{\left[\frac{1}{4}+\frac{1}{5}\right](\mathbf{21.25})(\mathbf{3.89})} = \mathbf{6.10}$$

h. Conclusiones

$\boldsymbol{Si} \ |\bar{\boldsymbol{X}}_{\boldsymbol{A}} - \bar{\boldsymbol{X}}_{\boldsymbol{P}}| = \mathbf{4.7} < \boldsymbol{DMS}$ **No h**ay diferencias significativas entre la preferencia entre acampar y pescar

$\boldsymbol{Si} \ |\bar{\boldsymbol{X}}_{\boldsymbol{A}} - \bar{\boldsymbol{X}}_{\boldsymbol{B}}| = \mathbf{11} > \boldsymbol{DMS}$ **H**ay diferencias significativas entre la preferencia entre acampar y pasear en bote

$\boldsymbol{Si} \ |\bar{\boldsymbol{X}}_{\boldsymbol{p}} - \bar{\boldsymbol{X}}_{\boldsymbol{B}}| = \mathbf{6.3} > \boldsymbol{DMS}$ **H**ay diferencias significativas entre la preferencia entre pasear en bote y pescar

i. Datos del computador

Descriptivos

Gastos

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Acampar	6	35,16667	2,71416	1,10805	32,31833	38,01500	32,000	38,000
Pescar	4	30,25000	4,11299	2,05649	23,70532	36,79468	25,000	35,000
Pasear en bote	5	24,20000	6,45755	2,88791	16,18189	32,21811	19,000	35,000
Total	15	30,20000	6,44981	1,66533	26,62822	33,77178	19,000	38,000

Prueba de homogeneidad de varianzas

Gastos

Estadístico de Levene	gl1	gl2	Sig.
1,192	2	12	,337

ANOVA

Gastos

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	328,017	2	164,008	7,737	,007
Intra-grupos	254,383	12	21,199		
Total	582,400	14			

Comparaciones múltiples

Variable dependiente: Gastos

DMS

(I) Preferencia	(J) Preferencia	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Acampar	Acampar					
	Pescar	4,91667	2,97200	,124	-1,55875	11,39209
	Pasear en bote	10,96667*	2,78798	,002	4,89218	17,04115
Pescar	Acampar	-4,91667	2,97200	,124	-11,39209	1,55875
	Pescar					
	Pasear en bote	6,05000	3,08859	,074	-,67945	12,77945
Pasear en bote	Acampar	-10,96667*	2,78798	,002	-17,04115	-4,89218
	Pescar	-6,05000	3,08859	,074	-12,77945	,67945
	Pasear en bote					

*. La diferencia entre las medias es significativa al nivel .05.


36
34
32
30
28
26
24
22
Acampar
Pescar
Pasear en bote
Preferencia

XXVI. REGRESION MULTIPLE

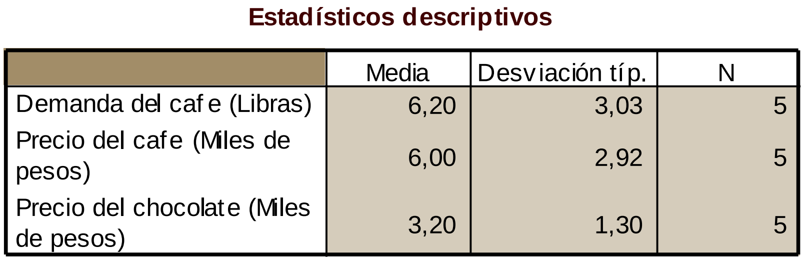
4.8. Análisis de un modelo de regresión múltiple

4.8.1. Ejemplo de la demanda de café

Y ~ Demanda en libras de café
X_1 ~ Precio del café en miles de pesos
X_2 ~ Precio del chocolate en miles de pesos

Y	10	8	5	6	2
X_1	3	5	4	8	10
X_2	5	4	3	2	2

4.8.2. Estadísticos descriptivos



Estadísticos descriptivos

	Media	Desviación típ.	N
Demanda del cafe (Libras)	6,20	3,03	5
Precio del cafe (Miles de pesos)	6,00	2,92	5
Precio del chocolate (Miles de pesos)	3,20	1,30	5

Análisis a través del coeficiente de variación

$$C.V = \frac{S}{\overline{X}} x100$$

4.8.3. Matriz de correlaciones

Correlaciones

		Demanda del cafe (Libras)	Precio del cafe (Miles de pesos)	Precio del chocolate (Miles de pesos)
Correlación de Pearson	Demanda del cafe (Libras)	1,000	-,792	,872
	Precio del cafe (Miles de pesos)	-,792	1,000	-,855
	Precio del chocolate (Miles de pesos)	,872	-,855	1,000
Sig. (unilateral)	Demanda del cafe (Libras)	,	,055	,027
	Precio del cafe (Miles de pesos)	,055	,	,032
	Precio del chocolate (Miles de pesos)	,027	,032	,
N	Demanda del cafe (Libras)	5	5	5
	Precio del cafe (Miles de pesos)	5	5	5
	Precio del chocolate (Miles de pesos)	5	5	5

Análisis de la relación entre las dos variables a través del coeficiente de correlación

$$\rho_{XY} = \frac{n\sum XY-(\sum X)(\sum Y)}{\sqrt{[n\sum X^2-(\sum X)^2][n\sum Y^2-(\sum Y)^2]}}$$

4.8.4. Coeficiente de determinación

Resumen del modelo[b]

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F	
1	,877[a]	,769	,538	2,06	,769	3,325	2	2	,231	3,180

a. Variables predictoras: (Constante), Precio del chocolate (Miles de pesos), Precio del cafe (Miles de pesos)

b. Variable dependiente: Demanda del cafe (Libras)

R^2 = 0.769; Esto indica que el modelo esta explicando un 76.9% de la demanda de café

4.8.5. Tabla ANOVA

$$H_0\text{: Las variables precio del cafe y precio del chocolate no estan aportando a la explicion de la demanda}$$

$$H_1\text{: Alguna de las variables precio del cafe y precio del chocolate esta aportando a la explicion de la demanda}$$

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	28,292	2	14,146	3,325	,231[a]
	Residual	8,508	2	4,254		
	Total	36,800	4			

a. Variables predictoras: (Constante), Precio del chocolate (Miles de pesos), Precio del cafe (Miles de pesos)

b. Variable dependiente: Demanda del cafe (Libras)

$Si\ Sig. < 0.05$ **; Se rechaza H_0**

Para este caso se puede decir que las variables precio del café y precio del chocolate no inciden en la demanda del café

4.8.6. Validación de los parámetros del modelo

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%		Estadísticos de colinealidad	
		B	Error típ.	Beta			Límite inferior	Límite superior	Tolerancia	FIV
1	(Constante)	1,849	8,692		,213	,851	-35,549	39,246		
	Precio del cafe (Miles de pesos)	-,177	,682	-,170	-,259	,820	-3,111	2,757	,269	3,717
	Precio del chocolate (Miles de pesos)	1,691	1,525	,727	1,109	,383	-4,870	8,252	,269	3,717

a. Variable dependiente: Demanda del cafe (Libras)

a. Hipótesis para el precio del café

$H_0: \boldsymbol{\beta}_{Precio\ cafe} = 0$ $H_1: \boldsymbol{\beta}_{Precio\ cafe} \neq 0$

H_0: ***El precio del cafe no aporta a la explicacion de la demanda del cafe***
H_1: ***El precio del cafe esta aportanda a la explicacion de la demanda del cafe***

Criterio de decisión

$Si\ Sig. < 0.05$ **; Se rechaza H_0**

Como el sig. = 0.82 > 0.05; se acepta la hipótesis nula, es decir el precio del café no tiene incidencia en la demanda de este

b. Hipótesis para el precio del chocolate

$H_0: \boldsymbol{\beta}_{Precio\ chocolate} = 0$ $H_1: \boldsymbol{\beta}_{Precio\ chocolate} \neq 0$

H_0: ***El precio del chocolate no aporta a la explicacion de la demanda del cafe***
H_1: ***El precio del chocolate esta aportanda a la explicacion de la demanda del cafe***

Criterio de decisión

$Si\ Sig. < 0.05$ **; Se rechaza H_0**

Como el sig. = 0.383 > 0.05; se acepta la hipótesis nula, es decir el precio del chocolate no tiene incidencia en la demanda del cafe

4.8.7. Estimación del modelo

Demanda de café = 1.849 – 0.177 (Precio del café) + 1.691 (Precio del chocolate)

4.8.8. Interpretación de parámetros

- .La demanda esperada café si el precio del café y el precio del chocolate permanecen constantes seria de 1.849 libras
- Si el precio de café se incrementa en 1000 pesos ; permaneciendo constante el precio del chocolate; a demanda de café decrecería en 0.177 libras
- Si el precio del chocolate se incremente en 1000 pesos , permaneciendo constante el precio del café; la demanda se incrementaría en 1.691 libras

5.1.9 Problemas de los modelos de regresión múltiple

5.1.9.1. Multicolinealidad

La multicolinealidad hace referencia al caso en el que dos o más variables explicativas del modelo de regresión guardan alguna una relación o correlacionadas entre sí, alcanzando una colinealidad imperfecta o perfecta dependiendo de la cercanía que tengan, lo que genera un dificultad para aislar sus efectos individuales sobre la variable dependiente

Algunas causas por las que se produce la multicolinealidad son:

- Restricciones del modelo o en la muestra de la población: Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso (X2) y el tamaño de las viviendas(X3), hay una restricción física en la población puesto que las familias con ingresos más altos generalmente tienen viviendas más grandes que las familias con ingresos más bajos.

- Problema de especificación del modelo: La adición de términos polinomiales a un modelo de regresión, especialmente cuando el intervalo de la variable X es pequeño.

- Modelo sobre determinado. Esto sucede cuando el modelo tiene más variables explicativas que el número de observaciones. Esto podría suceder en investigaciones médicas, donde puede haber un número reducido de pacientes sobre quienes se reúne información respecto de un gran número de variables.

¿Qué consecuencias pueden existir al estimar un modelo en presencia de multicolinealidad?

- En el caso de Multicolinealidad imperfecta los estimadores siguen siendo MELI, pero tienen varianzas y covarianzas grandes que hacen difícil la estimación precisa, mientras que en presencia de multicolinealidad perfecta no se puede obtener una solución única para los coeficientes de regresión individual, lo que produce varianzas y errores estándar infinitos, para mayor simpleza, observe que:

$$\beta_2 \;= 0.8 - 2\,\beta_3$$

Por tanto, es imposible conseguir o calcular los estimadores MCO de los parámetros ya que no habría valores únicos para los parámetros.

Esto surge como consecuencia de que dos o más variables son perfectamente colineales si una o más de las variables explicativas pueden expresarse como una combinación lineal de las otras variables. Por ejemplo, existe una multicolinealidad perfecta entre X1 y X2 si X1 = 2X2 ó X1 = 5 – (1/3) X2. Si dos o más variables explicativas están perfectamente correlacionadas linealmente, será imposible calcular los estimadores MCO de los parámetros (como se vio en el párrafo anterior) porque el sistema de ecuaciones normales incluirá dos o más ecuaciones que no son independientes.

En suma, si la multicolinealidad es perfecta, los coeficientes de regresión de las variables X son indeterminados y sus errores estándar son infinitos por lo que no es dable la estimación.

Si la multicolinealidad es elevada o menos que perfecta, los coeficientes de regresión aunque sean determinados poseen grandes errores estándar lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud.

- ¿Qué ocurre con los intervalos de confianza de acuerdo a lo mencionado en el punto 1? Es lógico pensar que los intervalos de confianza tenderán a ser mucho más amplios ya que si las varianzas son grandes o un número mayor, conllevará a que los errores estándar incrementen la amplitud de los intervalos de confianza. Por consiguiente, se propicia una aceptación más fácil de la hipótesis nula en detrimento de la hipótesis alterna. Lo que en definitiva es una estimación poco precisa.

- ¿Qué ocurriría con la significancia estadísticas de los estimadores en base al punto anterior? Como consecuencia de las grandes varianzas y sus respectivos errores estándar elevados, la razón t de uno o más coeficientes tenderán a ser insignificantes ya que los errores estándar estarían alejados de cero y por tanto el denominador será mucho mayor que el denominador en la prueba t. De ese modo, es plausible encontrar algunos modelos en que la multicolinealidad deje como insignificantes a todos nuestros parámetros.

- A pesar de la insignificancia de algunos parámetros en la prueba t, el R2 puede ser alto.

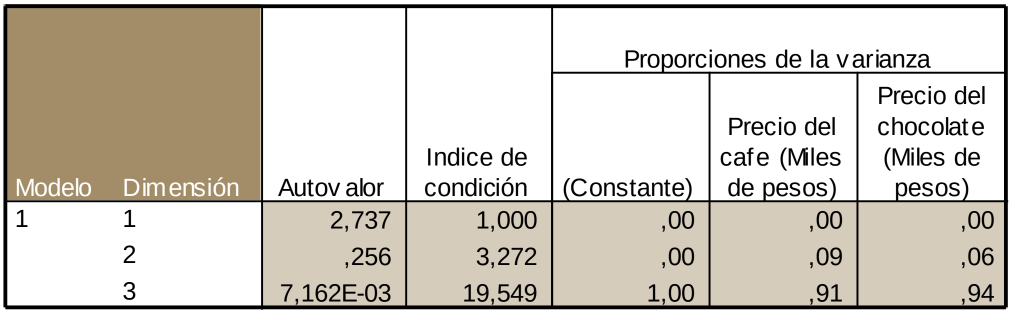
Detección: ¿Qué formas de detección existen?

- El clásico caso de de multicolinealidad se produce cuando ninguna de las variables explicativas de la regresión MCO es estadísticamente significativa (como se mencionó en el punto 3 anterior) y algunas pueden, incluso tener el signo equivocado, en cambio el coeficiente de determinación múltiple, R2, puede ser elevado (por ejemplo entre 0.7 y 1.0), lo que llevará a rechazar la hipótesis nula de que los parámetros son conjuntamente iguales a cero (prueba F). Se genera la disyuntiva entre significancia conjunta y significancia individual.

- En los casos menos evidentes, resulta más difícil detectar la multicolinealidad. A veces se utilizan elevados coeficientes de correlación parcial simple entre las variables explicativas como una medida de multicolinealidad. Sin embargo, es posible que exista una importante multicolinealidad a pesar que los coeficientes de correlación sean reducidos (menores que 0.5). Ergo, es plausible señalar que una correlación elevada son una condición suficiente pero no necesaria para la existencia de multicolinealidad debido a que la colinealidad puede existir en modelos con correlaciones simples muy bajas.

- Realizar las pruebas de FIV y TOL, aunque en la práctica son un poco criticadas.

Diagnósticos de colinealidad[a]



Modelo	Dimensión	Autovalor	Indice de condición	Proporciones de la varianza		
				(Constante)	Precio del cafe (Miles de pesos)	Precio del chocolate (Miles de pesos)
1	1	2,737	1,000	,00	,00	,00
	2	,256	3,272	,00	,09	,06
	3	7,162E-03	19,549	1,00	,91	,94

a. Variable dependiente: Demanda del cafe (Libras)

El índice de condicionamiento, es la raíz cuadrada del cociente entre el máximo y el mínimo autovalores. Si el **IC** es mayor que **30**, existe colinealidad elevada, si el **IC** es mayor que **10** y menor que **30**, la colinealidad es moderada, si el **IC** es menor que 10, no existe colinealidad.

- En resumen, para la detección de la multicolinealidad es importante recalcar que los métodos que se han descrito no son los únicos y no son perfectos. En el caso de la multicolinealidad no se puede hacer mucho al respecto, puesto que la multicolinealidad es un problema específico de una muestra dada sobre la cual el investigador puede no tener mucho control.

Corrección:

¿Qué formas de corrección podríamos deducir en base a lo revisado anteriormente?

No hacer nada puesto que la dificultad de superar la multicolinealidad es un problema en su esencia de deficiencia de los datos con los que se está realizando la regresión y en ocasiones no hay ninguna elección que podamos hacer para intentar corregir la colinealidad, por lo que el alumno no debe confundir el no hacer nada por simple pereza.

Para corregir la Multicolinealidad existen una serie de procedimientos prácticos que nos pueden ser de ayuda, sin embargo, no existen métodos seguros, solamente se pueden usar unas reglas como: 1) combinar datos de series de tiempo con corte transversal, 2) omitir variables y caer en el riesgo de sesgo de especificación, 3) obtener información adicional o nueva, entre otras. No obstante, tales métodos no serán analizados por este apunte.

5.1.9.2. Heterocedasticidad

Uno de los supuestos más importantes en la econometría es el de la homocedasticidad o igual varianza (o dispersión) de cada término de perturbación ui. Simbólicamente:

En contraste, en ocasiones cuando la varianza condicional de Y aumenta a medida que X aumenta, las varianzas de µi son distintas para cada observación, vale decir, las varianzas condicionales de µi dejan de ser constantes, simbólicamente se tiene:

Algunas causas por las que se produce la heterocedasticidad son:

- Curva de aprendizaje. Existen variables donde la adquisición del conocimiento y la acumulación de experiencia puede afectar (en este caso, disminuir) la cantidad de errores a cometer e incluso la magnitud de ellos. En estos casos, la varianza de los errores no es constante. Por ejemplo, para un nivel de producción, el número de errores promedio puede ir disminuyendo si los trabajadores son capacitados o sucede otro hecho importante que afecte el conocimiento y experiencia.

En ocasiones existen modelos en donde se incluyen variables explicativas que no necesariamente respetan la misma relación con la variable dependiente en cada nivel de X. Por ejemplo, a medida que el ingreso crece, las personas tienen más alternativas para elegir dónde destinar sus recursos, por lo tanto, existirán casos de ingresos altos donde la propensión marginal al ahorro se respetará (en promedio esto puede ser efectivamente así) pero seguramente existirán más casos donde no se cumpla. Esto genera variaciones en la varianza a medida que aumenta el ingreso. Por ejemplo, en la economía del deporte, los equipos de futbol que obtienen más ingresos (sea porque ganan campeonatos o tienen mayor número de hinchas) destinarán sus recursos de manera diferente que los otros equipos que recaudan menores ingresos. Es decir, existirá una variabilidad en el comportamiento de los equipos.

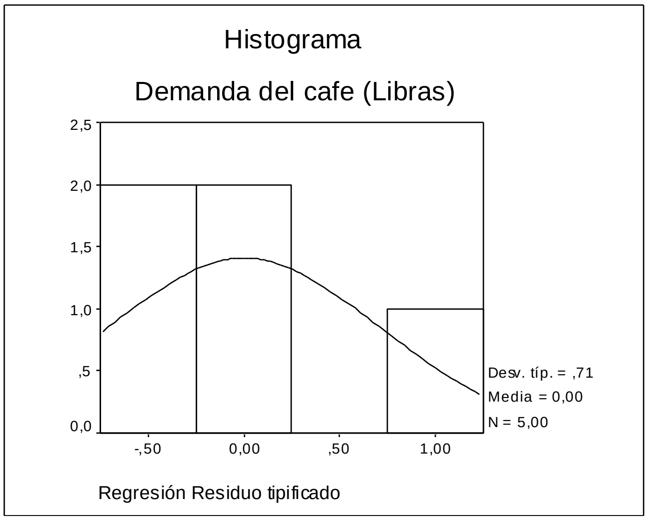
Mejores técnicas de recolección de datos y factores atípicos que están directamente relacionadas con las observaciones de la regresión, ya que pueden alterar sustancialmente los resultados.

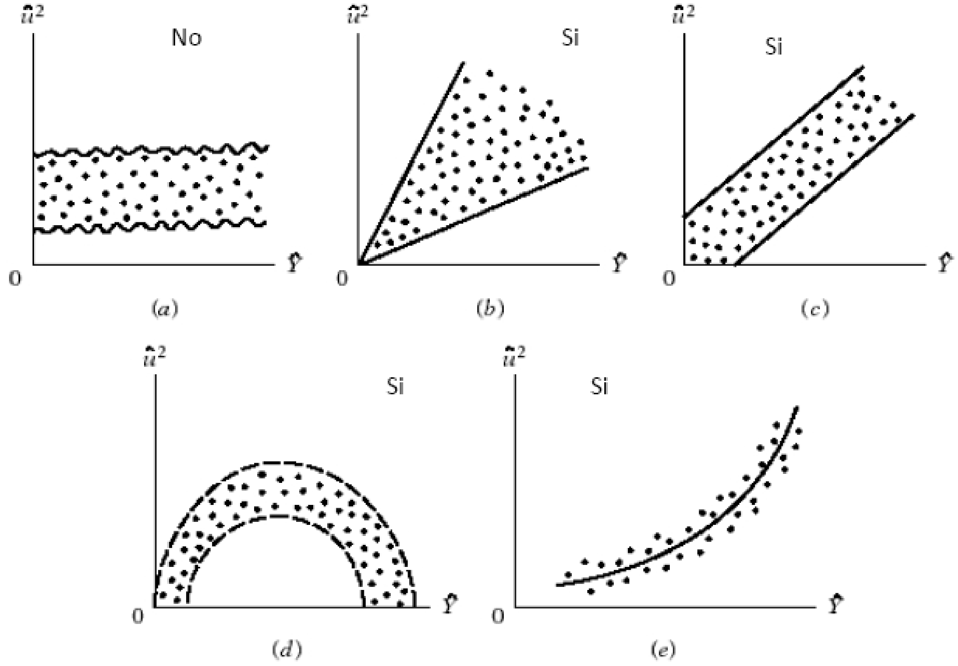
Error de especificación. En ocasiones puede deberse a omisión de variables.

Forma funcional incorrecta.

Detección

A continuación se muestran distintos diagramas de dispersión de los residuos estimados al cuadrado:





5.1.9.3. Autocorrelacion

La autocorrelación o correlación serial entre las perturbaciones (ui) se presenta tradicionalmente cuando se trabaja con series de tiempo, puesto que la autocorrelación hace referencia al caso en el que el término de error de un período está correlacionado con el término de error de cualquier otro período. Si el término de error de un período está correlacionado con el término de error del período anterior, estamos ante un caso de autocorrelación de primer orden. La mayoría de las aplicaciones de econometría implican autocorrelaciones de primer orden más que de segundo orden o de orden superior.

Además es posible que se produzca un caso de autocorrelación negativa. Aunque la mayoría de las series temporales económicas muestran casos de autocorrelación positiva.

Para comprender mejor el porqué la correlación se presenta más en los las series de tiempo, se presenta el siguiente caso. Por ejemplo, cuando un investigador toma datos de unidades transversales como de una casa o una empresa es difícil creer que exista una razón para que el término de error que presenta a un propietario de una empresa esté correlacionado con el término de error que presenta otro propietario. En cambio, las series de tiempo al seguir un ordenamiento a través del tiempo, es muy probable que las observaciones estén correlacionadas. Para diferenciar la correlación entre las perturbaciones de las unidades transversales de las de series de tiempo, a las primeras se les denomina correlación espacial y a la segunda, serial.

Simbólicamente, hay evidencia de correlación cuando el término de error del modelo está correlacionado consigo mismo a través del tiempo (o en el espacio en el caso de datos de corte transversal), esto es:

$E\ (\mu i, \mu j) \neq 0\ ;\ i \neq j$

Violando el supuesto que se mantenía en el principio:

E (µi, µj) = 0 ; i j

Ilustrando de mejor manera esta violación del supuesto, consideremos la tasa de delincuencia regresada sobre la tasa de detenciones en los distintos estados de los Estados Unidos. A priori, se puede pensar que las perturbaciones de cada unidad (en este caso un estado) no debiera tener una correlación con las perturbaciones de las otras unidades, lo que reflejaría el supuesto de no existencia de autocorrelación (espacial en este caso).

Sin embargo, a posteriori es factible encontrar una correlación entre las perturbaciones siendo que los incrementos de la tasa de delincuencia de un estado pueda inducir a algunas personas [de otro estado] a delinquir y por tanto, a aumentar la tasa de delincuencia de otro estado.

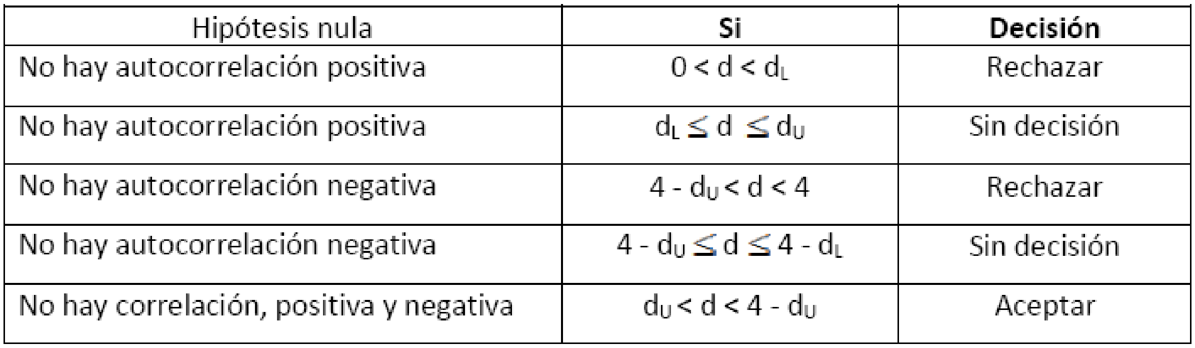
Entonces, la pregunta que surge es ¿qué tan sustancial es la existencia de autocorrelación en un modelo econométrico? Brevemente, en presencia de autocorrelación, las estimaciones MCO de los parámetros si bien siguen siendo insesgadas y consistentes, los errores estándar de los parámetros estimados de la regresión son sesgados, lo que da lugar a contrastaciones estadísticas incorrectas y a intervalos de confianza sesgados.

Detección

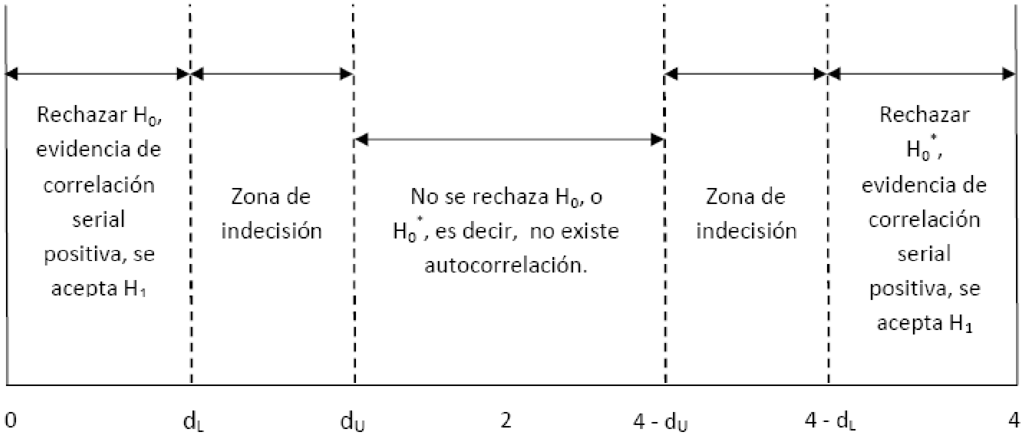
Prueba de Durbin-Watson (DW)

$$d \approx 2\,(1 - \hat{\rho})$$

$$0 \leq d \leq 4$$



Hipótesis nula	**Si**	**Decisión**
No hay autocorrelación positiva	$0 < d < d_L$	Rechazar
No hay autocorrelación positiva	$d_L \leq d \leq d_U$	Sin decisión
No hay autocorrelación negativa	$4 - d_U < d < 4$	Rechazar
No hay autocorrelación negativa	$4 - d_U \leq d \leq 4 - d_L$	Sin decisión
No hay correlación, positiva y negativa	$d_U < d < 4 - d_U$	Aceptar



H_0: No hay Autocorrelacion positiva

H_0*: No hay Autocorrelacion negativa

BIBLIOGRAFIA

BÁSICA: OROBIO Luís Eduardo. Notas de Clase realizadas por el docente, 2.000

Programas de Ingeniería

- Montgomery, Runger. Probabilidad y Estadística aplicada a la Ingeniería. 2ª Edición. Limusa Wiley.
- Walpole, Myers, Myers. Probabilidad y Estadística para Ingenieros. 6ª Edición. Pearson Educación.
- Jay, Devore. Probabilidad y Estadística para Ingenierías y Ciencias. 5ª Edición, Thomson.
- Ross. Probabilidad y Estadística para Ingenieros. 2ª Edición. McGraw-Hill.
- Milton, Arnold. Probabilidad y Estadística con aplicaciones para Ingeniería y Ciencias Computacionales
- Behar, Yepes. Estadística un Enfoque Descriptivo.2ª Edición. Universidad del Valle.
- Béhar,Grima. Eniístkd Apurada. Universidad de! Valle, Universidad Politécnica de Cataluña.

Otros programas

- MILLER, Irvin. Probabilidad y Estadística par Ingenieros. Prentice may
- BERENSON Mark L. , LEVINE David M., Estadística Básica en Administración , , , Editorial Prentice Hall, Hispanoamérica, S.A. México, 1997. Sexta Edición.
- WEBSTER Allen J. . Estadística Aplicada a los Negocios y la Economía;. Tercera edición. Editorial Mc Graw Hill. 2.000
- MONTIEL A. M. – RIUS F. Conceptos Básicos de Estadística Económica y Empresarial;. 1ª. Editorial Prentice Hall. 1.996
- MASÓN / LIND / MARCHAL. Estadística para Administración y Economía;. Décima edición. Editorial Alfa Omega. 2.000
- NEWBOLD Paúl. Estadística para Negocios y la Economía;. Cuarta edición. Editorial Prentice Hall. 1.997
- HANKE John. Estadística para Negocios ;. Segunda edición. Editorial McGraw Hill. 1.995 MENDENHALL William, Reinmuth. BEAVER J. Roberth. Introducción a la Probabilidad y Estadística, , Editorial Thompson, 2002, Mx.
- SPIEGEL, Murray R. Probabilidad y Estadística, Editorial Mc Graw-Hill, México, 1991.
- PEREZ Cesar. TÉCNICAS ESTADÍSTICAS CON SPSS;. Primera edición. Editorial Prentice Hall. 2.001
- KENNETH N. Berk – CAREY Patrick. Análisis de Datos con Microsoft EXCEL - Primera edición - Editorial Thompson. 2001. México.

Web grafía

- http://www.virtual.unal.edu.co/cursos/ciencias/2001065/index.html
- http://estadisticautil.es.fm/. El objetivo de esta página es dar a conocer la utilidad de la estadística para el mundo empresarial y para la vida cotidiana.
- http://www.hrc.es/bioest/estadis1.html. Aquí encontrará definiciones, explicaciones y ejercicios relacionados con la estadística descriptiva.
- http://engineering.uow.edu.au/courses/stats/. Página con definiciones y explicaciones relacionados con probabilidad y estadística.
- Http://ubmail.ubalt.edu/-harsham/businessstat/opre504.htm#rdischthought. Herramientas estadística en internet, también podrá encontrar por medio de una tabla interactiva los valores de las funciones chi-cuadrada, t-student, F y Normal.
- http://www.cortlan.edu/flteach/stats/stat-sp.html. Aquí encontrará definiciones, explicaciones y ejercicios relacionados con la estadística descriptiva y las probabilidades.
- http://www.cortland.edu/flteach/stats/links.html. El objetivo de esta página es dar a conocer una serie de enlaces en internet relacionados con ciencias básicas en los cuales encontrará algunos dedicados tanto a la estadística descriptiva como a la inferencial.