Universidad Continental

Estadística Inferencial

Sergio Jurado Chamorro

Datos de Catalogación Bibliográfica

Estadística Inferencial. Manual Autoformativo /
Sergio Jurado Chamorro
- Huancayo:
Universidad Continental. 2017. - 148 p.
Datos de catalogación del CENDOC UC

Estadística Inferencial. Manual Autoformativo
Autor: Sergio Jurado Chamorro
Primera edición
Huancayo, mayo de 2017
De esta edición
© Universidad Continental
Av. San Carlos 1980, Huancayo-Perú
Teléfono: (51 64) 481-430 anexo 7361
Correo electrónico: recursosucvirtual@continental.edu.pe
http://www.continental.edu.pe/
Versión e-book
Disponible en http://repositorio.continental.edu.pe/
ISBN electrónico n.° 978-612-4196-
Dirección: Emma Barrios Ipenza
Edición: Eliana Gallardo Echenique
Miguel Angel Cordova Solis
Asistente de edición: Andrid Kary Poma Acevedo
Asesor didáctico: Fabio Contreras Oré
Corrección de textos: Sara Maricruz Bravo Montenegro
Diseño y diagramación: Gerardo Favio Quispe Fernández

ÍNDICE

INTRODUCCIÓN

En el presente manual se presentan los contenidos a desarrollar y los aprendizajes esperados que se debe lograr en nuestra asignatura. El desarrollo de la asignatura requiere de conocimientos previos desarrollados en la asignatura de Probabilidad y Estadística.

Nuestra asignatura desarrolla procedimientos, estrategias, y cálculos especializados en el análisis de datos masivos, resumidos en una muestra. Lo que podamos determinar en la muestra será llevado hacia la población bajo ciertos márgenes de probabilidad. Este proceso de estudiar una muestra y luego trasladar los resultados, con ayuda de las distribuciones de probabilidades, a su población, constituye lo que técnicamente se conoce como inferencia.

En el proceso de las inferencias trabajamos con estadísticos con la finalidad de estimar los parámetros en una población, razón por la que es necesario tener muy presentes los conceptos y desarrollo de cálculos como la media, la mediana, la varianza, entre otras medidas estadísticas.

Se requiere, además, reconocer las propiedades de las distribuciones de probabilidades. Entre ellas, las que más nos interesan son la distribución normal y la binomial.

El manual se organiza en cuatro unidades y ha sido diseñado con la intención de generar el autoaprendizaje con descripción de la teoría, tomando como base el libro de Mario Triola, "Estadística" (10ª ed.). Se ha previsto el desarrollo de la solución de un ejercicio como mínimo, para cada uno de los temas tratados. Se han insertado algunas lecturas complementarias en los cuadros de texto de los lados en algunas páginas, con la finalidad de propiciar la reflexión sobre el uso de la estadística inferencial.

Sergio Jurado Chamorro

ORGANIZACIÓN DE LA ASIGNATURA

Resultado de aprendizaje de la asignatura

Al término de la asignatura, el estudiante será capaz de aplicar métodos y técnicas de la estadística inferencial y de pronóstico con el objetivo de proporcionar información pertinente y veraz que sirva para la toma de decisiones.

Unidades didácticas

UNIDAD I	UNIDAD II	UNIDAD III	UNIDAD IV
Muestreo y estimación de parámetros.	Prueba de hipótesis para una y dos muestras y análisis de varianza.	Experimentos multinomiales, tablas de contingencia y estadística no paramétrica.	Correlación y regresión. Series de tiempo.
Resultado de aprendizaje	**Resultado de aprendizaje**	**Resultado de aprendizaje**	**Resultado de aprendizaje**
Al finalizar la unidad, el estudiante será capaz de aplicar métodos de muestreo y calcular los parámetros poblacionales con datos provenientes de una o dos poblaciones.	Al finalizar la unidad, el estudiante será capaz de plantear y aplicar pruebas de hipótesis para la media, proporción y varianza a partir de situaciones estadísticas reales.	Al finalizar la unidad, el estudiante estará en la capacidad de realizar pruebas de hipótesis con experimentos multinomiales, pruebas de bondad o pruebas no paramétricas de acuerdo a la situación estadística planteada.	Al final de la unidad, el estudiante estará en la capacidad de realizar pronósticos utilizando el análisis de correlación y regresión, así como modelos de series de tiempo.

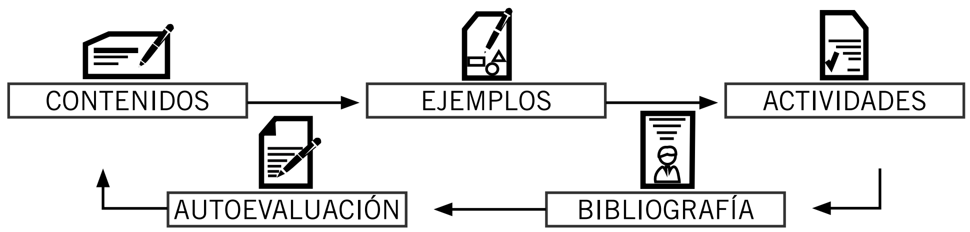
Tiempo mínimo de estudio

UNIDAD I	UNIDAD II	UNIDAD III	UNIDAD IV
Semana 1 y 2	Semana 3 y 4	Semana 5 y 6	Semana 7 y 8
16 horas	16 horas	16 horas	16 horas

UNIDAD I

MUESTREO Y ESTIMACIÓN DE PARÁMETROS

DIAGRAMA DE ORGANIZACIÓN DE LA UNIDAD I



ORGANIZACIÓN DE LOS APRENDIZAJES

Resultados del aprendizaje de la Unidad I: Al finalizar la unidad el estudiante será capaz de aplicar métodos de muestreo y calcular los parámetros poblacionales con datos provenientes de una o dos poblaciones.

CONOCIMIENTOS	HABILIDADES	ACTITUDES
Tema n.° 1: Diseño y técnicas de muestreo 1. Definiciones básicas. 2. Tipos de muestreo probabilístico. **Lectura seleccionada n.° 1** La pobreza en el Perú disminuyó en 1.2%. **Tema n.° 2: Estimación de parámetros** 1. Estimación de la proporción. 2. Estimación de la media. 3. Estimación de la varianza. 4. Determinación del tamaño muestral.	1. Distingue los métodos de muestreo. Observa las diapositivas animadas y elabora un organizador gráfico comparativo. 2. Planifica muestreos probabilísticos. Elabora una ficha técnica de muestreo. 3. Selecciona una muestra válida para realizar estimaciones de parámetros. 4. Identifica correctamente los valores críticos para el cálculo de intervalos de confianza. 5. Calcula intervalos de confianza para la media, proporción y varianza para una y dos muestras. **Actividad 1** Participa del foro de discusión sobre criterios de muestreo. **Actividad 2** Evaluación del tema n.° 1 y el tema n.° 2.	Valora la importancia del muestreo y de la estimación de parámetros e interpreta correctamente los resultados para una buena toma de decisiones.

Diseño y técnicas de muestreo

Tema n.° 1

1. DEFINICIONES BÁSICAS

1.1. Unidad elemental

Es el elemento mínimo en el que se pueden observar/medir las características (variables) de un estudio estadístico; es decir, es el elemento básico del que se puede obtener información (datos) para las variables.

Ejemplo:

- En un programa de control de calidad en una fábrica de calzado, la unidad elemental es cada zapato en algún punto específico del proceso de fabricación, por ejemplo, a la salida de la línea de fabricación.
- En estudios de satisfacción del cliente en una tienda de *retails*, la unidad elemental es cada consumidor.
- En un estudio sobre los estilos de aprendizaje en nuestra universidad, la unidad elemental es cada estudiante matriculado en el semestre actual.

Figura 1. Control de calidad. Tomada de Perú 21, 2014.

1.2. Población muestreada

Es el conjunto de todas las unidades elementales; es decir, todos aquellos elementos de quienes se puede obtener información. Constituye, entonces, el conjunto del que se puede extraer una muestra.

La población muestreada compuestos por personas, hogares, objetos, animales, medidas, los que se reconocen como unidades elementales.

Figura 2. Población. Tomada de: http://ngooipin.blogspot.pe/ , 2008.

1.3. Censo

Es el estudio de toda una población. Requiere de mucho tiempo y disponibilidad de recursos. Se estila a realizarse con poblaciones alcanzables; es decir, aquellas en las que es muy fácil acceder a cada una de las unidades elementales. Depende del espacio y el tiempo en el que se realiza, dado que ninguna población es estática.

1.4. Muestra

Es el subconjunto de unidades estadísticas o unidades elementales seleccionadas de la población de manera tal que se puede decir que es representativa de la población, es decir, que tiene las mismas características que la población en calidad y proporción de individuos.

1.5. Unidad de muestreo

Es cualquiera de las unidades estadísticas elegidas para formar parte de la muestra. Su elección constituye la parte más importante de un estudio estadístico. Se debe realizar siguiendo una secuencia de pasos que garanticen su representatividad.

Figura 3. Muestreo aleatorio. Tomado de Keyword & Suggestions, 2015.

En ocasiones una unidad de muestreo es la unidad elemental y es elegida de manera directa, como el caso de control de calidad mencionado líneas arriba; en otras, la unidad elemental está dentro de la unidad de muestreo, como el caso de una investigación en las que se eligen hogares como unidades de muestreo, pero se considera como unidad elemental, por ejemplo, a quienes aportan o sostienen económicamente estos hogares.

1.6. Marco muestral

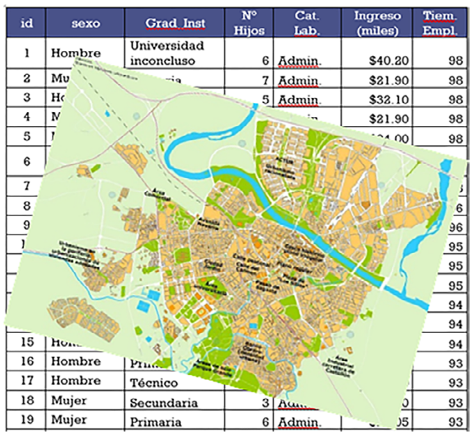


Figura 4. Marco muestral. Elaboración propia.

Un marco muestral constituye un documento actualizado que nos da la oportunidad de reconocer a todas las unidades estadísticas. Por ello, un marco muestral puede estar integrado de una lista, si se trata, por ejemplo, de un estudio sobre el uso de las tarjetas de débito en el mercado financiero, y/o un mapa en el caso de un estudio de hogares.

El marco muestral nos permite realizar el proceso de elección de las unidades muestrales que se denomina muestreo.

1.7. Muestreo

Es un conjunto de métodos y procedimientos estadísticos destinados a la selección de una o más muestras; es decir, es la técnica seguida para elegir muestras. El objetivo principal de un diseño de muestreo es proporcionar procedimientos para la selección de muestras que sean representativas de la población en estudio (INEI, 2006, p. 46).

El muestreo comprende por lo menos dos etapas:

- La selección de las unidades muestrales.
- El registro de las observaciones (Cerrón, 2013).

1.8. Ficha técnica

Una ficha técnica es el documento que obligatoriamente se presenta al presentar los resultados de una encuesta. En este documento se expone las características del estudio realizado que respaldan la coherencia de la información obtenida en la muestra.

Ejemplo 1:

Título del estudio: Encuesta de Opinión en Lima Metropolitana, noviembre 2009.

Objetivos del estudio: Evaluación y opinión sobre la situación económica.

Encuestadora: Pontificia Universidad Católica del Perú

N.º de registro: 0108 REE/JNE

Universo o población objetivo: Hombres y mujeres mayores de 18 años, habitantes de 31 distritos de Lima Metropolitana

Marco muestral: La selección de manzanas se hizo utilizando como marco muestral la cartografía digital del INEI del 2004 para los 31 distritos de Lima Metropolitana. Los distritos que no forman parte del marco muestral son los siguientes: Chaclacayo, Lurigancho, Cieneguilla y los distritos balnearios del Sur y del Norte de la Ciudad.

Representatividad: En los distritos que forman parte del universo y que están incluidos en el marco muestral se encuentra el 95.88% de la población electoral total de la provincia de Lima.

Tamaño de la muestra: 508 personas entrevistadas en Lima Metropolitana

Error y nivel de confianza estimados: ±4.32% con un nivel de confianza del 95%, asumiendo 50%-50% de heterogeneidad, bajo el supuesto de muestreo aleatorio simple.

Distritos que resultaron seleccionados en la muestra: La selección aleatoria de manzanas del marco muestral determinó que la encuesta se aplicará en 28 distritos de Lima Metropolitana (Cercado de Lima, Ate, Barranco, Breña, Carabayllo, Chorrillos, Comas, El Agustino, Jesús María, La Molina, La Victoria, Lince, Los Olivos, Magdalena del Mar, Pueblo Libre, Miraflores, Puente Piedra, Rímac, San Borja, San Juan de Lurigancho, San Juan de Miraflores, San Martín de Porres, San Miguel, Santa Anita, Santiago de Surco, Surquillo, Villa El Salvador y Villa María del Triunfo).

Procedimiento de muestreo: Se realizó una muestra probabilística polietápica. Dentro de Lima se estratificó la muestra de acuerdo con grandes zonas de la ciudad: cono norte, cono este, cono sur, centro, cono oeste-suroeste, y en cada estrato se seleccionó una muestra simple al azar de manzanas. Posteriormente se realizó un muestreo sistemático de viviendas en cada manzana seleccionada y se aplicaron cuotas de sexo y edad para la selección de personas al interior de las viviendas.

Ponderación: En Lima Metropolitana los datos se ponderaron en función del peso de los estratos en la población total.

Técnica de recolección de datos: Mediante entrevistas directas en las viviendas seleccionadas.

Supervisión de campo: Se supervisó el 30% de las entrevistas realizadas.

Fechas de aplicación: Entre los días 29 de octubre y 01 de noviembre de 2009.

Financiamiento: Pontificia Universidad Católica del Perú.

Página web: http://www.pucp.edu.pe

Email: iop@pucp.edu.pe

1.9. Plan de muestreo

Se considera plan de muestreo a la definición y sistematización de la secuencia de pasos y las consideraciones para definir el tamaño de una muestra y cómo obtenerla.

Entonces, como el objetivo es la obtención de una muestra, se emplean técnicas específicas, fórmulas y tablas.

1.10. Tipos de muestreo

Existen dos formas conocidas de realizar un muestreo: Probabilísticamente y no probabilísticamente.

Muestreo probabilístico: Desarrolla una serie de estrategias y procedimientos que incluyen la elección aleatoria de las unidades muestrales, de modo que procura que todas y cada una de las unidades elementales tengan la misma probabilidad de ser elegidas.

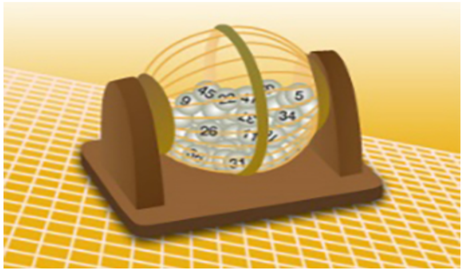


Figura 5. Muestreo aleatorio .
Tomada de EducaLab, 2016.

Muestreo no probabilístico: Este tipo de muestreo debe realizar una elección de unidades elementales, generalmente basada en el criterio del que investiga. No se puede asegurar, por tanto, que todas las unidades elementales tengan la misma probabilidad de ser elegidas.

Figura 6. Muestreo por criterio. Tomada de Graphic Resources LLC, 2015.

La diferencia fundamental entre estos tipos de muestreo radica en que en el muestreo probabilístico se puede medir el riesgo que se asume al muestrear, mientras que en el muestreo no probabilístico ello no es posible. En el estadístico, la fundamentación es teórica y las conclusiones pueden ser cuantitativas y precisas, mientras que en el no estadístico la fundamentación es práctica y las conclusiones pueden ser cualitativas y, en el mejor de los casos, razonables, además de perder ante determinados casos la posibilidad de ser extrapolables. Ante estas diferencias, ¿por qué recurrir al muestreo no probabilístico? En algunas circunstancias sucede que no es posible por falta de tiempo, por escasez de recursos, por limitaciones para acceder a la población u otras dificultades operativas llevar a cabo un muestreo probabilístico. Será preferible, entonces, acudir a determinadas pruebas específicas, para los que habrán de tenerse en cuenta ciertos cuidados mínimos y necesarios para sostener fundadamente las opiniones vertidas en un informe (Rubione, 2013, p. 4-5).

2. TIPOS DE MUESTREO PROBABILÍSTICO

2.1. Muestreo aleatorio simple

Es el método más simple para elegir la muestra. Básicamente es un sorteo realizado de tal manera que se garantice que cada una de las unidades elementales tenga la misma probabilidad conocida.

La elección, por lo tanto, es aleatoria y se puede llevar a cabo mediante un ánfora con balotas con números o identificadores de cada uno de elementos de la población. También, puede realizarse usando una tabla denominada Tabla de números aleatorios y la lista completa de los elementos de la población.

Figura 7. Muestreo por criterio. Tomada de Graphic Resources LLC, 2015.

Se puede mencionar dos cualidades:

- Independencia: La elección de cada unidad de muestreo es independiente de la elección de otra, dicho de otra manera, la elección de un elemento no influye en la probabilidad de elección del resto de unidades. Esto es posible si la elección se realiza con reemplazo.
- Representatividad: El conjunto de elementos elegidos es un fiel reflejo de su población.

Es muy eficiente con poblaciones pequeñas, desarrolla una muestra representativa sobre todo si la población es bastante homogénea. A cambio se puede asegurar que no es tan eficiente con poblaciones heterogéneas, dado que la elección aleatoria podría producir sesgo y no se puede realizar con poblaciones infinitas.

Ejemplo 2:

Los ingresos anuales, grado de instrucción, la categoría laboral de los 120 empleados de Mariana S. A., empresa dedicada a la producción de ropa para bebe, se muestran en el Apéndice A Tabla de datos de empleados de Mariana S. A.

Con esta base de datos, obtendremos una muestra aleatoria de tamaño 10 utilizando la Tabla de Números aleatorios, del Apéndice B.

En la tabla de números aleatorios, elegimos de manera aleatoria las columnas C7, C2, C11.

Datos:

N = 120 (tamaño de población)

n = 10 (tamaño de muestra)

De esta manera, seguimos buscando y los elegidos son los siguientes: 012; 014; 092; 007

1.° Tomamos la columna C7 y le añadimos 2 columnas consecutivas, C8 y C9.

Esto porque N = 120 es un número que tiene 3 dígitos.

2.° Buscamos en las tres columnas números menores o iguales a 120. Por ejemplo, 012 = 12.

El primer elemento elegido es el empleado n.° 12:

12	Mujer	Secundaria

Tabla n.ª 1. Números aleatorios.

C7	C8	C9
4	0	9
0	1	2
5	6	7
5	0	6
2	4	4
4	5	9
3	2	0
3	3	7
3	9	3
7	3	1
8	1	0

Fuente: Tomada de Manual autoformativo de Estadística II Cerrón, 2013.

De esta manera, seguimos buscando y los elegidos son los siguientes:

012; 014; 092; 007 Tabla 1.

3.° Hacemos lo mismo con la columna C2 a la que añadimos C3 y C4, y se obtiene lo siguiente:

036; 058

Con C11, añadimos C12 y C13:

086; 011; 088; 067

Por tanto, los elementos elegidos son 12, 14, 92, 7, 36, 58, 36, 58, 86, 11, 88, 67.

2.2. Muestreo sistemático

Se realiza mediante un proceso que se desarrolla mediante los siguientes pasos:

1.° Se calcula el tamaño de salto o intervalo a dar en la elección de los elementos que se muestran a continuación:

$K = \frac{N}{n}$, donde N es el tamaño de la población y n es el tamaño de la muestra a elegir.

2.° De manera aleatoria, se elige al primer elemento entre los "k primeros". Se puede usar la tabla de números aleatorios.

3.° Se elige a los siguientes dando "saltos" de k en k elementos.

Figura 8. Muestreo sistemático.

Este tipo de muestreo tiene la ventaja de una mayor posibilidad de elegir una muestra representativa. La elección de cada elemento es independiente de la probabilidad de elección de otros elementos.

En su contra, se puede mencionar que la población debe estar ordenada como en una lista. Si existe alguna forma de periodicidad en la lista, entonces el sesgo es mucho mayor que en el muestreo aleatorio simple. Esto puede producirse, por ejemplo, en datos de ventas que son muy susceptibles a las fechas de pago en las empresas.

Es muy sensible a la influencia de circunstancias sistemáticas, es decir, a sucesos que inciden en el desarrollo de los datos de manera indirecta como, por ejemplo, un engranaje defectuoso y que este defecto produzca cambios en la producción.

Ejemplo 3:

Del caso de la empresa Mariana S. A., desarrolle un muestreo para elegir una muestra de 10 personas de la lista de 120 empleados.

Datos:

N = 120

n = 10

1.º Calculamos el intervalo de selección así:

$$K = \frac{N}{n} = \frac{120}{10} = 12$$

2.º Elegimos el punto de partida.

Para ello, se debe elegir un elemento de los primeros k = 12. Luego, tomamos en la tabla de números aleatorios la columna C5 y C6 (aleatoriamente), y buscamos el primer número menor o igual a 12. El resultado fue el siguiente:

1.^{er} elegido = 04

La primera persona elegida es la que ocupa el cuarto lugar en la lista tal como se aprecia en la siguiente tabla:

Tabla n.º 2. Lista de datos de empleados (tabla de datos Mariana)

id	Sexo	Grand Inst	N.º de Hijos	Cat. Lab.	Ingreso (miles)	Tiemp. Empl.
1	Hombre	Universidad inconcluso	6	Admin.	$40.20	98
2	Mujer	Primaria	7	Admin.	$21.90	98
3	Hombre	Técnico	5	Admin.	$32.10	98
4	Mujer	Secundaria	2	Admin.	$21.90	98
5	Mujer	Secundaria	6	Admin.	$24.00	98
6	Mujer	Universidad inconcluso	4	Admin.	$30.30	98
7	Hombre	Técnico	3	Admin.	$27.75	98

3.° A partir de esta persona, se eligen los siguientes casos dando saltos con k = 12.

Tabla n.° 3. Tabla de datos Mariana.

id	Sexo	Grand Inst	N° de Hijos	Cat. Lab.	Ingreso (miles)	Tiemp. Empl.
1	Hombre	Universidad inconcluso	6	Admin.	$40.20	98
2	Mujer	Primaria	7	Admin.	$21.90	98
3	Hombre	Técnico	5	Admin.	$32.10	98
4	Mujer	Secundaria	2	Admin.	$21.90	98
5	Mujer	Secundaria	6	Admin.	$24.00	98
6	Mujer	Universidad inconcluso	4	Admin.	$30.30	98
7	Hombre	Técnico	3	Admin.	$27.75	98
8	Mujer	Primaria	5	Admin.	$31.35	96
9	Hombre	Técnico	2	Admin.	$31.35	96
10	Hombre	Secundaria	3	Admin.	$23.25	95
11	Mujer	Técnico	2	Admin.	$22.35	95
12	Mujer	Secundaria	8	Admin.	$30.00	95
13	Hombre	Secundaria	4	Admin.	$35.55	94
14	Hombre	Secundaria	8	Admin.	$25.05	94
15	Hombre	Primaria	7	Admin.	$22.50	94
16	Hombre	Primaria	6	Admin.	$21.90	93
17	Hombre	Técnico	4	Admin.	$41.10	93
18	Mujer	Secundaria	3	Admin.	$26.40	93
19	Mujer	Primaria	6	Admin.	$25.05	93
20	Mujer	Secundaria	4	Admin.	$28.50	92
21	Hombre	Técnico	8	Admin.	$33.45	90
22	Mujer	Técnico	2	Admin.	$32.55	90
23	Mujer	Secundaria	8	Admin.	$33.30	90
24	Hombre	Secundaria	4	Admin.	$27.30	90
25	Mujer	Técnico	2	Admin.	$26.65	88
26	Hombre	Universidad inconcluso	2	Admin.	$52.65	86
27	Hombre	Secundaria	7	Admin.	$26.70	86
28	Hombre	Técnico	1	Admin.	$37.50	84
29	Mujer	Secundaria	4	Admin.	$16.50	84
30	Mujer	Secundaria	8	Admin.	$24.75	84
31	Mujer	Secundaria	5	Admin.	$24.00	83
32	Mujer	Secundaria	8	Admin.	$20.40	83
33	Hombre	Técnico	2	Admin.	$30.15	82
34	Mujer	Secundaria	0	Admin.	$33.90	82
35	Hombre	Técnico	8	Admin.	$22.50	82
36	Mujer	Primaria	5	Admin.	$27.45	81
37	Mujer	Secundaria	3	Admin.	$27.30	81
38	Mujer	Secundaria	6	Admin.	$23.10	81
39	Mujer	Secundaria	2	Admin.	$23.10	81
40	Hombre	Secundaria	3	Admin.	$25.50	81
41	Hombre	Primaria	5	Admin.	$21.30	80
42	Mujer	Secundaria	3	Admin.	$23.40	80
43	Hombre	Técnico	2	Admin.	$28.65	79
44	Hombre	Universidad inconcluso	5	Admin.	$40.35	78
45	Hombre	Técnico	2	Admin.	$25.95	78
46	Hombre	Universidad inconcluso	3	Admin.	$26.55	78
47	Hombre	Técnico	6	Admin.	$30.75	77
48	Hombre	Técnico	6	Admin.	$34.60	77

Así resultan elegidos los siguientes casos:

4, 16, 28, 40, 52, 64, 76, 88, 100, 112.

2.3 Muestreo por estratos

Cuando las poblaciones son extensas y/o se quiere asegurar la representatividad de los grupos en la muestra, entonces la población se separa en estratos:

Los estratos se pueden formar de acuerdo con características (variables) muy importantes para el desarrollo de la investigación. Por ejemplo, en el caso de una encuesta de opinión, se estila separar a las personas en estratos como sus ingresos, nivel educativo, número de hijos.

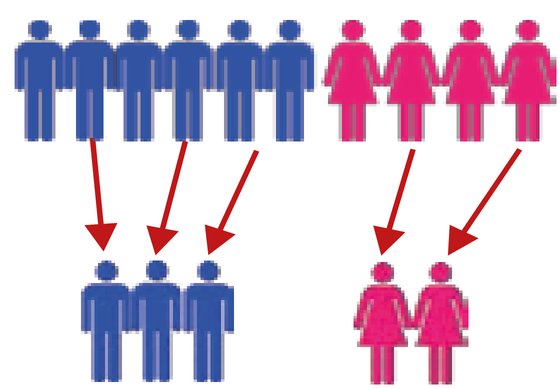


Figura 9. Muestreo por estratos.

Una vez que se han definido los estratos, se procede a un muestreo aleatorio simple o sistemático en cada estrato.

La elección puede darse tras obtener una muestra de igual tamaño de cada estrato o al tomar una cantidad proporcional de cada uno de ellos. Luego, se puede o no otorgar pesos relativos de acuerdo con la significancia de cada estrato.

Estas son algunas ventajas de este tipo de muestras:

- Una mayor representatividad.
- Una mejor precisión en las estimaciones.
- Menor costo en el recojo de datos.
- Ayuda a obtener estimaciones por estratos.

El procedimiento básico por excelencia es el siguiente:

a. Desarrolle una lista de variables que sean de importancia para definir los estratos.

b. Construya los estratos cuidando que sean mutuamente excluyentes; es decir, ningún elemento de la población puede pertenecer a dos estratos.

c. Determine el número de elementos a elegir en cada estrato. Si se trata de una elección proporcional, entonces se puede usar la siguiente fórmula:

$$n1 = n\left(\frac{N1}{N}\right)$$

$$n2 = n\left(\frac{N2}{N}\right)$$

$$n3 = n\left(\frac{N3}{N}\right)$$

donde

N = Tamaño de la población

N1, N2, N3... son los tamaños de los estratos en la población.

n = Tamaño de muestra

n1, n2, n3... son los tamaños de muestra a obtener de cada estrato.

d. Seleccione de preferencia muestras aleatorias de cada estrato.

Es recomendable tener en cuenta que muchos estratos generarían trabajo innecesario y redundante, y un número pequeño no ayudaría a reducir el sesgo. Se recomienda usar mínimo tres y máximo ocho estratos.

Ejemplo 4:

Con los datos del ejemplo n.° 1, se requiere realizar la elección de una muestra de 30 individuos.

Datos:

N = 120 (población)

n = 30 (muestra)

Resumiendo, los datos:

Tabla n.° 4. Muestra de individuos y categoría laboral.

Categoría Laboral	Ni	
Directivo	22	= N1
Administrativo	90	= N2
Seguridad	8	= N3
N =	**120**	

Tabla n.° 5. Categoría laboral y resultados de n.

Categoría Laboral	n $\left(\frac{N_i}{N}\right)$			
Directivo	$30\left(\frac{22}{120}\right)=$	5,499	➔	n1= 5
Administrativo	$30\left(\frac{90}{120}\right)=$	22,5	➔	n2= 23
Seguridad	$30\left(\frac{8}{120}\right)=$	2,000	➔	n3= 2

n = 30

Lo que nos indica que se tiene que realizar un muestreo aleatorio en cada estrato para obtener una muestra de 5 Directivos, 23 Administrativos y 2 de Seguridad, que sumarían en total 20 personas.

- Directivos (C3 y C4) → de los 22 casos se eligen los valores que se muestran a continuación: 13°, 8°, 5°, 18° y 14°.
- Administrativos (C7, C8) → de los 90 casos se eligen los siguientes: 35, 41, 16, 66, 27, 84, 40, 1, 56, 50, 24, 45, 32, 33, 39, 73, 81, 91, 59, 91, 39, 55, 18.
- Seguridad (C12) → de los 8 casos se eligen dos: 1° y 8°.

2.4. Muestreo por conglomerados

Un muestreo por conglomerados es el que se realiza en poblaciones muy extensas. La extensión se encuentra dividida y solo basta con reconocer los límites de cada división que se denomina conglomerado.

Además, se realiza un sorteo entre conglomerados para elegir uno o dos de ellos, los cuales constituirán la muestra.

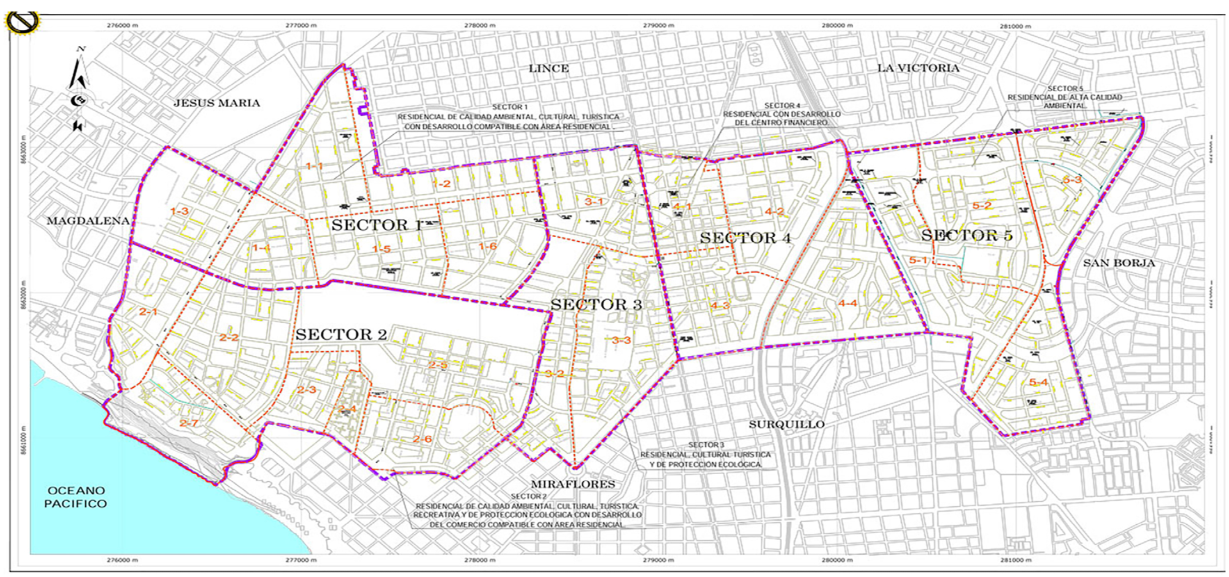


Figura 10. Muestreo por conglomerados (Municipalidad de San Isidro, 2015).

¿Conglomerados o estratos?

Alguien puede considerar similitudes entre conglomerados y estratos, pero son muy distantes; tan solo en el hecho de la formación de los grupos en estratos se deben usar variables para definir los estratos, mientras que en los conglomerados no se requieren de variables, sino que simplemente se toma la ubicación de los elementos y se identifican los límites.

Lectura seleccionada n.° 1

Diario Perú 21. (2015, 23 abril). INEI: Pobreza en el Perú disminuyó solo 1,2 puntos porcentuales en 2014. Perú 21. Lima. Disponible en: https://goo.gl/hXXKiq

Actividad n.º 1

Foro 1 - Criterios de muestreo y marco muestral

1. Ingrese al aula virtual a la Unidad 1, al Foro 1.
2. Lea con atención las indicaciones.
3. Escriba su respuesta a la pregunta planteada.
4. Se calificará con mayor puntaje las intervenciones propias respaldadas en su experiencia profesional y en lecturas adicionales (citas).

Estimación de parámetros

Tema n.º 2

Cuando hablamos de estimar, uno puede recurrir a un diccionario y encontrar lo siguiente:

estimar (Del lat. *aestimāre*.)

1. tr. Calcular o determinar el valor de algo.
2. tr. Atribuir un valor a algo. . .
4. tr. Creer o considerar algo a partir de los datos que se tienen" (Real Academia Española, 2016).

Por tanto, en nuestro desarrollo estas acepciones nos sirven como partida, ya que en estimación se desarrollan procedimientos que tienen como objetivo el estimar el valor de un parámetro.

Estimar entonces se convierte en una actividad dedicada a dar respuesta a la necesidad de averiguar el valor de una cualidad en la población, como puede ser la proporción (p), la media (μ), la varianza (σ^2) o la desviación estándar (σ).

Entiéndase que estos parámetros se pueden calcular en una población pequeña, pero en la mayoría de los casos las poblaciones son muy grandes y no se puede acceder al total de sus elementos, y mucho menos al total de sus mediciones. Por este motivo, sólo podemos acceder a una estimación.

De manera general, una estimación se puede realizar de dos formas: puntual y por intervalo.

1.1. Estimación puntual

Una estimación puntual es el cálculo de un estadístico o estimador. Por ejemplo, se calcula una medida como la proporción o la media en una muestra y se afirma que esta medida es la misma que el parámetro o medida en la población.

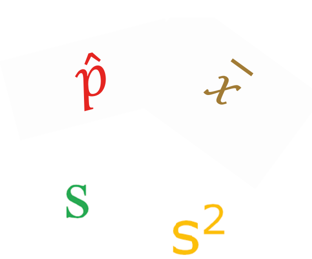


Figura 11. Estadísticos.

Esta estimación se fundamenta en que se ha obtenido una muestra que es aleatoria e idéntica a su población en todas las características de interés para lo que se quiere medir; es decir, la muestra es representativa de su población y, por tanto, todo lo que se calcule en ella debe ser igual o semejante a lo que verdaderamente se da en la población.

Entonces se puede asegurar:

Parámetro = valor del estadístico

1.2. Estimación por intervalo

Una estimación por intervalo requiere del cálculo de un límite inferior y un límite superior:

Límite inferior < parámetro < Límite superior

Dentro de estos límites se tienen la confianza de que se encuentra el verdadero valor del parámetro.

Los límites inferior y superior dependen de cálculo del estadístico al que le sumamos y restamos el margen de error (E):

Límite inferior = Estadístico – E

Límite superior = Estadístico + E

Daremos inicio a esta unidad desarrollando las estimaciones para la proporción poblacional (p).

Recuerde el teorema del límite central

El margen de error no es otra cosa que un múltiplo de la desviación estándar de una distribución muestral.

1.° Distribución muestral

Si usted pudiera obtener todas las muestras de una población y pudiera calcular la media y la desviación estándar de todas sus medias muestrales, encontraría que:

$\bar{\bar{x}} = \mu$	Donde la media de las medias: $x = \dfrac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5 + \ldots}{\text{Cantidad de muestras}}$
$\sigma_{\bar{x}}$ = Error estándar	$\sigma_{\bar{x}}$ = Desviación estándar de las medias

2.° Normalidad

La otra razón por la que acudimos al teorema del límite central es para respaldar nuestros cálculos en la normalidad.

Ninguno de los procedimientos puede darse por cierto si no se trabaja con poblaciones normales, lo que es una dificultad, ya que no todas las variables provienen de poblaciones normales. En consecuencia, ¿cómo haremos estimaciones si no tenemos normalidad en los datos? El teorema del límite central nos dará una solución.

Teorema del límite central

Sea x1, x2, x3, un conjunto de variables aleatorias, independientes y con la misma distribución con media μ y varianza $\sigma^2 \neq 0$, Por lo tanto, si n es suficientemente grande: $\bar{X}_i = \dfrac{\sum x_i}{n}$, tendrá una distribución normal.

En palabras sencillas, si n > 30, entonces la distribución muestral de las muestras de tamaño n será normal, es decir, será simétrica[1].

Entonces, si se cumple el teorema del límite central (n>30), podemos tener la seguridad de que nuestra muestra pertenece a un conjunto de muestras cuya distribución es normal o casi normal y podremos aplicar todas las fórmulas que se mostrarán en los cálculos de estimación por intervalo.

1 Uno puede encontrar una simulación del profesor Francisco Javier Barón en el vídeo Teorema Central de Limite.

1.3. Estimación de la proporción

Realizamos esta estimación en el caso de que nos preocupe averiguar cuál es valor de la proporción de una cualidad en la población, como:

- Se desea saber la proporción de artículos defectuosos en un proceso de fabricación.
- El porcentaje de personas que consumen cierta marca de ropa.
- La proporción de nacimientos con problemas de peso.
- La proporción de cultivos atacados por una plaga luego del uso de un plaguicida ecológico.

1.3.1. Estimación puntual

Como se explicó líneas arriba, una estimación puntual se puede realizar tan solo calculando la proporción en una muestra $\left(\hat{p}\right)$.Este valor sería el valor de la proporción en la población:

$$p = \hat{p}$$

El mejor estimador de la proporción poblacional es la proporción muestral.

p = proporción poblacional

$\hat{p}$: proporción muestral

La proporción muestral se calcula de x = número de éxitos y n = tamaño de la muestra:

$$\hat{p} = \frac{x}{n}$$

Ejemplo 5:

En la fábrica de chocolates Winters, se desea determinar la proporción de productos que tienen defectos en el empaquetado. Una muestra de 1350 chocolates se toma a la salida de la línea de producción, de los cuales se descubre que 6 han sido mal empaquetados. Determine la proporción verdadera de defectuosos en toda la producción de la fábrica de manera puntual.

Datos:

n = 1350

x = 6

$$\hat{p} = \frac{x}{n}$$

$$\hat{p} = \frac{6}{1350}$$

$$\hat{p} = 0{,}0044$$

Entonces, la verdadera proporción de defectuosos en la población será la siguiente:

p = 0.0044

Es decir, traducido a porcentaje, sería 0.44 %

1.3.2. Estimación por intervalo

Es necesario tener en cuenta que se requieren las siguientes consideraciones:

1° La muestra es aleatoria simple...

2° Se tiene un número fijo de ensayos, los cuales son independientes. Además, existen dos categorías de resultados y las probabilidades permanecen constantes para cada ensayo.

3° Existen al menos 5 éxitos y al menos 5 fracasos. Así la distribución normal es una aproximación adecuada para la distribución binomial (Triola, 2013).

En una estimación por intervalo, se debe calcular el valor de límite inferior y el del límite superior.

El límite inferior	El límite superior
LI = $\hat{p}$ − E	LS = $\hat{p}$ + E

De lo anterior, se desprende lo siguiente:

El margen de error E = $Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

El error estándar $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Se obtiene la puntuación $Z_{\alpha/2}$ de la forma que se observa en la tabla 6 (ver tabla A-2 en el Apéndice C). Este valor depende del nivel de confianza con el que se desea realizar la estimación.

Nivel de confianza: es la probabilidad 1 - α (a menudo expresada como el valor de porcentaje equivalente) de que el intervalo de confianza realmente contenga el parámetro poblacional, suponiendo que el proceso de estimación se repite un gran número de veces (Triola, 2013, p. 330).

Entonces, un intervalo para la proporción se puede expresar tal como se aprecia:

$\hat{p}$ - E < p < $\hat{p}$ + E	($\hat{p}$ − E ; $\hat{p}$ + E)
$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$	$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

Recuerde lo siguiente:

$\hat{p} = \frac{x}{n}$ y $\hat{q}$ = 1- $\hat{p}$

n: tamaño de muestra; y x: número de éxitos

Ejemplo 6:

En la fábrica de chocolates Winters, se desea determinar la proporción de productos que tienen defectos en el empaquetado. Una muestra de 1350 chocolates se toma a la salida de la línea de producción, de los que se descubre que 6 han sido mal empaquetados. Estime la proporción verdadera de defectuosos en toda la producción de la fábrica al 95% de confianza.

2.° Realizamos el cálculo del margen de error E.

Datos:

n = 1350

x = 6

$\hat{p} = \frac{x}{n} = \frac{6}{1350} = 0.00444$

$\hat{q} = 1 - \hat{p} = 0.99556$

$$E = Z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$E = 1.96 \sqrt{\frac{(0.00444)(0.99556)}{1350}}$$

$$E = 0.00355$$

3° Calculamos el intervalo.

NC = 95% y $\alpha = 0.05$

$Z_{\alpha/2} = 1.96$

$$\hat{p} - E < p < \hat{p} + E$$

$$0.00444 - 0.00355 < p < 0.00444 + 0.00355$$

$$0.000896 < p < 0.00799$$

4° Interpretamos el intervalo.

Tenemos la confianza del 95% para afirmar que el verdadero valor de la proporción de envoltorios defectuosos se encuentra entre 0.000896 y 0.00799. Es decir, si pudiéramos sacar muchas muestras, en el 95% de los casos acertaríamos.

2. Estimación de la media

La media es una medida representativa de una población; es más representativa cuanto más simétrica es la distribución, es decir, más normal.

2.1. Estimación puntual

En este caso, se calcula la media muestral (estadístico;) y se asume que este valor es el mismo que tiene el verdadero valor de la media poblacional (parámetro; μ).

$$\mu = \bar{x}$$

μ: media poblacional

$\hat{p}$: media muestral

La proporción muestral se calcula del promedio de los datos en la muestra tal como se representa a continuación:

$$\bar{x} = \frac{\sum x_i}{n}$$

El mejor estimador de la media poblacional es la media muestral.

Ejemplo 7:

"Los peruanos tienen un gasto per cápita promedio de US$ 289 al año en salud, que incluye el costo de insumos y servicios médicos" señaló el presidente del Gremio de Salud de la Cámara de Comercio de Lima (CCL), Mario Mongilardi (Andina noticias, 2013). Para determinar si en nuestra región se produce el mismo resultado se realiza una encuesta a 32 hogares y se obtiene lo siguiente:

Tabla n.° 7. Gasto en salud de los peruanos.

210	280	129	290	310	260	250
180	170	220	300	330	257	290
230	240	190	300	180	180	300
260	160	200	250	260	290	310
234	310	240	280			

Determine el promedio verdadero de gasto en salud de manera puntual.

Datos:

n = 32
$\bar{x}$ = 246.5625

$$\bar{x} = \frac{\sum x_i}{n} = \frac{210+280+129+290+310+\ldots+280}{32}$$

$$\bar{x} = \frac{7890}{32}$$

$$\bar{x} = 246.5625$$

Entonces, la verdadera media de los gastos en salud en la población será la siguiente:

$$\mu = \$246.5625$$

2.2. Estimación por intervalo

A continuación, se presenta una serie de requisitos:

- La muestra debe ser aleatoria.
- Si la muestra es mayor a 30, se puede usar la distribución normal estándar (se cumple el teorema del límite central).
- Si la muestra es menor a 30 y la distribución de la población es normal, se puede usar la distribución t (*student*).

Los límites serán los siguientes:

El límite inferior	El límite superior
LI = $\bar{x}$ − E	LS = $\bar{x}$ + E

2.2.1 Estimación por intervalo para la media con muestras grandes n ≥ 30

Si se cuenta con una muestra grande n ≥ 30, entonces se cumple que la muestra pertenece a una distribución muestral que tiene distribución normal; por tanto, se puede usar una puntuación z para realizar la estimación.

Entonces, se obtiene lo siguiente:

El margen de error $E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$

El error estándar $\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}}$

La puntuación $z_{\alpha/2}$ se obtiene de la tabla 8 (ver Apéndice C, tabla A-2). Este valor depende del nivel de confianza con el que se desea realizar la estimación.

La desviación estándar puede ser poblacional (σ) o muestral (s) si la muestra es grande (n≥30).

Por tanto, un intervalo para la media se puede expresar tal como se muestra a continuación:

$\bar{x}$ - E < μ < $\bar{x}$ + E	($\bar{x}$− E ; $\bar{x}$+ E)
$\bar{x} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$	$\bar{x} - Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$

Ejemplo 8:

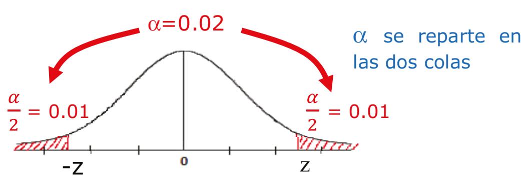
En el caso expuesto en el ejercicio 7, se nos informa que el gasto promedio per cápita en salud en el Perú es de $289 al año.

Ahora se desea estimar el verdadero valor per cápita de gastos en salud de la población en esta ciudad. Al 98% de confianza ¿podríamos decir que el gasto en salud es menor al que muestran las estadísticas a nivel nacional?

1.° Obtenemos el valor crítico de $z_{\alpha/2}$.

Datos:
n = 32
$\bar{x}$ = 246.5625
s = 52.3202
NC = 98%
α = 0.02

En la curva normal estándar:



En la Tabla 8, obtenemos el valor de $Z_{\alpha/2}$ con $\frac{\alpha}{2}$ =0.01

Interpolando:

b
a
2.32 z 2.33
0.0102 0.0099
0.01
A
B

Tabla n.° 8. Valor crítico obtenido por interpolación

Z	0	0.01	0.02	0.03	0.0
-2.9	0.0019	0.0018	0.0018	0.0017	0.00
-2.8	0.0026	0.0025	0.0024	0.0023	0.00
-2.7	0.0035	0.0034	0.0033	0.0032	0.00
-2.6	0.0047	0.0045	0.0044	0.0043	0.00
-2.5	0.0062	0.0060	0.0059	0.0057	0.00
-2.4	0.0082	0.0080	0.0078	0.0075	0.00
-2.3	0.0107	0.0104	0.0102	0.0099	0.00
-2.2	0.0139	0.0136	0.0132	0.0129	0.01
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

$$\frac{a}{A} = \frac{b}{B} \rightarrow \frac{z-2.32}{0.01-0.0102} = \frac{2.33-2.32}{0.0099-0.0102}$$

Despejando z = 2.3267

2.° Calculamos el margen de error E.

$$E = Z_{\alpha/2} * \frac{s}{\sqrt{n}}$$

$$E = 2.3267 * \frac{52.3202}{\sqrt{32}}$$

$$E = 21.5196$$

3.° Calculamos el intervalo tal como se observa a continuación:

$$\bar{x} - E < \mu < \bar{x} + E$$

$$246.5625 - 21.5196 < \mu < 246.5625 + 21.5196$$

$$225.0429 < \mu < 268.0821$$

4.° Interpretamos el intervalo de la siguiente manera:

Tenemos la confianza del 98% para afirmar que el verdadero valor de la media per cápita de gasto en salud se encuentra entre $225.0429 y $268.0821. Es decir, si pudiéramos sacar muchas muestras, en el 98% de los casos acertaríamos.

Respuesta:

($225.0422; $268.0821) < $289

Todo el intervalo es menor que $289; por tanto, se puede afirmar que el gasto per cápita en la ciudad es menor a lo que muestran las estadísticas a nivel nacional.

2.2.2. Estimación por intervalo para la media muestras pequeñas n < 30

En el caso de tener una muestra pequeña, se puede usar la desviación estándar de la muestra, pero es necesario que la población tenga una distribución normal o casi normal.

Estas condiciones permiten trabajar con un valor puntual "t", que es la distribución normal estándar modificada para trabajar con varianzas diferentes de 1.

Entonces, se tiene lo siguiente:

El margen de error $E = t_{\alpha/2} * \frac{s}{\sqrt{n}}$

El error estándar $\sigma_{\hat{p}} = \frac{s}{\sqrt{n}}$

La puntuación $t_{\alpha/2}$ se obtiene de la tabla 9 que está más adelante (ver Apéndice C, tabla A-3). Este valor depende del nivel de confianza con el que se desea realizar la estimación.

La desviación estándar puede debe ser muestral (s) si la muestra es pequeña (n < 30).

Entonces, un intervalo para la media se puede expresar como se muestra a continuación:

$\bar{x} - E < \mu < \bar{x} + E$	$(\bar{x} - E ; \bar{x} + E)$
$\bar{x} \pm Z_{\alpha/2} * \frac{s}{\sqrt{n}}$	$\bar{x} - Z_{\alpha/2} * \frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} * \frac{s}{\sqrt{n}}$

Ejemplo 9:

Un sondeo de opinión sobre el nivel de IGV que se debe aplicar a las transacciones comerciales indaga sobre cuánto debería ser según la ciudadanía. De 26 personas entrevistadas se pudo obtener un promedio de 16,7% con una desviación estándar de 4.23. ¿Se puede decir que los ciudadanos desean que se rebaje el impuesto del nivel actual de 18%?

1.° Obtenemos el valor crítico de $t_{\alpha/2}$, a dos colas con α = 0.05 y gl = n-1 = 25.

Tabla n.° 9. Valor crítico de t.

Datos:

n = 26
$\bar{x}$ = 16.7
s = 4.23

NC = 95 %
α = 5 % = 0.05
gl = 25g

2 colas	0.30	0.20	0.10	0.05	0.02	0.01	0.00
1 cola	0.15	0.1	0.05	0.025	0.01	0.005	0.00
gl	Valores t						
19	1.066	1.328	1.729	**2.093**	2.539	**2.861**	3
20	1.064	1.325	1.725	**2.086**	2.528	**2.845**	3
21	1.063	1.323	1.721	**2.080**	2.518	**2.831**	3
22	1.061	1.321	1.717	**2.074**	2.508	**2.819**	3
23	1.060	1.319	1.714	**2.069**	2.500	**2.807**	3
24	1.059	1.318	1.711	**2.064**	2.492	**2.797**	3
25	1.058	1.316	1.708	**2.060**	2.485	**2.787**	3

Entonces, el valor de la puntuación $z_{\alpha/2}$ es el valor positivo tal como se observa a continuación:

$t_{\alpha/2}$ = 2.060

2.° Calculamos el margen de error.

$$E = t_{\alpha/2} * \frac{s}{\sqrt{n}}$$

$$E = 2.060 * \frac{4.23}{\sqrt{26}}$$

E = 1.7089

3.° Calculamos el intervalo.

$$\bar{x} - E < \mu < \bar{x} + E$$

$$16.7 - 1.7089 < \mu < 16.7 + 1.7089$$

$$14.99 < \mu < 18.41$$

4.° Interpretamos el intervalo.

Tenemos la confianza del 95% para afirmar que el verdadero valor de la media del IGV estimada por la población se encuentra entre 14.99 y 18.41. Es decir, si pudiéramos sacar muchas muestras, en el 95% de los casos acertaríamos.

Respuesta:

18 ϵ (14.99; 18.41)

El intervalo contiene el 18 %; por tanto, no se puede afirmar que la ciudadanía desee que el impuesto se rebaje.

3. Estimación de la varianza

Las varianzas muestrales tienen una distribución que se ajusta muy bien a una distribución Chi cuadrada (χ^2); por ello, el intervalo de confianza se calcula así:

Para la varianza

$$\frac{s^2(n-1)}{x_D^2} < \sigma^2 < \frac{s^2(n-1)}{x_I^2}$$

Para la desviación estándar

$$\sqrt{\frac{s^2(n-1)}{x_D^2}} < \sigma < \sqrt{\frac{s^2(n-1)}{x_I^2}}$$

Donde:

n = tamaño de muestra

s^2 = varianza muestral

x_D^2 y x_D^2 son valores o puntuaciones X^2 que se obtienen de la tabla 10 (ver Apéndice C, tabla A-4[2]).

Ejemplo 10:

La contaminación generada por la actividad humana tiene diferentes formas de las cuales algunas como la sonora es la que más afecta a los propios seres humanos en las ciudades. Un estudio realizado por los estudiantes de la Universidad Continental logró una muestra de mediciones en horas punta, en los 38 puntos de más congestión de la ciudad. Ellos encontraron una media de 72dB con una desviación estándar de 12.3dB. Desarrolle un intervalo de confianza para la desviación estándar del nivel de ruido al 98% de confianza.

1.° Obtenemos los valores críticos de X^2, a dos colas con α = 0.02 y gl = n-1 = 37.

Datos:

n = 38
s = 12.3

NC = 98 %
α = 2 % = 0.02
gl = 37

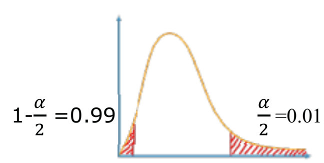


Tabla 10. Valores críticos de Chi cuadrado

	α					
gl	0.990	0.975	0.025	0.020	0.010	0.0
31	15.655	17.539	48.232	49.226	52.191	55.0
33	17.074	19.047	50.725	51.743	54.776	57.6
35	18.509	20.569	53.203	54.244	57.342	60.2
37	19.960	22.106	55.668	56.730	59.893	62.8
40	22.164	24.433	59.342	60.436	63.691	66.7
50	29.707	32.357	71.420	72.613	76.154	79.4

Fuente: elaboración propia

2 Si desea saber cómo se obtienen los valores X^2, puede revisar la presentación en PREZI: Valores Críticos X^2.

Entonces, los valores de la puntuación X² son los siguientes:

$$x_I^2 = 19.960 \quad \text{y} \quad x_I^2 = 59.893$$

2.° Calculamos el intervalo, así como se muestra a continuación:

$$\sqrt{\frac{s^2(n-1)}{x_D^2}} < \sigma < \sqrt{\frac{s^2(n-1)}{x_I^2}}$$

$$\sqrt{\frac{12.3^2(38-1)}{59.893}} < \sigma < \sqrt{\frac{12.3^2(38-1)}{19.960}}$$

$$9.668 < \sigma < 16.747\text{dB}$$

3.° Interpretamos el intervalo de la siguiente manera:

Tenemos la confianza del 98% para afirmar que el verdadero valor de la desviación estándar del ruido en la población se encuentra entre 9.668dB y 16.747dB. Es decir, si pudiéramos sacar muchas muestras, en el 98% de los casos acertaríamos al estimar la desviación estándar del ruido en la ciudad.

4. Intervalos con poblaciones finitas

Todos los cálculos anteriores se refieren a muestras obtenidas de poblaciones infinitas. En algunas ocasiones nos enfrentamos a situaciones con muestras que resultan de poblaciones finitas.

Una población es finita si

$\frac{n}{N} > 0.05$

En el caso de poblaciones finitas se debe utilizar un factor de corrección en el cálculo del margen de error:

Tabla n.° 11. Margen de error en intervalos con poblaciones finitas.

Proporción		$E = Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} * \sqrt{\frac{N-n}{N-1}}$
Media	n≥30	$E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$
	n<30	$E = t_{\alpha/2} * \frac{s}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$

Fuente: Elaboración propia.

Ejemplo 11: e tiene una población de 400 sujetos de la que se obtiene aleatoriamente una muestra de 32. En la muestra se contabilizan 7 éxitos. Desarrolle un intervalo de confianza al 90% de confianza.

Datos:
N = 400
n = 32
x = 7
$\hat{p} = \frac{7}{32} = 0.2188$
$\hat{q} = 1-\hat{p} = 0.7812$
NC = 90 %
$\alpha = 0.1$
z = 1.645

1.° Averiguamos si se trata de una población finita:

$$\frac{n}{N} = \frac{32}{400} = 0.08 > 0.05$$

Es una población finita.

2.° Obtenemos el margen de error.

$$E = Z\alpha/_2 \sqrt{\frac{\hat{p}\hat{q}}{n}} * \sqrt{\frac{N-n}{N-1}}$$

$$E = 1.645 * \sqrt{\frac{(0.2188)(0.7812)}{32}} * \sqrt{\frac{400-32}{400-1}}$$

$$E = 0.1155$$

Calculamos el intervalo:

$$\hat{p} - E < p < \hat{p} + E$$

$$0.2188 - 0.1155 < p < 0.2188 + 0.1155$$

$$0.1033 < p < 0.3343$$

3.° Interpretación del intervalo:

Tenemos la confianza del 90% para afirmar que el verdadero valor de la proporción está entre 10.33% y 33.43%.

5. Determinación del tamaño muestral

Cuando queremos realizar un estudio estadístico se debe tener un tamaño de muestra apropiado. Para ello contamos con las siguientes fórmulas, derivadas de la relación parámetro = estadístico ± margen de error:

Tabla n.° 12. Tamaño de muestra.

Población	Tamaño muestral para estimar la proporción	Tamaño muestral para estimar la media
Infinita	$n = \frac{pqz^2}{E^2}$	$n = \frac{\sigma^2 z^2}{E^2}$
Finita	$n = \frac{Npqz^2}{(N-1)E^2 + z^2 pq}$	$n = \frac{N\sigma^2 z^2}{(N-1)E^2 + \sigma^2 z^2}$
	p y **q** se toman de otros estudios, la opinión de un experto. **p*q = 0.25** si no se tiene información. **E** se puede tomar en el rango de 0.01 a 0.10.	**σ** se asume de otros estudios o de la estimación de un experto. **E** se toma de la opinión de un experto.

Fuente: Elaboración propia.

Ejemplo 12: Carlos Burgos desarrolla una tesis sobre el mercado de las tarjetas de crédito. Requiere de una muestra aleatoria. Desarrolla un sondeo y logra determinar que de 50 personas 13 tienen una tarjeta de crédito. ¿Cuál es el tamaño de muestra necesario para realizar una estimación al 95% de confianza y si se quiere un error máximo de ±4%?

Datos:
$p = \frac{13}{50} = 0.26$
$q = 1-p = 0.74$
NC = 95%
$\alpha = 0.05$
$z = 1.96$
$E = 4\% = 0.04$

Se asume una población infinita dado que se trata de una investigación de mercado en una ciudad como la nuestra

$$n = \frac{pqz^2}{E^2}$$

$$n = \frac{(0.26)(0.74)(1.96)^2}{0.04^2}$$

$n = 461.9524$

$n = 462$ (siempre se redondea a más)

Glosario de la Unidad I

A

Aleatorio. Que tiene la cualidad de producirse de manera inesperada al azar.

C

Coeficiente de confianza. Probabilidad de que un parámetro de población esté contenido dentro de un intervalo de confianza particular; también se denomina nivel de confianza o grado de confianza.

Curva de densidad. Gráfica de una distribución de probabilidad continua.

D

Datos. Información o números que describen alguna característica.

Datos continuos. Datos que se obtienen de un número infinito de valores posibles, que corresponden a puntos de una escala continua que abarca un rango de valores sin huecos, saltos ni interrupciones.

Datos cualitativos. Datos que pueden dividirse en diferentes categorías que se distinguen por alguna característica no numérica.

Datos cuantitativos. Datos que consisten en números que representan conteos o mediciones.

Datos de atributo. Datos que pueden dividirse en diferentes categorías que se distinguen por alguna característica no numérica.

Datos discretos. Datos con la propiedad de que el número de valores posibles es un valor finito o que puede contarse, que resulta en 0 posibilidades, 1 posibilidad o 2 posibilidades, etcétera.

Datos numéricos. Datos que consisten en números que representan conteos o mediciones.

Desviación estándar. Medida de variación igual a la raíz cuadrada de la varianza.

Distribución muestral. Distribución de las medias o las proporciones muestrales y que por el teorema del límite central se determina normal.

Distribución normal. Distribución de probabilidad con forma de campana, descrita algebraicamente con la fórmula:

$$P(a < x < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{\frac{(x-\mu)^2}{2\sigma^2}} . dx \rightarrow x \in R$$

Distribución normal estándar. Distribución normal con una media igual a 0 y una desviación estándar igual a 1.

Distribución t. Distribución normal que suele estar asociada con datos muestrales de una población con una desviación estándar desconocida.

Distribución t de Student. Véase distribución t.

E

Estadístico. Medida que se calcula o identifica con los datos de una muestra como la proporción muestral (), la media muestral () y la desviación estándar muestral (s), etcétera.

Estimación. Calcular o determinar el valor de un parámetro a partir de los datos que se tienen en una muestra.

M

Margen de Error. Es el error máximo (E) que se puede cometer al realizar una estimación. Está supeditado al valor puntual de z o t según el nivel de confianza utilizado en la estimación.

P

Parámetro. Medida que se calcula o identifica con los datos de una población como la proporción poblacional (p), la media poblacional (μ) y la desviación estándar (σ), etcétera.

V

Valor crítico. Es una puntuación de alguna distribución como la normal estándar, t, Chi cuadrada, F u otra, que separa el área de las puntuaciones más probables de aquellos menos probables.

Bibliografía de la Unidad I

Andina noticias. (10 de setiembre de 2013). CCL: Peruanos gastan en promedio US$ 289 al año en salud. Recuperado de http://gestion.pe/economia/ccl-peruanos-gastan-promedio-us289-al-ano-salud-2075795#comentarios

Cerrón, C. (2013). *Manual autoformativo del curso de Estadística II*. Huancayo, Junín, Perú.

Díaz, A. (2013). *Estadística aplicada a la administración y la economía*. México D. F.: Mc Graw Hill.

Frepik. (2015). Concepto de elegir a la persona correcta Vector Gratis (página web). Recuperado de: http://www.freepik.es/vector-gratis/concepto-de-elegir-a-la-persona-correcta_771040.htm

INEI. (mayo de 2006). Instituto Nacional de Estadística e Informática INEI. Recuperado de https://www.inei.gob.pe/estadisticas/metodologias/

Perú21. (23 de abril de 2015). INEI: Pobreza en el Perú disminuyó solo 1,2 puntos porcentuales en 2014. Recuperado de http://peru21.pe/economia/inei-pobreza-peru-disminuyo-solo-12-2014-2217321keyword & suggestions. (marzo de 2015). Recuperado de http://www.keyword-suggestions.com/bXVlc3RyYQ/

MINSA. (10 de octubre de 2012). Cuatro de cada diez peruanos sufre al algún problema mental. Recuperado de http://elcomercio.pe/sociedad/lima/cuatro-cada-diez-peruanos-sufren-algun-problema-mental-advirtio-minsa-noticia-1480838

Municipalidad de San Isidro. (2015). *Municipalidad de San Isidro*. Recuperado de http://www.msi.gob.pe/portal/wp-content/uploads/2014/01/3.PLANO-DE-VALORES-ARANCELARIOS.pdf

Ng Ooi Pin. (28 de setiembre de 2008). Production of the built environment (part II) [blog post] . Recuperado de http://ngooipin.blogspot.pe/

Perú21. (27 de mayo de 2014). Mayor demanda se registra en febrero y abril. (USI). La industria del calzado: un negocio que pisa firme. Recuperado de http://cde.peru21.pe/ima/0/0/2/1/3/213397.jpg

Real Academia Española. (RAE). (2016). *Diccionario de la Lengua Española*. España: RAE. Disponible en http://dle.rae.es/?id=GsZ6PSI

Rubione. M. (2011). Técnicas de muestreo para auditorías: Guía teórico – práctica. Buenos Aires, Argentina: Gerencia de Control de la Deuda Pública, Departamento de Control de Operaciones de Crédito Público y Sustentabilidad de la Auditoría General de la Nación.

Triola, M. (2013). *Estadística* (11.ª ed.) (trad. L. Pineda). México DF: Pearson Education.

Autoevaluación n.° 1

Lea con atención los enunciados. Repase en el manual el tema relacionado a la pregunta e intente una respuesta que se ajuste a su lectura y al criterio de aplicación de esta teoría.

1. **Se realiza un estudio con objeto de determinar el tiempo de permanencia en un mismo trabajo de los empleados en las empresas de la ciudad. Para ello, de las empresas existentes, se selecciona aleatoriamente 5 de ellas, y se elige una muestra aleatoria de colaboradores, atendiendo al tipo de puesto de trabajo. El muestreo realizado es**

   a. Sistemático

   b. Aleatorio y por conglomerados

   c. Aleatorio

   d. Estratificado y aleatorio

   e. Por conglomerados

2. **¿Cuál de los siguientes tipos de errores en encuestas ejemplificaría la obtención de demasiados varones en su muestra?**

   a. Error de muestreo

   b. Error de cobertura o sesgo en la selección

   c. Error o sesgo de no respuesta

   d. Error de medición

   e. Error de toma de datos

3. **El ministerio de Turismo e Integración tomó muestras de las personas que viajan al Cusco en viaje de turismo, para estimar la proporción de compatriotas que realizan este tipo de viajes. Calcule el intervalo de confianza del 96% para la proporción de turistas nacionales (Allen, 2000) si 1098 de los 3769 turistas entrevistados eran peruanos. ¿Cuál es el valor del margen de error y el límite superior del intervalo?**

   a. 0.015 - 30.65%

   b. 0.015 - 30.58%

   c. 0.019 - 31.04%

   d. 0.012 - 30.35%

   e. 0.019 - 30.58%

4. **Los valores críticos para un intervalo de confianza para la varianza poblacional con un NC de 90% y 9 grados de libertad son**

   a. 16.919 y 3.325

   b. 18.307 y 3.940

c. 34.4 y 9.0

d. 4.678 y 234.0

e. 15.507 y 2.733

5. De forma precisa, un intervalo de confianza es

a. Un nivel de confianza que permite conocer la probabilidad.

b. Un rango dentro del cual está la proporción.

c. Una estimación puntual, con nivel de confianza.

d. Un mínimo y un máximo en un nivel de confianza.

e. Un rango de datos para estimar un parámetro.

6. ¿Por qué es necesario un intervalo de confianza si se puede tener una estimación puntual?

a. No se sabe con exactitud qué es un estimador puntual.

b. No se conoce el margen de error y los valores mínimo y máximo del estimador.

c. Se conoce el parámetro para calcularlo, por ello se requiere de un estimador.

d. No se sabe qué tan bueno es nuestro mejor estimado.

e. Se ha tomado mal los datos y se debe corregir con el intervalo.

7. En el proceso de estimación de la media, marque V o F:

Se utiliza una muestra aleatoria simple.	
Se utiliza la distribución z si se conoce la desviación estándar muestral.	
Se puede usar una estimación puntual con un nivel de significancia adecuado.	
Se utiliza la distribución t porque se desconoce σ.	
Si la muestra es menor a 30, la población debe ser normal.	

a. VVFFV b. FVVFF c. FVFVV d. VVVFF e. VFFVV

8. Dado el siguiente intervalo de confianza al 98%, 0.21<p< 0.29, la interpretación correcta es:

I. Existe un 98 % de probabilidad de que el valor real del parámetro esté entre 0.21 y 0.29.

II. Tenemos una confianza del 98% de que el intervalo de 0.21 a 0.29 realmente contiene el valor verdadero del parámetro.

III. Si seleccionáramos muchas muestras diferentes de tamaño n y construimos los intervalos de confianza correspondientes, el 98% de ellos incluirían realmente el valor del parámetro.

IV. Con probabilidad de 98% el verdadero valor del parámetro está incluido en el intervalo 0.21<p<0.29.

V. Se tiene un 95% de confianza de que el verdadero valor del parámetro está contenido en el intervalo de 0.21 a 0.29.

a. Solo II b. I y V c. II, IV, V d. I y III e. II, III

9. El administrador de una tienda de pinturas "Arco Iris" desea estimar la cantidad real de pintura contenida en las latas de 1 Gl compradas a un fabricante de renombre. Según las especificaciones de fábrica, se sabe que la desviación estándar de la cantidad de pintura es igual a 0.02 Gl. Se selecciona una muestra aleatoria de 50 latas y la cantidad promedio de pintura por lata es de 0.995 Gl. Establezca una estimación del intervalo de confianza de 99% para el promedio poblacional de la cantidad de pintura incluida en una lata de 1 galón.

a. $0.8888 < \mu < 2.0051$

b. $0.9876 < \mu < 1.0024$

c. $0.9895 < \mu < 1.0005$

d. $0.9877 < \mu < 1.0023$

e. $1.0025 < \mu < 2.0051$

10. Para determinar una medida estándar del diámetro de los tacos de madera, un fabricante de muebles mide 20 espigas seleccionadas al azar y encuentra una media de 10.32 cm y una desviación estándar de 0.16 cm. Encuentre el intervalo de confianza del 95% considerando que la población está levemente sesgada.

a. $10.25 < p < 10.39$

b. $10.245 < \mu < 10.395$

c. $10.250 < \mu < 10.390$

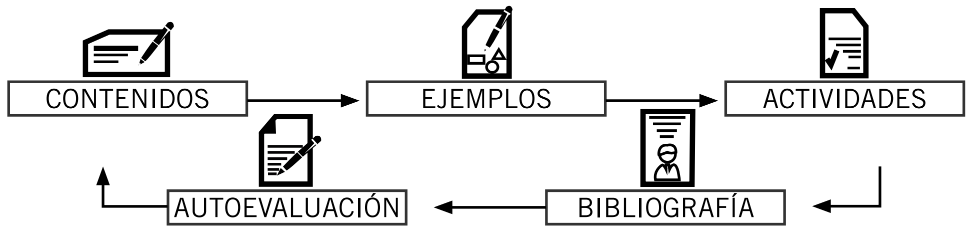
d. $0.1025 < p < 0.1040$

e. $1.02 < \mu < 1.04$

UNIDAD II

PRUEBA DE HIPÓTESIS Y ANÁLISIS DE LA VARIANZA

DIAGRAMA DE ORGANIZACIÓN DE LA UNIDAD II



ORGANIZACIÓN DE LOS APRENDIZAJES

Resultados del aprendizaje de la Unidad II: Al finalizar la unidad, el estudiante será capaz de plantear y aplicar pruebas de hipótesis para la media, proporción y varianza a partir de situaciones estadísticas reales.

CONOCIMIENTOS	HABILIDADES	ACTITUDES
Tema n.º 1: Pruebas de hipótesis 1. Definiciones básicas 2. Tipos de hipótesis. 3. Tipos de pruebas de hipótesis. 4. Métodos para realizar pruebas de hipótesis. **Tema n.º 2: Análisis de varianza** 1. Análisis de varianza. 2. ANOVA de 1 factor. 3. ANOVA de 2 factores.	1. Identifica las clases de hipótesis. 2. Plantea pruebas de hipótesis. 3. Identifica correctamente los valores críticos para la aplicación de las pruebas de hipótesis. 4. Determina el procedimiento pertinente de la prueba de hipótesis. 5. Realiza la interpretación del resultado de la prueba de hipótesis. **Actividad 1** Participa del foro de discusión "Importancia de las pruebas de hipótesis". **Actividad 2** Evaluación del tema n.º 1 y el tema n.º 2.	Valora la importancia de las pruebas de hipótesis e interpreta correctamente los resultados para una buena toma de decisiones.

Prueba de hipótesis

Tema n.° 1

1. Definiciones básicas

1.1. Hipótesis

Una hipótesis se constituye como toda afirmación que se hace sobre una propiedad de una población con la intención de probar si es cierta.

Ejemplos:

- "Cuatro de cada diez peruanos sufren de algún problema mental", advirtió el Minsa (2012).
- El gasto per cápita en diversión en la región Junín es menor a S/. 780 mensuales.
- La variación en las medidas de diámetros de pernos de 10 mm φ no supera 0.003 mm.
- Altomayo ocupa más del 13.3% del mercado regional.

En el proceso de prueba de hipótesis se pueden distinguir dos tipos de hipótesis, las cuales veremos a continuación.

1.2. Tipos de hipótesis

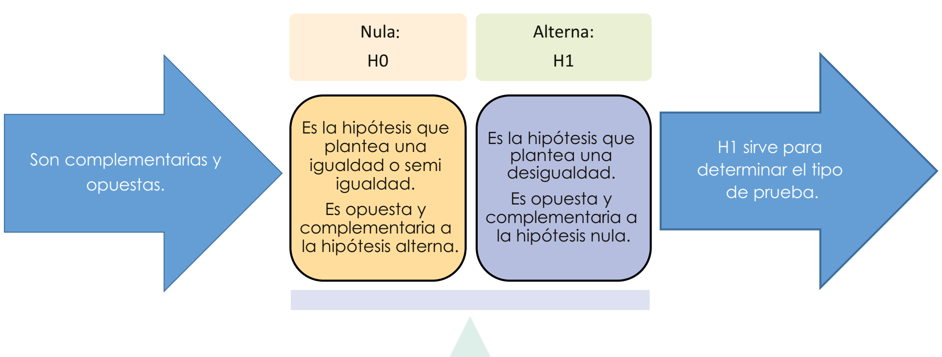


Figura 12: Tipos de pruebas de hipótesis. Fuente: Elaboración propia.

Tabla 13. Formas de hipótesis

Formas de plantear hipótesis		
H0: $\phi = A$ H1: $\phi \neq A$	H0: $\phi \leq A$ H1: $\phi > A$	H0: $\phi \geq A$ H1: $\phi < A$

Fuente: Elaboración propia.

Ejemplo 13:

- "Cuatro de cada diez peruanos sufren de algún problema mental", advirtió el Minsa (Minsa, 2012).

  H0 $p = 4/10$

  H1 $p \neq 4/10$

- El gasto per cápita en diversión en la región Junín es menor a S/. 780 mensuales.

  H0: $\mu \geq$ S/780 mensuales

  H1: $\mu <$ S/780 mensuales

- La variación en las medidas de diámetros de pernos de 10mm φ no supera 0.003mm.

  H0: $\sigma \leq 0.003$ mm

  H1: $\sigma > 0.003$ mm

- Altomayo ocupa más del 13.3% del mercado regional.

  H0: $p \leq 0.133$

  H1: $p > 0.133$

No existe regla sobre cuál de las hipótesis se plantea primero; aunque generalmente se plantee H1, también se puede plantear H0 inicialmente.

1.3. Estadístico de prueba

Un estadístico de prueba es una puntuación calculada según:

Tabla 14. Estadístico de prueba

Parámetro	Muestra	Estadístico	
Proporción		$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$	El valor de la proporción p se obtiene de H0.
Media	Muestras grandes	$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$	El valor de la media μ se obtiene de H0.
	Muestras pequeñas	$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$	El valor de la proporción μ se obtiene de H0
Varianza		$X^2 = \frac{s^2(n-1)}{\sigma}$	El valor de la proporción σ se obtiene de H0.

Fuente: Elaboración propia.

El estadístico de prueba puede ser positivo o negativo para z o t, y en el caso de X^2 y F los valores son siempre positivos.

1.4. Prueba de hipótesis

Es un procedimiento estándar que desarrolla una secuencia de procesos que sirven para determinar el valor de verdad de la hipótesis nula.

Inicia con la suposición de que H0 es verdadera y puede desarrollarse en pasos como los siguientes:

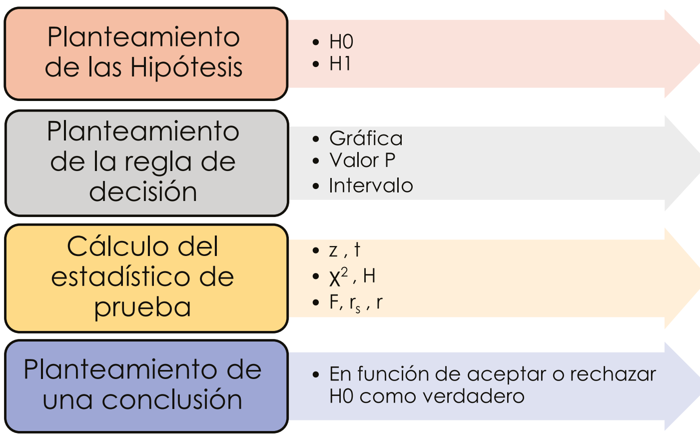
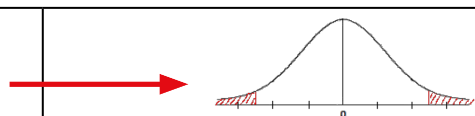
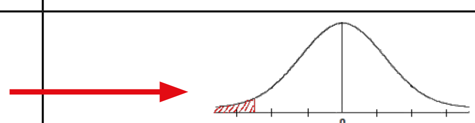
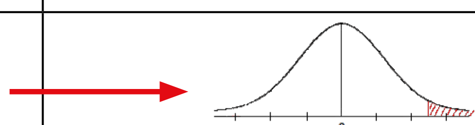


Figura 13: Prueba de hipótesis Fuente: Elaboración propia.

1.5. Tipos de pruebas

Las pruebas pueden ser de dos tipos:

Tabla 15. Tipos de prueba

Dos colas		H0: $\phi = A$ H1: $\phi \neq A$	
Una cola	Izquierda	H0: $\phi \geq A$ H1: $\phi < A$	
	Derecha	H0: $\phi \leq A$ H1: $\phi > A$	

Fuente: Elaboración propia.

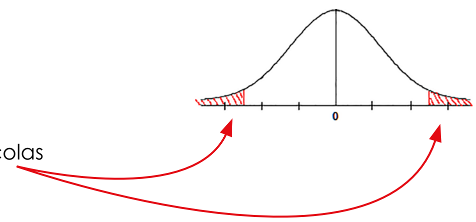
Las pruebas son de una o dos colas de acuerdo a la hipótesis alterna.

Ejemplo14:

- Cuatro de cada diez peruanos sufren de algún problema mental, advirtió el Minsa (2012).

H0 p = 0.40

H1 p ≠ 0.40

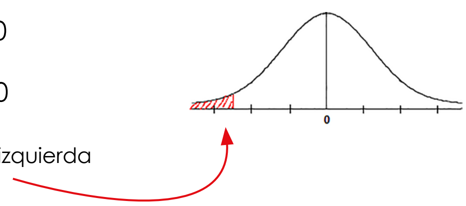


Dos colas

- El gasto per cápita en diversión en la región Junín es menor a S/.780 mensuales.

H0: μ ≥ S/780

H1: μ < S/780

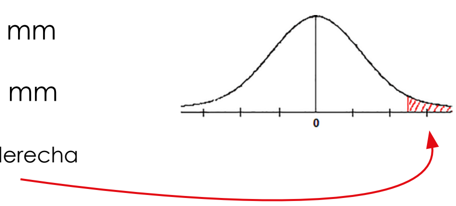


Cola izquierda

- La variación en las medidas de diámetros de pernos de 10 mm ϕ no supera 0.003 mm.

H0: σ ≤ 0.003 mm

H1: σ > 0.003 mm

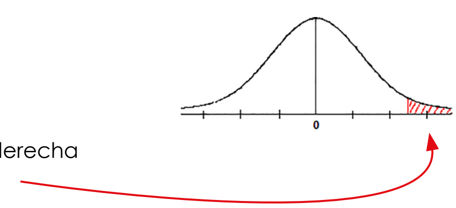


Cola derecha

- Altomayo ocupa más del 13.3% del mercado regional.

H0: p ≤ 0.133

H1: p > 0.133



Cola derecha

Cada gráfica incluye dos áreas, por ejemplo:

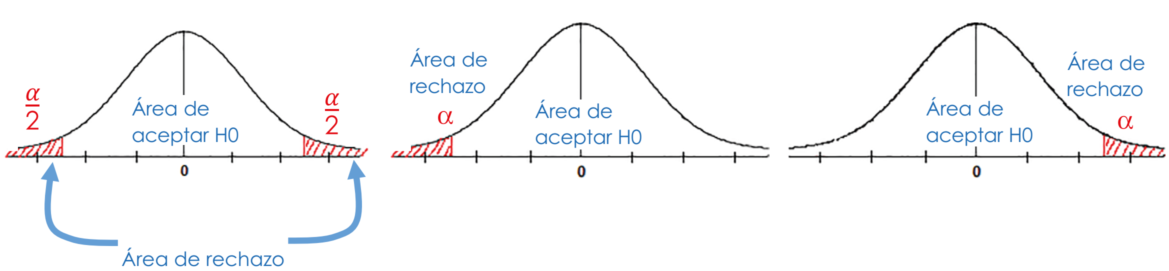


Figura 14: Tipos de pruebas. Fuente: Elaboración propia.

El límite de las áreas es un valor z, t, x^2 … que depende del valor de a que es conocido como el nivel de significancia, que es el valor del área de rechazo.

Área de rechazo

El área de rechazo inicia en el valor crítico y puede conocerse como región crítica. Si algún valor de prueba cae debajo de esta área, nos obligará a rechazar H0 como verdadera.

Los valores comunes de a son 0.01, 0.05 y 0.10.

> Normalidad y muestras pequeñas:
>
> Se debe recordar que las muestras pequeñas no se pueden respaldar en el teorema del límite central; por tanto, es obligatorio constatar que la población es normal. En realidad, no es estricto que sea normal, sino que es aceptable que sea casi normal, que no existan datos atípicos, o que no sea sesgada.

Los métodos para realizar una prueba de hipótesis son dos: el método tradicional y el método del valor P.

2. Métodos para realizar una prueba de hipótesis

2.1. Método tradicional

2.1.1 Prueba de hipótesis de dos colas para la proporción

Ejemplo 15: "Cuatro de cada diez peruanos sufren de algún problema mental, advirtió el Minsa" (Minsa, 2012), este es el titular de un diario al publicar las estadísticas del Ministerio de Salud. Un grupo de estudiantes de la escuela de Psicología desarrolla una investigación para probar si esta proporción se repite en nuestra región. Se toma una muestra de 300 personas y se les aplica un test de estabilidad de Eysenck y se determina que 116 tienen algún problema emocional. ¿Esto prueba que en la región se tiene una proporción diferente a lo que se registra a nivel nacional? Utilice un nivel de significancia de 0.05.

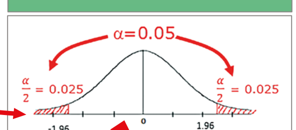
Datos:

n = 300
x = 116
$\hat{p} = \frac{116}{300} = 0.3867$
$\alpha = 0.05$
z = 1.96

1.° Planteamiento de Hipótesis

H0: p = 0.40
H1: p ≠ 0.40

2.° Planteamiento de regla de decisión



3.° Cálculo del estadístico de prueba

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$z = \frac{0.3867-0.40}{\sqrt{\frac{0.40(1-0.40)}{300}}} = -04702$$

4.° Conclusión

Aceptar H0 como verdadera
H0: p = 0.40 **V**
H1: p ≠ 0.40 **F**

Respuesta:

No existe evidencia muestral suficiente para afirmar que en la región se tiene una proporción diferente a la que se registra a nivel nacional.

2.1.2 Prueba de hipótesis de una cola para la media con n ≥ 30

Ejemplo 16: Una publicación en un diario local menciona que el gasto per cápita en diversión en la región Junín es menor a S/.780 mensuales. Esta afirmación es sometida a prueba al nivel de significancia del 1%. Se toma una muestra de 150 individuos y se logra una media en gastos de S/.768 con una desviación estándar de S/.63. ¿Se puede asegurar que la afirmación del diario es cierta?

Datos:

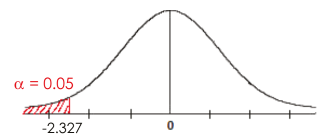
n = 150
x = 768
s = 63
α = 0.01
z = 2,327 (una cola)

El valor crítico z se obtiene de la tabla 8 por interpolación. Ver ejemplo 8.

H0: μ ≥ S/.780
H1: μ < S/.780

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$z = \frac{768-780}{\frac{123}{\sqrt{150}}} = -2.333$$



Rechazar H0 como verdadera
H0: μ ≥ S/.780 **F**
H1: μ < S/.780 **V**

Existe evidencia muestral suficiente para afirmar que el gasto per cápita en diversión en la región Junín es menor a S/.780 mensuales.

2.1.3 Prueba de hipótesis una cola para la media n < 30

Ejemplo 17. En su tesis de grado, Carlos Carhuapoma desarrolla la teoría de que el promedio de vida útil de una vivienda unifamiliar es más de 25 años, considerando que todos los elementos constructivos e instalaciones funcionan correctamente. Toma una muestra de 18 viviendas y obtiene los datos de la tabla adjunta. Realice una prueba al 0.05 de significancia, considerando que no existen datos atípicos.

Tabla 16. Datos de vida útil de viviendas

25	24	29	30
22	20	24	28
26	36	34	26
21	23	17	14
25	35		

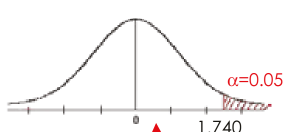
Datos:

n = 18
x = 25.5
s = 5.8937
α = 0.05
gl = 17
t = 1.740 (una cola Tabla 16)

H0: μ ≤ 25 años
H1: μ > 25 años

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{25.5-25}{\frac{5.8937}{\sqrt{18}}} = 0.36$$



No rechazar H0 como verdadera
H0: μ ≤ 25 **V**
H1: μ > 25 **F**

Tabla 17. Prueba de hipótesis una cola

1 cola	0.15	0.1	0.05	0.025
gl	Valores t			
14	1.076	1.345	1.761	2.1
15	1.074	1.341	1.753	2.1
16	1.071	1.337	1.746	2.1
17	1.069	1.333	1.740	2.1

No existe evidencia muestral suficiente para afirmar que el promedio vida útil una vivienda unifamiliar es más de 25 años.

2.1.4. Prueba de hipótesis para la varianza de una cola

Ejemplo 18. Corona S.A. es una fábrica de pernos de larga trayectoria. Implementa un sistema de control de calidad y entre otros datos a trabajar se verifica que la variación en las medidas de diámetros de pernos de 10 mmØ no supere 0.003 mm. Una muestra de 36 pernos reporta una media de 10.002mmØ con una varianza de $7.3x10^{-6}$. La evidencia comprueba que se cumple con el requisito. Utilice un nivel de significancia de 0.05.

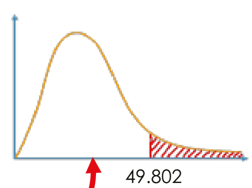
Datos:
n = 36 $s^2 = 7.3x10^{-6}$ $\alpha = 0.05$ gl = 35 Valor crítico en la tabla 18

H0: σ ≤ 0.003 mm
H1: σ > 0.003 mm

$$\chi^2 = \frac{s^2(n-1)}{\sigma^2}$$

$$\chi^2 = \frac{7.3x10^{-6}(36-1)}{0.003^3}$$

$$\chi^2 = 28.389$$



No rechazar H0 como verdadera
H0: σ ≤ 0.003 **V**
H1: σ > 0.003 **F**

Tabla 18. Valor crítico

	α			
gl	**0.100**	**0.075**	**0.050**	**0.**
29	39.087	40.573	42.557	**45.**
30	40.256	41.762	43.773	**46.**
31	41.422	42.948	44.985	**48.**
33	43.745	45.311	47.400	**50.**
35	46.059	47.663	49.802	**53.**

Existe evidencia muestral suficiente para afirmar que la variación en las medidas de diámetros de pernos de 10 mm no supera 0.003 mm.

2.2. Método del valor P

La regla para decidir si se rechaza H0 con este método se basa en la comparación de áreas en la curva normal. Si la probabilidad o área de la hipótesis nula es menor al del nivel de significancia, entonces se rechaza H0:

Si valor P≤ α,

se rechaza H0 como verdadera

Caso contrario se aceptará H0 como verdadera.

El valor P se ubica en la tabla 19 (ver Apéndice C, tabla A-2) con el valor del estadístico de prueba. Si la prueba es a una cola, el valor P es el área hallada en la tabla.

Si la prueba es a dos colas, el valor P es igual al doble del área encontrada en la tabla.

En el desarrollo de método del valor P, se procede de la siguiente manera:

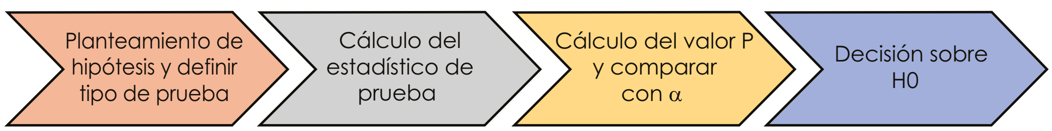


Figura 15. Proceso con el valor P. Fuente: Elaboración propia.

2.2.1. Prueba de hipótesis para la proporción - valor P

Ejemplo 19. En China, un fabricante de juguetes afirma que solo 10% o menos del total de osos de peluche parlantes que produce están defectuosos. Se sometieron a prueba en forma aleatoria a 400 de estos juguetes y se encontró que 50 estaban defectuosos. Compruebe la afirmación del fabricante con un nivel de significación de 5% (Díaz, 2013).

Datos:

$n = 400$
$x = 50$
$\hat{p} = \frac{50}{400} = 0.125$
$\alpha = 0.05$

1° Planteamiento de hipótesis

H0: $p \leq 0.10$
H1: $p > 0.10$ (cola derecha)

2° Estadístico de prueba

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$z = \frac{0.125-0.10}{\sqrt{\frac{0.10(0.90)}{400}}} = 1.67$$

Tabla 19. Valor P

3° Valor P y α

Z	0.05	0.06	0.07
-1.9	0.0256	0.0250	0.0244
-1.8	0.0322	0.0314	0.0307
-1.7	0.0401	0.0392	0.0384
-1.6	0.0495	0.0485	0.0475
-1.5	0.0606	0.0594	0.0582

Por tanto, el valor P=0.0475
Valor P < α
Rechazamos HO

4° Conclusión

No rechazar H0 como verdadera
H0: $p \leq 0.10$ **V**
H1: $p > 0.10$ **F**

Respuesta:

No existe evidencia muestral suficiente para confirmar que solo 10% o menos del total de osos de peluche parlantes que produce están defectuosos.

Se utiliza el lado negativo de los z porque la curva normal es simétrica. Un valor de área con -1.67 o 1.67 es igual en la izquierda como en la derecha de la curva.

2.2.2. Prueba de hipótesis para la media - valor P

Ejemplo 20. Un ejercicio planteado en el libro de Triola (2013) se ha modificado para este caso:

El total de los pesos individuales de la basura desechada por 62 hogares en una semana tiene una media de 27.443 libras. Suponga que la desviación estándar de los pesos es de 12.458 libras. Utilice un nivel de significancia de 0.05 para someter a prueba la afirmación de que la población de hogares tiene una media de 30 libras, que es la cantidad máxima que puede manejar el sistema actual de eliminación de desperdicios. ¿Hay alguna razón para preocuparse?

Datos:

n = 62
$\bar{x}$ = 27.443lb
s =12.458lb
α = 0.05

H0: μ = 30lb
H1: μ ≠ 30lb (dos colas)

No rechazar HO como verdadera
H0: μ = 30 **V**
H1: μ = 30 **F**

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$z = \frac{27.443-30}{\frac{12.458}{\sqrt{62}}} = -1.62$$

Tabla 20. Valor crítico

z	0.100	0.075	0.050
-1.9	0.0287	0.0281	0.0274
-1.8	0.0359	0.0351	0.0344
-1.7	0.0446	0.0436	0.0427
-1.6	0.0548	0.0537	0.0526
-1.5	0.0668	0.0655	0.0643

Por tanto, el valor P = 2(0.0526)
Valor P = 0.1092 y α = 0.05
Valor P > α
No rechazamos HO como verdadera.

Existe evidencia muestral suficiente para confirmar que la población de hogares tiene una media de 30 libras de basura desechada.

3. Inferencias con dos poblaciones

Cuando hablamos de “inferencias”, nos referimos a dos procesos en el campo de la estadística: intervalos y prueba de hipótesis.

El proceso de inferencia con dos muestras tiene que ver con la comparación de los parámetros mediante una diferencia en la que se puede observar si estos parámetros son iguales, o uno menor o mayor que el otro.

Se deben tener en cuenta las condiciones que ya se utilizaron en los capítulos anteriores, como la normalidad de las poblaciones o el teorema del límite central. Las muestras grandes nos darán oportunidad de dejar de lado la observación de las distribuciones de las poblaciones.

En esta sección desarrollaremos el subtema de los procesos que se deben llevar a cabo. De manera directa expondremos las fórmulas en organizadores que nos permitan ver cómo y cuándo emplear un determinado método de solución.

3.1. Inferencias con dos proporciones

Tabla 21. Inferencias con dos proporciones

Intervalo de confianza	$\hat{p}_1 = \frac{x_1}{n_1}$; $\hat{p}_2 = \frac{x_1}{n_2}$ E=$Z_{\alpha/2} * \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_1}}$	
	$(\hat{p}_1 - \hat{p}_2)$ - E<$(p_1 - p_2)$<$(\hat{p}_1 - \hat{p}_2)$ +E	
Prueba de Hipótesis	H0: $p_1 = p_2$ H1: $p_1 \neq p_2$	H0: $p_1 = p_2$ H1: $p_1 > p_2$
	H0: $p_1 = p_2$ H1: $p_1 < p_2$	
	z = $\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$	

Fuente: Elaboración propia.

3.1.1. Intervalo de confianza para dos proporciones

Ejemplo 21: Un gerente de finanzas está analizando el comportamiento de sus cuentas por pagar y ha obtenido muestras de cuentas del mes de mayo de 2 años consecutivos. En mayo del año 1, con una muestra de 1 300 cuentas por pagar, descubrió 50 que no habían sido liquidadas en el plazo convenido, mientras que, en una muestra de 1 000 cuentas de mayo del año 2, había 50 que se pagaron a tiempo. Con un nivel de confianza de 95%, ¿puede afirmarse que ha habido un aumento en la proporción de cuentas por pagar que caen en la morosidad? (Díaz, 2013).

<u>Datos</u>

Año 1	Año 2
$n_1 = 1300$	$n_1 = 1000$
$x_1 = 40$	$x_1 = 65$
$\hat{p}_1 = \frac{40}{1300}$	$\hat{p}_2 = \frac{55}{1000}$
$\hat{p}_1 = 0.0308$	$\hat{p}_2 = 0.065$

1.° Calculamos el margen de error:

$$E = Z_{\alpha/2} * \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$E = 1.96 * \sqrt{\frac{(0.0308)(0.9692)}{1300} + \frac{(0.065)(0.935)}{1000}}$$

$$E = 0.0333$$

NC = 95%

$\alpha = 0.05$

$Z_{\alpha/2} = 1.96$

2.° Cálculo del intervalo:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.0308 - 0.065) - 0.0333 < (p_1 - p_2) < (0.0308 - 0.065) + 0.0333$$

$$-0.0675 < (p_1 - p_2) < -0.00089$$

3.° Interpretamos lo siguiente:

Tenemos la confianza de que el verdadero valor de la diferencia de proporciones está entre -6.75% y -0.089%.

Como los límites del intervalo son negativos, esto quiere decir que el verdadero valor de la diferencia de proporciones es negativo; por tanto, si:

$$(p_1 - p_2) = \text{Negativo}$$

Significa que la proporción correspondiente al 2.° *año es mayor al del primer año*; por tanto, se puede asegurar que ha habido un aumento en la proporción de cuentas por pagar que caen en la morosidad.

Interpretación de intervalos para dos poblaciones

En general se pueden aplicar las siguientes reglas para interpretar los resultados:

Tabla 22. Interpretación de intervalos

+ < Ø1 – Ø2 < +	La diferencia siempre es positiva. El primer parámetro es mayor que el segundo.
- < Ø1 – Ø2 < -	La diferencia siempre es negativa. El primer parámetro es menor que el segundo.
- < Ø1 – Ø2 < +	La diferencia siempre es un valor entre los valores negativos y positivos. La diferencia tiene mucha probabilidad de ser cero. El primer parámetro es igual que el segundo.

Fuente: Elaboración propia.

3.1.2. Prueba de hipótesis para dos proporciones

Ejemplo 22: Se obtiene una muestra aleatoria simple de ocupantes del asiento delantero involucrados en choques de automóviles. De 2823 ocupantes que no usaban el cinturón de seguridad, 31 murieron. De los 7765 ocupantes que usaban el cinturón de seguridad, 16 murieron (según datos de *Who Wants Airbags?*, de Meyer y Finney, Chance 18 (2)). ¿Qué sugiere el resultado acerca de la eficacia de los cinturones de seguridad? Con un nivel de significancia de 0.05 someta a prueba la afirmación de que la tasa de mortalidad es más alta entre los individuos que no usan cinturones de seguridad.

<u>Datos</u>

Sin Cinturón	Con Cinturón
$n_1 = 2823$	$n_1 = 7765$
$x_1 = 31$	$x_1 = 16$
$\hat{p}_1 = \frac{31}{2832}$	$\hat{p}_2 = \frac{16}{7765}$
$\hat{p}_1 = 0.0110$	$\hat{p}_2 = 0.0021$

1.° Planteamos las hipótesis:

H0: p1 > p2

H1: p1 > p2 (una cola a la derecha)

2.° Cálculo del estadístico de prueba:

$$\bar{p} = \frac{x_1+x_2}{n_1+n_2}$$

$$\bar{p} = \frac{31+16}{2823+7765} = 0.0029$$

$$\bar{q} = 0.9971$$

NC = 95%

$\alpha = 0.05$

$$z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$z = \frac{(0.0110-0.0021)-0}{\sqrt{0.0029(0.9971)\left(\frac{1}{2823}+\frac{1}{7765}\right)}}$$

$$z = 7.53$$

3.° Valor P:

z tiene un valor muy alto 7.53, por tanto, el valor P = 0.0001

Valor P < a

Rechazamos H0

Respuesta:

Existe evidencia muestral suficiente para respaldar la idea de que la tasa de mortalidad es más alta entre los individuos que no usan cinturón de seguridad.

3.2. Inferencias con dos medias

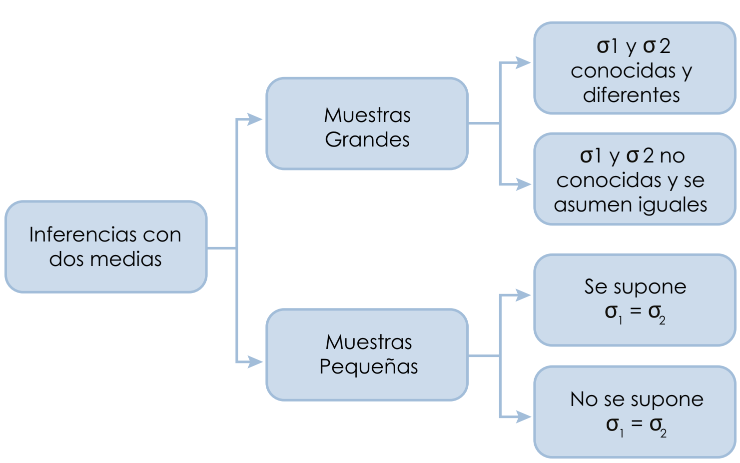


Figura 16. Pruebas dos medias. Fuente: Elaboración propia.

Tabla 23. Inferencias para dos medias

		σ1 ≠ σ2	σ1 = σ2
Intervalo	n ≥ 30	E = $z_{\propto/2} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	E = $t_{\propto/2} * \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
	n < 30	E = $t_{\propto/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ gl = el menor	E = $t_{\propto/2} * \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ gl = $n_1 + n_2 - 2$
Prueba de hipótesis	n ≥ 30	Estadístico de prueba: z = $\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1+\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	z = $\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1+\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
	n < 30	t = $\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1+\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ gl = el menor	t = $\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1+\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ gl = $n_1 + n_2 - 2$

Fuente: Elaboración propia.

En todos los casos: Si $\frac{s^2 mayor}{s^2 menor} < 3 \Rightarrow$ se asume σ1= σ2. (Mendehall, 2010)

3.2.1. Prueba de hipótesis para dos medias n1 ≥ 30 y n2 ≥ 30

Ejemplo 23: La Comisión Nacional del Pisco informó de un nuevo record de exportación de nuestra bebida de bandera situado en 9.5 millones de litros (CONAPISCO, 2016). Un estudio independiente realizado el 2013 obtiene la información de exportación de treinta y seis empresas mostrando una media de 101.75 mil litros con una desviación estándar de 20.35 mil litros. Treinta y dos empresas entrevistadas este año muestran una media de 116.28 mil litros exportados con una desviación estándar de 23.26 mil litros.

¿Esta información será suficiente para afirmar que la exportación de pisco se ha incrementado significativamente?

Datos

2013	2016
$n_1 = 36$	$n_1 = 32$
x = 101.75	x = 116.28
s = 20.35	s = 23.26

Prueba de igualdad de σ1 y σ2:

$\frac{s^2_{mayor}}{s^2_{menor}} < 3 \Rightarrow$ se asume σ1= σ2

$\frac{23.26^2}{20.35^2} = 1.306 < 3 \Rightarrow$ σ1= σ2

1.° Planteamos las hipótesis:

H0: μ1 = μ2

H1: μ1 < μ2 (una cola a la izquierda)

2.° Cálculo varianza conjunta:

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$s_p^2 = \frac{(36-1)20.35^2-(32-1)23.26^2}{36+32-2} = 473.7294$$

NC = 95 %

α = 0.05

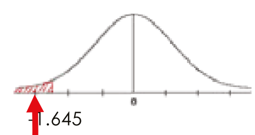
z = 1.645

(n1 y n2 > 30)

3.° Cálculo estadístico de prueba:

$$z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_1)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$z = \frac{(101.75-116.20)-0}{\sqrt{473.7294\left(\frac{1}{36}+\frac{1}{32}\right)}} = -2.748$$



4.° Rechazar H0 como verdadera:

Existe evidencia muestral suficiente para confirmar que sí hubo un incremento significativo en las exportaciones.

3.2.2. Pruebas de hipótesis para dos medias n1< 30 y n2 < 30

Ejemplo 24: El contenido medio de alquitrán en una muestra aleatoria simple de 25 cigarrillos tamaño grande sin filtro es de 21.1 mg, con una desviación estándar de 3.2 mg. El contenido medio de alquitrán de una muestra aleatoria simple de 25 cigarrillos de 100 mm con filtro es de 13.2 mg, con una desviación estándar de 3.7 mg. Utilice un nivel de significancia de 0.05 para someter a prueba la afirmación de que los cigarrillos tamaño grande sin filtro tienen un contenido medio de alquitrán mayor que el de los cigarrillos de 100 mm con filtro. ¿Qué sugiere el resultado acerca de la eficacia de los filtros de los cigarrillos? (Triola, 2013)

Datos

Sin Filtro	Con filtro
$n_1 = 25$	$n_2 = 25$
x_1= 21.1mg	x_2 = 13.2mg
s_1 = 3.2mg	s_2 = 3.7mg

Prueba de igualdad de σ1 y σ2:

$\frac{s^2_{mayor}}{s^2_{menor}} < 3$ ⇒ se asume σ1= σ2

$\frac{3.7^2}{3.2^2} = 1.337 < 3$ ⇒ σ1= σ2

1.° Planteamos las hipótesis:

H0: μ1 = μ2

H1: μ1 > μ2 (una cola a la derecha)

2.° Cálculo varianza conjunta:

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$s_p^2 = \frac{(25-1)3.2^2-(25-1)3.7^2}{25+25-2} = 11.965$$

Valor crítico

NC = 95%

α = 0.05

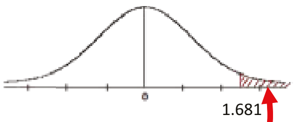
gl=25+25-2

gl = 48

t = 1.681

(n1 y n2 < 30)

3.° Cálculo estadístico de prueba:



$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_1)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$z = \frac{(21.1-13.2)-0}{\sqrt{11.965\left(\frac{1}{25}+\frac{1}{25}\right)}} = 8.075$$

4.° Rechazar H0 como verdadera:

Existe evidencia muestral suficiente para confirmar que los cigarrillos sin filtro tienen un mayor contenido de alquitrán.

3.3. Prueba de hipótesis para dos varianzas

Prueba de hipótesis

Estadístico de prueba:

$$F = \frac{s^2_{mayor}}{s^2_{menor}}$$

Valor crítico:

Tabla A-5 con el valor de α

$gl_{numerador} = n_{numerador}-1$

$gl_{denominador} = n_{denominador}-1$

Ejemplo 25: Un caso de reclamo se presenta ante SUNAS por el caso de la fluctuación excesiva en los voltajes entre la zona residencial y una zona comercial. Se toma una muestra de 17 días en la zona residencial, y en el mismo periodo se toma otra muestra de 16 días en la zona comercial. Los datos se muestran a continuación. Al nivel del 0.05, ¿se puede asegurar que la variación de los voltajes es mayor en la zona comercial?

Datos

Residencial	Comercial
$n_2 = 17$	$n_1 = 16$
$\bar{x}_2 = 210$v.	$\bar{x}_1 = 208$v
$s_2 = 5$v	$s_1 = 7$v.

s =7v. es la **desv. est. Mayor,** por tanto, es el **numerador** y la muestra con 5v. de desv. Est. el denominador.

Valor Crítico

$\alpha = 0.05$

$gl_{numerador} = 16-1 = 15$

$gl_{denominador} = 17-1 = 16$

1.° Planteamos las hipótesis:

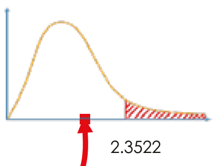
H0: $\sigma_1 = \sigma_2$

H1: $\sigma_1 > \sigma_2$

2.° Cálculo del estadístico de prueba:

$$F = \frac{s_1^2}{s_2^2}$$

$$F = \frac{7^2}{5^2} = 1.96$$



3.° No rechazar H0 como verdadera:

No existe evidencia muestral suficiente para probar que exista una mayor variación en los voltajes en la zona comercial.

Análisis de varianza – ANOVA

Tema n.° 2

ANOVA es un método de prueba de hipótesis cuando se tiene 3 o más poblaciones o muestras. Se supone que los datos proceden de poblaciones normales, y las varianzas no son muy diferentes.

El método resuelve las siguientes hipótesis:

H0: μ1 = μ2 = μ3 = . . .

H1: Por lo menos una media es diferente.

Las poblaciones se obtienen de dividir una población en bloques o estratos que comparten características comunes. Cada división es diferenciada por una o más variables o factores.

La prueba se desarrolla comparando las variaciones producidas por el factor o tratamiento, con la variación en cada muestra (error).

El estadístico de prueba es: $F = \dfrac{\textit{Variación entre muestras}}{\textit{Variación dentro de las muestras}}$

1. ANOVA de un factor

Se utiliza ANOVA de un factor cuando:

- Las poblaciones se distinguen por medio de un factor o vía.
- Se tiene tres o más poblaciones normales.
- Las varianzas entre poblaciones no son muy diferentes (homogeneidad de varianzas).

1.1. Tabla ANOVA de un factor

Dependiendo del autor las columnas pueden cambiar de posición, pero todas las tablas ANOVA tienen los mismos componentes:

Tabla 24. ANOVA de una vía

Origen de la variación	Suma de cuadrados SC	gl	Cuadrados medios CM	Estadístico de prueba F
Por el tratamiento	$SCT = \sum \frac{T_j^2}{n_j} - G$	K-1	$CMT = \frac{SCT}{K-1}$	$F = \frac{CMT}{CME}$
Por el error	SCE = SS-SCT	N-K	$CME = \frac{SCE}{N-K}$	
Total	$SS = \sum x_{IJ}^2 - G$	N-1		

Fuente: Elaboración Propia.

Donde:

N: Total de datos

n_j: Tamaño de la muestra j

K: Número de muestras o poblaciones

T_j es la sumatoria de datos de la muestra j.

$$G = \frac{\left(\sum x_{ij}\right)^2}{N}$$

Tabla 25. Datos recibidos de laboratorios

Lab1	Lab2	Lab3
120.1	98.3	103.0
110.7	112.1	108.5
108.9	107.7	101.1
104.2	107.9	110.0
100.4	99.2	105.4
111.4		

Ejemplo 26: Un ejemplo modificado del libro de Mendehall (2010). Los médicos dependen de resultados de exámenes de laboratorio cuando manejan problemas médicos como diabetes o epilepsia. En un examen de uniformidad para tolerancia a la glucosa, a tres laboratorios diferentes se les enviaron muestras de sangre idénticas de una persona que había bebido 50 miligramos (mg) de glucosa disuelta en agua. Los resultados de laboratorio (en mg/dl) son los mostrados en la tabla. ¿Los datos indican una diferencia en el promedio de lecturas para los tres laboratorios? Use α = 0.05.

H0: μ1 = μ2 = μ3

H1: Por lo menos una media es diferente.

Usando un programa estadístico como SPSS obtenemos:

Tabla 26. ANOVA – nivel medido de glucosa

	Suma de cuadrados	Gl	Media cuadrática	F	Sig.
Entre grupos	59.689	2	29.845	.907	.428
Dentro de grupos	427.840	13	32.911		
Total	487.529	15			

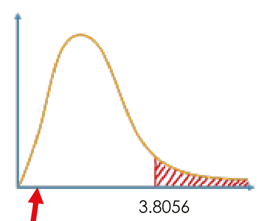
Fuente: Elaboración propia.

Buscamos el valor crítico en la tabla 27 (ver Apéndice C, tabla A-5) con $\alpha = 0.05$ $gl_{num} = 2$ $gl_{denom} = 13$:

Tabla 27. Valor crítico del nivel de glucosa

		Grados de libertad del numerador (gl1			
		1	2	3	
minador (gl2)	11	4.8443	3.9823	3.5874	3.35
	12	4.7472	3.8853	3.4903	3.25
	13	4.6672	3.8056	3.4105	3.17
	14	4.6001	3.7389	3.3439	3.11
	15	4.5431	3.6823	3.2874	3.05

F = 3.8056



El estadístico de prueba de la tabla ANOVA de la tabla es F = 0.907.

Aceptamos H0 como verdadera:

No existe evidencia muestral para afirmar que hay una diferencia en el promedio de lecturas para los tres laboratorios.

Resolución usando el valor:

Para este problema se puede usar el valor P de la tabla ANOVA.

Nivel medido de Glucosa

Tabla 28. Valor P del nivel medido de glucosa

	Suma de cuadrados	**gl**	**Media cuadrática**	**F**	**Sig.**
Entre grupos	59.689	2	29.845	.907	**.428**
Dentro de grupos	427.840	13	32.911		
Total	487.529	15			

Valor P = 0.428 > α 0.05

No rechazamos H0 como verdadera. No existe evidencia muestral para afirmar que hay una diferencia en el promedio de lecturas para los tres laboratorios.

2. ANOVA de dos factores

El ANOVA de dos factores se emplea cuando:

- La población está dividida o caracterizada por dos variables o factores.
- Las poblaciones son normales y no existe mucha diferencia entre sus varianzas.

Existe dos formas de llevar a cabo este tipo de pruebas como veremos a continuación.

2.1. Modelo aditivo

Cuando se analizan las poblaciones de manera separada por cada factor, se debe tener en cuenta que para este tipo de prueba cada casilla solo se tiene una observación.

Tabla ANOVA de dos factores o dos vías:

Tabla 29. ANOVA dos vías sumativo

Origen de la Variación	Suma de Cuadrados SC	Gl	Cuadrados Medios CM	Estadístico de prueba F
Columnas o tratamiento	SCC = $\sum \frac{T_j^2}{n_j} - G$	K-1	CMC = $\frac{SCC}{K-1}$	F = $\frac{CMC}{CME}$
Filas o muestras	SCF = $\sum \frac{T_i^2}{n_i} - G$	L-1	CMF = $\frac{SCF}{K-1}$	
Por el **Error**	SCE = SS-(SCC+SCF)	(K-1)(L-1)	CME = $\frac{SCE}{N-K}$	
Total	SS = $\sum x_{IJ}^2 - G$	N-1		

Donde:

N: Total de datos

n_j: Tamaño de la muestra j (columna).

n_i: Tamaño de la muestra i (fila).

K: Número de muestras o poblaciones en las columnas.

L: Número de muestras o poblaciones en las filas.

G = $\frac{\left(\sum x_{ij}\right)^2}{N}$

T_j es la sumatoria de datos de la muestra en la columna j.

T_i es la sumatoria de datos de la muestra en la fila i.

Ejemplo 27: Este ejemplo ha sido modificado de Díaz (2013). Una empresa que fabrica y vende purificadores de agua para el comercio y la industria tiene 3 zonas de ventas: centro, norte y sur. Como la labor de venta tiene aspectos tanto técnicos como financieros, la empresa tiene 4 vendedores: un químico, un licenciado en administración, un ingeniero mecánico y un ingeniero industrial. La gerencia de la empresa muestra interés en saber si las 3 zonas tienen un potencial equivalente y si los 4 vendedores tienen igual capacidad. La variable dependiente es el volumen de ventas. La información pertinente se da en la tabla adjunta.

Tabla 30. Datos de venta por regiones y vendedores

Vendedor	Centro	Sur	Norte
Ingeniero químico	506	528	513
Licenciado en administración	529	496	508
Ingeniero mecánico	518	504	520
Ingeniero industrial	510	505	520

Fuente: Elaboración propia.

Solución: Debe notarse que existe una observación por cada casilla, lo que configura el trabajo con un ANOVA aditivo.

Un *software* como Excel nos da la siguiente tabla:

Tabla 31. ANOVA Excel

ANÁLISIS DE VARIANZA						
Origen de las variaciones	**Suma de cuadrados**	**Grados de libertad**	**Promedio de los cuadrados**	**F**	**Probabilidad**	**Valor crítico para F**
Filas	41.5833	3	13.8611	0.0886	0.9637	4.7571
Columnas	140.6667	2	70.3333	0.4495	0.6578	5.1433
Error	938.6667	6	156.4444			
Total	1120.9167	11				

Fuente: Elaboración propia.

Se producen dos pruebas:

Filas (vendedores):

H0: μ1 = μ2 = μ3 = μ4

H1: Por lo menos uno de los vendedores tiene una media diferente.

Valor P = 0.9637 > α = 0.05

No rechazamos H0 como verdadera.

No existe evidencia muestral suficiente para demostrar diferencias entre vendedores.

Columnas (regiones):

H0: μ1 = μ2 = μ3

H1: Por lo menos una de las regiones tiene una media diferente.

Valor P = 0.6578 > α = 0.05

No rechazamos H0 como verdadera.

No existe evidencia muestral suficiente para demostrar diferencias entre regiones.

2.2. Modelo con interacción

En este método, además de analizar las poblaciones por cada factor por separado, también se hace un análisis por la interacción de los factores en la población. Por cada celda de datos existe más de uno; es decir, existe iteraciones o repeticiones.

Tabla ANOVA de dos factores con interacción:

Tabla 32. ANOVA 2 vías con interacción

Origen de la variación	Suma de cuadrados SC	Gl	Cuadrados medios CM	Estadístico de prueba F
Columnas o tratamiento	SCC	K-1	CMC=$\frac{SCC}{K-1}$	F = $\frac{CMC}{CME}$
Filas o muestras	SCF	L-1	CMF=$\frac{SCF}{L-1}$	F = $\frac{CMF}{CME}$
Interacción	SCI	(K-1)(L-1)	CMI=$\frac{SCI}{(K-1)(L-1)}$	F = $\frac{CMI}{CME}$
Por el **Error**	SCE	(N-1)-Σgl	CME=$\frac{SCE}{(N-1)-\Sigma gl}$	
Total	SS	N-1		

Fuente: Elaboración propia.

Donde:

N: Total de datos

n_j: Tamaño de la muestra j (columna)

n_i: Tamaño de la muestra i (fila)

K: Número de muestras o poblaciones en las columnas

L: Número de muestras o poblaciones en las filas

Ejemplo 28: El problema del ejemplo 27 completo. Una empresa que fabrica y vende purificadores de agua para el comercio y la industria tiene 3 zonas de ventas: centro, norte y sur. Como la labor de venta tiene aspectos tanto técnicos como financieros, la empresa tiene 3 vendedores: un químico, un licenciado en administración y un ingeniero mecánico. La gerencia de la empresa muestra interés en saber si las 3 zonas tienen un potencial equivalente si los 3 vendedores tienen igual capacidad y si es indistinto el trabajo de los vendedores en cualquier zona o existen diferencias. La variable dependiente es el volumen de ventas. La información pertinente se da en la tabla.

Solución. Como se puede apreciar, existen 2 observaciones en cada casilla por lo que los datos involucran un ANOVA con interacción.

Tabla 33. Datos de venta por regiones y vendedor

Profesión del vendedor	**Zona centro**	**Zona sur**	**Zona norte**
Químico	506 512	528 534	513 495
Licenciado en administración	529 525	496 498	508 500
Ingeniero mecánico	500 518	512 504	528 520

Fuente: Elaboración propia.

Un *software* como MINITAB nos da a siguiente tabla ANOVA:

Tabla 34. ANOVA Minitab

Source	**DF**	**SS**	**MS**	**F**	**P**
Profession	2	96.444	48.222	0.931	0.429
Zone	2	59.111	29.556	0.571	0.584
Interaction	4	2072.889	518.222	10.009	0.002
Error	9	466	51.778		
Total	17	2694.444			

Fuente: Elaboración propia.

Se producen tres pruebas:

Filas (profesión):

H0: μ1 = μ2 = μ3

H1: Por lo menos uno de los vendedores tiene una media diferente.

Valor P = 0.429 > α = 0.05

No rechazamos H0 como verdadera.

No existe evidencia muestral suficiente para demostrar diferencias entre vendedores.

Columnas (zona):

H0: μ1 = μ2 = μ3

H1: Por lo menos uno de las zonas tiene una media diferente.

Valor P = 0.584 > α = 0.05

No rechazamos H0 como verdadera.

No existe evidencia muestral suficiente para demostrar diferencias entre vendedores.

Interacción (profesión y zona):

H0: Las medias son iguales

H1: Por lo menos uno de las medias es diferente.

Valor P = 0.002 > α = 0.05

Rechazamos H0 como verdadera.

Existe evidencia muestral suficiente para demostrar que existen diferencias en el trabajo de los vendedores por zona.

En el caso del requisito de normalidad, este no es muy estricto, pues basta que en las muestras no existan datos atípicos.

De similar forma, respecto a la igualdad de varianzas o desviaciones estándar poblacionales, el especialista en estadística George E. P. Box demostró que, siempre y cuando los tamaños de muestra sean iguales (o casi iguales), las varianzas pueden diferir de tal forma que la más grande sea hasta nueve veces el tamaño de la más pequeña, y los resultados del ANOVA continúan siendo confiables en esencia (Triola, 2013).

3. Errores en la prueba de hipótesis

3.1. Error tipo I

Se comete un error de tipo I si al finalizar una prueba de hipótesis el proceso nos obliga a *rechazar H0 como verdadera cuando en la población es verdadera.*

La probabilidad de cometer el error tipo I es a (nivel de significancia).

3.2. Error tipo II

Se comete un error de Tipo II si al finalizar una prueba de hipótesis el proceso nos obliga a *no rechazar H0 como verdadera cuando en la población es falsa.*

La probabilidad de cometer el error tipo II es β.

α y β no son complementarios, pero guardan una relación inversa; es decir, si α se incrementa, β disminuye. En todo caso, si queremos disminuir ambas probabilidades, entonces el tamaño de la muestra debe incrementarse.

Lectura seleccionada n.° 2

Perúeconomico.com. (2011). Una mirada a los programas sociales. Disponible en: https://goo.gl/11il9l

Actividad n.° 2

FORO 2: ¿CUÁL ES LA IMPORTANCIA DE LAS PRUEBAS DE HIPÓTESIS?

- Ingrese al aula virtual, a la pestaña de la Unidad 2 >> Foro 2.
- Lea la consigna y las instrucciones.
- Escriba su respuesta a la pregunta tomando en cuenta las indicaciones de la consigna.

Glosario de la Unidad II

D

Distribución F. Distribución de las probabilidades continúa de una variable aleatoria que tiene un comportamiento muy cercano a la distribución muestral de las varianzas.

N

Nivel de significancia. Valor del área crítica o de rechazo de H0; es también el valor de la probabilidad de cometer un error de tipo I.

P

Potencia de una prueba. Es la medida de la probabilidad de n cometer el error de tipo II. Es decir, es igual a 1-β.

Proporción. Es la relación entre la cantidad de éxitos y el total muestreado. Se puede leer como un porcentaje si se le multiplica por 100. Puede existir una proporción de "éxitos" como complementariamente una proporción de "fracasos".

Bibliografía de la Unidad II

Alfaro, D., & Macera, D. (2011). Una mirada a los programas sociales [página web]. Recuperado de: goo.gl/e6AR6

Burdorf, A., van Riel, M., van Wingerden, J.P., van Wingerden, S., & Snijders C. (1995). Isodynamic evaluation of trunk muscles and low-back pain among workers in a steel factory. *Ergonomics, 3*(10), 2107-2117. doi: 10.1080/0014 0139508925254

Producción de Pisco marcó récord histórico de 9.5 millones de litros (5 de febrero de 2016). Recuperado de http://gestion.pe/economia/produccion-pisco-marco-record-historico-95-millones-litros-2154088

Díaz, A. (2013). *Estadística aplicada a la administración y la economía*. México D. F.: Mc Graw Hill.

Mendehall, W., Beaver, R. & Beaver, M. (2010). *Introducción a la probabilidad estadística* (13.ª ed.). México D. F.: CENGAGE Learning. Recuperado de: goo.gl/yFE3QX

Diario El Comercio. (10 de octubre de 2012). Cuatro de cada diez peruanos sufre algún problema mental. Diario El Comercio [en línea]. Recuperado de: goo.gl/eK8Gan

Triola, M. (2013). *Estadística* (11.ª ed.). (L. Pineda, Trad.). México D. F.: Pearson Education.

Autoevaluación n.º 2

Lea con atención los enunciados. Repase en el en manual el tema relacionado a la pregunta e intente una respuesta que se ajuste a su lectura y al criterio de aplicación de esta teoría.

1. **El valor de la región crítica en una prueba de hipótesis es**
   a. La potencia de la prueba
   b. El nivel de confianza
   c. El nivel de significancia
   d. El margen de error
   e. El error estándar

2. **Halle el/los valor(es) crítico(s) si n = 33 (la población parece estar distribuida normalmente), = 23.6 s = 3.56, H1: σ ≠ 0.36 al 0.05 de significancia.**
   a. ±1.96
   b. 50.725-19.047
   c. -2.2622
   d. ±2.037
   e. 47.4-20.867

3. **“Un investigador afirma que las cantidades de paracetamol en una determinada marca de tabletas para la gripe tiene una desviación estándar diferente de 2,5 mg indicado por el fabricante” (Triola, 2011). Suponiendo que se ha llevado a cabo una prueba de hipótesis y que la conclusión es que no se rechaza la hipótesis nula, ¿cuál es la conclusión?**
   a. No hay pruebas suficientes para apoyar la afirmación de que la desviación estándar es diferente de 2,5 mg.
   b. Hay pruebas suficientes para apoyar la afirmación de que la desviación estándar es diferente de 2,5 mg.
   c. No hay pruebas suficientes para apoyar la afirmación de que la desviación estándar es igual a 2,5 mg.
   d. Hay pruebas suficientes para apoyar la afirmación de que la desviación estándar es igual a 2,5 mg.
   e. Hay evidencias suficientes para rechazar la hipótesis alterna de que la desviación estándar es igual a 2.5 mg.

4. **De las siguientes afirmaciones, ¿cuáles son ciertas respecto de la hipótesis nula?**

Es la hipótesis denominada de trabajo en una investigación.	
Se prueba de forma directa, dado que inicialmente se supone que es verdadera.	
Es la afirmación de que el parámetro tiene un valor que, de alguna manera, difiere del valor en la hipótesis alterna.	

Es la afirmación de que el valor de un parámetro de población es igual a un valor aseverado.

Con base en los datos muéstrales es rechazada o no.

a. FVVFF
b. VVFVV
c. VFVFF
d. FFVVF
e. FVFVV

5. Una hipótesis en estadística es

a. Una afirmación que se hace respecto de una muestra para probar una cualidad en una población.
b. La conclusión que se realiza sobre los datos muestrales para afirmar una aseveración en una muestra.
c. Una afirmación acerca de una propiedad o cualidad de una población y que se prueba con datos muestrales.
d. La aseveración realizada de acuerdo con los datos muestrales sobre la población.
e. Una afirmación sobre cualidades de una población o más.

6. Un investigador trabaja para probar si un fármaco a base de ketorolaco tiene efectos secundarios sobre el colon en el 15% de los pacientes hipertensos. Al final de una prueba con 300 individuos concluye lo siguiente: "No existe suficiente evidencia muestral para sostener la aseveración de que el fármaco a base de ketorolaco cause efectos secundarios en el colon del 15% de los pacientes hipertensos". En este caso se puede estar

a. Afirmando una conclusión falsa.
b. Cometiendo un error de tipo I.
c. Cometiendo un error muestral.
d. Cometiendo un error de tipo II.
e. Afirmando una conclusión verdadera.

7. Joan Carrasco es un ingeniero industrial en FAMIA Industrial y le gustaría determinar si se producen más unidades en el turno nocturno que en el matutino. Suponga que la desviación estándar de la población para el número de unidades producidas en el turno matutino es 21 y 28 en el nocturno. Una muestra de 54 trabajadores del turno matutino reveló que el número medio de unidades producidas fue 345. Una muestra de 60 trabajadores del turno nocturno reveló que el número medio de unidades producidas fue 351. Con un nivel de significación de 0.05, ¿es mayor el número de unidades producidas en el turno nocturno?

a. El valor crítico es -1.645, H0 no se rechaza; por tanto, $\mu d > \mu n$.
b. El valor de prueba es -1.96, se acepta H0, las medias son iguales.
c. Se rechaza H1, existe evidencia suficiente para afirmar $\mu d > \mu n$.
d. El valor de prueba es -1.3021, no se rechaza H0 y no se puede afirmar que $\mu d < \mu n$.
e. Se rechaza Ho y no se puede afirmar que sea mayor el número de unidades en el turno noche.

8. Dos muestras son independientes si

a. Se tiene la misma cantidad de datos por muestra.

b. Los datos de las muestras no están relacionados de alguna manera.

c. Las muestras provienen de poblaciones relacionadas, pero independientes.

a. Los datos están condicionados a un suceso que define su naturaleza.

e. Los datos se mantienen en igualdad de número por cada unidad de observación.

9. Se desea saber si existe relación entre el número de horas que pasan los alumnos en la biblioteca y su nivel de comprensión de lectura. Se estipulan 4 niveles de comprensión lectora y se toman muestras aleatorias. El número de horas no tiene datos extremos y las varianzas poblacionales no son muy diferentes. Un análisis ANOVA brinda la siguiente tabla:

Fuente de variación	Suma de Cuadrados	Gl	Cuadrados Medios	F	Valor P
Entre grupos	87.29	3	29.10	1.15	0.3349
Dentro de los grupos	2361.80	93	25.40		
	2449.09	96			

¿Las horas de estadía en la biblioteca incidirán en el nivel de comprensión lectora?

a. No se rechaza Ho; por tanto, no existe relación alguna entre las variables.

b. No se rechaza Ho; entonces, a mayor tiempo en la biblioteca mejor nivel de comprensión.

c. Se rechaza Ho; en consecuencia, existe evidencia para afirmar que existe influencia entre el número de horas en la biblioteca y el nivel.

d. Se rechaza Ho; entonces, existe algún tipo de relación las horas dependen del nivel.

e. No rechaza Ho; por tanto, los datos muestrales confirman que existe una relación significativa.

10. Dolor deespalda. El dolor de espalda baja (DEB) es un serio problema de salud en muchos entornos industriales. El artículo "Isodynamic Evaluation of Trunk Muscles and Low-Back Pain Among Workers in a Steel Factory" (Burdorf et al., 1995) reportó los datos adjuntos sobre rango lateral de movimiento (grados) para una muestra de trabajadores sin antecedentes de dolor de espalda baja y otra muestra con antecedentes de esta dolencia (Triola, 2011).

Condición	Tamaño de muestra	Media muestral	s
Sin DEB (dolor espalda baja)	28	91.5°	5.5°
Con DEB	31	88.3°	7.8°

Si no se tienen datos extremos, ¿estos resultados sugieren que el movimiento lateral medio difiere en las dos condiciones a un nivel de confianza de 95%?

a. $z = -1.645$, se rechaza H0, existe diferencia entre las dos condiciones.

b. $E = 3.419$, $-0.219 < \mu 1 = \mu 2 < 6.619$, no existe diferencia entre las dos condiciones.

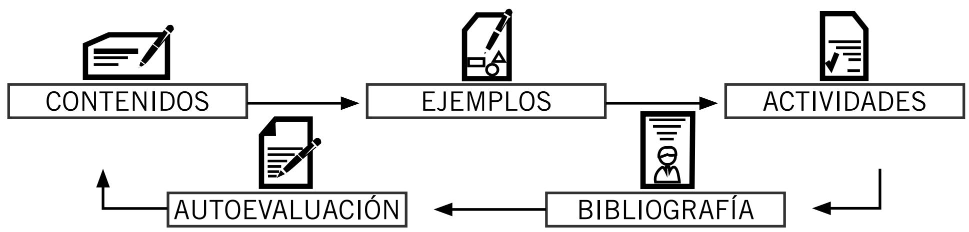
c. H1: $\mu 1 \neq \mu 2$, $z = 1.30$, no se rechaza H0, no existe diferencia.

d. H1: $\mu 1 < \mu 2$, $z = -1.30$, no se rechaza H0, no existen diferencias.

e. $E = 3.419$, $3.032 < \mu 1 = \mu 2 < 15.032$, no existe diferencia entre las dos condiciones.

UNIDAD III
ESTADÍSTICA NO PARAMÉTRICA

DIAGRAMA DE ORGANIZACIÓN DE LA UNIDAD III



ORGANIZACIÓN DE LOS APRENDIZAJES

Resultados del aprendizaje de la Unidad III: Al finalizar la unidad, el estudiante estará en la capacidad de realizar pruebas de hipótesis con experimentos multinomiales, pruebas de bondad o pruebas no paramétricas de acuerdo a la situación estadística planteada.

CONOCIMIENTOS	HABILIDADES	ACTITUDES
Tema n.° 1: Experimentos multinomiales y tablas de contingencia 1. Experimentos multinomiales. 2. Pruebas de bondad, tablas de contingencia y pruebas de independencia y homogeneidad. **Tema n.° 2: Pruebas no paramétricas** 1. Prueba de signos. 2. Prueba de rangos con signo-Wilcoxon. 3. Kruskal Wallys. 4. Correlación de rangos. 5. Prueba de rachas.	1. Identifica las clases de hipótesis. 2. Plantea pruebas de hipótesis. 3. Identifica correctamente los valores críticos para la aplicación de las pruebas de hipótesis. 4. Determina el procedimiento pertinente de la prueba de hipótesis. 5. Realiza la interpretación del resultado de la prueba de hipótesis. **Actividad 1** Participa del foro de discusión estimaciones en la empresa. **Actividad 2** Evaluación del tema n.° 1 y el tema n.° 2.	Valora reflexivamente la importancia de las pruebas de bondad y no paramétricas en la toma de decisiones.

Experimentos multinomiales y tablas de contingencia

Tema n.º 1

Los experimentos multinomiales se refieren al contraste de hipótesis con variables categóricas, con múltiples categorías. Su desarrollo implica cálculos con las frecuencias de las categorías en la muestra a lo que se denominará frecuencias observadas (O) y que se comparan con otros que responden a las hipótesis, se denominarán frecuencias esperadas (E).

Las pruebas pueden hacerse sobre una variable (bondad de ajuste) o con dos variables (independencia / homogeneidad).

1. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste comparan las frecuencias observadas (O) en la muestra con las que hipotéticamente deberían producirse para la distribución de frecuencias en la población. Tratamos, entonces, de determinar en función de las frecuencias muestrales (O) si la población se ajusta a una distribución específica (E).

Se puede tener tres formas generales de aplicación:

- Frecuencia esperada uniforme.
- Frecuencia esperada no uniforme (proporciones diferentes).
- Frecuencia esperada normal, binomial, de Poisson, etc.

Las pruebas se realizan usando la distribución Ji cuadrado (X^2).

Estadístico de prueba:

$$x^2 = \sum \frac{(O-E)^2}{E}$$

Valor crítico:

La prueba es de cola derecha siempre y los valores críticos se pueden encontrar en la tabla A-3 (ver Apéndice C) con

gl = K -1 y α

k es el número de categorías de la variable.

1.1. Prueba con frecuencias uniformes:

Ejemplo 29: Trabajaremos una adaptación del ejercicio de Díaz (2013), p. 335.

En la UC quiere evaluarse si el número de descomposturas de computadoras es igual todos los días de la semana. Para hacerlo se obtuvieron datos de descomposturas promedio por cada día de la semana. Los resultados se muestran a continuación:

Tabla 35. Datos de descomposturas promedio por cada día de la semana.

Día	Máquinas descompuestas
Lunes	13
Martes	14
Miércoles	6
Jueves	10
Viernes	8
Sábado	7
Domingo	12

Pruebe la hipótesis con un nivel de significación de 0.05.

Solución:

Datos:

n = 70

K = 7 (días)

gl = 6

α = 0.05

H0: O = E. La frecuencia de máquinas malogradas es igual en cualquier día.

H1: O ≠ E. La frecuencia de máquinas malogradas no es igual en cualquier día.

1.° Las frecuencias esperadas (E) son de acuerdo a H1 y la frecuencia total debe repartirse por igual entre las categorías: n/k = 70/7 = 10.

2.° Los datos muestrales toman el nombre de frecuencias observadas (O) ❶, mientras que las frecuencias 10 son las frecuencias esperadas (E) ❷.

Tabla 36. Solución del ejemplo 29.

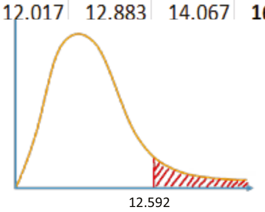
	❶	❷	❸	
Día	O	E	$\frac{(O-E)^2}{E}$	
Lunes	13	10	0.9	$=\frac{(13-10)^2}{10}$
Martes	14	10	1.6	$=\frac{(14-10)^2}{10}$
Miércoles	6	10	1.6	$=\frac{(6-10)^2}{10}$
Jueves	10	10	0	$=\frac{(10-10)^2}{10}$
Viernes	8	10	0.4	$=\frac{(8-10)^2}{10}$
Sábado	7	10	0.9	$=\frac{(7-10)^2}{10}$
Domingo	12	10	0.4	$=\frac{(12-10)^2}{10}$
	70	70	5.8	$= X^2$

3.° Calculamos el estadístico de prueba con cada par de frecuencias ❸. La suma de estos resultados es el estadístico de prueba.

4.° Regla de decisión con valor crítico en la tabla 37 (ver Apéndice C, tabla A-3) con gl = 6 y α= 0.05

Tabla 37. Regla de decisión con valor crítico

	α			
gl	0.100	0.075	0.050	0.0
3	6.251	6.905	7.815	9.3
4	7.779	8.496	9.488	11.1
5	9.236	10.008	11.070	12.8
6	10.645	11.466	12.592	14.4
7	12.017	12.883	14.067	16.0



Conclusión y respuesta:

Como 5.8 < 12.592 no rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para afirmar que el número de descomposturas de computadoras es igual todos los días de la semana.

Nota: Cuando se afirma que el número de descomposturas es igual en todos los días de la semana, en realidad se está hablando de una distribución uniforme.

1.2. Prueba con frecuencias no uniformes

Tabla 38. Tipo de vivienda

Tipos de vivienda particular	**2015**
Casa independiente	86.5%
Departamento en edificio	6.3%
Vivienda en quinta	1.6%
Vivienda en casa de vecindad (callejón, solar o corralón)	4.3%
Choza o cabaña	1.3%

Fuente: INEI (2016).

Ejemplo 30: El porcentaje encontrado para el tipo de vivienda que ocupan las familias a nivel nacional fue publicado por el INEI y es mostrado en la tabla adjunta (INEI, 2016).

Un estudio en nuestra región revela que, de una muestra de 150 hogares, 125 son casas independientes; 6, departamentos en edificio; 9, viviendas en quinta; 8, viviendas en vecindad, y 2, chozas o cabañas.

¿Al nivel del 5% se puede asegurar que las proporciones en nuestra región son diferentes a lo informado por el INEI?

Solución:

Datos:

$n = 150$

$K = 5$

$\alpha = 0.05$

$gl = 4$

H0: O = E Las proporciones regionales no son diferentes a lo informado por el INEI.

H1: O ≠ E Las proporciones regionales son diferentes a lo informado por el INEI.

1.° Desarrollamos la tabla de frecuencias. Las frecuencias O ❶ se toman de los datos de la muestra: 125, 6, 9, 8 y 2. Las frecuencias E ❷ se obtienen usando los porcentajes de la encuesta nacional del INEI.

Tabla 39. Tipos de vivienda: frecuencia observada y frecuencia esperada

Tipo vivienda	❶ O	❷ E		❸ $\frac{(O-E)^2}{E}$	
Casa independiente	125	129.75	=86.5% de 150	0.174	$=\frac{(125-129.75)^2}{129.75}$
Departamento	6	9.45	=6.3% de 150	1.260	$=\frac{(6-9,45)^2}{9.45}$
Vivienda en quinta	9	2.40	=1.6% de 150	18.150	$=\frac{(9-2.40)^2}{2.40}$
Vivienda en casa de vecindad.	8	6.45	=4.3% de 150	0.372	$=\frac{(8-6.45)^2}{6.45}$
Choza o cabaña	2	1.95	=1.3% de 150	0.001	$=\frac{(2-1.95)^2}{1.95}$
	150	150.00		19.957	$=\chi^2$

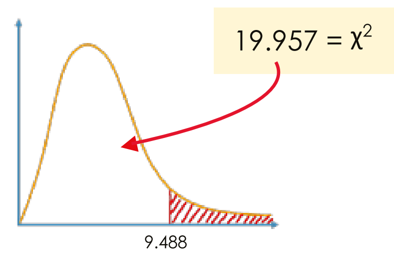
Fuente: Elaboración propia.

2.° Calculamos el estadístico de prueba ❸ y buscamos el valor crítico en la tabla 40(ver Apéndice C, tabla A-3) con gl = 4 y α = 0.05.

Tabla 40. Valor crítico con gl y a

	α		
gl	0.100	0.075	0.050
2	4.605	5.181	5.991
3	6.251	6.905	7.815
4	7.779	8.496	9.488
5	9.236	10.008	11.070

Fuente: Elaboración propia



Conclusión y respuesta:

Como 19.957 > 9.488 rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para afirmar que las proporciones en nuestra región son diferentes a lo informado por el INEI.

1.3. Prueba ajuste a una distribución estadística

Ejemplo 31: Trabajaremos un ejercicio adaptado de Mendehall (2010).

Tabla 41. Peso de paquetes de mensajería

Peso g	fo
150 a 200	8
200 a 250	15
250 a 300	15
300 a 350	15
350 a 400	10
400 a 450	1
Totales	64

Una compañía de servicios de mensajería registra el peso de 70 paquetes elegidos al azar. Determine si el peso de los paquetes se ajusta a una distribución normal con un nivel de significación de 0.01.

Solución:

HO: O = E El peso de los paquetes se ajusta a una distribución normal.

H1: O ≠ E El peso de los paquetes no se ajusta a una distribución normal.

Datos:

n = 64

k = 6

gl = 5

α = 0.01

1.° Para poder calcular las frecuencias esperadas, calculamos la media y la desviación estándar de la muestra:

$$\overline{X} = 280.469 \text{ y } s = 66.139$$

Tabla 42. Peso de paquetes de la mensajería: cálculos.

Peso g.	fo	❶ z	❷ Áreas Acum.	❸ Proporción
150 a 200	8	-1.22	0.1112	0.1112
200 a 250	15	-0.46	0.3228	0.2116
250 a 300	15	0.30	0.6179	0.2951
300 a 350	15	1.05	0.8531	0.2352
350 a 400	10	1.81	0.9649	0.1118
400 a 450	1	2.56	0.9948	0.0052
Totales	64			≈ 1

Entonces, la distribución normal a la que se ajustan los datos tiene una media y desviación estándar como las calculadas.

Si convertimos los pesos a "z" (distribución normal estándar), por cada intervalo existe entonces una proporción (área) en la curva normal estándar:

$$z = \frac{xi - \bar{x}}{s}$$ ❶

$$z = \frac{200 - 280.469}{66.139}$$

$$z = 1.22$$

Buscamos en la tabla el área acumulada:

A = 0.1112

Se procede igual con el resto.

2.° Cálculo de las proporciones ❸. En este caso, como se trata de áreas acumuladas se restan cada una con la anterior.

A la primera se le resta 0 → 0.1112-0= 0.1112

Para la segunda 0.3228 – 0.1112 = 0.2116

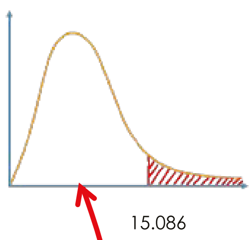
Tercera 0.6179 – 0.3228 = 0.2951

Así con todos los demás excepto el último, ya que se trata de un área cola derecha. Entonces, la proporción buscada es 1 – 0.9948 = 0.0052.

3.° Cálculo de frecuencias E ❷: Utilizaremos las proporciones halladas:

Tabla 43. Frecuencia esperada de peso

Peso g.	❶ O	❷ E		❸ $\frac{(O-E)^2}{E}$
150 a 200	8	7.1168	=0.1112 * 64	0.1096
200 a 250	15	13.5424	=0.2116 * 64	0.1569
250 a 300	15	18.8864	=0.2951 * 64	0.7997
300 a 350	15	15.0528	=0.2352 * 64	0.0002
350 a 400	10	7.1552	=0.1118 * 64	1.1310
400 a 450	1	0.3328	=0.0052 * 64	1.3376
Totales	64		χ^2 =	3.5351



4º Se calcula el estadístico de prueba ❸ y buscamos el valor crítico en la tabla 44 (ver Apéndice C, A-3) con gl = 5 y α = 0.01.

Tabla 44. Estadístico de prueba

gl	0.050	0.025	0.020	0.010	0.0
3	7.815	**9.348**	9.837	11.345	12.8
4	9.488	**11.143**	11.668	13.277	14.8
5	11.070	**12.833**	13.388	15.086	16.7
6	12.592	**14.449**	15.033	16.812	18.5

Conclusión y respuesta:

Como 3.5351 < 15.086 no rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para afirmar que el peso de los paquetes se ajusta a una distribución normal.

2. Pruebas de independencia y homogeneidad

Una prueba de independencia u homogeneidad trabaja sobre datos organizados en función de dos variables en una tabla de doble entrada que se conoce como tabla de contingencia.

Tabla 45. Tabla de contingencia

		Variable B		
		Cat. B1	**Cat. B2**	**Cat. B3**
Variable A	**Cat. A1**	f_{11}	f_{22}	f_{33}
	Cat. A2	f_{44}	f_{55}	f_{66}

Fuente: elaboración propia.

2.1. Pruebas de independencia

La finalidad de una prueba de este tipo es averiguar si existe alguna forma relación entre las variables A y B. Se someten a prueba las siguientes Hipótesis:

H0: O = E. La variable A es independiente de la variable B.

H1: O ≠ E. La variable A no es independiente de la variable B.

Se aplica a variables de tipo categórico, generalmente cualitativas.

Si se llega a rechazar H0 y se confirma que A no es independiente de B, entonces solo se puede afirmar que existe relación entre ellas, pero no se puede decir si existe dependencia tanto como en un análisis de correlación lineal.

Los datos son conteos, es decir, la frecuencia de ocurrencia del cruce de categorías de la variable A y B.

Estadístico de prueba:

$$x^2 = \sum \frac{(O-E)^2}{E}$$

Valor crítico:

La prueba es de cola derecha siempre y los valores críticos se pueden encontrar en la tabla A-3 (ver Apéndice C) con:

gl = (f -1)(c-1) y α

f: es el número de filas de datos y **c** es el número de columnas de datos.

La frecuencia esperada se calcula con:

$$E = \frac{(Total\ fila)\ (Total\ columna)}{Total}$$

Ejemplo 32: Desarrollaremos un ejercicio de Triola (2013).

La tabla siguiente incluye datos obtenidos de víctimas de crímenes elegidas al azar (según datos del Departamento de Justicia de Estados Unidos). ¿Qué podemos concluir? Aplique una prueba de independencia al 0.05 de significancia.

Tabla 46. Datos sobre criminales y sus delitos cometidos.

	Homicidio	**Robo**	**Asalto**
El criminal era un extraño.	12	379	727
El criminal era conocido o pariente.	39	106	642

Fuente: Triola, 2013 (datos del Departamento de Justicia de Estados Unidos).

Solución:

H0: O = E. La variable tipo de delito es independiente de la variable tipo de criminal.

H1: O ≠ E. La variable tipo de delito no es independiente de la variable tipo de criminal.

1.° Calculamos los totales en la tabla O:

Tabla 47. Tabla O.

	Homicidio	**Robo**	**Asalto**	**Totales**
Criminal es extraño	12	379	727	1118
Criminal es conocido	39	106	642	787
Totales	51	485	1369	1905

2.° Para determinar los valores de la tabla de frecuencias esperadas, calculamos las frecuencias $E = \frac{\textit{(Total columna) (Total fila)}}{\textit{Total}}$, copiamos la tabla con los totales y procedemos por cada casilla:

Tabla 48. Cálculo de frecuencias esperadas

	Homicidio	**Robo**	**Asalto**	**Totales**
Criminal es extraño	$\frac{(51)(1118)}{1905}$	$\frac{(485)(1118)}{1905}$	$\frac{(1369)(1118)}{1905}$	1118
Criminal es conocido	$\frac{(51)(787)}{1905}$	$\frac{(485)(787)}{1905}$	$\frac{(1369)(787)}{1905}$	787
Totales	51	485	1369	1905

Y obtenemos la tabla de frecuencias E:

Tabla 49. Tabla de frecuencias esperadas

	Homicidio	**Robo**	**Asalto**	**Totales**
Criminal es extraño	29.9307	284.6352	803.4341	1118
Criminal es conocido	21.0693	200.3648	565.5659	787
Totales	51	485	1369	1905

3.° Con los datos de la tabla de frecuencias O y la tabla de frecuencias $\frac{(O-E)^2}{E}$ calculamos para cada par de datos:

Tabla 50. Tabla de cálculo de FO y FE.

	Homicidio	Robo	Asalto
Criminal es extraño	12	379	727
Criminal es conocido	39	106	642

Tabla de frecuencias o

	Homicidio	Robo	Asalto
Criminal es extraño	29.9307	284.6352	803.4341
Criminal es conocido	21.0693	200.3648	565.5659

Tabla de frecuencias E

Tabla 51. Resultados del cálculo FO y FE.

	Homicidio	Robo	Asalto
Criminal es extraño	$\frac{(12-29.9307)^2}{29.9307}$	$\frac{(379-200.3648)^2}{200.3648}$	$\frac{(727-803.4341)^2}{803.4341}$
Criminal es conocido	$\frac{(39-21.0693)^2}{21.0693}$	$\frac{(106-200.3648)^2}{200.3648}$	$\frac{(642-565.5659)^2}{565.5659}$

4.° Sume todos los resultados y hallará el estadístico de prueba:

10.7418	31.2847	7.2715	
15.2596	44.4425	10.3298	
		$\chi^2 =$	119.3299

5.° Determine el valor crítico en la tabla 51 con gl = (f-1)(c-1) y α = 0.05

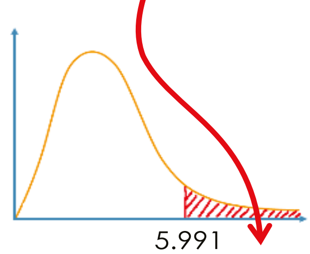
La tabla tiene f=2 filas y c= 3 columnas de datos, por tanto:

gl = (2-1)(3-1) = 2

α = 0.05

Tabla 52. Valor crítico

gl	α 0.100	0.075	0.050	0.02
1	2.706	3.170	3.841	5.02
2	4.605	5.181	5.991	7.37
3	6.251	6.905	7.815	9.34



Conclusión y respuesta:

Como 119.33 > 5.991 rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para afirmar que el tipo de delito no es independiente del tipo de criminal.

Pruebas no paraméticas

Tema n.° 2

Las pruebas no paramétricas son pruebas de hipótesis desarrolladas con métodos que aplican a muestras obtenidas de poblaciones no normales. Por ello el nombre más apropiado sería "pruebas de libre distribución". Las pruebas que revisaremos en esta unidad son las siguientes:

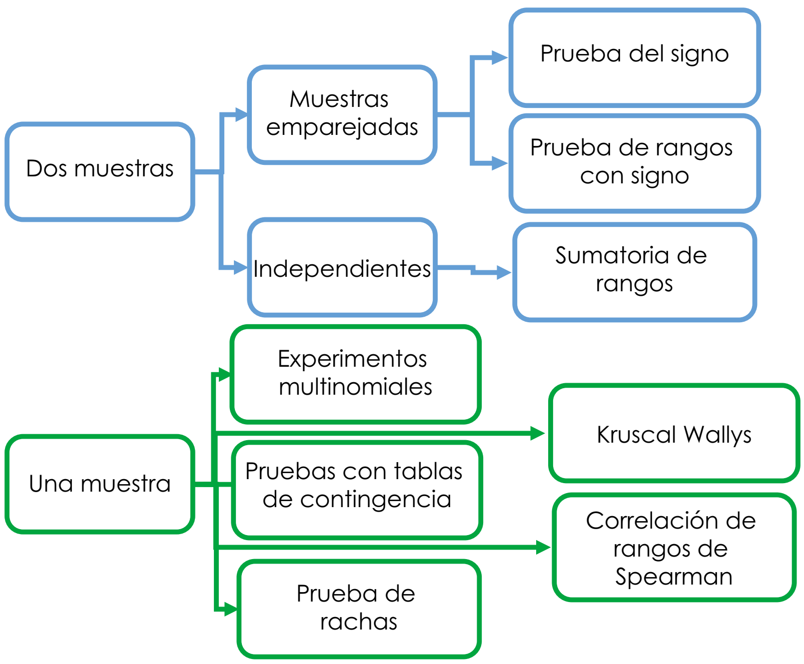


Figura 17: Pruebas no paramétricas. Fuente: Elaboración propia

1. Prueba del signo

Una prueba de signo es utilizada en dos oportunidades: con muestras emparejadas, y cuando se quiere probar el valor de una mediana.

La prueba del signo requiere:

- Muestras aleatorias.
- No se conoce la distribución de la población o no es requisito que sea normal.
- Se puede identificar dos agrupaciones, las que tomarán los signos (+) y (-).
- Los valores críticos se encuentran en la tabla A-7.
- Si $n \leq 25$, el estadístico de prueba será T; el número de veces que se repite el signo menos frecuente.
- Si $n > 25$, el estadístico de prueba es $z = \dfrac{(T+0.5)-\frac{n}{2}}{\frac{\sqrt{n}}{2}}$

Tabla 53. Número de fusibles defectuosos

Día	Línea A	Línea B
1	170	201
2	164	179
3	140	159
4	184	195
5	174	177
6	142	170
7	191	183
8	179	179
9	161	170
10	220	212

Ejemplo 33: Los números de fusibles eléctricos defectuosos producidos por dos líneas de producción, A y B, se registraron a diario durante un periodo de 10 días, con los resultados mostrados en la tabla. La variable de respuesta, el número de fusibles defectuosos, tiene una distribución binomial exacta con un gran número de fusibles producidos por día. Aun cuando esta variable tendrá aproximadamente una distribución normal, el supervisor de planta preferirá una prueba estadística rápida y fácil para determinar si una línea de producción tiende a producir más fusibles defectuosos que la otra. Use la prueba del signo para probar la hipótesis apropiada (Mendehall, 2010).

Solución:

H0: $Med_{dif} = 0$

H1: $Med_{dif} \neq 0$

1.° Se debe tomar las muestras como emparejadas, ya que por cada día se tiene dos datos (A y B).

Tabla 54. Número de fusibles con signos

Día	Línea A	Línea B	Signos
1	170	201	-
2	164	179	-
3	140	159	-
4	184	195	-
5	174	177	-
6	142	170	-
7	191	183	+
8	179	179	0
9	161	170	-
10	220	212	+

2.° Se restan los datos en pares y solo se anotan los signos de los resultados.

3.° El estadístico de prueba es el número de signos menos frecuente: T = 2 (solo dos positivo).

4.° n = 9; porque se eliminan los ceros. Solo toman en cuenta las diferencias que no son cero.

5° Como n ≤ 25, la regla de decisión es
Si T ≤ Valor crítico, entonces se rechazar H0

6° El valor crítico estará en la tabla 55.

Tabla 55. Valor crítico

n	0.005	0.01	0.025	0.05	1 Cola
	0.01	0.02	0.05	0.1	2 colas
6	•	•	0	0	
7	•	0	0	0	
8	0	0	0	1	
9	0	0	1	1	

Conclusión y respuesta:

Como T = 2 > Valor crítico = 1 No rechazamos H0 como verdadera.

Por tanto, no existe evidencia muestral suficiente para afirmar que una línea de producción tiende a producir más fusibles defectuosos que la otra.

Ejemplo 34: CPP es una empresa dedicada a la producción de pinturas. Intenta desarrollar un nuevo compuesto que prolongue el tiempo de vida útil de sus pinturas para exteriores. Prueba 35 especímenes de los que obtiene una duración en días presentada a continuación:

Tabla 56. Duración en días de pinturas para exteriores.

722	723	700	742	710	712	732	756	756	720
756	734	690	756	720	743	737	745	768	698
756	706	705	710	712	745	730	729	735	745
708	695	700	727	730					

¿El nuevo compuesto en realidad aumenta el tiempo de vida útil de las pinturas? Realice la prueba de signos si la mediana actual de la vida útil de las pinturas es de 720 días.

Solución:

H0: Med = 720
H1: Med ≠ 720 (prueba a dos colas)

1.° A todos los valores obtenidos en la muestra se resta el valor de 720 y se obtienen:

21 signos (+)
12 signos (–)

2 diferencias = 0

2.° El valor T = 12 y n = 33 > 25:

$$Z = \frac{(T+0.5)-\frac{n}{2}}{\frac{\sqrt{n}}{2}}$$

$$Z = \frac{(12+0.5)-\frac{33}{2}}{\frac{\sqrt{33}}{2}} = -1.393$$

3.° Buscamos el valor P en la tabla 57 (ver tabla A-2, en el Apéndice C):

Tabla 55. Valor P

Z	0.4	0.05	0.06
-1.6	0.0505	0.0495	0.0485
-1.5	0.0618	0.0606	0.0594
-1.4	0.0749	0.0735	0.0721
-1.3	0.0901	0.0885	0.0869
-1.2	0.1075	0.1056	0.1038

4.° El valor P es el doble del valor que aparece en la tabla:

Valor P = 2(0.0869) = 0.1738

Valor P > α = 0.05

Conclusión y respuesta:

Como valor P > α no rechazamos H0 como verdadera.

Por tanto, no existe evidencia muestral suficiente para afirmar que el nuevo compuesto en realidad aumenta el tiempo de vida útil de las pinturas.

2. Prueba de rangos con signo de Wilcoxon para datos apareados

En este caso se trabaja con los rangos de las diferencias. Cada diferencia es ordenada de menor a mayor sin considerar sus signos. Los signos son dos: positivos y negativos y vuelven a considerarse para poder agrupar dos sumas. La suma menor en valor absoluto es el valor o estadístico T. Para esta prueba se requiere:

- Muestras aleatorias emparejadas.
- No es necesario que las poblaciones sean normales.

Tabla 58. Datos requeridos para trabajar prueba de rangos con signo de Wilcoxon

Tamaño de muestra	Estadístico de prueba	Valor crítico VC	Rechazar H0
n ≤ 30	T = Suma menor	Tabla A - 8	Si T ≤ VC
n > 30	$z = \frac{T - \mu_2}{\sigma_T}$	Tabla A - 2	Si Valor P ≤ α

Ejemplo 35: **MUESTRAS PEQUEÑAS**: Trabajaremos un ejemplo modificado de (Díaz, 2013). Para evaluar si la participación de psiquiatras en el proceso de recuperación de alcohólicos mejoraba su patrón de conducta, en un estudio médico se evaluó una muestra de 12 pacientes. Al principio del tratamiento y al cabo de 2 meses, se obtuvieron las siguientes calificaciones:

Solución:

H0: $Med_{dif} = 0$

H1: $Med_{dif} < 0$ (una cola)

Son muestras emparejadas: por cada paciente se tiene dos datos (inicial y final).

1.° Con los datos de calculan las diferencias ❶, se retiran los signos ❷.

2.° Asigne rangos ❸; es decir, decimos qué diferencia es primero, segundo, tercero, etc.

3.° Tome en cuenta que en la lista de diferencias ❷ se repiten el 1 y el 3 (se les dice empates), por lo que se debe promediar los rangos para lograr un valor crítico real ❹. Así, en el caso del número 1 que se repite 5 veces se promedian sus rangos (1+2+3+4+5)/5 =3 y se cambia todos estos rangos por 3 en ❹. Luego, trabaje lo mismo con los rangos del número 3 que se repite 2 veces. Promediamos los rangos (6+7)/2 = 6.5 y este es el rango asignado a este empate en ❹.

4° Sume por separado los rangos negativos y positivos ❺. El signo corresponde a la diferencia original ❶.

Tabla 59. Datos de pacientes de alcoholismo

	Calificación		1	2	3	4	5	
Pacien.	**Inicial**	**Final**	**Dif.**	**Dif. s7 signos**	**Rangos**	**Rangos prom.**	**Rangos signo (-)**	**Rangos signo (+)**
1	31	32	**-1**	**1**	**1**	**3**	3	
2	40	39	**1**	**1**	**2**	**3**		3
3	39	42	-3	3	6	6.5	6.5	
4	88	99	-11	11	10	10	10	
5	57	73	-16	16	11	11	11	
6	56	57	**-1**	**1**	**3**	3	3	
7	65	60	5	5	8	8		8
8	35	45	-10	10	9	9	9	
9	68	71	-3	3	7	6.5	6.5	
10	71	70	**1**	**1**	**4**	**3**		3
11	80	80	0					
12	84	85	**-1**	**1**	**5**	**3**	-1	
							52	14

5° Como n = 11 (se eliminó una diferencia 0) es una muestra pequeña (n<25); por lo tanto, nuestro estadístico de prueba es la suma menor:

T = 14

6° Busque el valor crítico en la tabla A-8 con n = 11 y α = 0.05.

n	α				
	0.005	0.01	0.025	0.05	1 cola
	0.01	0.02	0.05	0.1	2 colas
9	2	3	6	8	
10	3	5	8	11	
11	5	7	11	14	
12	7	10	14	17	

Conclusión y respuesta:

Como T = Valor crítico rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para afirmar que existe mejora el patrón de conducta, con la participación de psiquiatras en el proceso de recuperación de alcohólicos.

3. Pruebas de suma de rangos con signo de Wilcoxon para muestras independientes

Es una prueba no paramétrica aplicada a muestras independientes. Las muestras independientes se distinguen porque los datos pertenecen a dos poblaciones diferentes. La prueba requiere de:

- Dos muestras independientes.
- Las poblaciones pueden tener cualquier tipo de distribución.
- "Ordene todas las n1 y n2 observaciones de menor a mayor.
- Encuentre T_1, la suma de rangos para las observaciones de la muestra 1. Éste es el estadístico de prueba para la prueba de cola izquierda.

- Encuentre T_{1*} = n1(n1+n2+1) - T_1, la suma de los rangos de las observaciones de la población 1 si los rangos asignados se hubieran invertido de grandes a pequeños. (El valor de T_{1*} no es la suma de los rangos de las observaciones de la muestra 2.) Este es el estadístico T de prueba para una prueba de cola derecha.

- "El estadístico de prueba para una prueba de dos colas es T, la mínima de T_1 y T_{1*}" (Mendenhal, 2010).

- H0 es rechazada si el estadístico de prueba observado es menor o igual al valor crítico hallado usando la tabla A-8 (ver Apéndice C).

- Se puede emplear una transformación normal:

$$z = \frac{T - \mu_T}{\sigma_T}$$

si $\mu_T = \dfrac{n_1(n_1 + n_2 + 1)}{2}$ y $\sigma_T^2 = \dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}$

Ejemplo 36: En el desarrollo de nuevas formas de construcción para reparar edificios históricos hechos en barro y quincha, se ha agregado fibra de papel a la argamasa de barro para probar su resistencia al paso del tiempo. Los datos obtenidos de resistencia al golpe se presentan a continuación para un mortero común y otro con fibra de papel. ¿Se puede verificar una mejora en la resistencia? Utilice un nivel de significancia de 0.01.

Tabla 60. Resistencia de mortero común y mortero con fibra de papel

Resistencia en kg/cm²									
Común	93	104	92	96	97	108	104	96	89
	88	90	102	105	97	96	101	108	
Fibra de papel	89	90	106	97	109	107	106	96	100
	108	110	100	98	97	101	99		

Solución:

H0: Med1 = Med2 Las medianas de resistencia son iguales.

H1: Med1 < Med2 El mortero común tiene una mediana de resistencia menor al que tienen fibra de papel.

1° Asignamos rangos a todos los datos como si se tratase de una sola muestra. Recuerde que si existen empates como 89 (dos veces), se debe promediar los rangos (2+3)/2 = 2.5.

Tabla 61. Rangos para resistencia de mortero común y mortero con fibra de papel

	Resistencia en kg/cm²								
Común	93	104	92	96	97	108	104	96	**89**
	7	23.5	6	9.5	13.5	30	23.5	9.5	2.5
	88	90	102	105	97	96	101	108	
	1	4.5	22	25	13.5	9.5	20.5	30	
Fibra de papel	**89**	90	106	97	109	107	106	96	100
	2.5	4.5	26.5	13.5	32	28	26.5	9.5	18.5
	108	110	100	98	97	101	99		
	30	33	18.5	16	13.5	20.5	17		

2° Como se trata de una prueba de una cola a la izquierda hallamos el **estadístico** de prueba T_1:

- T_1 = suma de rangos de la muestra 1
- T_1 = 7+23.5+6+9.5+13.5+30+23+ . . . . . .+20.5+30 = **251**

3° Transformamos el estadístico en un valor z:

$$\mu_T = \frac{n_1(n_1+n_2+1)}{2} = \frac{17*(17+16+1)}{2} = \mathbf{289}$$

$$\sigma_T^2 = \frac{n_1 n_2(n_1+n_2+1)}{12} = \frac{17(16)(17+16+1)}{12} = \frac{\mathbf{2312}}{\mathbf{3}}$$

$$z = \frac{T-\mu_T}{\sigma_T} = \frac{251-289}{\sqrt{\frac{2312}{3}}} = -1.37$$

4° Hallamos el valor P en la tabla A-2 (ver Apéndice C).

z	0.05	0.06	0.07
-1.6	0.0495	0.0485	0.0475
-1.5	0.0606	0.0594	0.0582
-1.4	0.0735	0.0721	0.0708
-1.3	0.0885	0.0869	0.0853
-1.2	0.1056	0.1038	0.1020

Valor P = 0.0853

Conclusión y respuesta:

Como Valor P > α no rechazamos H0 como verdadera.

Por tanto, existe evidencia muestral suficiente para verificar una mejora en la resistencia. En todo caso sigue igual.

4. Prueba de Kruskal Wallis

Es una prueba no paramétrica que utiliza rangos de muestras aleatorias simples de tres o más poblaciones independientes. Se utiliza para someter a prueba la hipótesis nula de que las poblaciones tienen medianas iguales. (La prueba es una prueba alternativa a la prueba ANOVA).

Para aplicar la prueba de Kruskal-Wallis, calculamos el estadístico de prueba H, el cual tiene una distribución que puede aproximarse por medio de la distribución chi cuadrada, siempre y cuando cada muestra tenga al menos cinco observaciones. Cuando utilizamos la distribución chi cuadrada en este contexto, el número de **grados de libertad es k - 1**, donde k es el número de muestras (para una revisión rápida de las características clave de la distribución chi cuadrada).

El estadístico de prueba H es básicamente una medida de la varianza de las sumas de rangos R1, R2,…, Rk. Si los rangos están distribuidos de forma equitativa entre los grupos muestrales, entonces H debe ser un número relativamente pequeño. Si las muestras son muy diferentes, entonces los rangos serán excesivamente bajos en algunos grupos y altos en otros, con el efecto neto de que H será grande. En consecuencia, solo los valores grandes de H nos conducen al rechazo de la hipótesis nula de que las muestras provienen de poblaciones idénticas. La prueba de Kruskal-Wallis es, por lo tanto, una prueba de cola derecha (Triola, 2013).

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \ldots\right] - 3(N+1)$$

N: Total de datos en todas las muestras.

R1 , R2, R3,… suma de rangos de las muestras 1 2 3… respectivamente.

n1, n2, n3,… Tamaño de las muestras 1, 2, 3… respectivamente.

Las hipótesis a trabajar son:

H0: Las medianas de las poblaciones son iguales.

H1: Las medianas de las poblaciones no son iguales

Las condiciones necesarias para aplicar este procedimiento son:

- Se tiene 3 o más poblaciones.
- Las poblaciones pueden no ser normales.
- Las muestras son aleatorias.
- Cada muestra tiene como mínimo 5 observaciones ($n_i \geq 5$).

Tabla 62. Modelo de gestión.

Por procesos	Funcional	EFQM
77	52	67
64	53	72
67	44	52
62	51	54
42	44	52
	51	66
	54	

Ejemplo 37: En su trabajo de tesis, Franco Alvarado desarrollo la hipótesis de que dependiendo del Modelo de Gestión las empresas obtienen calificaciones en desempeño diferentes.

La tabla adjunta es una muestra a aleatoria de sus datos. En base a esta información ¿se puede afirmar alguno de los modelos es mejor que los otros?

Desarrollo una prueba al 0.10 de significancia.

Tabla 63. Modelo de gestión

Por procesos		**Funcional**		**EFQM**	
77	18	52	7	67	15.5
64	13	53	9	72	17
67	15.5	44	2.5	52	7
62	12	521	4.5	54	10.5
42	1	44	2.5	52	7
		51	4.5	66	14
		54	10.5		
	59.5		40.5		71

Solución:

Datos:

N = 18

k = 3

gl = 2

α = 0.10

n_1 = 5

n_2 = 7

n_2 = 6

H0: Las medianas de las poblaciones son iguales.

H1: Las medianas de las poblaciones no son iguales

1° Asignamos rangos a todos los datos a todo como si fuese una sola muestra.

2° Sume los rangos en de cada muestra para hallar R1, R2 y R3:

R1 = 59.5

R2 = 40.5

R3 = 71

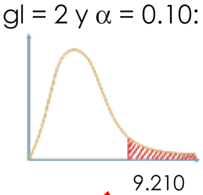
3° Calcule el estadístico de prueba:

$$H=\frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1}+\frac{R_2^2}{n_2}+\frac{R_3^2}{n_3}\right]-3(N+1)$$

$$H=\frac{12}{18(18+1)}\left[\frac{59.5^2}{5}+\frac{40.5^2}{7}+\frac{71^2}{6}\right]-3(18+1)$$

H = 5.545

4° El valor crítico de la tabla A-3 con gl = 2 y α = 0.10:



Conclusión y respuesta:

Como **H <** 9.210 no rechazamos H0 como verdadera.

Por tanto, no existe evidencia muestral suficiente afirmar que alguno de los modelos es mejor que los otros. En todo caso son calificados por igual.

5. Correlación de rangos de Spearman

5.1. Correlación

Se refiere al hecho de comprobar que existe algún tipo de relación entre los datos de dos variables. Es decir que si de alguna manera al cambiar los valores de los datos, en alguna medida cambian los valores de la otra variable.

Medida de la correlación

No todas las variables tienen el mismo grado de correlación, por ello es indispensable medir cuán relacionadas están. La medida de la correlación se calcula mediante el coeficiente de correlación de Spearman.

Coeficiente de Spearman

ρ_s es el coeficiente de correlación de Spearman para la población.

$\mathbf{r_s}$ es el coeficiente de correlación de Spearman para la muestra.

Este coeficiente tiene valores que varían desde **-1 hasta 1**.

El valor del coeficiente se puede interpretar de la siguiente manera:

- $|rs| = 1$ La correlación es perfecta
- $0.6 < |rs| < 1$ Fuerte correlación
- $|rs| \approx 0$ Incorrelación (no existe correlación)

Se entiende entonces que cuanto más cerca de -1 o 1 la correlación es más fuerte.

5.2. Prueba de hipótesis para la correlación

Hipótesis

Se prueban las siguientes hipótesis:

H0: $\rho_s = 0$ No existe correlación entre las variables A y B.

H1: $\rho_s \neq 0$ Existe correlación entre las variables A y B.

Estadístico de prueba:

Para elegir el estadístico de prueba:

No existen empates en los rangos. $$r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)}$$

Existen empates en los rangos. $$r_s = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

Valor crítico:

Depende del tamaño de muestra o n = número de pares de datos:

n < 30 Tabla A-9

n ≥ 30 $rs = \dfrac{\pm z}{\sqrt{n-1}}$

Regla de decisión:

Si $|r_s|$ > valor crítico; rechazar H0

Ejemplo 38 El artículo “Objective Measurement of the Stretchability of Mozzarella Cheese” (J. of Texture Studies, 1992, 185-194) reportó sobre un experimento para investigar la variación del comportamiento del queso mozzarella con la temperatura. Considere los datos adjuntos sobre temperatura y alargamiento (%) en el momento de la falla del queso (Devore, 2008).

Tabla 64. Temperatura y alargamiento del queso

Temp. °F	59	63	68	72	74	78	83
% Alarg	118	182	247	208	197	135	132

Desarrolle una prueba, si no se conoce la distribución de las poblaciones, para sustentar la aseveración de que existe algún tipo de relación entre estas dos variables:

Solución:

H0: $\rho_s = 0$ No existe correlación entre las variables temperatura y alargamiento.

H1: $\rho_s \neq 0$ Existe correlación entre las variables temperatura y alargamiento.

1° Como ninguna de las dos variables tiene datos expresados en rangos, debemos iniciar con asignar rangos por separado.

1	2	3	4	5	6	7	
Temp. °F	59	63	68	72	74	78	83
% Alarg	118	182	247	208	197	135	132
1	4	7	6	5	3	2	

2° No existen empates en las variables, así que el estadístico de prueba se calcula con:

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)}$$

Los rangos asignados en cada variable no ayudan a obtener :

x	1	2	3	4	5	6	7	
y	1	4	7	6	5	3	2	suma
d	0	-2	-4	-2	0	3	5	0

$\Sigma d^2 = 0$, y reemplazando en la fórmula $\Rightarrow r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)} = 1 - \frac{6(0)}{7\left(7^2-1\right)}$

$r_s = 1$

3° El valor crítico obtenido de la tabla A-6 con n = 7 y $\alpha = 0.05$ es: 0.786

Conclusión y respuesta:

Como $r_s > 0.786$ rechazamos H0 como verdadera.
Por tanto, existe evidencia muestral suficiente afirmar que existe correlación entre las variables temperatura y alargamiento.

6. Prueba de rachas

Llamada también prueba de aleatoriedad, es un procedimiento para determinar si un conjunto de datos secuenciales es aleatorio. La prueba requiere:

- La muestra esté ordenada de manera secuencial, es decir los datos se encuentren tal y como fueron obtenidos.
- Los datos están caracterizados en dos categorías.

Hipótesis:

H0: La secuencia es aleatoria.

H1: La secuencia no es aleatoria.

Estadístico de prueba:

Si $n_1 \leq 20$ y $n_2 \leq 20$ y $\alpha = 0.05$ G (número de rachas)

Si $n_1 > 20$ y $n_2 > 20$ $z = \frac{G - \mu_G}{\sigma_G}$

$$\mu_G = \frac{2n_1n_2}{n_1+n_2} + 1 \ \text{ y } \ \sigma_G = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

El valor crítico:

Si $n_1 \leq 20$ y $n_2 \leq 20$ y $\alpha = 0.05$	Tabla A-10 con n_1 y n_2
Si $n_1 > 20$ y $n_2 > 20$	Tabla A-2

Ejemplo 39 Se ha planteado una investigación sobre la preferencia de ladrillos artesanales (A) y ladrillos industriales (I). Una muestra de 15 hogares presenta la siguiente secuencia de resultados:

A A I A A A I I I A A A I A A

Al nivel del 5 % de significancia, la secuencia de datos es aleatoria[1].

Solución:

Datos:

$n_1 = 10$ (A)	H0: La secuencia de datos es aleatoria.
$n_2 = 5$ (I)	H1: La secuencia de datos no es aleatoria.

1° Determine el número de rachas (G) contando las agrupaciones que se han producido en los datos de cada una de las dos categorías:

A A I A A A I I I A A A I A A

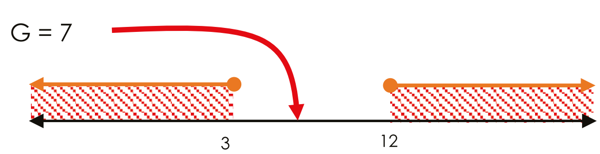
1 2 3 4 5 6 7

Existe 7 agrupaciones consecutivas entonces: G = 7 rachas

2° Determine los valores críticos en la tabla A-10 con $n_1 = 10$ y $n_2 = 5$.

		Valores n2				
		2	3	4	5	6
Valores n1	9	1	2	3	3	4
		6	8	10	12	13
	10	1	2	3	3	4
		6	8	10	12	13

Con estos valores se puede configurar una regla:



1 Si desea ver otro ejemplo con datos numéricos revise las diapositivas en: https://goo.gl/ccIIRy

Conclusión y respuesta:

Como G = 7 caen la zona de no rechazo, aceptamos H0 como verdadera. Por tanto, existe evidencia muestral suficiente afirmar que la secuencia de datos es aleatoria.

Lectura seleccionada n.° 3

Quinteros, P. (2009). La importancia de la evaluación del desempeño – II parte, [artículo en blog]. Disponible en: goo.gl/1iAWhH

Glosario de la Unidad III

E

Experimento. Actividad realizada con el propósito de obtener datos que permitan probar una teoría o hipótesis.

M

Muestras emparejadas. Revise muestras relacionadas.

Muestras independientes. Son dos o más muestras que tienen elementos que no tienen ninguna relación con otros en otras muestras. De ninguna manera forman parejas o se asocian a valores de otras muestras.

Muestras relacionadas. Dos o más muestras que tienen elementos que forman pares o asociaciones con los datos de otras muestras y que en general pertenecen a una misma unidad de análisis.

Multinomial. Se refiere a una variable que tiene múltiples categorías como puesto de trabajo: gerente, jefe de área, secretaria.

P

Prueba paramétrica. Procedimiento estándar para realizar una prueba de hipótesis con una población que tiene una distribución normal o casi normal.

Prueba no paramétrica. Procedimiento para realizar una prueba de hipótesis cuando no se conoce el tipo de distribución de las poblaciones.

R

Racha. Secuencia de sucesos iguales como bueno, bueno, bueno . . . o malo, malo, malo . . .

Bibliografía de la Unidad III

APTiTUS. (julio de 2009). La importancia de la evaluación del desempeño - II parte [blog post]. Recuperado de: goo.gl/pw04QD

Devore, J. (2008). *Probabilidad y estadística para ingeniería y ciencias*. México D. F.: CENGAGE Learning.

Díaz, A. (2013). *Estadística aplicada a la administración y la economía*. México D. F.: Mc Graw Hill.

INEI. (12 de mayo de 2016). Índice temático: población y vivienda. Recuperado de: goo.gl/3fRuUA

Levin, R., & Rubin, D. (2004). *Estadística para administración y economía* (7ma. ed.). México: Pearson Education.

Mendehall, W. (2010). *Introducción a la probabilidad y estadística* (13.ª ed.) (trad. Jorge Humberto Romo Muñoz). México D. F.: CENGAGE Learning. Recuperado de: goo.gl/zADbwr

Triola, M. (2013). *Estadística* (11.ª ed.). México D. F.: Pearson Education.

Autoevaluación n.º 3

Lea con atención los enunciados. Repase en el en manual el tema relacionado a la pregunta e intente una respuesta que se ajuste a su lectura y al criterio de aplicación de esta teoría.

1. **En el desarrollo de un nuevo método para la determinación de niveles de alcohol en la sangre, se analizó cinco veces una muestra de sangre, con los resultados siguientes: 64.5, 66.0, 63.9, 65.1 y 64 mg/100 ml. El método de análisis estándar aplicado a la misma muestra proporciona los siguientes resultados: 66.2, 65.8, 66.3, 65.6 y 66.5 mg/100ml. Utilizando la prueba de suma de rangos de Wilcoxon con muestras apareadas, pruebe si los métodos difieren significativamente. (Determine también el estadístico de prueba con a = 10%) (Universidad de Granada, 2012)**
   a. No se rechaza Ho, x = 5, no existe no diferencia significativa.
   b. No se rechaza Ho, T = 5, los métodos no difieren significativamente.
   c. No se rechazar Ho, T = 1, los métodos difieren significativamente.
   d. No se rechaza Ho, T = 10, los métodos no difieren significativamente.
   e. No se rechaza Ho, T = 1, los métodos son iguales significativamente.

2. **Una prueba de signo se dice no paramétrica porque se quiere probar una aseveración que**
   a. No usa estadísticos o estadígrafos.
   b. Solo usa datos muestrales.
   c. No usa parámetros.
   d. Solo emplea mediana en algunos casos.
   e. No emplea la media, proporción o varianza poblacionales.

3. **En una prueba de suma de rangos con signo de Wilcoxon es cierto:**
   a. Se cuentan el número de signos negativos y se igualan a R.
   b. Si existen empates, se asigna el promedio de los rangos a los empates.
   c. Se suman los signos negativos y positivos, y la menor suma es T.
   d. Se cuentan el número de rangos positivos y negativos, el menor es R.
   e. No usa parámetros. Se suman los signos y rangos para probar medianas iguales.

4. **"Si rechazamos un valor hipotético porque difiere de un estadístico de la muestra en más de 1.75 errores estándar, ¿cuál es la probabilidad de que hayamos rechazado una hipótesis que de hecho es cierta?" (Levin & Rubin, 2004, p. 323)**
   a. 0,05
   b. 0,005
   c. 0,06

d. 0,080

e. 0.10

5. Una distribución Chi cuadrado se caracteriza por

I. El nivel de significancia es diferente para cada grado de libertad.

II. Sus valores pueden ser cero o positivos.

III. Es diferente para cada número de grados de libertad.

IV. La distribución chi cuadrada no es simétrica.

V. A mayor grado de libertad su forma se hace más normal.

a. II y IV

b, I, II y III

c. II, III y IV

d. III y V

e. No es I

6. En una prueba de bondad de ajuste, el tamaño de la muestra n = 60, y las frecuencias están repartidas en una tabla de contingencia de cuatro filas y cinco columnas. El valor crítico al 0.01 de significancia es

a. 1.6706

b. 26.217

c. 13.277

d. 30.578

e. 27.688

7. Un sondeo de empleados de una compañía aseguradora se ocupaba de las relaciones entre trabajadores y supervisores. Una frase de evaluación era la siguiente: "No estoy seguro de lo que mi supervisor espera". Los resultados del sondeo se presentan en la siguiente tabla de contingencia. ¿Podemos rechazar la hipótesis de que "las respuestas a la frase y los años de empleo son independientes" al nivel de significación de 0,10? (Elemental, 2012).

Años de empleo	No estoy seguro de lo que mi supervisor espera	
	Verdadero	Falso
Menos de 1 año	18	13
De 1 a 3 años	20	8
De 3 a 10 años	28	9

a. x^2 = 4.605, se rechaza Ho. Los años de empleo y las respuestas son independientes.

b. El valor crítico es 2.567. Se rechaza Ho y las variables son dependientes.

c. x^2 = 96, se rechaza Ho. Los años de empleo y las respuestas no son independientes.

d. x^2 = 2.567, no se rechaza Ho. Los años de empleo y las respuestas son independientes.

e. El valor crítico es 4.605 y los grados de libertad 2. Se rechaza Ho.

8. El director de seguridad de la empresa Honda, de Estados Unidos, tomó muestras al azar del archivo de accidentes menores y los clasificó de acuerdo al tiempo en que tuvo lugar cada uno (Allen, 2000):

Hora	8 a 9 am	9 a 10 am	10a11 am	11a12 pm	1 a 2 pm	2 a 3 pm	3 a 4 pm	4 a 5 pm
Número de accidentes	6	6	20	8	7	8	13	6

Al nivel de significancia de 1%, ¿se puede afirmar que el número de accidentes no depende del horario de trabajo?

a. Se rechaza Ho, x^2 = 18.475 (prueba), el número de accidentes depende de la hora de trabajo.

b. No se rechaza Ho, x^2 = 18.324 (crítico), el número de accidentes es independiente de la hora de trabajo.

c. No se rechaza Ho, x^2 = 18.324 (prueba), el número de accidentes y la hora de trabajo son independientes.

d. Se rechaza Ho, x^2 = 18.324 (crítico), el número de accidentes y la hora de trabajo no son independientes.

e. Se rechaza Ho, x^2 = 18.345 (prueba), el número de accidentes no depende de la hora de trabajo.

9. En una prueba CHi cuadrado se plantean las hipótesis, de las cuales serían las formas correctas de Ho:

I. Las frecuencias observadas son iguales a las esperadas.

II. Las frecuencias observadas son semejantes a las esperadas.

III. Las variables son independientes.

IV. Las variables son dependientes.

V. p1=p2=p3= . . .

a. I y II

b. I, III, V

c. II y IV

d. III, IV, Y

e. IV y V

10. La información que sigue resultó de un experimento para comparar los efectos de la vitamina C de jugo de naranja y de ácido ascórbico sintético, sobre la duración de odontoblastos en conejillos de Indias en un periodo de seis semanas

Jugo de naranja	8.2	9.4	9.6	9.7	10.0	14.5
	15.2	16.1	17.6	21.5	13.2	
Ácido ascórbico	4.2	5.2	5.8	6.4	7.0	7.3
	10.1	11.2	11.3	11.5	8.0	

(Nutrition, J. (1947). *The Growth of the Odontoblasts of the Incisor Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig*, 491-504). Utilice la prueba Wilcoxon de prueba de rangos al nivel 0.01 para determinar si la mediana verdadera de duración difiere para los dos tipos de ingesta de vitamina C. Calcule también un valor p apropiado (Jay, 2008).

a. H1: Med1 ≠ Med2, z = 2.575, el estadístico de prueba es 2.659. Las medianas son diferentes.

b. Ho: Med1 = Med2, valor crítico 5, T = 11, n = 11. Existe diferencia entre las medianas.

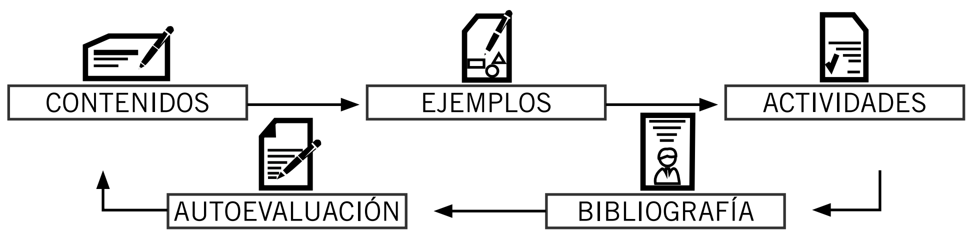
c. H1: Med1 ≠ Med2, z = 3.537, el estadístico de prueba es 2.575. Las medianas son diferentes.

d. H1: Mediana de las diferencias es cero, z = 2.575, T = 11. La mediana es diferente.

e. H1: μ 1 ≠ μ2, T = 5, el estadístico de prueba es 3.537. Las medias son diferentes.

UNIDAD IV
CORRELACION, REGRESIÓN Y CONTROL ESTADÍSTICO DE PROCESOS

DIAGRAMA DE ORGANIZACIÓN DE LA UNIDAD IV



ORGANIZACIÓN DE LOS APRENDIZAJES

Resultados del aprendizaje de la Unidad IV: Al finalizar la unidad el estudiante estará en la capacidad de realizar pronósticos utilizando el análisis de correlación y regresión y modelos de series de tiempo.

CONOCIMIENTOS	HABILIDADES	ACTITUDES
Tema n.° 1: Correlación y regresión 1. Prueba de hipótesis de correlación. 2. Prueba de hipótesis para coeficientes. 3. Construcción del modelo lineal de regresión. 4. Intervalos de confianza y predicción. 5. Regresión múltiple. Análisis de multicolinealidad. 6. Validación de modelos . **Tema n.° 2: Series temporales** 1. Modelos de series de tiempo. 2. Promedios móviles y suavizamiento exponencial. 3. Análisis de tendencia.	1. Analiza y valida la correlación entre variables. 2. Propone y formula modelos lineales. 3. Calcula el intervalo de predicción para la estimación de valores pronosticados. 4. Identifica modelos de regresión múltiple y los interpreta. 5. Realiza la suavización exponencial. 6. Construye modelos de series de tiempo y analiza la tendencia y estacionalidad. 7. Interpreta los modelos de series de tiempo. **Actividad n.° 1** Participa del foro de discusión sobre criterios de muestreo. **Actividad n.° 2** Evaluación del tema n° 1 y el tema n.° 2.	Valora reflexivamente la importancia de la interpretación de los modelos de predicción y de series de tiempo en la toma de decisiones.

Correlación y regresión

Tema n.º 1

1. Correlación y regresión lineal simple

1.1. Análisis de correlación lineal simple

Correlación

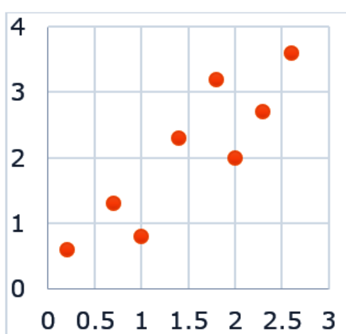
En la unidad anterior se adelantó el concepto de correlación, que se trata de la existencia de algún tipo de relación entre los datos de dos o más variables.

En esta unidad trataremos el tema desde el caso de tener solo dos variables involucradas, por ello el nombre de lineal simple.

Diagrama de dispersión

Es un diagrama en el que se visualizan los datos pareados simbolizados con puntos, los que permiten tener una idea de la existencia de correlación lineal entre las variables.

X	*Y*
2.3	2.7
1.0	0.8
0.2	0.6
2.0	2.0
1.4	2.3
2.6	3.6
1.8	3.2
0.7	1.3



X	*Y*
0.7	2.7
1.8	3.2
2.6	0.8
1.4	2
2	2.3
0.2	3.8
1	3.2
2.3	1.3

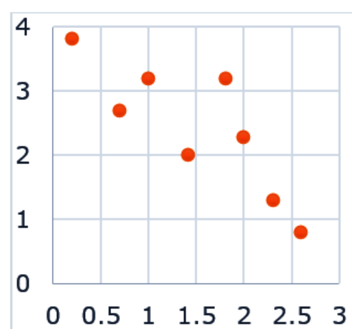


Figura 18. Diagrama de dispersión.

Correlación lineal directa: Los valores de una de las variables aumentan, y los valores de la otra también.

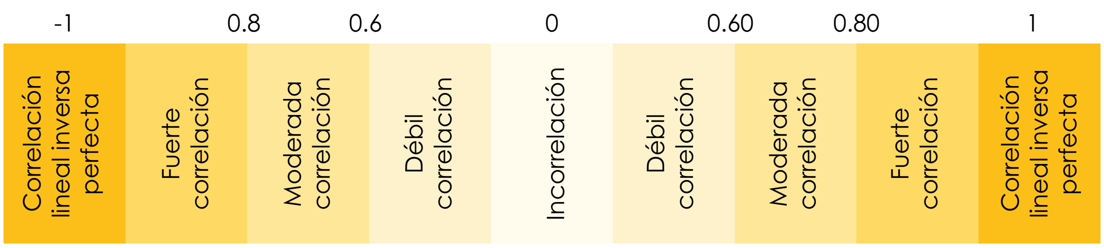
En los dos diagramas se puede apreciar que los puntos tienden a acomodarse en una forma que se acercan bastante a una línea recta, por ello se puede afirmar que existe correlación lineal entre los datos de las variables en la muestra.

Medida de la correlación

La correlación entre dos variables se mide con el coeficiente de correlación lineal o de Pearson:

> ***ρ** es el coeficiente de correlación lineal en la población.*
>
> ***r** es el coeficiente de correlación lineal en la muestra.*

Los valores del coeficiente varían dentro del intervalo [-1; 1] y pueden interpretarse de la siguiente manera:



1.2. Prueba de hipótesis

La prueba es a dos colas. Se prueba si existe correlación lineal entre las variables mediante el coeficiente de correlación lineal.

Hipótesis:

H0: ρ = 0 No existe correlación lineal entre las variables X e Y.

H1: ρ ≠ 0 Existe correlación lineal entre las variables X e Y.

Estadístico de prueba:

Prueba simplificada (Triola, 2013)

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Prueba T

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Valor crítico:

Prueba simplificada (Triola, 2013) — Con n, número de pares de datos en la tabla A-6, si |r| ≥ valor crítico, rechace H0 como verdadera.

Prueba T — Con gl = n - 2 y α en la tabla A-3, si |t| ≥ valor crítico, rechace H0 como verdadera.

Tabla 65. Resistencia de suelos con contenido de grava

% Grav.	Resist kg/cm²
8	10
22	25
10	15
3	7
12	16
15	18
30	42
18	21
25	37

Ejemplo 40:

En el cálculo de la resistencia de suelos tienen mucha importancia los componentes del suelo estudiado. La tabla 65 resume las resistencias encontradas en 9 terrenos con diferentes contenidos de arena o grava. ¿Al nivel de 5% de significancia se puede decir que el contenido de arena en el suelo influye en su resistencia?

Solución:

H0: $\rho = 0$. No existe correlación lineal entre las variables resistencia y cantidad de arena del suelo.

H1: $\rho \neq 0$. Existe correlación lineal entre las variables resistencia y cantidad de arena del suelo.

1.° Realice un gráfico de dispersión que le permita definir si existe la posibilidad de la correlación lineal entre las variables:

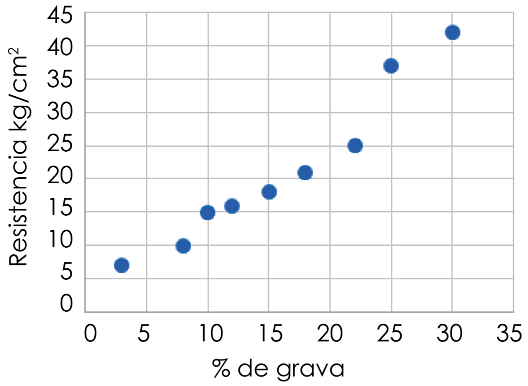


Figura 19. Gráfico de dispersión de resistencia de suelos

El gráfico nos dice que existe una marcada tendencia lineal en los datos. Podemos asegurar que existe correlación lineal entre los datos de las variables en la muestra.

2.° Calculamos el estadístico de prueba:

Tabla 66. Estadístico de prueba de resistencia del suelo.

%Grav. X	Resist kg/cm² Y	X^2	Y^2	XY
8	10	64	100	80
22	25	484	625	550
10	15	100	225	150
3	7	9	49	21
12	16	144	256	192
15	18	225	324	270
30	42	900	1764	1260
18	21	324	441	378
25	37	625	1369	925
143	191	2875	5153	3826

Calculamos las sumatorias:

$\Sigma X = 143$

$\Sigma Y = 191$

$\Sigma X^2 = 2875$

$\Sigma Y^2 = 5153$

$\Sigma XY = 3826$

Reemplazando en la fórmula:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{9(3826) - (143)(191)}{\sqrt{9(2875) - (143)^2}\sqrt{9(5153) - (191)^2}} = 0.9718$$

3.° El valor crítico de la tabla A-6, si n = 6 y $\alpha = 0.05$:

Valor crítico = 0.666

Conclusión y respuesta:

Como |r| > valor crítico rechazamos H0 como verdadera.

Existe evidencia muestral suficiente para confirmar que existe correlación lineal entre las variables resistencia y contenido de arena del suelo.

1.3. Análisis de regresión lineal simple

Cuando se ha descubierto que existe correlación lineal entre las variables, entonces se puede averiguar por un modelo matemático que se ajuste mejor a los datos. Este modelo debe dibujar una recta, la cual será aquella que se acerca más a todos los puntos en el gráfico de dispersión.

Encontrar este modelo matemático es realizar una regresión.

Modelo lineal simple

El modelo lineal es $\hat{y} = b_0 + b_1 x$

- Pendiente:

$$b_1 = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}}$$

- Intercepto:

$$b_0 = \bar{y} - b_1 \bar{x}$$

Si la pendiente es positiva, la relación es directa, y si la pendiente es negativa, la relación es inversa.

Utilidad del modelo

Cuando conseguimos el modelo lineal de regresión, de inmediato nos preguntamos si este es aplicable a los datos de la población.

Entonces podemos poner a prueba el valor de la pendiente, ya que sabemos que existe correlación lineal. Cabe la posibilidad de que la pendiente sea cero, entonces la ecuación de regresión se convertiría en una relación horizontal igual a una constante.

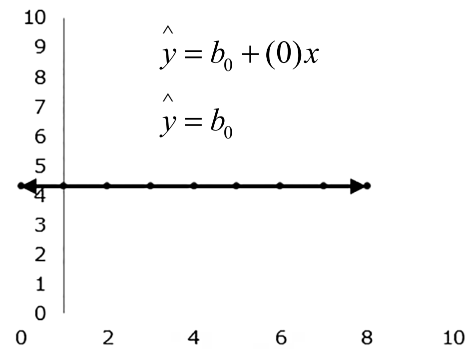


Figura 20. Relación horizontal de una ecuación de regresión.

Prueba para la pendiente del modelo de regresión lineal

H0: $\beta_1 = 0$ La pendiente es cero. La ecuación es constante.

H1: $\beta_1 \neq 0$ La pendiente no es cero. La ecuación no es constante.

Estadístico de prueba

$$t = \frac{b_1 + \beta_1}{\frac{S_e}{\sqrt{s_x * (n-1)}}}$$

Donde $S_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$ y $S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ Desv. Est. de los x

Valor crítico

Tabla A-3 con gl = n-2 y α

Ejemplo 41: Con los datos del ejemplo 35, buscaremos el modelo de regresión y probaremos que la ecuación es válida.

Datos:
n = 9
gl = 7
$\Sigma X = 143$
$\Sigma Y = 191$
$\Sigma X^2 = 2875$
$\Sigma Y^2 = 5153$
$\Sigma XY = 3826$
$\bar{y} = 21.222$
$\bar{x} = 15.889$

Pendiente:

$b_1 = \frac{n\sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$

$b_1 = \frac{9(3826) - (143)(191)}{9(2875) - (143)^2}$

$b_1 = 1.3124$

Pendiente:

$b_0 = \bar{y} - b_1\bar{x}$

$b_0 = 21.2222 - 1.3124(15.8889)$

$b_0 = 0.3699$

El mejor modelo de regresión es $\hat{y} = 0.3699 + 1.3124X$.

Prueba para la pendiente:

H0: $\beta_1 = 0$ La pendiente es cero. La ecuación es constante.

H1: $\beta_1 \neq 0$ La pendiente no es cero. La ecuación no es constante.

Estadístico de prueba

$$t = \frac{b_1 - \beta_1}{\frac{S_e}{\sqrt{s_x * (n-1)}}}$$

Donde:

$$S_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 8.6811$$

n = 9

$$S_e = \sqrt{\frac{5153 - (0.3699)(191) - (1.3124)(3826)}{n-2}}$$

$S_e = 8.7036$

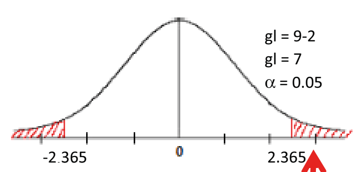
Reemplazando:

$$t = \frac{b_1 - \beta_1}{\frac{S_e}{\sqrt{s_x * (n-1)}}}$$

$$t = \frac{1.3124 - 0}{\frac{8.7036}{\sqrt{8.6811 * 8}}}$$

$t = 3.6999$

2 colas	0.20	0.10	0.05	0.02
gl	Valores t			
5	1.476	2.015	2.571	3.36
6	1.440	1.943	2.447	3.14
7	1.415	1.895	2.365	2.99
8	1.397	1.860	2.306	2.8



Conclusión y respuesta:

Como |t| > Valor crítico rechazamos H0 como verdadera.

Existe evidencia muestral para confirmar que la ecuación no es constante, es válida.

1.4. Utilizar el modelo para realizar pronósticos

Si queremos usar el modelo hallado para hacer pronósticos, es necesario tener en cuenta:

- Los valores deben tomarse del intervalo de datos de la muestra o no muy alejados de este intervalo.
- El resultado se debe tomar como un promedio de todos los posibles resultados cuando x toma un valor particular.
- Si no existiese correlación, la mejor proyección es la media de Y.

Ejemplo 42. Continuamos con el ejercicio 40: ¿Cuál sería el valor de la resistencia de un suelo con un contenido de 28 % de grava?

Si el mejor modelo de regresión es $\hat{y}$ = 0.3699 + 1.3124

X = 28

$\hat{y}$ = 0.3699 + 1.3124

$\hat{y}$ = 0.3699 + 1.3124(28)

$\hat{y}$ = 37.1171 kg/cm^2

1.5. Bondad de ajuste

Se denomina así a la capacidad que tiene un modelo para explicar las variaciones en la variable dependiente (Y). Su valor se calcula con coeficiente de determinación.

Coeficiente de determinación:

$$R^2 = r^2$$

R^2 se expresa en porcentaje generalmente.

Ejemplo 43: Con los datos del ejemplo 35, encontramos el valor de r = 0.9718 y el modelo hallado es $\hat{y}$ = 0.3699 + 1.3124

¿Cuál es la bondad de ajuste del modelo? ¿Qué proporción de la variación de la resistencia del suelo puede explicarse por el modelo?

$R^2 = 0.9718^2$

R^2 = 0.94 44

Conclusión y respuesta:

R^2 = 94.44% es la bondad de ajuste del modelo, es la proporción de la variación de la resistencia del suelo (Y) que puede explicar la relación contenido de arena-resistencia del suelo, es decir, el modelo y $\hat{y}$ = 0.3699 + 1.3124X.

1.6. Intervalos de confianza y predicción

Cuando realizamos un cálculo de predicción de Y usando la ecuación de regresión, obtenemos un valor que es uno de muchos otros resultados probables. Recuerde que el modelo de correlación que se tiene es un modelo probabilístico; es decir que, por ejemplo, si habláramos del peso y de la estatura de las personas, y tenemos una ecuación que predice el peso (y) en función de la estatura (x), no podríamos asegurar de que para cualquier persona que mide 1.67 m el peso sea siempre 69 kg como lo calcularía la ecuación de regresión.

Por ello, el resultado que se obtiene al usar la ecuación de regresión se puede entender como un promedio, una estimación puntual del verdadero valor de "y" cuando "x" toma un valor determinado.

Si deseamos más precisión, deberemos emplear un intervalo:

$$\hat{y} - E < y < \hat{y} + E$$

El margen de error E:

$$E = s_e * t_{\alpha/2} * \sqrt{1 + \frac{1}{n} + \frac{n(x-\bar{x})^2}{n\sum x^2 - (\sum x)^2}}$$

Se: Error estándar de la estimación.

$t_{\alpha/2}$: Valor crítico a dos colas en la tabla A-3 con gl = n-2

Ejemplo 44. Con los datos del ejercicio de la unidad (ejemplo 40), determine un intervalo de confianza al 95% de confianza para la resistencia del terreno cuando la proporción de arena es de 28%.

Datos:
$r = 0.9718$
$x = 28$
$\bar{x} = 15.8889$
$\sum x = 143$
$\sum x^2 = 2875$

$\hat{y} = 0.3699 + 1.3124X$
$\hat{y} = 0.3699 + 1.3124\,(28)$
$\hat{y} = 37.1171$ kg/cm²

$S_e = 8.7036$
$t_{\alpha/2} = 2.365$
gl = 7
$\alpha = 0.05$

$$E = s_e * t_{\alpha/2} * \sqrt{1 + \frac{1}{n} + \frac{n(x-\bar{x})^2}{n\sum x^2 - (\sum x)^2}}$$

$$E = 8.7036\,(2.365) * \sqrt{1 + \frac{1}{9} + \frac{9(28-15.8889)^2}{9(2875) - 143^2}}$$

E = 23.9554

Intervalo de confianza:

$$\hat{y} - E < y < \hat{y} + E$$

$$37.1171 - 23.9554 < y < 37.1171 + 23.9554$$

$$13.162 < y < 61.073 \text{ kg/cm}^2$$

Conclusión y respuesta:

Tenemos la confianza del 95 % de que el verdadero valor de la resistencia de suelo a la compresión esta entre 13.162 y 61.073 kg/cm2 cuando la proporción de arena es de 28 %.

2. Correlación y regresión lineal múltiple

En este caso se trata de averiguar si un conjunto de variables independientes está relacionado con otro de variables dependientes.

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \ldots$$

2.1. Análisis de correlación múltiple

Se trata de averiguar en primer lugar si existe correlación lineal múltiple entre las variables independientes (x_1, x_2, x_3, …) y la variable dependiente (y).

Grado de correlación múltiple

El coeficiente de correlación múltiple es el coeficiente de determinación lineal modificado, debido a que cuanto más variables independientes se agreguen al modelo este coeficiente se incrementa sin reflejar el verdadero valor del grado de correlación existente entre las variables. Por ello se emplea:

Coeficiente de ajustado de determinación:

$$R^2 = 1 - \frac{n-1}{n-(k-1)}(1-R^2)$$

Si n = tamaño de muestra, k = número de variables independientes

Análisis de correlación

H0: ρ = 0. No existe correlación lineal múltiple entre las variables.

H1: ρ ≠ 0. Existe correlación lineal múltiple entre las variables.

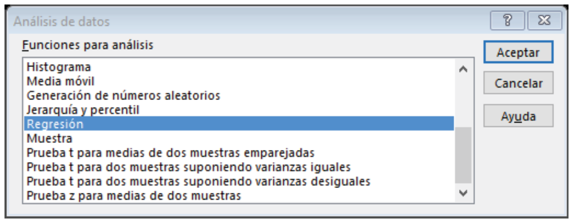
El proceso se lleva a cabo mediante ANOVA y se obtiene fácilmente usando un *software* como Excel, SPSS o Minitab.

Ejemplo 45. La siguiente tabla incluye las longitudes del muslo, las circunferencias del brazo y las estaturas de hombres elegidos al azar. Todas las medidas están en centímetros. Utilice esos datos para resolver: la estatura promedio a la alcanzada por una persona con longitud de muslo de 42 cm, circunferencia de brazo de 32 cm.

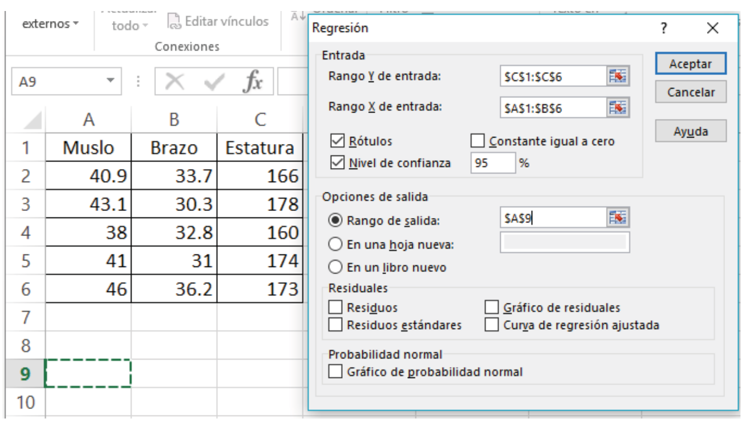
Tabla 68. Estatura promedio de una persona en relación a la longitud de su muslo y la circunferencia de su brazo.

Muslo	40.9	43.1	38.0	41.0	46.0
Brazo	33.7	30.3	32.8	31.0	36.2
Estatura	166	178	160	174	173

En Excel >> Datos >> Análisis de datos



Elegimos la opción Regresión y luego clic en el botón Aceptar.

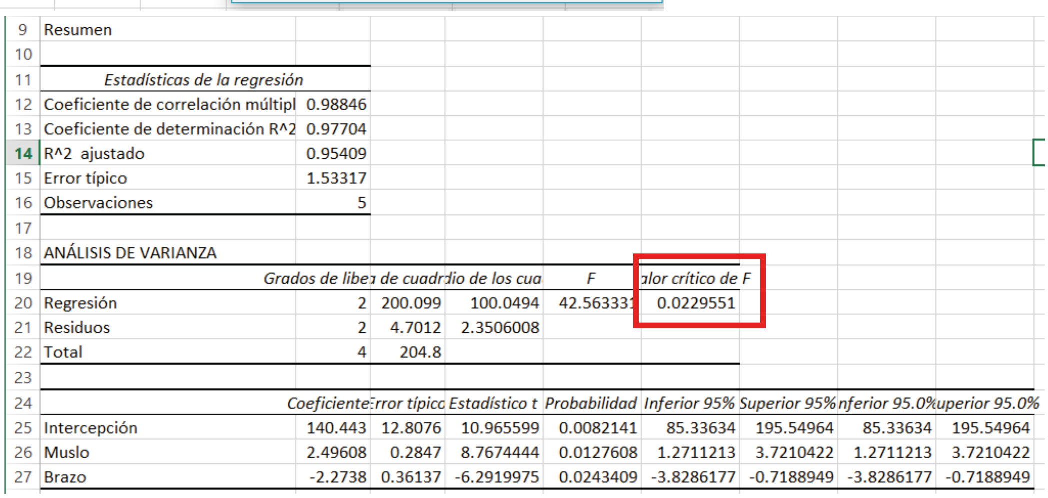


Elija los datos y columna estatura.

Seleccione los datos x (dos columnas primeras).

Active rótulos y nivel de confianza.

Elija la celda de inicio de salida de resultados y clique en Aceptar:



Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltipl	0.98846
Coeficiente de determinación R^2	0.97704
R^2 ajustado	0.95409
Error típico	1.53317
Observaciones	5

ANÁLISIS DE VARIANZA

	Grados de libe	de cuadr	dio de los cua	F	alor crítico de F
Regresión	2	200.099	100.0494	42.563331	0.0229551
Residuos	2	4.7012	2.3506008		
Total	4	204.8			

	Coeficiente	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	nferior 95.0%	uperior 95.0%
Intercepción	140.443	12.8076	10.965599	0.0082141	85.33634	195.54964	85.33634	195.54964
Muslo	2.49608	0.2847	8.7674444	0.0127608	1.2711213	3.7210422	1.2711213	3.7210422
Brazo	-2.2738	0.36137	-6.2919975	0.0243409	-3.8286177	-0.7188949	-3.8286177	-0.7188949

De la primera tabla: R2 ajustado = 0.95409.

Con la tabla ANOVA respondemos a las hipótesis:

H0: ρ = 0. No existe correlación lineal múltiple entre las variables.

H1: ρ ≠ 0. Existe correlación lineal múltiple entre las variables.

Tabla 69. ANOVA de la relación entre estatura y longitud de muslo y contorno del brazo.

ANÁLISIS DE VARIANZA					
	Grados de libertad	**Suma de cuadrados**	**Promedio de los cuadrados**	**F**	**Valor crítico de F**
Regresión	2	200.0988	100.0493	42.5633	**0.02295509**
Residuos	2	4.701202	2.3506		
Total	4	204.8			

El valor P = 0.02295 < α = 0.05. Rechazamos H0 como verdadera. Existe correlación lineal múltiple entre estatura y longitud de muslo y contorno de brazo.

La tercera tabla brinda la oportunidad de determinar si todas las variables consideradas para este análisis debieron ser consideradas o no:

Tabla 70. Variables que debieron ser consideradas o no.

	Coeficientes	**Error típico**	**Estadístico t**	**Probabilidad**	**Inferior 95 %**	**Superior 95 %**
Intercepción	140.443	12.8076	10.9656	0.0082	85.3363	195.5496
Muslo	2.4960	0.2846	8.7674	**0.0128**	1.2711	3.7210
Brazo	-2.2737	0.3613	-6.2920	**0.0243**	-3.8286	-0.7189

Si la ecuación de regresión en la población es

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

H0: $\beta_1 = 0$	H0: $\beta_2 = 0$
H1: $\beta_1 \neq 0$	H1: $\beta_2 \neq 0$
Valor P = 0.0128 < α = 0.05	Valor P = 0.0243 < α = 0.05
Rechazamos H0, por tanto, x_1 sí contribuye en la ecuación y debe mantenerse.	Rechazamos H0, por tanto, x_2 sí contribuye en la ecuación y debe mantenerse.

Como las dos pruebas ratifican la necesidad de considerar las dos variables "x", entonces la ecuación de regresión es

$$\hat{y} = 140.443 + 2.4960x_1 - 2.2737x_2$$

Para los datos del problema $x_1 = 42$ y $x_2 = 32$

$$\hat{y} = 140.443 + 2.4960(42) - 2.2737(32)$$

$$\hat{y} = 172.5166 \text{ cm}$$

3. Correlación no lineal: construcción de modelos

No todas las relaciones entre las variables pueden acomodarse a una forma lineal. Algunas correlaciones se ajustan mejor a una forma curvilínea, como lo muestra el gráfico de dispersión:

En estos casos se tiene varios modelos que podrían muy bien acomodarse a los datos.

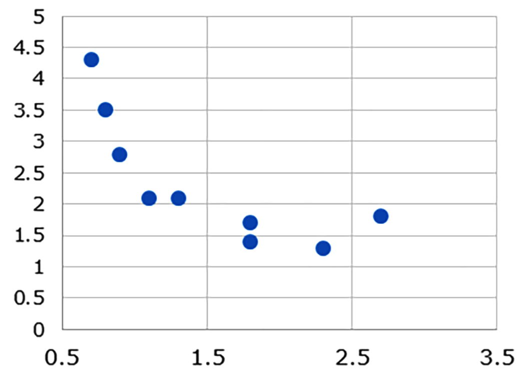


Figura 21. Forma curvilínea de algunas correlaciones.

3.1. Modelos de regresión no lineal

Denominación	**Ecuación**
Logarítmico	$\hat{y} = b_0 + b_1 Ln(x)$
Exponencial	$\hat{y} = b_0 * e^{b_1 x}$
Potencial	$\hat{y} = b_0 x^{b_1}$
Inverso	$\hat{y} = b_0 + \frac{b_1}{x}$
Cuadrático	$\hat{y} = b_0 + b_1 x + b_2 x^2$

Estos modelos son apropiados para la regresión por cuanto pueden linealizarse mediante métodos matemáticos.

3.2. Coeficiente de determinación R^2

Este coeficiente mide el nivel de bondad de ajuste de un modelo; por tanto, podemos usarlo como medida de comparación de la eficiencia del modelo.

Para calcularlo se pueden usar *software* como Excel, SPSS, Minitab o calculadoras como la casio fx82s.

3.3. Buscar el mejor modelo

- Busque un patrón en la gráfica. Utilice los datos muestrales para construir una gráfica (por ejemplo, un diagrama de dispersión). Luego, compare el patrón básico con las gráficas genéricas conocidas de las funciones lineales, cuadráticas, logarítmicas, exponenciales y potencia.
- Calcule y compare valores de R^2 para cada modelo que considere. Debe utilizar programas de cómputo o una calculadora para hallar el valor del coeficiente de determinación R^2. Los valores de R^2 se pueden interpretar aquí de la misma forma que se interpretaron en la sección de regresión lineal: seleccione funciones que den como resultado valores

más grandes de R^2, ya que corresponden a funciones que se ajustan mejor a los puntos observados.

- Reflexione. Aplique el sentido común. No utilice un modelo que conduzca a valores predichos que son poco realistas. Utilice el modelo para calcular valores futuros, valores pasados y valores faltantes; luego, determine si los resultados son realistas y lógicos.

Ejemplo 46: Un experimento para una clase de física implica dejar caer una pelota de golf y registrar la distancia (en metros) que cae en diferentes tiempos (en segundos) después de ser soltada. Los datos se incluyen en la siguiente tabla. Proyecte la distancia para un tiempo de 12 segundos, considerando que la pelota de golf se dejó caer de un edificio con una altura de 50 m.

Tabla 71. Distancia que cae una pelota de golf en diferentes tiempos

Tiempo	0.5	1	1.5	2	2.5	3	4	5
Distancia	2	5.4	7	18	30.5	44	100	260

Solución:

1.° Realice el gráfico de dispersión:

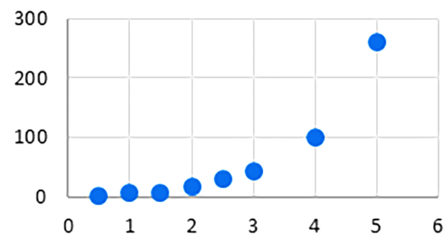


Figura 22. Forma curvilínea de correlación no lineal entre el tiempo y la distancia a la que cae una pelota de golf.

Se puede ver que los puntos, tienen una forma curvilínea. Por tanto, la correlación es no lineal.

Tabla 72. Datos de la correlación entre el tiempo y la distancia a la que cae una pelota de golf

Modelo	**R^2**
Log.	0.7135
Exp.	0.9792
Pot.	0.9480
Inv.	0.2572

2.° Ingresamos los datos en la calculadora Casio fx82s y obtenemos los coeficientes de determinación:

Como se puede apreciar el modelo que tiene un coeficiente de determinación más cercano a 1 es el modelo exponencial.

3.º Obtenga la ecuación de regresión. La ecuación general es:

$\hat{y} = b_0 * e^{b_1x}$, en su calculadora A = b_0 = 1.6962 y B = b_0 = 1.0439, entonces la ecuación de regresión será:

$$\hat{y} = 1.6962e^{1.0439x}$$

4.º Pronóstico cuando el tiempo es x = 12 s.

$\hat{y} = 1.6962e^{1.0439x} = 1.6962e^{(1.0439(12))}$

$\hat{y} = 467516.7495$ m.

Series temporales

Tema n.º 2

Una serie temporal se produce cuando se tiene una variable que varía conforme pasa el tiempo. Es probable que las observaciones cambien en función del desarrollo de sucesos, acontecimientos o fenómenos que se comportan de manera dialéctica.

Las ventas, los gastos, el PBI, la producción de una fábrica, los clientes atendidos y el desarrollo de factores económicos son ejemplos de aplicación del análisis de series temporales. En todos ellos los datos suceden en el tiempo y varían aleatoriamente.

Por esto, una serie temporal se conoce como un suceso estocástico de variable discreta-discreta y discreta-continua. Se compone de los datos aleatorios tomados sucesivamente lo que configura una gráfica como la siguiente

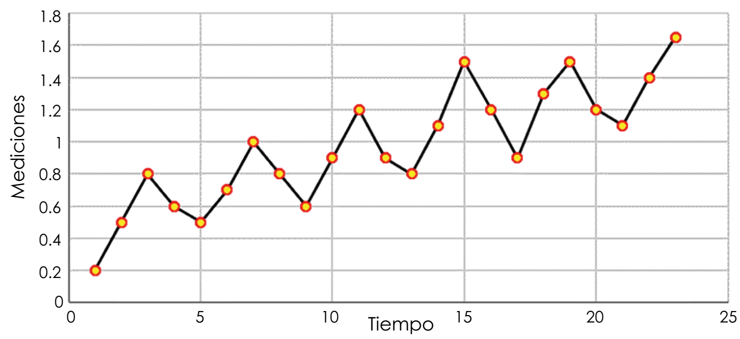


Figura 23. Datos aleatorios en el tiempo.

Una serie temporal es, en realidad, una unión de componentes que son los que configuran el comportamiento de la serie. Estos componentes son la tendencia, la estacionalidad, la variación cíclica, y la aleatoria.

1. Componentes de una serie temporal

1.1. Tendencia

La tendencia en una serie temporal causa que la serie se mueva de manera ascendente, descendente o de forma constante. Su cálculo se realiza empleando los métodos de la regresión vistos en el tema 1.

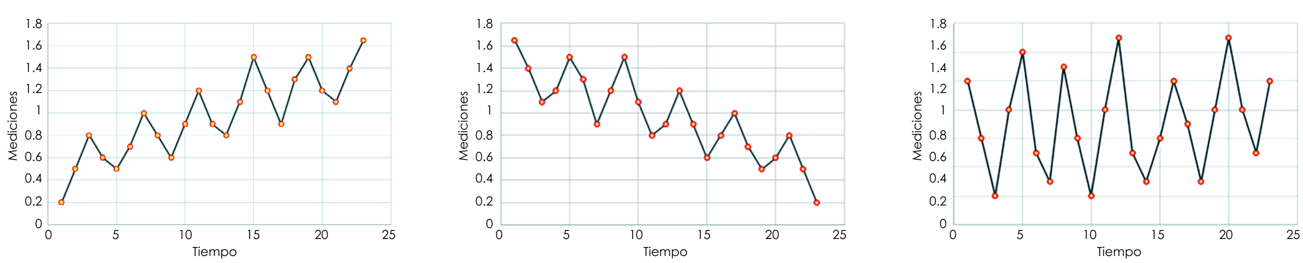


Figura 24. Componente tendencia de una serie temporal.

1.2. Estacional

La estacionalidad en una serie temporal ocasiona que una serie se mueva de arriba hacia debajo de manera periódica en periodos cortos, generalmente en periodos de un año. Es decir, las variaciones suceden año a año en las mismas fechas, como las estaciones. Por ello, este componente recibe el nombre de estacionalidad.

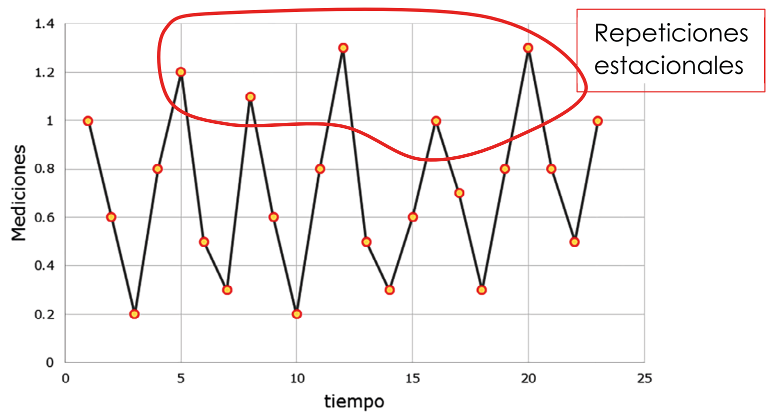


Figura 25. Componente estacional de una serie temporal.

Se puede analizar por medio de métodos de suavizamiento como promedios móviles.

1.2.1. Promedio móviles

Los promedios móviles pueden ser de dos tipos: pares e impares.

a. Promedios móviles impares

Los promedios se toman de 3, 5, 7 . . . datos sucesivos.

Ejemplo 47: Calcule los promedios móviles de amplitud 3.

Tabla 73. Promedios móviles PM3

t	*y*	*PM3*	
1	345		
2	342	341.667	= (345+342+338)/3
3	338	342.333	= (342+338+347)/3
4	347	343.000	
5	344	344.000	
6	341	345.000	
7	350		

Se le dice técnica de suavizamiento porque el efecto de graficar los promedios produce la línea naranja en el gráfico que se muestra a continuación.

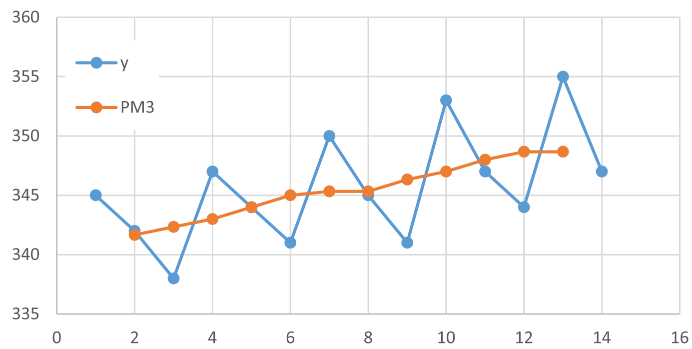


Figura 26. Técnica de suavizamiento

b. Promedios móviles pares

En este caso los promedios son tomados de 2, 4, 6, … datos sucesivos:

Ejemplo 48: Calcule los promedios móviles de amplitud 4:

Se realiza en dos etapas: ❶ se calcula el promedio móvil, para lo cual se toman 4 datos y luego se procede a calcular; ❷ se calculan los promedios móviles de 2 de los promedios de 4:

Tabla 74. Promedios móviles PM4

t	*y*	❶	❷ *PM4*
1	345		
2	342	343.00	
3	338	342.75	342.875
4	347	342.50	342.625
5	344	345.50	344
6	341	345.00	345.25
7	350	344.25	344.625
8	345	347.25	345.75
9	341	346.50	346.875
10	353	346.25	346.375
11	347	349.75	348
12	344	348.25	349
13	355		
14	347		

Los promedios móviles no son centrados; es decir, cuando se calculan no corresponden a ningún periodo particular de tiempo (t). Por ejemplo, el primer promedio móvil es 343.00 y se sitúa entre el segundo y tercer dato.

Si se vuelve a calcular el promedio móvil de 2 en 2, entonces este promedio sí corresponde con un periodo. Por ejemplo, 342.875 corresponde a t = 3.

La gráfica resultante es la siguiente:

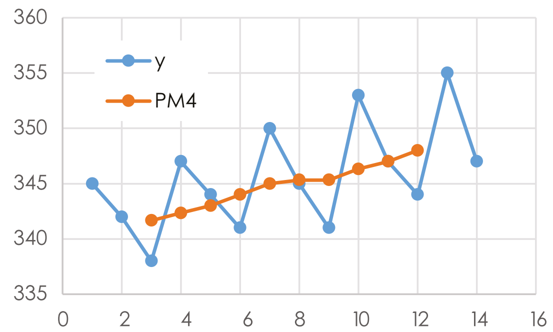


Figura 27. Promedios móviles.

Elegir un tipo de promedio móvil

El periodo o amplitud del promedio móvil se elige en función de la periodicidad que se pueda aprecia en el gráfico. Ejemplo:

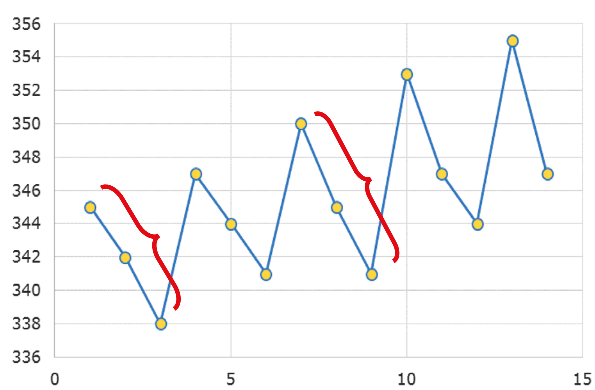


Figura 28: Estacionalidad. Fuente: propia.

Se puede apreciar en este gráfico una secuencia repetida de 3 puntos antes de un pico; por tanto, el promedio móvil que debe aplicarse debería ser de amplitud 3 (PM3).

Componente estacional:

$$E_i = \frac{\sum \frac{Y}{PM}}{k} * f$$

1.3. Cíclico

En las series temporales, este componente se refiere a movimientos de amplitud muy grande (por ejemplo, de 4 a más años) y que hacen que la serie cambie su tendencia a largo plazo.

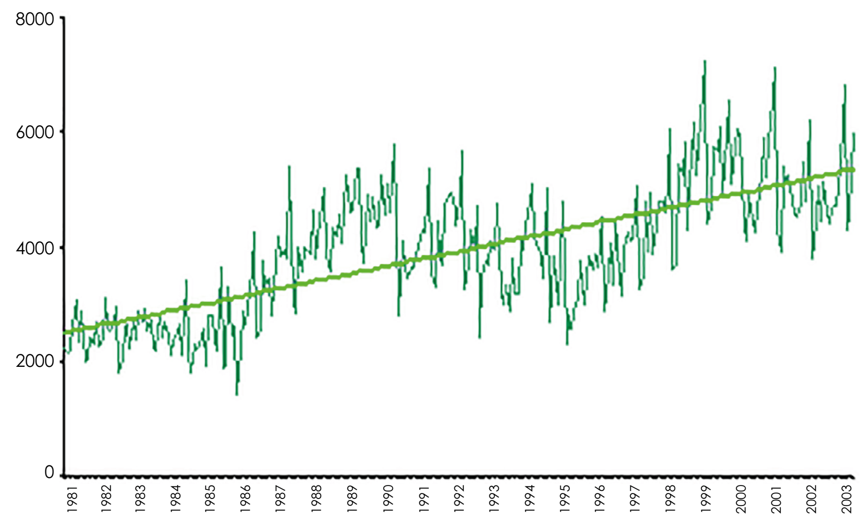


Figura 29. Componente cíclico. Tomada de Universidad de Valladolid, 2012.

Ejemplo de las causas de estas variaciones son los cambios de gobiernos democráticos, el fenómeno del niño, etcétera.

1.4. Irregular

Este componente es producto de sucesos que de manera aleatoria influyen en el desarrollo de la serie causando variaciones atípicas. No se puede saber cuándo ocurrirán, ni en qué magnitud afectarán un suceso, por ello es difícil de analizarlos y calcularlos.

Ejemplos de estos sucesos son las guerras, la muerte de líderes políticos, cambios climáticos producto del calentamiento global, crisis financieras, etcétera.

1.5. Modelos de series temporales

Las series temporales pueden componerse por medio de

Modelo aditivo

Figura 30. Modelo aditivo. Fuente: Elaboración propia.

Modelo multiplicativo

Figura 31. Modelo multiplicativo. Fuente: Elaboración propia.

En nuestro desarrollo, emplearemos el modelo multiplicativo.

2. Análisis de series temporales

Para explicar este proceso realizaremos la solución de un caso:

Ejemplo 49: Los siguientes datos corresponden a las ventas trimestrales (en miles de pesos) alcanzadas por una compañía en el transcurso de 10 años (Díaz, 2013).

Solución:

Tabla 75. Ventas trimestrales en 10 años

Año	T1	T2	T3	T4
1	747	927	783	1215
2	882	1089	909	1386
3	1089	1305	1143	1539
4	1233	1440	1278	1728
5	1566	1773	1620	2079
6	1638	1845	1674	2160
7	1800	1998	1845	2259
8	2007	2259	2061	2493
9	2223	2421	2259	2682
10	2322	2583	2340	2898

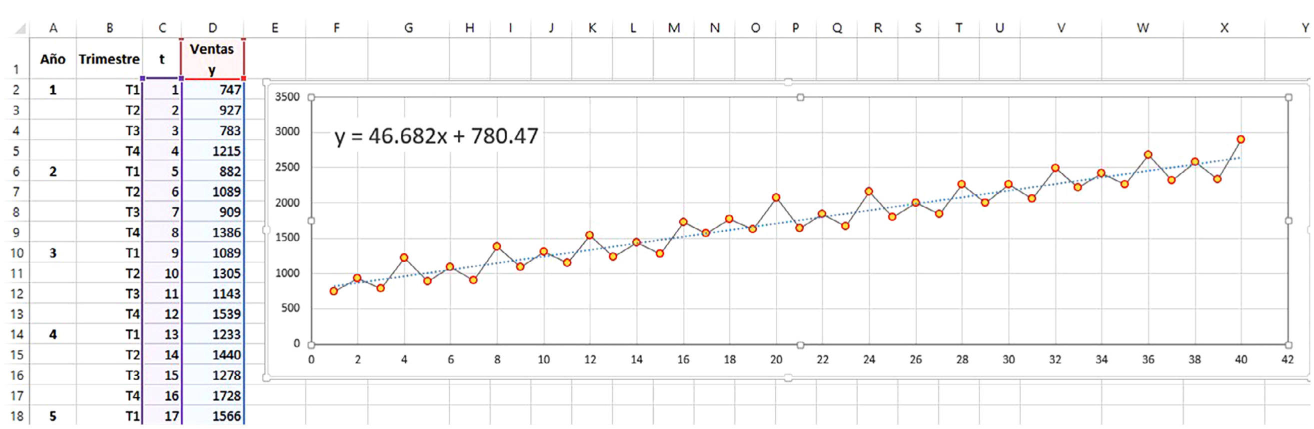
a. Calcule el componente estacional o ecuación de regresión.

b. Determine los índices estacionales.

c. Determine un pronóstico de las ventas trimestrales para el año 11.

a. Análisis de tendencia

Con los datos, haciendo uso de una calculadora o de Excel y ordenando los datos de manera vertical, obtenemos la ecuación de regresión.



Año	Trimestre	t	Ventas y
1	T1	1	747
	T2	2	927
	T3	3	783
	T4	4	1215
2	T1	5	882
	T2	6	1089
	T3	7	909
	T4	8	1386
3	T1	9	1089
	T2	10	1305
	T3	11	1143
	T4	12	1539
4	T1	13	1233
	T2	14	1440
	T3	15	1278
	T4	16	1728
5	T1	17	1566

Figura 32. Ecuación de regresión de las ventas trimestrales en el transcurso de 10 años.

Por tanto, la componente de tendencia se calculará con la ecuación de regresión lineal:

$$\hat{y} = 780.4731 + 46.6818x$$

b. Índices estacionales

Como se puede ver en el gráfico, los puntos se mueven arriba – abajo y el pico más alto se produce cada 4 puntos. Por tanto, el promedio móvil a calcular es de amplitud 4 (PM4).

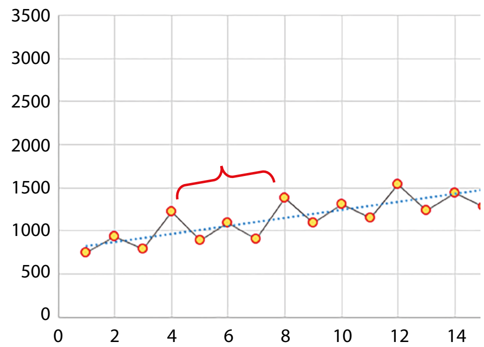


Figura 33. Promedio móvil de amplitud 4 .

Calculamos los promedios móviles PM4:

Tabla 76. Cálculo de los PM4.

Año	Trimestre	t	Ventas y	1	2 PM4
1	T1	1	747		
	T2	2	927	918.00	
	T3	3	783	951.75	934.875
	T4	4	1215	992.25	972.000
2	T1	5	882	1023.75	1008.000
	T2	6	1089	1066.50	1045.125
	T3	7	909	1118.25	1092.375
	T4	8	1386	1172.25	1145.250
3	T1	9	1089	1230.75	1201.500
	T2	10	1305	1269.00	1249.875
	T3	11	1143	1305.00	1287.000
	T4	12	1539	1338.75	1321.875
4	T1	13	1233	1372.50	1355.625
	T2	14	1440	1419.75	1396.125
	T3	15	1278	1503.00	1461.375
	T4	16	1728	1586.25	1544.625
5	T1	17	1566	1671.75	1629.000
	T2	18	1773	1759.50	1715.625
	T3	19	1620	1777.50	1768.500
	T4	20	2079	1795.50	1786.500
6	T1	21	1638	1809.00	1802.250
	T2	22	1845	1829.25	1819.125
	T3	23	1674	1869.75	1849.500
	T4	24	2160	1908.00	1888.875
7	T1	25	1800	1950.75	1929.375
	T2	26	1998	1975.50	1963.125
	T3	27	1845	2027.25	2001.375
	T4	28	2259	2092.50	2059.875
8	T1	29	2007	2146.50	2119.500
	T2	30	2259	2205.00	2175.750
	T3	31	2061	2259.00	2232.000
	T4	32	2493	2299.50	2279.250
9	T1	33	2223	2349.00	2324.250
	T2	34	2421	2396.25	2372.625
	T3	35	2259	2421.00	2408.625
	T4	36	2682	2461.50	2441.250
10	T1	37	2322	2481.75	2471.625
	T2	38	2583	2535.75	2508.750
	T3	39	2340		
	T4	40	2898		

De la ecuación general:

$\hat{y}$ = T * E * C * I despejando $E = \dfrac{Y}{T*C*I} = \dfrac{Y}{PM}$

Determinamos el valor de los coeficientes estacionales E_0 =Y/PM4

Tabla 77. Coeficientes estacionales

Año	Trimestre	t	Ventas y	❷ PM4	Y/PM4 E_O
1	T1	1	747		
	T2	2	927		
	T3	3	783	934.875	0.8375
	T4	4	1215	972.000	1.2500
2	T1	5	882	1008.000	0.8750
	T2	6	1089	1045.125	1.0420
	T3	7	909	1092.375	0.8321
	T4	8	1386	1145.250	1.2102
3	T1	9	1089	1201.500	0.9064
	T2	10	1305	1249.875	1.0441
	T3	11	1143	1287.000	0.8881
	T4	12	1539	1321.875	1.1643
4	T1	13	1233	1355.625	0.9095
	T2	14	1440	1396.125	1.0314
	T3	15	1278	1461.375	0.8745
	T4	16	1728	1544.625	1.1187
5	T1	17	1566	1629.000	0.9613
	T2	18	1773	1715.625	1.0334
	T3	19	1620	1768.500	0.9160
	T4	20	2079	1786.500	1.1637
6	T1	21	1638	1802.250	0.9089
	T2	22	1845	1819.125	1.0142
	T3	23	1674	1849.500	0.9051
	T4	24	2160	1888.875	1.1435
7	T1	25	1800	1929.375	0.9329
	T2	26	1998	1963.125	1.0178
	T3	27	1845	2001.375	0.9219
	T4	28	2259	2059.875	1.0967
8	T1	29	2007	2119.500	0.9469
	T2	30	2259	2175.750	1.0383
	T3	31	2061	2232.000	0.9234
	T4	32	2493	2279.250	1.0938
9	T1	33	2223	2324.250	0.9564
	T2	34	2421	2372.625	1.0204
	T3	35	2259	2408.625	0.9379
	T4	36	2682	2441.250	1.0986
10	T1	37	2322	2471.625	09395
	T2	38	2583	2508.750	1.0296
	T3	39	2340		
	T4	40	2898		

Los valores de la última columna son valores no estandarizados, por ello se debe proceder a estandarizarlos. Copiamos la última columna y la ordenamos por trimestres y años en una tabla de doble entrada:

	1	2	3	4	5	6	7	8	9	10
T1		0.8750	0.9064	0.9095	0.9613	0.9089	0.9329	0.9469	0.9564	0.9395
T2		1.0420	1.0441	1.0314	1.0334	1.0142	1.0178	1.0383	1.0204	1.0296
T3	0.8375	0.8321	0.8881	0.8745	0.9160	0.9051	0.9219	0.9234	0.9379	
T4	1.2500	1.2102	1.1643	1.1187	1.1637	1.1435	1.0967	1.0938	1.0986	

Se debe sacar un promedio de cada trimestre:

	Promed.
T1	0.9263
T2	1.0301
T3	0.8930
T4	1.1488
	3.9982

Los índices deben corregirse multiplicando cada promedio por el factor de corrección fc:

fc = 4/3.9982

Los resultados obtenidos son los índices o componentes estacionales:

	Promed.	Prom*fc
T1	0.9263	0.9267
T2	1.0301	1.0306
T3	0.8930	0.8933
T4	1.1488	1.1493
	3.9982	

	Comp. Estacional
T1	$E_1 = 0.9267$
T2	$E_2 = 1.0306$
T3	$E_3 = 0.8933$
T4	$E_4 = 1.1493$

c. La proyección para el año 11

$T = \hat{y} = 780.4731+46.6818x$

$T = \hat{y} = 780.4731+46.6818(11)$

$T = 1293.9729$

$E_{T1} = 0.9267$

$E_{T2} = 1.0306$

$E_{T3} = 0.8933$

$E_{T4} = 1.1493$

Proyecciones:

Trimestre 1 $\hat{y} = T*E_1 = 1293.9729*0.9267 = 1199.124686$

Trimestre 2 $\hat{y} = T*E_2 = 1293.9729*1.0306 = 1333.568471$

Trimestre 3 $\hat{y} = T*E_3 = 1293.9729*0.8933 = 1155.905992$

Trimestre 4 $\hat{y} = T*E_4 = 1293.9729*1.1493 = 1487.163054$

Lectura seleccionada n.º 4

Vidal-Beneyto, J. (2004, 14 de febrero). Las cuentas secuestradas. El País. Recuperado de: https://goo.gl/SzuRgy

Glosario de la Unidad IV

B

Bondad de ajuste. Medida de la capacidad de un modelo de regresión para explicar las variaciones en la variable dependiente. Lo que también nos lleva decir que es la medida de lo bien que se ajusta un modelo a los datos de dos variables correlacionadas.

L

Linealizar. Método matemático que permite transformar una ecuación exponencial, logarítmica, etcétera, en una ecuación lineal con la finalidad de aplicar el análisis de correlación y regresión lineal.

S

Suavizamiento. Proceso del método de análisis de series temporales que permite quitar la componente estacional de la serie.

V

Variable dependiente. Es la variable con valores que cambian en alguna medida al cambiar los valores de la variable independiente.

Variable independiente. Variable que tiene la cualidad de tener datos que cuando se producen, provocan que la variable dependiente los produzca.

Bibliografía de la Unidad IV

Cerrón, C. (2013). *Manual autoformativo del curso de Estadística II*. Huancayo, Junín, Perú.

Díaz, A. (2013). *Estadística aplicada a la administración y la economía*. México D. F.: Mc-Graw Hill.

INEI. (Mayo de 2006). Índice temático. Recuperado de: goo.gl/B4142e

Triola, M. (2013). *Estadística*. (11.ª ed.). México D. F.: Pearson Education.

Universidad de Valladolid. (13 de abril de 2012). Probabilidad y estadística orientada a la economía y la empresa [página web]. Recuperado de: goo.gl/41Ck0B

Vidal-Beneyto, J. (14 de febrero de 2004). *Las cuentas secuestradas*. El País. Recuperado de: goo.gl/ZnBkSp

Autoevaluación n.º 4

Lea con atención los enunciados. Repase en el manual el tema relacionado a la pregunta e intente una respuesta que se ajuste a su lectura y al criterio de aplicación de esta teoría.

1. **Una serie temporal es**
   a. Resultados de un proceso estocástico de variable aleatoria y tiempo discretos.
   b. Resultados de un proceso estocástico de variable numérica aleatoria y tiempo discreto.
   c. Resultados de un proceso aleatorio variable y el tiempo secuencial.
   d. Una variable cuantitativa continua que produce resultados en el tiempo.
   e. Secuencia de resultados estocásticos en el tiempo.

2. **Si al calcular el coeficiente de correlación de dos variables X e Y se tiene r = -0.890 (lineal), ocurre que**
   a. El modelo lineal de regresión explica el 89% de la varianza de una variable cualquiera en función de la otra.
   b. Existe correlación lineal y la pendiente de la recta de regresión es grande.
   c. X e Y correlacionadas, aun cuando X decrece, pero Y tiene tendencia a crecer.
   d. Existe correlación lineal y la pendiente de la recta de regresión es pequeña.
   e. Existe correlación lineal directa entre las variables X e Y.

3. **Tiempo de respuesta: El servicio de serenazgo de la ciudad para ciertos lugares es un problema. El jefe de esta oficina está preocupado por el tiempo de respuesta a las llamadas de emergencia. Ordena una investigación para determinar si la distancia del lugar de la llamada medida en kilómetros puede explicar el tiempo de respuesta, medido en minutos. Basándose en 37 emergencias se recolectaron los siguientes datos:**

Distancia (km)	4	5	6	7	9	10	12	14
Tiempo (mint)	8	6	13	23	30	24	28	42

Al 0.05 de significancia, ¿existe alguna razón para afirmar que la distancia no tiene que ver con el tiempo de demora?

M1	M2	M3
78	92	81
89	81	91
83	70	85
86	82	87
89	84	93
88		96
		81

a. Las muestras no son aleatorias, r = 0.9305, se rechaza Ho, distancia y tiempo son independientes.

b. d^2 = 8, el valor del estadístico de prueba es 0.738, no se rechaza H0 y no existe relación.

c. rs = 0.738, no se rechaza H0; la distancia no es el factor que influye en el tiempo de demora.

d. rs = 0.9048, se rechaza H0; existe evidencia muestral para afirmar lo contrario.

e. Rs^2 = 0.9048, existe una correlación no lineal. Los tiempos están afectados por las distancias.

4. Se tiene un proceso de trabajo en una empresa de construcción de maquinaria para la industria pesquera. Este proceso puede realizarse de tres formas diferentes. Se capacita a los trabajadores sobre estos métodos en grupos separados. Una evaluación final arroja los siguientes resultados en puntajes de 0 a 100. Realiza una prueba de Kruskal Wallis. ¿Se puede afirmar que alguno de los métodos presenta una diferencia significativa?

a. El valor crítico es 5.991, no se rechaza Ho; por tanto, ninguno de los métodos marca diferencia.

b. Estadístico de prueba = 6.045, se rechaza Ho. Entonces, por lo menos uno de los métodos es diferente.

c. El valor crítico es 6.045, se rechaza Ho; en consecuencia, las medias no son iguales.

d. El estadístico de prueba es 5.991. Entonces, no existe evidencia muestral para firmar que sean iguales.

e. El valor crítico es 7.915, Ho no se rechaza; por tanto, no existe diferencia en los métodos.

5. ¿Qué no es cierto sobre el coeficiente de correlación de Spearman?

I. Si se tiene un valor de 1,23, entonces existe una supercorrelación.

II. Si las variables están asociadas, entonces el valor de rs es menor al de la tabla.

III. Si rs es tal que supera en valor absoluto al valor de la tabla, se rechaza Ho.

IV. En todo caso, rs pertenece al intervalo [-1; 1]

V. Si las variables dependen una de la otra, entonces se ha rechazado Ho.

a. I, II y IV

b. III y IV

c. III, IV y V

d. I y V

e. I, II, V

6. A los siguientes datos:

Intereses (S/. miles)	6	16	9	102	56	24
Tiempo (años)	84	100	120	12	18	41

Se ajusta mejor un modelo:

a. Lineal:
b. Exponencial:
c. Potencial:
d. Inverso:
e. Logarítmico:

7. **¿Qué es cierto sobre la pendiente de una recta en una función de regresión lineal $Y=b_0+b_1X$?**
   I. Representa el incremento de Y por cada unidad de incremento de X.
   II. Tiene el mismo signo que r.
   III. Es el valor de la variable Y cuando X = 0.
   IV. Tiene el mismo valor que r.
   V. Es b1 el valor de la inclinación de la recta de regresión.
   a. II, III, V
   b. I, II, V
   c. III, V
   d. IV, V
   e. I, III, V

8. **Dos variables numéricas se encuentran incorrelacionadas, entonces:**
   a. Las variables tienen un diagrama de puntos no lineal.
   b. El modelo lineal de regresión solo propone el valor de $\bar{y}$ como predicción de Y.
   c. $\rho \neq 0$ al nivel de significancia de 0.05.
   d. La nube de puntos no presenta aspecto creciente.
   e. La variación de Y es igual para cualquier valor de x.

9. **Una serie temporal es la interacción de**
   a. La tendencia, la aleatoriedad, la secuencia y la variación cíclica.
   b. La variación estacional, variación cíclica, la secuencial y la irregular.
   c. La variación estacional, variación secuencial, la correlación lineal y la irregular.
   d. La tendencia, la variación estacional, variación cíclica y la irregular.
   e. Variación cíclica, la irregular, secuencial y la variación estacional.

10. **¿Cuáles de las siguientes no son razones para estudiar tanto tendencias seculares como variación estacional?**

a. Permite la separación de los componentes de la serie temporal.
b. Describe patrones pasados.
c. Describe el comportamiento del tiempo en la variable.
d. Proyecta patrones pasados hacia el futuro.
e. Desarrolla un modelo matemático ajustado a los patrones de la variable.

Apéndices

APÉNDICE A. Datos de empleados de Mariana S. A.

ID	SEXO	GRAD_INST	N.° HIJOS	CAT. LAB.	INGRESO (MILES)	TIEM.
1	Hombre	Universidad inconcluso	6	Admin.	$40.20	98
2	Mujer	Primaria	7	Admin.	$21.90	98
3	Hombre	Técnico	5	Admin.	$32.10	98
4	Mujer	Secundaria	2	Admin.	$21.90	98
5	Mujer	Secundaria	6	Admin.	$24.00	98
6	Mujer	Universidad inconcluso	4	Admin.	$30.30	98
7	Hombre	Técnico	3	Admin.	$27.75	98
8	Mujer	Primaria	5	Admin.	$31.35	96
9	Hombre	Técnico	2	Admin.	$31.35	96
10	Hombre	Secundaria	3	Admin.	$23.25	95
11	Mujer	Técnico	2	Admin.	$22.35	95
12	Mujer	Secundaria	8	Admin.	$30.00	95
13	Hombre	Secundaria	4	Admin.	$35.55	94
14	Hombre	Secundaria	8	Admin.	$25.05	94
15	Hombre	Primaria	7	Admin.	$22.50	94
16	Hombre	Primaria	6	Admin.	$21.90	93
17	Hombre	Técnico	4	Admin.	$41.10	93
18	Mujer	Secundaria	3	Admin.	$26.40	93
19	Mujer	Primaria	6	Admin.	$25.05	93
20	Mujer	Secundaria	4	Admin.	$28.50	92
21	Hombre	Técnico	8	Admin.	$33.45	90
22	Mujer	Técnico	2	Admin.	$32.55	90
23	Mujer	Secundaria	8	Admin.	$33.30	90
24	Hombre	Secundaria	4	Admin.	$27.30	89
25	Mujer	Técnico	2	Admin.	$26.55	88
26	Hombre	Universidad inconcluso	2	Admin.	$52.65	86
27	Hombre	Secundaria	7	Admin.	$26.70	86
28	Hombre	Técnico	1	Admin.	$37.50	84
29	Mujer	Secundaria	4	Admin.	$16.50	84
30	Mujer	Secundaria	8	Admin.	$24.75	84
31	Mujer	Secundaria	5	Admin.	$24.00	83
32	Mujer	Secundaria	8	Admin.	$20.40	83
33	Hombre	Técnico	2	Admin.	$30.15	82
34	Mujer	Secundaria	0	Admin.	$33.90	82
35	Hombre	Técnico	8	Admin.	$22.50	82
36	Mujer	Primaria	5	Admin.	$27.45	81
37	Mujer	Secundaria	3	Admin.	$27.30	81
38	Mujer	Secundaria	6	Admin.	$23.10	81
39	Mujer	Secundaria	2	Admin.	$23.10	81
40	Hombre	Secundaria	3	Admin.	$25.50	81
41	Hombre	Primaria	5	Admin.	$21.30	80
42	Mujer	Secundaria	3	Admin.	$23.40	80
43	Hombre	Técnico	2	Admin.	$28.65	79
44	Hombre	Universidad inconcluso	5	Admin.	$40.35	78
45	Hombre	Técnico	2	Admin.	$25.95	78
46	Hombre	Universidad inconcluso	3	Admin.	$26.55	78
47	Hombre	Técnico	6	Admin.	$30.75	77
48	Hombre	Técnico	6	Admin.	$34.50	77
49	Mujer	Secundaria	5	Admin.	$34.50	77
50	Hombre	Secundaria	3	Admin.	$27.75	77
51	Hombre	Secundaria	7	Admin.	$22.05	76
52	Mujer	Secundaria	4	Admin.	$22.05	76
53	Hombre	Técnico	3	Admin.	$27.30	75
54	Mujer	Secundaria	2	Admin.	$24.45	75
55	Mujer	Secundaria	4	Admin.	$26.10	74
56	Mujer	Primaria	7	Admin.	$15.90	74
57	Mujer	Primaria	8	Admin.	$21.75	74

ID	SEXO	GRAD_INST	N.° HIJOS	CAT. LAB.	INGRESO (MILES)	TIEM.
58	Hombre	Secundaria	5	Admin.	$35.70	73
59	Mujer	Técnico	5	Admin.	$22.95	73
60	Mujer	Secundaria	7	Admin.	$23.10	73
61	Hombre	Técnico	6	Admin.	$28.35	72
62	Mujer	Primaria	7	Admin.	$17.70	72
63	Mujer	Secundaria	8	Admin.	$29.40	72
64	Hombre	Universidad inconcluso	6	Admin.	$35.70	72
65	Mujer	Secundaria	5	Admin.	$24.75	72
66	Mujer	Secundaria	3	Admin.	$21.00	70
67	Hombre	Técnico	0	Admin.	$30.15	69
68	Mujer	Secundaria	2	Admin.	$27.90	69
69	Mujer	Primaria	7	Admin.	$29.10	69
70	Mujer	Secundaria	5	Admin.	$22.65	69
71	Mujer	Secundaria	8	Admin.	$20.85	69
72	Mujer	Secundaria	6	Admin.	$22.95	69
73	Mujer	Secundaria	8	Admin.	$22.80	69
74	Hombre	Técnico	3	Admin.	$30.30	68
75	Hombre	Técnico	7	Admin.	$27.15	67
76	Mujer	Técnico	4	Admin.	$31.35	67
77	Hombre	Primaria	5	Admin.	$26.25	67
78	Mujer	Primaria	8	Admin.	$24.15	66
79	Mujer	Técnico	3	Admin.	$23.85	66
80	Mujer	Secundaria	8	Admin.	$24.45	66
81	Hombre	Técnico	3	Admin.	$49.00	66
82	Mujer	Secundaria	4	Admin.	$16.35	66
83	Hombre	Técnico	6	Admin.	$28.50	65
84	Hombre	Técnico	2	Admin.	$24.45	65
85	Mujer	Primaria	5	Admin.	$21.60	65
86	Mujer	Técnico	8	Admin.	$20.70	65
87	Mujer	Universidad inconcluso	4	Admin.	$32.85	64
88	Mujer	Doctorado	4	Admin.	$36.00	45
89	Hombre	Técnico	4	Admin.	$26.00	48
90	Hombre	Técnico	7	Admin.	$28.00	57
91	Hombre	Universidad inconcluso	2	Direct.	$103.75	97
92	Hombre	Maestría	0	Direct.	$110.63	96
93	Hombre	Universidad inconcluso	2	Direct.	$45.25	93
94	Hombre	Maestría	2	Direct.	$68.75	92
95	Hombre	Maestría	1	Direct.	$78.25	91
96	Hombre	Maestría	1	Direct.	$91.25	91
97	Hombre	Universidad	1	Direct.	$68.75	89
98	Hombre	Doctorado	4	Direct.	$59.38	89
99	Hombre	Universidad inconcluso	2	Direct.	$66.00	86
100	Hombre	Universidad inconcluso	4	Direct.	$45.63	86
101	Hombre	Universidad inconcluso	2	Direct.	$40.20	81
102	Hombre	Maestría	3	Direct.	$75.00	81
103	Mujer	Universidad inconcluso	4	Direct.	$54.38	81
104	Hombre	Maestría	2	Direct.	$52.13	80
105	Mujer	Universidad inconcluso	1	Direct.	$43.00	79
106	Hombre	Universidad inconcluso	3	Direct.	$48.75	76
107	Hombre	Secundaria	3	Direct.	$59.40	74
108	Hombre	Maestría	1	Direct.	$66.88	69
109	Hombre	Maestría	2	Direct.	$66.25	67
110	Hombre	Universidad inconcluso	1	Direct.	$100.00	66
111	Hombre	Maestría	3	Direct.	$47.55	64
112	Hombre	Doctorado	4	Direct.	$158.00	46
113	Hombre	Secundaria	3	Segur.	$30.75	95
114	Hombre	Secundaria	1	Segur.	$30.75	94
115	Hombre	Secundaria	2	Segur.	$30.75	91
116	Hombre	Primaria	2	Segur.	$30.75	87
117	Hombre	Primaria	1	Segur.	$31.95	85
118	Hombre	Secundaria	3	Segur.	$35.25	78
119	Hombre	Primaria	1	Segur.	$29.55	76
120	Hombre	Primaria	2	Segur.	$30.00	67

Fuente: Propia.

APÉNDICE B. Tabla de números aleatorios.

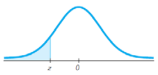
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
4	8	2	4	6	6	3	5	4	5	6	0	5	2	6	9	8	0	0	9
9	2	9	8	1	4	4	1	9	8	5	1	1	9	7	9	8	5	9	0
0	2	1	3	3	9	1	6	2	9	7	1	2	6	6	0	7	5	6	4
9	6	0	8	3	5	6	6	6	4	0	8	6	3	4	8	1	8	5	4
1	6	4	1	6	5	2	7	7	2	9	9	9	9	7	4	1	5	4	9
2	9	0	5	5	0	8	4	8	7	4	6	2	1	7	0	1	5	8	7
6	1	2	9	5	0	4	0	9	8	2	0	2	6	8	7	0	1	9	7
1	3	1	8	9	9	0	1	2	6	3	7	1	9	6	1	7	9	9	8
4	5	8	1	1	4	5	6	7	9	9	9	2	1	3	2	3	7	7	9
0	0	3	6	9	6	5	0	6	4	7	9	8	1	2	4	4	8	3	6
7	2	4	5	4	1	2	4	4	6	9	2	6	6	6	5	2	0	0	4
4	9	3	4	4	2	4	5	9	0	8	7	4	8	4	2	1	2	5	4
6	1	2	8	1	3	3	2	0	2	6	0	7	2	7	9	1	4	6	5
9	3	4	0	8	1	3	3	7	3	2	4	8	6	7	9	0	6	2	8
1	8	7	1	3	4	3	9	3	1	7	8	3	7	3	3	0	8	3	5
0	2	1	4	7	5	7	3	1	1	9	3	3	8	7	4	8	0	2	5
3	6	3	4	1	9	8	1	0	9	0	1	1	0	9	3	6	8	6	0
9	4	6	7	6	7	9	1	2	2	7	2	3	9	3	4	6	9	8	1
5	9	9	8	4	4	5	9	1	5	4	7	3	0	6	8	1	6	8	1
8	1	8	8	2	3	9	1	4	2	4	9	1	4	0	6	0	3	2	8
0	5	3	8	0	4	3	9	4	6	0	8	8	3	8	7	1	2	2	3
9	7	1	4	2	7	5	5	2	8	6	6	3	5	5	9	9	0	6	8
6	9	5	9	4	9	1	8	2	0	2	5	3	9	1	2	0	3	0	8
7	4	9	1	4	8	8	6	6	8	5	9	4	8	5	7	7	9	6	7
3	8	1	2	2	4	0	1	4	5	7	7	4	0	4	8	9	4	7	0
9	9	9	7	8	0	0	9	3	2	7	0	5	0	2	7	8	7	3	6
4	8	1	5	8	5	5	1	4	9	6	4	4	4	7	4	5	7	5	0
8	6	7	3	6	1	7	1	1	3	5	5	7	4	4	7	6	7	2	8
4	7	1	4	0	3	6	2	4	4	4	4	0	3	6	3	4	1	2	8
6	5	5	8	8	4	3	4	8	9	0	6	7	6	0	0	8	6	8	4
9	2	0	9	8	2	8	3	4	3	2	8	9	4	8	7	9	4	9	4
1	3	7	9	4	8	3	7	0	8	6	6	6	8	4	1	1	3	1	3
3	3	2	5	6	7	6	1	6	6	1	7	6	5	8	1	6	2	2	7
9	9	9	8	2	8	8	1	9	1	6	2	7	5	1	8	6	1	4	4
1	7	5	4	0	9	5	7	8	7	5	0	8	6	6	2	5	3	2	3
2	7	1	7	8	8	3	8	6	9	9	2	7	4	5	9	5	6	6	6
6	0	9	2	6	1	5	1	2	3	1	8	1	2	0	8	6	4	4	0
3	3	6	3	4	9	6	4	4	9	8	5	7	3	3	4	2	3	2	8
0	1	9	7	9	7	9	4	4	1	6	6	7	7	0	7	9	8	6	8
4	7	1	5	3	7	0	9	2	5	2	1	0	0	4	0	4	6	8	8
7	8	9	9	6	8	5	6	8	1	9	2	7	5	1	7	0	1	5	5
2	2	3	3	1	8	1	9	8	4	2	8	5	2	8	1	7	6	4	6
2	6	6	4	1	4	8	1	0	6	0	1	3	4	0	9	1	2	8	6
5	1	9	0	3	9	1	6	1	7	8	8	2	8	0	7	8	4	8	0
9	0	5	8	4	9	2	2	3	9	8	5	9	5	7	8	4	9	9	4
8	6	1	9	2	5	0	0	7	9	0	0	7	4	5	4	8	6	2	3
1	9	1	0	9	7	5	1	2	7	1	9	4	8	4	8	9	6	6	9
5	6	0	6	1	3	3	5	2	1	0	1	9	2	8	0	2	6	6	3
8	6	9	9	8	0	8	1	8	2	6	6	8	4	0	7	8	2	5	1
3	1	6	1	0	5	7	5	7	0	6	3	0	4	1	4	0	3	0	8

Fuente: Cerrón, 2013.

APÉNDICE C. Tablas de valores Z de la distribución normal estándar

Tabla A2. Distribución Normal de Valores - z

$$P_{(a\leq z)} = \frac{1}{2\pi}\int_{-\infty}^{z} e^{\frac{x^2}{2}}.dx$$



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-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	0.0013	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	0.0228	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	0.1587	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	0.2296	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	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Nota: Para valores de z menores que -3.49 utilice 0.0001 como área.

*Utilice valores comunes que resultan por interpolación.

z = -1.645 Área = 0.0500

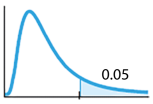
z = -2.575 Área = 0.0050

Valores comunes críticos	
Nivel de confianza	Valor crítico
0.90	1.645
0.95	1.96
0.99	2.575

Tabla A3. Distribución “t”: Valores críticos t

2 colas	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001	α
1 cola	0.45	0.4	0.35	0.25	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005	α
gl	Valores t													
1	0.158	0.325	0.510	1.000	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619	
2	0.142	0.289	0.445	0.816	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599	
3	0.137	0.277	0.424	0.765	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924	
4	0.134	0.271	0.414	0.741	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	0.132	0.267	0.408	0.727	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	0.131	0.265	0.404	0.718	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	0.130	0.263	0.402	0.711	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408	
8	0.130	0.262	0.399	0.706	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	
9	0.129	0.261	0.398	0.703	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	
10	0.129	0.260	0.397	0.700	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587	
11	0.129	0.260	0.396	0.697	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	
12	0.128	0.259	0.395	0.695	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	
13	0.128	0.259	0.394	0.694	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	
14	0.128	0.258	0.393	0.692	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	
15	0.128	0.258	0.393	0.691	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	
16	0.128	0.258	0.392	0.690	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	
17	0.128	0.257	0.392	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	
18	0.127	0.257	0.392	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	
19	0.127	0.257	0.391	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	
20	0.127	0.257	0.391	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	
21	0.127	0.257	0.391	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819	
22	0.127	0.256	0.390	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	
23	0.127	0.256	0.390	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768	
24	0.127	0.256	0.390	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	
25	0.127	0.256	0.390	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	
26	0.127	0.256	0.390	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707	
27	0.127	0.256	0.389	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690	
28	0.127	0.256	0.389	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674	
29	0.127	0.256	0.389	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659	
30	0.127	0.256	0.389	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	
31	0.127	0.256	0.389	0.682	1.054	1.309	1.696	2.040	2.453	2.744	3.022	3.375	3.633	
32	0.127	0.255	0.389	0.682	1.054	1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622	
34	0.127	0.255	0.389	0.682	1.052	1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601	
36	0.127	0.255	0.388	0.681	1.052	1.306	1.688	2.028	2.434	2.719	2.990	3.333	3.582	
38	0.127	0.255	0.388	0.681	1.051	1.304	1.686	2.024	2.429	2.712	2.980	3.319	3.566	
40	0.126	0.255	0.388	0.681	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	
45	0.126	0.255	0.388	0.680	1.049	1.301	1.679	2.014	2.412	2.690	2.952	3.281	3.520	
50	0.126	0.255	0.388	0.679	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496	
55	0.126	0.255	0.387	0.679	1.046	1.297	1.673	2.004	2.396	2.668	2.925	3.245	3.476	
60	0.126	0.254	0.387	0.679	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	
65	0.126	0.254	0.387	0.678	1.045	1.295	1.669	1.997	2.385	2.654	2.906	3.220	3.447	
70	0.126	0.254	0.387	0.678	1.044	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435	
75	0.126	0.254	0.387	0.678	1.044	1.293	1.665	1.992	2.377	2.643	2.892	3.202	3.425	
80	0.126	0.254	0.387	0.678	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416	
90	0.126	0.254	0.387	0.677	1.042	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402	
100	0.126	0.254	0.386	0.677	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390	
200	0.126	0.254	0.386	0.676	1.039	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340	
300	0.126	0.254	0.386	0.675	1.038	1.284	1.650	1.968	2.339	2.592	2.828	3.118	3.323	
500	0.126	0.253	0.386	0.675	1.038	1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310	
750	0.126	0.253	0.385	0.675	1.037	1.283	1.647	1.963	2.331	2.582	2.815	3.101	3.304	
1000	0.126	0.253	0.385	0.675	1.037	1.282	1.646	1.962	2.330	2.581	2.813	3.098	3.300	
2000	0.126	0.253	0.385	0.675	1.037	1.282	1.646	1.961	2.328	2.578	2.810	3.094	3.295	
Grande	0.126	0.253	0.385	0.675	1.036	1.282	1.645	1.960	2.327	2.576	2.808	3.091	3.291	

Tabla A4. Distribución chi cuadrada (X^2)



	α											
gl	0.995	0.990	**0.975**	0.950	0.900	0.100	0.075	0.050	**0.025**	0.010	0.005	0.001
1	0.000	0.000	**0.001**	0.004	0.016	2.706	3.170	3.841	**5.024**	6.635	7.879	10.828
2	0.010	0.020	**0.051**	0.103	0.211	4.605	5.181	5.991	**7.378**	9.210	10.597	13.816
3	0.072	0.115	**0.216**	0.352	0.584	6.251	6.905	7.815	**9.348**	11.345	12.838	16.266
4	0.207	0.297	**0.484**	0.711	1.064	7.779	8.496	9.488	**11.143**	13.277	14.860	18.467
5	0.412	0.554	**0.831**	1.145	1.610	9.236	10.008	11.070	**12.833**	15.086	16.750	20.515
6	0.676	0.872	**1.237**	1.635	2.204	10.645	11.466	12.592	**14.449**	16.812	18.548	22.458
7	0.989	1.239	**1.690**	2.167	2.833	12.017	12.883	14.067	**16.013**	18.475	20.278	24.322
8	1.344	1.646	**2.180**	2.733	3.490	13.362	14.270	15.507	**17.535**	20.090	21.955	26.124
9	1.735	2.088	**2.700**	3.325	4.168	14.684	15.631	16.919	**19.023**	21.666	23.589	27.877
10	2.156	2.558	**3.247**	3.940	4.865	15.987	16.971	18.307	**20.483**	23.209	25.188	29.588
11	2.603	3.053	**3.816**	4.575	5.578	17.275	18.294	19.675	**21.920**	24.725	26.757	31.264
12	3.074	3.571	**4.404**	5.226	6.304	18.549	19.602	21.026	**23.337**	26.217	28.300	32.909
13	3.565	4.107	**5.009**	5.892	7.042	19.812	20.897	22.362	**24.736**	27.688	29.819	34.528
14	4.075	4.660	**5.629**	6.571	7.790	21.064	22.180	23.685	**26.119**	29.141	31.319	36.123
15	4.601	5.229	**6.262**	7.261	8.547	22.307	23.452	24.996	**27.488**	30.578	32.801	37.697
16	5.142	5.812	**6.908**	7.962	9.312	23.542	24.716	26.296	**28.845**	32.000	34.267	39.252
17	5.697	6.408	**7.564**	8.672	10.085	24.769	25.970	27.587	**30.191**	33.409	35.718	40.790
18	6.265	7.015	**8.231**	9.390	10.865	25.989	27.218	28.869	**31.526**	34.805	37.156	42.312
19	6.844	7.633	**8.907**	10.117	11.651	27.204	28.458	30.144	**32.852**	36.191	38.582	43.820
20	7.434	8.260	**9.591**	10.851	12.443	28.412	29.692	31.410	**34.170**	37.566	39.997	45.315
21	8.034	8.897	**10.283**	11.591	13.240	29.615	30.920	32.671	**35.479**	38.932	41.401	46.797
22	8.643	9.542	**10.982**	12.338	14.041	30.813	32.142	33.924	**36.781**	40.289	42.796	48.268
23	9.260	10.196	**11.689**	13.091	14.848	32.007	33.360	35.172	**38.076**	41.638	44.181	49.728
24	9.886	10.856	**12.401**	13.848	15.659	33.196	34.572	36.415	**39.364**	42.980	45.559	51.179
25	10.520	11.524	**13.120**	14.611	16.473	34.382	35.780	37.652	**40.646**	44.314	46.928	52.620
26	11.160	12.198	**13.844**	15.379	17.292	35.563	36.984	38.885	**41.923**	45.642	48.290	54.052
27	11.808	12.879	**14.573**	16.151	18.114	36.741	38.184	40.113	**43.195**	46.963	49.645	55.476
28	12.461	13.565	**15.308**	16.928	18.939	37.916	39.380	41.337	**44.461**	48.278	50.993	56.892
29	13.121	14.256	**16.047**	17.708	19.768	39.087	40.573	42.557	**45.722**	49.588	52.336	58.301
30	13.787	14.953	**16.791**	18.493	20.599	40.256	41.762	43.773	**46.979**	50.892	53.672	59.703
31	14.458	15.655	**17.539**	19.281	21.434	41.422	42.948	44.985	**48.232**	52.191	55.003	61.098
33	15.815	17.074	**19.047**	20.867	23.110	43.745	45.311	47.400	**50.725**	54.776	57.648	63.870
35	17.192	18.509	**20.569**	22.465	24.797	46.059	47.663	49.802	**53.203**	57.342	60.275	66.619
37	18.586	19.960	**22.106**	24.075	26.492	48.363	50.005	52.192	**55.668**	59.893	62.883	69.346
40	20.707	22.164	**24.433**	26.509	29.051	51.805	53.501	55.758	**59.342**	63.691	66.766	73.402
50	27.991	29.707	**32.357**	34.764	37.689	63.167	65.030	67.505	**71.420**	76.154	79.490	86.661
60	35.534	37.485	**40.482**	43.188	46.459	74.397	76.411	79.082	**83.298**	88.379	91.952	99.607
70	43.275	45.442	**48.758**	51.739	55.329	85.527	87.680	90.531	**95.023**	100.425	104.215	112.317
80	51.172	53.540	**57.153**	60.391	64.278	96.578	98.861	101.879	**106.629**	112.329	116.321	124.839
90	59.196	61.754	**65.647**	69.126	73.291	107.565	109.969	113.145	**118.136**	124.116	128.299	137.208
100	67.328	70.065	**74.222**	77.929	82.358	118.498	121.017	124.342	**129.561**	135.807	140.169	149.449

Fuente: Triola, 2013.

Tabla A6. Valores críticos para el coeficiente de correlación de Pearson r

n	α=0.05	α=0.01
4	0.95	0.999
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
9	0.666	0.798
10	0.632	0.765
11	0.602	0.735
12	0.576	0.708
13	0.553	0.684
14	0.532	0.661
15	0.514	0.641
16	0.497	0.623
17	0.482	0.606
18	0.468	0.59
19	0.456	0.575
20	0.444	0.561
25	0.396	0.505
30	0.361	0.463
35	0.335	0.43
40	0.312	0.402
45	0.294	0.378
50	0.279	0.361
60	0.254	0.33
70	0.236	0.305
80	0.22	0.286
90	0.207	0.269
100	0.196	0.256

NOTA:
Para una prueba Ho: $\rho = 0$ contra H1: $\rho \neq 0$,
rechace Ho si el valor absoluto de r es mayor que el valor crítico en la tabla.

Tabla A7. Valores críticos para la prueba de rangos con signo de Wilcoxon

	α				
n	**0.005**	**0.01**	**0.025**	**0.05**	**1 cola**
	0.01	**0.02**	**0.05**	**0.1**	**2 colas**
1	*	*	*	*	
2	*	*	*	*	
3	*	*	*	*	
4	*	*	*	*	
5	*	*	*	0	
6	*	*	0	0	
7	*	0	0	0	
8	0	0	0	1	
9	0	0	1	1	
10	0	0	1	1	
11	0	1	1	2	
12	1	1	2	2	
13	1	1	2	3	
14	1	2	2	3	
15	2	2	3	3	
16	2	2	3	4	
17	2	3	4	4	
18	3	3	4	5	
19	3	4	4	5	
20	3	4	5	5	
21	4	4	5	6	
22	4	5	5	6	
23	4	5	6	7	
24	5	5	6	7	
25	5	6	7	7	

1. * indica que no es posible obtener un valor en la región crítica. No rechace Ho.
2. Rechace la hpótesis nula si el número del signo menos frecuente (x) es menor o igual al valor crítico de la tabla.
3. Para valores de n > 25, se utiliza una aproximación normal con:

$$z = \frac{(x+0.5)-\frac{n}{2}}{\frac{\sqrt{n}}{2}}$$

Tabla A8. Valores críticos para la prueba de fuego

	α				
n	**0.005**	**0.01**	**0.025**	**0.05**	**1 cola**
	0.01	**0.02**	**0.05**	**0.1**	**2 colas**
5	*	*	*	1	
6	*	*	1	2	
7	*	0	2	4	
8	0	2	4	6	
9	2	3	6	8	
10	3	5	8	11	
11	5	7	11	14	
12	7	10	14	17	
13	10	13	17	21	
14	13	16	21	26	
15	16	20	25	30	
16	19	24	30	36	
17	20	28	35	41	
18	28	33	40	47	
19	32	38	46	54	
20	37	43	52	60	
21	43	49	59	68	
22	49	56	66	75	
23	55	62	73	83	
24	61	69	81	92	
25	68	77	90	101	
26	76	85	98	110	
27	84	93	107	120	
28	92	102	117	130	
29	100	111	127	141	
30	109	120	137	152	

1. * indica que no es posible obtener un valor en la región crítica. No rechace Ho.
2. Rechace la hipótesis nula si el estadístico de prueba (T) es menor o igual que *el valor crítico en la tabla.*
3. Para valores de n > 30, se utiliza una aproximación normal con:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Tabla A9. Valores críticos para la prueba de rangos de Spearman

n	α=0.10	α=0.05	α=0.02	α=0.01
5	0.9	----	----	----
6	0.829	0.886	0.943	----
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.678	0.727
13	0.484	0.56	0.648	0.703
14	0.464	0.538	0.626	0.679
15	0.446	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.680	0.433	0.475
30	0.306	0.362	0.425	0.467

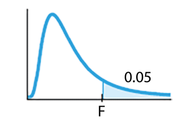
NOTA:
Para n > 30 utilice rs = $\pm \frac{z}{\sqrt{n-1}}$ donde z corresponde a nivel de significancia.
Por ejemplo, si α = 0.05, entonces z = 1.96

Universidad Continental

Tabla A5. Distribución F $\alpha = 0{,}025$

Grados de libertad del denominador (gl2)	Grados de libertad del numerador (gl1)																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**30000**
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	976.71	984.87	993.10	997.25	1001.41	1005.60	1009.80	1014.02	1018.24
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.415	39.431	39.448	39.456	39.465	39.473	39.481	39.490	39.498
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731	14.4189	14.3366	14.2527	14.1674	14.1241	14.0805	14.0365	13.9921	13.9473	13.9022
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565	8.5599	8.5109	8.4613	8.4111	8.3604	8.3092	8.2575
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277	6.3286	6.2780	6.2269	6.1750	6.1225	6.0693	6.0155
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687	5.1684	5.1172	5.0652	5.0125	4.9589	4.9044	4.8493
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678	4.4667	4.4150	4.3624	4.3089	4.2544	4.1989	4.1426
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012	3.9995	3.9472	3.8940	3.8398	3.7844	3.7279	3.6704
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694	3.6669	3.6142	3.5604	3.5055	3.4493	3.3918	3.3331
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217	3.4185	3.3654	3.3110	3.2554	3.1984	3.1399	3.0800
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299	3.2261	3.1725	3.1176	3.0613	3.0035	2.9441	2.8830
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2773	3.1772	3.0728	3.0187	2.9633	2.9063	2.8478	2.7874	2.7252
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1532	3.0527	2.9477	2.8932	2.8372	2.7797	2.7204	2.6590	2.5957
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0502	2.9493	2.8437	2.7888	2.7324	2.6742	2.6142	2.5519	2.4875
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.9633	2.8621	2.7559	2.7006	2.6437	2.5850	2.5242	2.4611	2.3956
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.8890	2.7875	2.6808	2.6252	2.5678	2.5085	2.4471	2.3831	2.3165
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8249	2.7230	2.6158	2.5598	2.5020	2.4422	2.3801	2.3153	2.2477
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.7689	2.6667	2.5590	2.5027	2.4445	2.3842	2.3214	2.2558	2.1872
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7196	2.6171	2.5089	2.4523	2.3937	2.3329	2.2696	2.2032	2.1336
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.6758	2.5731	2.4645	2.4076	2.3486	2.2873	2.2234	2.1562	2.0856
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977	2.7348	2.6368	2.5338	2.4247	2.3675	2.3082	2.2465	2.1819	2.1141	2.0425
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6017	2.4984	2.3890	2.3315	2.2718	2.2097	2.1446	2.0760	2.0035
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9023	2.8077	2.7313	2.6682	2.5699	2.4665	2.3567	2.2989	2.2389	2.1763	2.1107	2.0415	1.9680
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5411	2.4374	2.3273	2.2693	2.2090	2.1460	2.0799	2.0099	1.9356
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5149	2.4110	2.3005	2.2422	2.1816	2.1183	2.0516	1.9811	1.9058
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.4908	2.3867	2.2759	2.2174	2.1565	2.0928	2.0257	1.9545	1.8784
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309	2.5676	2.4688	2.3644	2.2533	2.1946	2.1334	2.0693	2.0018	1.9299	1.8530
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4484	2.3438	2.2324	2.1735	2.1121	2.0477	1.9797	1.9072	1.8295
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919	2.5286	2.4295	2.3248	2.2131	2.1540	2.0923	2.0276	1.9591	1.8861	1.8075
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4120	2.3072	2.1952	2.1359	2.0739	2.0089	1.9400	1.8664	1.7870
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.2882	2.1819	2.0677	2.0069	1.9429	1.8752	1.8028	1.7242	1.6375
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.1692	2.0613	1.9445	1.8817	1.8152	1.7440	1.6668	1.5810	1.4826
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.0548	1.9450	1.8249	1.7597	1.6899	1.6141	1.5299	1.4327	1.3110
#####	5.0240	3.6889	3.1162	2.7859	2.5666	2.4083	2.2876	2.1919	2.1137	2.0484	1.9448	1.8326	1.7085	1.6402	1.5660	1.4836	1.3884	1.2685	1.0173

Tabla A5. Distribución F $\alpha = 0{,}05$



		Grados de libertad del numerador (gl1)																		
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**30000**
Grados de libertad del denominador (gl2)	**1**	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
	2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
	3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5265
	4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6282
	5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3651
	6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6690
	7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2299
	8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9277
	9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7068
	10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5381
	11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4046
	12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2964
	13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2066
	14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1309
	15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0660
	16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0098
	17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9606
	18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9171
	19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8782
	20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8434
	21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8119
	22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7833
	23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7572
	24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7333
	25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7112
	26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6908
	27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6719
	28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6543
	29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6379
	30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6225
	40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5092
	60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3896
	120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2543
	#####	3.8415	2.9958	2.6050	2.3720	2.2141	2.0986	2.0096	1.9385	1.8799	1.8308	1.7522	1.6664	1.5706	1.5173	1.4592	1.3940	1.3181	1.2215	1.0145

Tabla A10. Valores críticos para el número de Rachas G

	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
2	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6
3	1 6	1 8	1 8	1 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	3 8	3 8	3 8	3 8	3 8	3 8
4	1 6	1 8	1 9	2 9	2 9	2 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	4 10	4 10	4 10	4 10	4 10
5	1 6	1 8	2 9	2 10	3 10	3 11	3 11	3 12	3 12	4 12	4 12	4 12	4 12	4 12	4 12	4 12	5 12	5 12	5 12
6	1 6	2 8	2 9	3 10	3 11	3 12	3 12	4 13	4 13	4 13	4 13	5 14	5 14	5 14	5 14	5 14	5 14	6 14	6 14
7	1 6	2 8	2 10	3 11	3 12	3 13	4 13	4 14	5 14	5 14	5 14	5 15	5 15	6 16	6 16	6 16	6 16	6 16	6 16
8	1 6	2 8	3 10	3 11	3 12	4 13	4 14	5 14	5 15	5 15	6 16	6 16	6 16	6 16	7 17	7 17	7 17	7 17	7 17
9	1 6	2 8	3 10	3 12	4 13	4 14	5 14	5 15	5 16	6 16	6 16	6 17	7 17	7 18	7 18	7 18	8 18	8 18	8 18
10	1 6	2 8	3 10	3 12	4 13	5 14	5 15	5 16	6 16	6 17	7 17	7 18	7 18	7 18	8 19	8 19	8 19	8 20	9 20
11	1 6	2 8	3 10	4 12	4 13	5 14	5 15	6 16	6 17	7 17	7 18	7 19	8 19	8 19	8 20	9 20	9 20	9 21	9 21
12	2 6	2 8	3 10	4 12	4 13	5 14	6 16	6 16	7 17	7 18	7 19	8 19	8 20	8 20	9 21	9 21	9 21	10 22	10 22
13	2 6	2 8	3 10	4 12	5 14	5 15	6 16	6 17	7 18	7 19	8 19	8 20	9 20	9 21	9 21	10 22	10 22	10 23	10 23
14	2 6	2 8	3 10	4 12	5 14	5 15	6 16	7 17	7 18	8 19	8 20	9 20	9 21	9 22	10 22	10 23	10 23	11 23	11 24
15	2 6	2 8	3 10	4 12	5 14	6 16	6 16	7 18	7 18	8 19	8 20	9 21	9 22	10 22	10 23	11 23	11 24	11 24	12 25
16	2 6	2 8	4 10	4 12	5 14	6 16	6 17	7 18	8 19	8 20	9 21	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 25
17	2 6	2 8	4 10	4 12	5 14	6 16	7 17	7 18	8 19	9 20	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 26	13 26
18	2 6	2 8	4 10	5 12	5 14	6 16	7 17	8 18	8 19	9 20	9 21	10 22	10 23	11 24	11 25	12 25	12 26	13 26	13 27
19	2 6	2 8	4 10	5 12	6 13	6 16	7 17	8 18	8 20	9 21	10 22	10 23	11 23	11 24	12 25	12 26	13 26	13 27	13 27
20	2 6	2 8	4 10	5 12	6 14	6 16	7 17	8 18	9 20	9 21	10 22	10 23	11 24	12 25	12 25	13 26	13 27	13 27	14 28

1. Los valores en esta tabla son valores críticos G, suponiendo una prueba de dos colas con un nivel de significancia de 50.05.
2. La hipótesis nula de aleatoriedad se rechaza si el número total de rachas G es menor que o igual al valor más bajo, o si es mayor que o igual al valor más alto.

De "Tables for testing randomness of groupings in a sequence of alternatives", *The annals of mathematical statistics,* vol. 14, n.° 1. Reprroducido con permiso del Institute of Mathematical Statistics.

Tabla 14-2. Constantes de una gráfica de control

n: Número de observaciones en subgrupo	$\bar{X}$		S		R	
	A_2	**A_3**	**B_3**	**B_4**	**D_3**	**D_4**
2	1.880	2.659	0.000	3.267	0.000	3.267
3	1.023	1.954	0.000	2.568	0.000	2.574
4	0.729	1.628	0.000	2.266	0.000	2.282
5	0.577	1.427	0.000	2.089	0.000	2.114
6	0.483	1.287	0.030	1.970	0.000	2.004
7	0.419	1.182	0.118	1.882	0.076	1.924
8	0.373	1.099	0.185	1.815	0.136	1.864
9	0.337	1.032	0.239	1.761	0.184	1.816
10	0.308	0.975	0.284	1.716	0.223	1.777
11	0.285	0.927	0.321	1.679	0.256	1.744
12	0.266	0.886	0.354	1.646	0.283	1.717
13	0.249	0.850	0.382	1.618	0.307	1.693
14	0.235	0.817	0.406	1.594	0.328	1.672
15	0.223	0.789	0.428	1.572	0.347	1.653
16	0.212	0.763	0.448	1.552	0.363	1.637
17	0.203	0.739	0.466	1.534	0.378	1.622
18	0.194	0.718	0.482	1.518	0.391	1.608
19	0.187	0.698	0.497	1.503	0.403	1.597
20	0.180	0.680	0.510	1.490	0.415	1.585
21	0.173	0.663	0.523	1.477	0.425	1.575
22	0.167	0.647	0.534	1.466	0.434	1.566
23	0.162	0.633	0.545	1.455	0.443	1.557
24	0.157	0.619	0.555	1.445	0.451	1.548
25	0.153	0.606	0.565	1.435	0.459	1.541

Fuente: Adaptado del *ASTM Manual on the Presentation of Data and Control Chart Analysis,* © 1976 ASTM, pp. 134-136.
Reproducido bajo permiso de American Society of Testing and Materials.

Anexos

UNIDAD I

Pregunta	Respuesta
1	b
2	b
3	a
4	a
5	e
6	d
7	e
8	a
9	d
10	b

UNIDAD II

Pregunta	Respuesta
1	c
2	b
3	a
4	e
5	c
6	d
7	d
8	b
9	a
10	b

UNIDAD III

Pregunta	Respuesta
1	b
2	e
3	b
4	e
5	e
6	b
7	d
8	c
9	b
10	a

UNIDAD IV

Pregunta	Respuesta
1	b
2	c
3	d
4	a
5	e
6	d
7	b
8	b
9	d
10	c

Huancayo

Av. San Carlos 1980 - Huancayo

Teléfono: 064 - 481430

Lima

Jr. Junín 355 - Miraflores

Teléfono: 01 - 2132760

Cusco

Av. Collasuyo S/N Urb. Manuel Prado - Cusco

Teléfono: 084 - 480070

Arequipa

Calle Alfonso Ugarte 607 - Yanahuara

Oficina administrativa: Calle San José 308 2° piso - Cercado

Teléfono: 054 - 412030