PRUEBA DE HIPÓTESIS ESTADÍSTICA CON EXCEL

Dr. Rafael Hernández Rodríguez
Mtra. Teresa Nohemi Cárdenas Arriaga
Mtra. Nancy Araceli Hernández Rodríguez

amat Editorial

ISBN 841831323-4
9 788418 313233

Tamaño final abierto: 28,4 cm × 21,5 cm | Ancho total [formato cerrado]: 14 cm | Lomo: 0,4 cm

PRUEBA DE HIPÓTESIS ESTADÍSTICA CON EXCEL

PRUEBA DE HIPÓTESIS ESTADÍSTICA CON EXCEL

Dr. Rafael Hernández Rodríguez
Mtra. Teresa Nohemi Cárdenas Arriaga
Mtra. Nancy Araceli Hernández Rodríguez

amat Editorial
Guadalajara, México, 2020

La presentación y disposición en conjunto de:

PRUEBA DE HIPÓTESIS
ESTADÍSTICA CON EXCEL

Es propiedad del autor.

Guadalajara, México, 2020

ISBN: 978-84-18313-23-3

Impreso en México / Printed in Mexico.

ÍNDICE

Introducción

La estadística es una rama de las matemáticas cuyo objetivo es reunir información cuantitativa de grupos o muestras de ellos donde, mediante el análisis de estos datos, da significados precisos o predictivos.

La palabra *estadística* proviene del vocablo latín *status* que significa posición, estado. Así la estadística era conocida como la parte de las matemáticas que permitía al gobierno predecir la cantidad de impuestos a cobrar al hacer suposiciones después de contabilizar el número de habitantes. Ahora no solo es usada por los gobiernos, sino en todas las ciencias ya que permite organizar, analizar, presentar e interpretar datos y en muchas ocasiones predecir con solo conocer una parte representativa válida para toda la población y con esto obtener conclusiones.

En la toma de decisiones, la estadística recopila, presenta, analiza y usa los datos para resolver problemas. En todos los ámbitos de la vida, sobre todo en lo profesional se está en contacto con múltiples datos diariamente con los cuales se deba sacar conclusiones o incluso tomar decisiones por lo que es útil hacer uso de herramientas estadísticas.

Los eruditos de esta disciplina la dividen en dos partes: *estadística descriptiva y estadística inferencial.* En este libro solo se revisarán algunos temas selectos de la estadística inferencial pues el objetivo es plantear y probar hipótesis para ser usadas en futuras investigaciones.

Capítulo 1. Hipótesis estadísticas

Existen muchas situaciones en las cuales, no es pertinente calcular el valor de un parámetro, sino que es necesario decidir a cerca de una suposición relativa a este parámetro: es verdadera o es falsa. Para tomar esta decisión será necesario probar una hipótesis relativa a dicho parámetro.

Entre los temas que toca la estadística inferencial (o inferencia estadística) se encuentra la prueba de hipótesis (también llamado contraste de hipótesis o test de significación) que es un procedimiento para juzgar las propiedades de una **población** a partir de observaciones tomadas a una parte de ella llamada **muestra** significativa.

La palabra hipótesis proviene del griego *hypothesis* que significa suposición, es el conjunto de datos a partir del cual se intenta demostrar una proposición (Hipótesis en Larousse, 2009).

La prueba de hipótesis no es un procedimiento que intenta explorar diversos panoramas, más bien es una forma de confirmar o refutar la relación entre parámetros.

Existen diferentes métodos cada uno con sus fórmulas y sus características que ayudan a tomar decisiones acerca de las hipótesis, la elección de estas está relacionada a:

- Tamaño de la muestra
- Conocimiento o no de las varianzas
- Proporciones
- Si son una o dos medias
- Una o dos proporciones

En los temas y capítulos presentados a continuación serán presentados de manera que el lector los comprenda y aprenda de manera simple apoyándose con el software Excel.

Excel es un programa informático creado por Microsoft que cuenta con hojas de cálculo, herramientas gráficas, tablas y un lenguaje de programación macro para aplicaciones.

López (Sin Fecha) asegura que la hoja de cálculo es una herramienta poderosa para el calculo de gran cantidad de datos. Es excelente para crear ambientes de aprendizaje que enriquezcan el modelado, comprensión y solución de problemas.

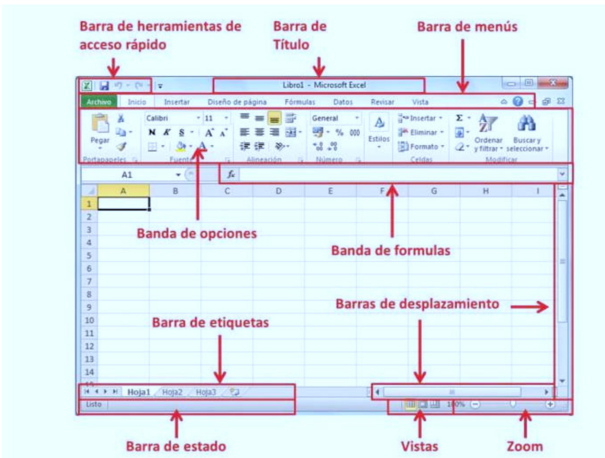


Ilustración 1 Entorno de la hoja de Excel

En la Ilustración 1 se muestra el entorno básico de una hoja de cálculo, esta es una herramienta importante dentro de la probabilidad y la estadística. Está ultima es una asignatura de suma importancia dentro de los programas educativos pues ayuda en la toma de decisiones, precisamente el objetivo de este libro.

Actualmente es imprescindible usar un programa que manipule gran cantidad de datos de forma que se tenga una precisión en los cálculos, cosa que sería tardada, tediosa e inexacta al hacerlo en una calculadora simple o a lápiz y papel.

1. Elección de la muestra

Diariamente en nuestra vida tomamos muestra de infinidad de cosas, por ejemplo, al escuchar el fragmento de una nueva canción para decidir si nos gustó y la seguimos escuchando o la cambiamos; en el supermercado ofrecen prueba de un nuevo queso para decidir si lo compramos o no.

En estadística es muy útil realizar un buen muestreo pues en la mayoría de las investigaciones y experimentaciones sería imposible trabajar con toda la población, así que antes de determinar la mejor manera de realizar un muestreo se presentan las definiciones de estos dos conceptos.

- **Población:** Conjunto de individuos u objetos a estudiar que poseen características comunes. Esta puede ser infinitas (no en el sentido estricto de la palabra infinito, sino que no desde el punto de vista de su manejabilidad) que no es posible analizar o contabilizar el total de elementos ya que se desconoce el primero o el último o finita cuando se conozca el total de la población.
- **Muestra:** Un subconjunto de la población, una parte representativa en la cual se llevará a cabo la medición con la intención de inferir o generalizar los resultados para toda la población.

Estudiar a todos los elementos de una población resulta contraproducente en cuanto a recursos se refiere por eso es importante realizar un buen muestro con el fin de hacer inferencia de toda la población por lo que existen diversos métodos para obtener estas muestras. No hay uno mejor que otro, todo depende de la población y la manera en que esta esté distribuida. Un buen muestreo asegura objetividad en la investigación.

Muestreo Aleatorio Simple

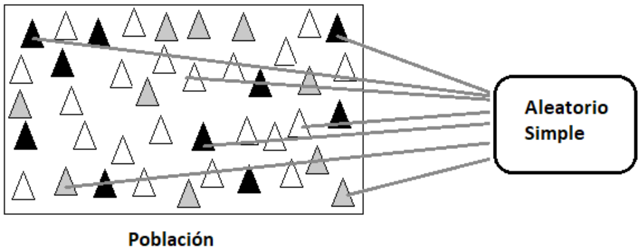


Ilustración 2 Muestreo Aleatorio Simple

Es el más simple pues consiste tomar la cantidad de muestras completamente al azar, es decir, sin orden ni patrón alguno como se muestra en la Ilustración 2.

Por ejemplo: Se tiene un directorio telefónico de los maestros que imparten las materias de Estadística en un centro universitario y se desea que 5 de ellos contesten una encuesta sobre el uso de la biblioteca virtual. Si el total de profesores en el directorio es de 37, se seleccionan 5 completamente al azar (Ilustración 3).

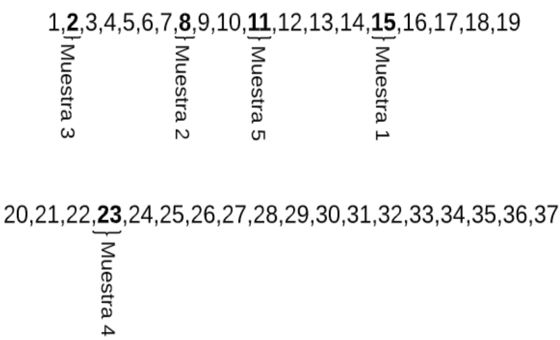
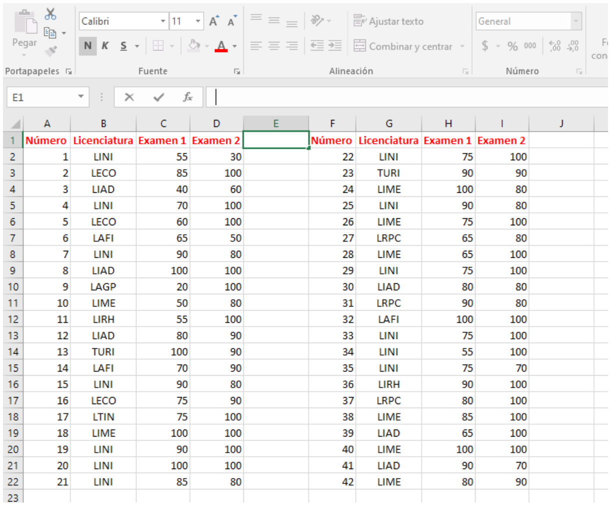


Ilustración 3 Ejemplo Muestreo Aleatorio Simple

A continuación, se muestra en una hoja de Excel, una lista de 42 alumnos de una materia de segundo semestre de licenciatura.



Número	Licenciatura	Examen 1	Examen 2		Número	Licenciatura	Examen 1	Examen 2
1	LINI	55	30		22	LINI	75	100
2	LECO	85	100		23	TURI	90	90
3	LIAD	40	60		24	LIME	100	80
4	LINI	70	100		25	LINI	90	80
5	LECO	60	100		26	LIME	75	100
6	LAFI	65	50		27	LRPC	65	80
7	LINI	90	80		28	LIME	65	100
8	LIAD	100	100		29	LINI	75	100
9	LAGP	20	100		30	LIAD	80	80
10	LIME	50	80		31	LRPC	90	80
11	LIRH	55	100		32	LAFI	100	100
12	LIAD	80	90		33	LINI	75	100
13	TURI	100	90		34	LINI	55	100
14	LAFI	70	90		35	LINI	75	70
15	LINI	90	80		36	LIRH	90	100
16	LECO	75	90		37	LRPC	80	100
17	LTIN	75	100		38	LIME	85	100
18	LIME	100	100		39	LIAD	65	100
19	LINI	90	100		40	LIME	100	100
20	LINI	100	100		41	LIAD	90	70
21	LINI	85	80		42	LIME	80	90

Ilustración 4 Lista de Alumnos

Excel tiene una función que puede ayudar en este proceso. Se presenta el ejemplo con los alumnos de un grupo de segundo semestre. Para hacer uso de las funciones del programa es necesario escribir un igual (=) en la banda de fórmulas, así al comenzar a escribir se irán desplegando las funciones o formulas ya establecidas; en este caso se necesitan las de aleatorio.

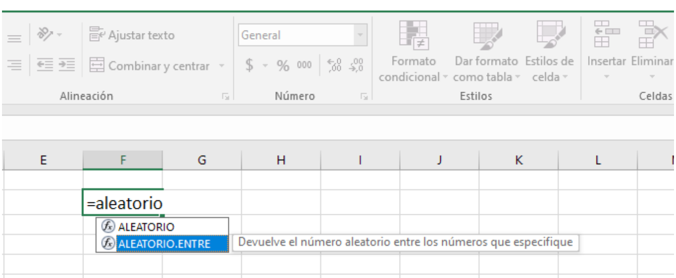


Ilustración 5 Funciones Aleatorio

Las funciones aleatorio muestran números al azar de manera aleatoria son:

- ALEATORIO devuelve un número mayor o igual a 0 y menor que 1
- ALEATORIO.ENTRE devuelve números aleatorios dentro del rango especifico.

Para el muestreo aleatorio simple es necesario usar la función ALEATORIO.ENTRE , esta se utiliza escribiendo dentro del paréntesis la posición donde se encuentra el primer valor (inferior), una coma (,), la posición del valor final (o el superior).

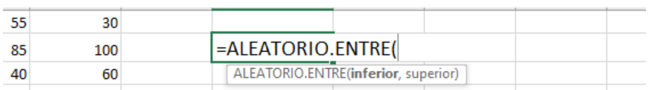


Ilustración 6 ALEATORIO.ENTRE(

Una manera más fácil de especificar el rango del que se pretenden obtener el muestreo aleatorio simple es seleccionando la celda donde se ubica el primer elemento y desplazarse hacia abajo usando las teclas shift y flecha hacia abajo del teclado.

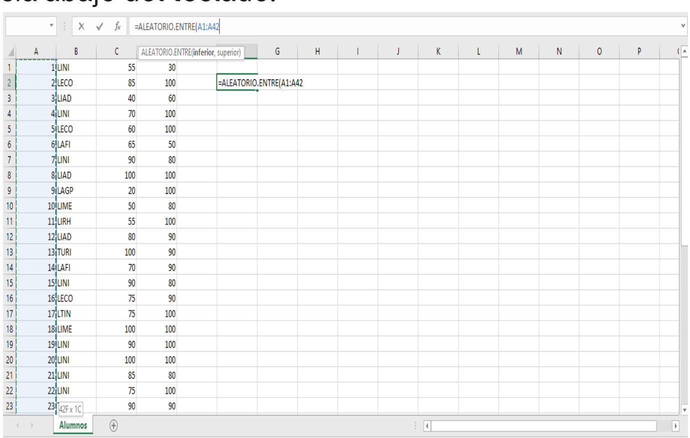
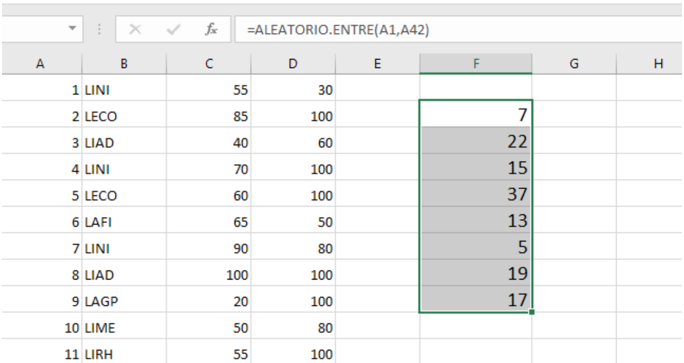


Ilustración 7 Selección de celdas

De esta manera cada que se dé un enter el programa le mostrará un número de manera aleatoria dentro del rango seleccionado.



=ALEATORIO.ENTRE(A1,A42)

A	B	C	D	E	F	G	H
1	LINI	55	30				
2	LECO	85	100		7		
3	LIAD	40	60		22		
4	LINI	70	100		15		
5	LECO	60	100		37		
6	LAFI	65	50		13		
7	LINI	90	80		5		
8	LIAD	100	100		19		
9	LAGP	20	100		17		
10	LIME	50	80				
11	LIRH	55	100				

Ilustración 8 MAS

En el ejemplo de los alumnos, los ocho que son seleccionados en este muestreo son 7,22,15,37,13,5,19,17 (mostrado en la Ilustración 8) obteniendo así una muestra de 8 de los 42 en total.

Muestreo Estratificado

En este tipo de muestreo la población está dividida naturalmente por estratos y cada individuo pertenece solo a uno. Entre los estratos más usados esta sexo o edad. Muchas veces en cuanto a edad, se considera rangos de esta, por ejemplo: 0-19 años, 20 a 39 años, de 40 a 59 años, de 60 a 79 años y de 80 a 99 años. Así cada rango de edad es un estrato.

El método se utiliza separando a la población por estratos, luego se sacan individuos de manera aleatoria de cada estrato hasta completar la muestra.

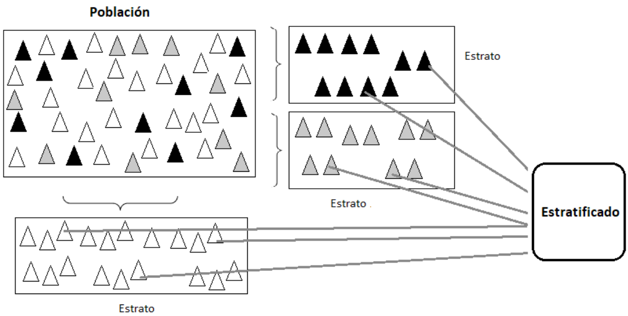


Ilustración 9 Muestreo Estratificado

Muestreo Sistemático

Para poder realizar este tipo de muestreo la población debe estar debidamente ordenada, con ella se crea una lista y de ahí se sigue una secuencia.

Para el mismo ejemplo del directorio telefónico de los profesores de Estadística, el total de profesores es de 37, así que para hacer el muestreo sistemático es necesario realizar la división del total de profesores entre la cantidad de encuestados

Solo consideraremos los enteros ya que se trata de personas. Se elige un número aleatorio de entre el 1 y el 7 para comenzar ahí la secuencia. En este caso se eligió el 3 así que a partir del 3 iremos sumando 7 hasta completar la muestra. Para este ejemplo se obtiene:

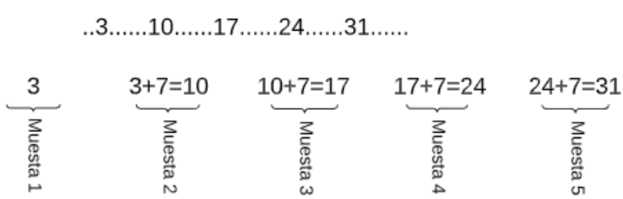


Ilustración 10 Muestro Sistemático

Muestreo por conglomerados

La población se clasifica en unidades ampliar de agrupación física o de espacio. Cada conglomerado es una representación de toda la población, pero ubicada en una región física más cercana.

Este método de muestreo requiere menor tiempo, no se necesita tener la población ordenada ni en lista y se tiene una visión general de la población puesto que es imposible verificar de manera física todos los elementos de la población además de resultar prohibitivo.

Ejercicios:

1. Un distribuidor de motocicletas desea obtener una muestra de 20 opiniones referente al modelo de lujo, entre los 189 últimos clientes que han adquirido tal modelo. Explique de qué manera puede obtener dicha muestra y cuáles son las ventajas de esta elección.
2. Se desea conocer la opinión de los estudiantes de una universidad sobre el uso de tecnologías limpias. ¿Cuál será el método de muestreo adecuado y por qué?
3. ¿Cuál es el método adecuado de muestreo si se desea conocer la satisfacción de los estudiantes de la distribución del estacionamiento en un campus universitario si se tiene conocimiento de que 1500 estudiantes hacen uso de él?

2. Cálculo del tamaño de muestra

Cuando el tamaño de la población total se desconoce, pero se supone que presenta una distribución normal, existe una manera a partir de los valores críticos que se obtienen según el nivel de confianza y el error de estimación deseado. Para obtener estos valores será útil una representación gráfica del concentrado de probabilidad, esta tiene forma acampanada. Está campana es una curva simétrica respecto al eje central (en muchos casos la media aritmética) y representa la probabilidad de ocurrencia. El área bajo la curva (bajo la campana de Gauss) es igual a 1.

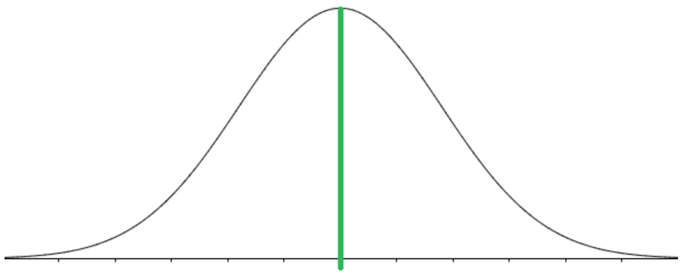


Ilustración 11 Campana de Gauss

Para esto primero se obtendrán los valores críticos. El **valor crítico** z_α es el valor normal estándar de madera que a su derecha el área bajo la curva es .
Una distribución puede tener dos valores críticos $-z\alpha/2$ y $z\alpha/2$ uno negativo del otro. El área que está bajo la curva y entre estos valores se le llama **nivel de confianza**.

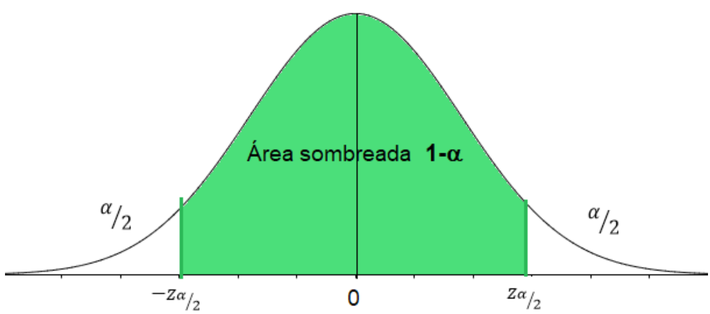


Ilustración 12 Valores Críticos

Ejemplos:

1. Encontrar el valor crítico para 95% de confianza.

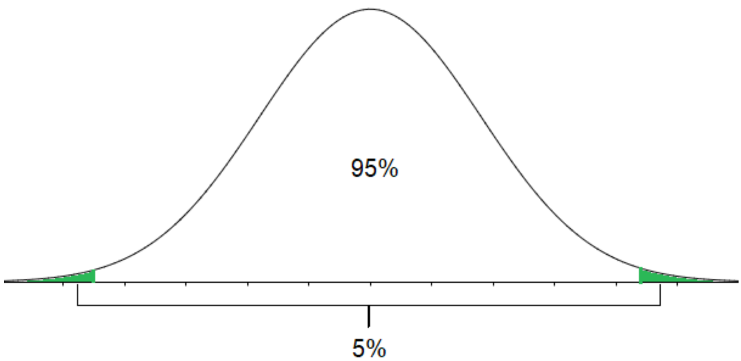


Ilustración 13 Nivel de confianza de 95%

Solución

Comenzaremos reescribiendo el 95% en forma de un número decimal, es sabido que el 100% es el total del área bajo la curva ósea 1, entonces 95% es igual a 0.95. Se determina el valor de α con la fórmula $1-\alpha=0.95$ entonces $\alpha=0.05$ ó 5% y como la región que determina α está dividida en dos será necesario dividir también el valor, así $\alpha/2=0.025$. Este valor se busca en la tabla de distribución normal **Z**. En el Apéndice 1 de este libro encontrará las indicaciones para generarla en Excel, aquí solo se muestra el fragmento necesario para este ejemplo.

El valor 0.025 se ubica en la fila correspondiente al -1.9 y la columna 0.06 (Ilustración 13) y como se trata de áreas se dice que el valor crítico de 95% es 1.96

	A	B	C	D	E	F	G	H	I	J	
1	**z**	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.
2	-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.
3	-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.
4	-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.
5	-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.
6	-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.
7	-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.
8	-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.
9	-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.
10	-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.
11	-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.
12	-2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.
13	-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.

Ilustración 14 Valor z correspondiente a 95%

2. Encontrar el valor crítico para 99% de confianza.

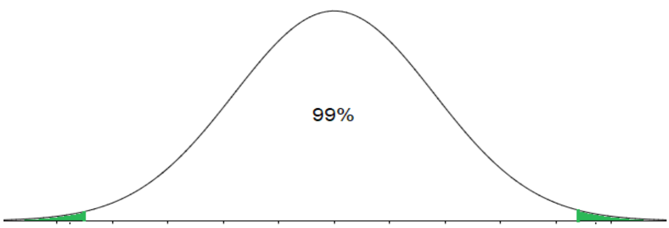


Ilustración 15 Nivel de confianza del 99%

Solución

Reescribir el 99% en forma decimal al ser igual a 0.99. Determinar el valor de α mediante la fórmula 1-α=0.99 así así α=0.01 y α⁄2=0.0050.

Al revisar la tabla de distribución normal Z (Ilustración 16) no se encuentra el valor de 0.0050 pero si dos muy cercanos 0.0051 y 0.0049. El 0.0051 en 2.570 y 0.0049 en 2.580 entonces se suman los números y dividen entre dos.

	A	B	C	D	E	F	G	H	I	J	K	L
1	z	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090	
2	-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010	
3	-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	
4	-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019	
5	-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	
6	-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036	
7	-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048	
8	-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064	
9	-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084	
10	-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110	
11	-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143	
12	-2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183	
13	-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233	
14	-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294	
15	-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367	
16	-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455	

Ilustración 16 Valor z correspondiente a 99%

Por lo que el valor crítico positivo de z para el 99% es 2.575

Ejercicios

1. Encontrar el valor crítico para 90% de confianza.
2. Encontrar el valor crítico para 85% de confianza.
3. Encontrar el valor crítico para 93.4% de confianza.

Ahora que ya es claro cómo obtener los valores críticos, continuamos con el cálculo de tamaño de la muestra.

Cuando se elige el tamaño de la muestra adecuado se tiene un equilibrio entre el muestreo y la precisión. A mayor tamaño de la muestra, mayor precisión, pero en muchos casos resulta con costos más altos y tedioso. Es recomendable también asegurar una muestra heterogénea.

Se encuentran diversas maneras de encontrar el tamaño ideal de una muestra, entre ellas están:

$$n=\frac{N\sigma^2}{(N-1)\frac{E^2}{4}+\sigma^2}$$	(1)	n=tamaño de la muestra N=tamaño de la población σ^2=varianza poblacional (si se desconoce es posible surtirla por S^2) E=error de estimación
$$n=\frac{Np(1-p)}{(N-1)\frac{E^2}{4}+p(1-p)}$$	(2)	n=tamaño de la muestra N=tamaño de la población p=proporción de éxitos de la muestra E=error de estimación
$$n=\frac{Z^2\sigma^2}{E^2}$$	(3)	n=tamaño de la muestra Z=valor de para el nivel de confianza σ^2=varianza poblacional (si se desconoce es posible surtirla por S^2) E=error de estimación
$$n=\frac{Z^2p(1-p)}{E^2}$$	(4)	n=tamaño de la muestra Z=valor de para el nivel de confianza p=proporción de éxitos de la muestra E=error de estimación

Ejemplos:

1. En un centro universitario se pretende contratar un camión que transporte a los alumnos desde una estación del tren al centro universitario. El centro cuenta con 8000 estudiantes. Como sería difícil entrevistar a cada uno de los alumnos, se necesita obtener una muestra que permita que el error sea máximo del 5%. No existe información previa respecto a esta situación. ¿De qué tamaño debe ser la muestra?

$$n=\frac{Np(1-p)}{(N-1)\frac{E^2}{4}+p(1-p)}$$

Solución

Como no existe información previa se asigna el valor de p=0.5, el problema proporciona los datos E=5%=0.05 N=8000 entonces será necesario utilizar la formula (2)

$$n = \frac{Np(1-p)}{(N-1)\frac{E^2}{4} + p(1-p)}$$

$$n = \frac{8000(0.5)(1-0.5)}{(8000-1)\frac{(0.05)^2}{4} + 0.5(1-0.5)}$$

Como no es posible entrevistar a 380.9977 personas se toma el número entero siguiente que son 381 alumnos para saber la proporción de ellos que están a favor de contratar el transporte para tener un máximo de error de estimación del 5%

2. Una empresa de presencia en todo el país desea determinar el ingreso medio de guardias de seguridad. El error al estimar la media debe ser menor que $120 con un nivel de confianza del 95%. El departamento de finanzas estima una desviación estándar de $1000, Determinar el tamaño adecuado de la muestra.

Solución

Valor de $z_{\alpha/2}$=1.96 E=120 pesos σ=1000

Al sustituir los datos en la formula (3) se obtiene

$$n = \frac{Z^2\sigma^2}{E^2}$$

$$n = \frac{(1.96)^2(1000)^2}{(120)^2}$$

$$n = 266.7777$$

Entonces será necesario una muestra de 267 guardias de seguridad para llevar a cabo el estudio.

3. Prueba de una hipótesis estadística

¿Qué es una hipótesis? Una creencia que se tiene sobre un parámetro, este puede ser:

- Media
- Varianza
- Proporción

Una **hipótesis estadística** es una afirmación sobre un modelo de probabilidad. El procedimiento para juzgar la veracidad de la hipótesis se llama prueba de hipótesis. La **prueba de hipótesis** es un método estadístico que se utiliza para tomar decisiones de supuestos formados acerca de una población a partir de una muestra de ella.

El procedimiento consiste en formular dos hipótesis: **Hipótesis nula e Hipótesis alternativa**

- **Hipótesis Nula (H_0)** se lee *H* subcero o *H*. La letra *H* hace referencia a la hipótesis y el subíndice "0" a que no hay diferencia. Es la cual el investigador está dispuesto a sostener como estimable, a menos que haya evidencias estadísticamente significativa en su contra.
- **Hipótesis Alternativa (H_0)** o **(H_A)** se lee *H* subuno o *H* "A". La letra *H* hace referencia a la hipótesis y el subíndice "1 o A" a lo contrario que se afirma en la hipótesis nula. Dicho en otras palabras: es la negación de la hipótesis nula. Está hipótesis solo se aceptará si los datos muestran evidencia de que la hipótesis nula es falsa.

4. Tipo de Error

Un error es un concepto o juicio equivocado, sin embargo, en estadística hace referencia a la variación entre muestras.

Los errores dependen de diversos factores:

- Un mal muestreo. Si en una encuesta realizada sobre la percepción del transporte público en la ciudad se realiza en el estacionamiento de un centro comercial no se tendrán datos reales pues muchos entrevistados no son usuarios asiduos de este tipo de transporte.
- Un error no muestral. Surge al tomar las muestras, específicamente cuando la información no es verídica, los instrumentos de medición están mal calibrados o se obtienen respuestas falsas.
- Un error muestral. Es la variación natural que existe en las muestras de una población. Este se puede medir. El más usado es el error estándar.

Cuando se lleva a cabo una prueba de hipótesis en una investigación, desgraciadamente, no siempre se obtienen las conclusiones correctas y, por lo tanto, se cometen errores. Estos clasifican como Error tipo I y Error tipo II

	H_0 es verdadera	H_0 es falsa
Se rechaza H_0	Error Tipo I	Decisión CORRECTA
No se rechaza H_0	Decisión CORRECTA	Error Tipo II

El error Tipo I consiste en rechazar la hipótesis nula cuando es cierta.

El error Tipo II consiste en NO rechazar la hipótesis nula aun cuando sea falsa.

5. Hipótesis unilaterales y bilaterales

Al inicio de una investigación se tiene noción sobre lo que se quiere investigar y para eso es primordial establecer las hipótesis **antes** de comenzar la investigación y el análisis de los datos.

Para realizar el estudio y análisis, el investigador debe seguir los pasos:

1. Formular la Hipótesis nula H_0.
2. Establecer la Hipótesis alternativa H_1.
3. Determinar el estadístico de la prueba adecuado a los datos conocidos
4. Delimitar la Región de rechazo

La hipótesis nula es el enunciado relativo al valor de un parámetro poblacional que se formula con el fin de probar evidencia numérica a su favor. En el planteamiento de la hipótesis al escribirlo con símbolos se tiene

$$H_0: \; = \, , \, \leq \, , \, \geq$$

En resumen, la hipótesis nula es la que se contrasta y solo los datos numéricos pueden refutarla.

La hipótesis alternativa es el enunciado que se acepta si los datos muestran suficiente evidencia para rechazar la hipótesis nula.

$$H_1: \; \neq \, , \, > \, , \, <$$

Así que está niega y para hacer aceptada los datos pueden mostrar evidencia a favor.

Ejemplos

Plantear la hipótesis correspondiente en cada situación

1. El salario medio de un cajero en la recaudadora estatal en 2019 fue de $19,044.00. Los empleados del estado de Jalisco sospechan que están por debajo.

Solución

En la afirmación se encuentra una palabra clave: *debajo* esto hace referencia a un símbolo de menor en la hipótesis alternativa $H_1: \mu < \$19,044.00$ de manera que la zona de decisión queda de la siguiente manera

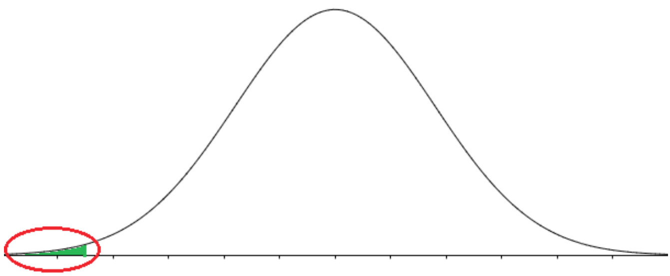


Ilustración 17 Prueba de una cola izquierda

Así las hipótesis quedan $H_0 >: \mu \$19,044.00$

$H_1: \mu < \$19,044.00$

2. Una muestra de cierta marca de tequila mostró que en promedio contenían 39° de concentración de alcohol, ¿indica esto que la concentración promedio siempre es mayor a 35°?

Solución

En la pregunta que se encuentra en el ejemplo hay una palabra que es la clave para la construcción de la hipótesis alternativa: *mayor* de manera que se usará el símbolo de > quedando de la siguiente manera

$$H_0: \mu \leq 35°$$
$$H_1: \mu > 35°$$

Y la representación gráfica es

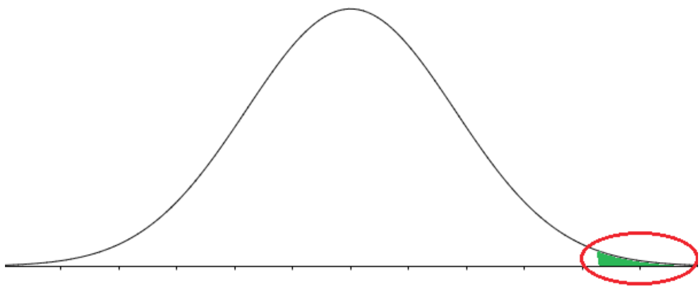


Ilustración 18 Prueba de una cola derecha

3. Se cree que la edad promedio de los estudiantes inscritos a la Universidad de Guadalajara es de 19.8. Con base en una muestra tomada el calendario pasado ¿habrá cambiado el promedio de edad de los estudiantes?

Solución

En la pregunta que se presenta en el ejercicio la palabra clave para determinar el tipo de prueba que se necesita está en *habrá cambiado*, esto indica que en la hipótesis alternativa es necesario el símbolo de diferente por lo que la zona de rechazo está dividida en dos siendo una prueba de dos colas o bilateral con hipótesis de la forma

$$H_0: \mu = 19.8$$
$$H_1: \mu \neq 19.8$$

La siguiente gráfica queda de la siguiente manera

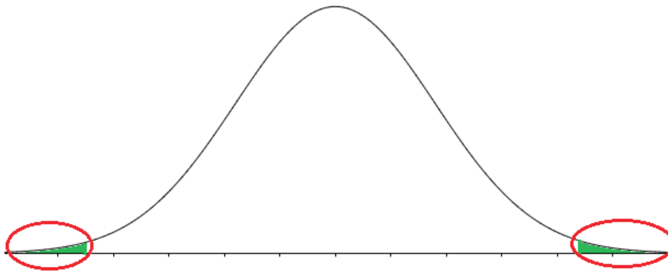


Ilustración 19 Prueba de dos colas

Ejercicios

Determinar las hipótesis y grafica correspondiente en cada ejercicio.

1. La media de un productor de maíz fue de 7 toneladas por hectárea el año pasado. Este año el productor ensayó un nuevo tipo de fertilizante en una muestra aleatoria de parcelas, está interesado en saber si la producción ha incrementado.
2. Se rumora que en el estacionamiento de la Universidad que no hay suficientes lugares para los autos de los alumnos. Para decidir si se debe ampliar el estacionamiento el encargado de servicios tomará una muestra aleatoria y si la media del tiempo de espera para encontrar lugar es menor a 12 minutos no se realizará la ampliación.
3. El encargado de la biblioteca de una preparatoria manifiesta que el número de lectores diarios es de 350, se instala un sistema electrónico de conteo para verificar si la información otorgada es correcta.

Capítulo 2. Prueba de hipótesis sobre una media

Una prueba de hipótesis se utiliza en la toma de decisiones a partir de muestras. Cuando se trata de una sola muestra se separan en muestras grandes ($n \geq 30$) o muestras pequeñas ($n<30$) en ambos casos resulta una prueba nada difícil de realizar haciendo referencia a una hipótesis para media poblacional.

En el capítulo anterior se explicó cómo obtener los valores críticos para $Z_{\alpha/2}$ sin embargo en las pruebas de hipótesis los valores cambian dependiendo de las colas de la prueba. A continuación, se presenta una tabla con los valores más usuales.

α	Nivel de confianza	$-Z_{\alpha}$	Z_{α}	$-Z_{\alpha/2}$	$Z_{\alpha/2}$
0.01	99%	−2.325	2.325	−2.575	2.575
0.02	98%	−2.055	2.055	−2.325	2.325
0.03	97%	−1.885	1.885	−2.170	2.170
0.04	96%	−1.755	1.755	−2.055	2.055
0.05	95%	−1.645	1.645	−1.96	1.96
0.06	94%	−1.555	1.555	−1.885	1.885
0.07	93%	−1.475	1.475	−1.815	1.815
0.08	92%	−1.405	1.405	−1.755	1.755
0.09	91%	−1.345	1.345	−1.695	1.695
0.10	90%	−1.285	1.285	−1.645	1.645

1. Con muestras grandes ($n \geq 30$)

Para esta prueba además de considerar el caso en que la muestra de la población es igual o mayor a 30 se supone que la desviación estándar poblacional (σ) conocida.

Como toda prueba de hipótesis primero se debe establecer las hipótesis nula e hipótesis alternativa

$$H_0: \mu = X$$
$$H_1: \mu \neq X$$

$$H_0: \mu \geq X$$
$$H_1: \mu < X$$

$$H_0: \mu \leq X$$
$$H_1: \mu > X$$

Después determinar el valor de significancia (α) y seleccionar el estadístico de prueba $z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

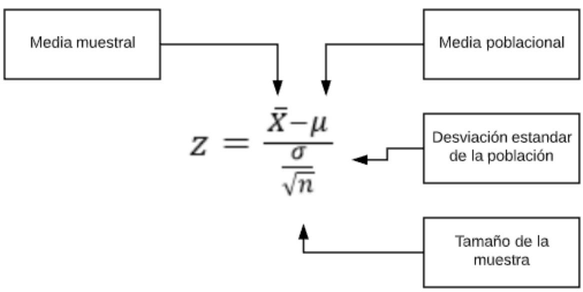


Ilustración 21 Estadístico cuando se conoce la desviación estándar poblacional

Ejemplos:

1. En un estudio realizado por la Secretaria de Educación Pública indica que los niños de educación primaria ven un promedio de 23.5 horas de televisión a la semana, una asociación de padres de familias cree que están equivocados así que realizan un estudio en 60 niños que arroja un promedio de 25.5 horas con una desviación estándar de 6.1. ¿Deberá la Secretaria rechazar lo que dice la investigación si el experimento se realiza con ?

Solución:

Los datos obtenidos en el problema son:

n=60 σ=6.1 X=25.5 μ=23.5

Las hipótesis quedan $H_0 : \mu=23.5$
$H_1 : \mu\neq 23.5$)

Al sustituir los valores en el estadístico $z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$z = \frac{25.5 - 23.5}{\frac{6.1}{\sqrt{60}}}$$

$$z = 2.5396$$

Ahora se realiza el comparativo en la campana correspondiente

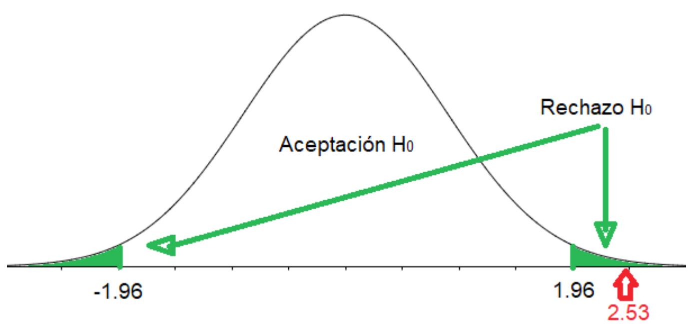


Ilustración 22 Decisión respecto a las hipótesis

Debido a que el valor calculado de Z (2.53 al truncar el valor) en el estadístico se ubica en la zona de rechazo de H_0 delimitada por los valores críticos Z de tablas ±1.96 se afirma que existe evidencia estadísticamente significativa para *rechazar la hipótesis nula* afirmarmando que los niños ven un promedio diferente a 23.5 horas de televisión a la semana (de hecho es más pues el valor se ubicó en la cola derecha).

$$\cancel{H_0: \mu = 23.5}$$
$$H_1: \mu \neq 23.5$$

Por lo que la Secretaria de Educación Pública no debería rechazar la investigación.

1. Una muestra aleatoria de 100 muertes registradas el mes pasado en el estado de Jalisco muestra que el promedio de vida fue de 75.98 años con una desviación estándar poblacional de 8.72 años. ¿Esto indica que la esperanza de vida en el estado es igual que (edad que indica el INEGI)? Realizar una prueba de hipótesis con un α=0.1

Solución:

Los datos obtenidos en el problema son:

n=100 σ=8.72 X=75.98 μ=73.58 α=0.1

X=valor que se obtiene de la muestra

μ=valor supuesto por el INEGI

Las hipótesis quedan H_0:μ=73.58
H_1:μ≠73.58

Al sustituir los valores en el estadístico $z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$z = \frac{75.98 - 73.58}{\frac{8.72}{\sqrt{100}}}$$

$$z = 2.7523$$

Para tomar la decisión de aceptar o rechazar la hipótesis nula se realiza el comparativo en una campana de Gauss

Otra manera de obtener el valor crítico de Z ahora con Excel es utilizando la función

=DISTR.NORM.ESTAND.INV(*nivel de confianza* + $\alpha/2$)

Lo primero es fijar una celda con el valor de significancia α, en este caso queda en la celda B1 con un valor de 0.1 tal y como lo menciona el problema del ejemplo. Después se calcula el nivel de confianza al hacer la simple resta del 100% menos el nivel de significancia, resultado que se ubica en la celda B2

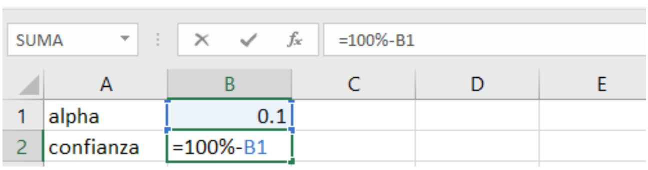


Ilustración 23 Nivel de confianza

Para obtener el valor crítico de Z se utiliza la función de distribución normal estándar inversa que tiene como argumento la probabilidad en este caso es el nivel de confianza más el valor de $\alpha/2$.

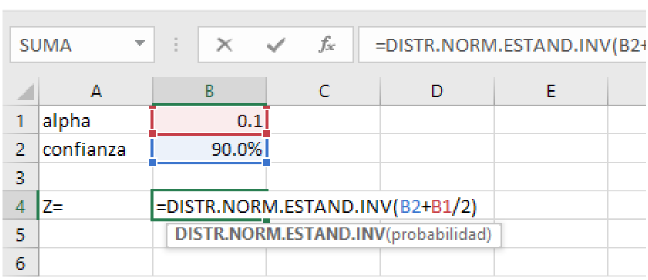
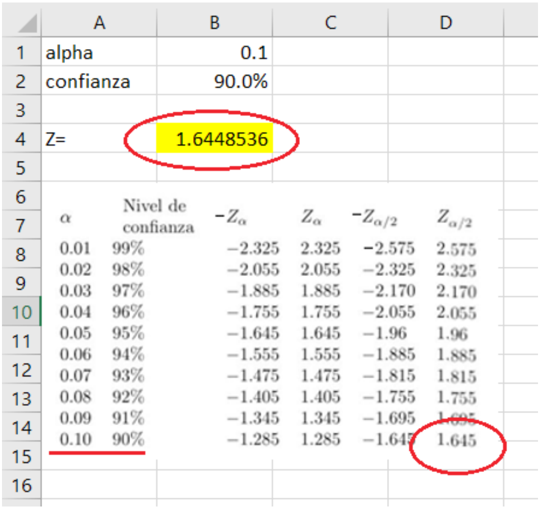


Ilustración 24 Valor crítico para Z

De esta manera se obtiene el valor de z=1.6448 que al redondearlo a tres decimales es z=1.645 mismo que se obtiene al utilizar el procedimiento mostrado en el capítulo anterior cuando se obtuvo el tamaño de la muestra.



	A	B
1	alpha	0.1
2	confianza	90.0%
4	Z=	1.6448536

α	Nivel de confianza	$-Z_\alpha$	Z_α	$-Z_{\alpha/2}$	$Z_{\alpha/2}$
0.01	99%	−2.325	2.325	−2.575	2.575
0.02	98%	−2.055	2.055	−2.325	2.325
0.03	97%	−1.885	1.885	−2.170	2.170
0.04	96%	−1.755	1.755	−2.055	2.055
0.05	95%	−1.645	1.645	−1.96	1.96
0.06	94%	−1.555	1.555	−1.885	1.885
0.07	93%	−1.475	1.475	−1.815	1.815
0.08	92%	−1.405	1.405	−1.755	1.755
0.09	91%	−1.345	1.345	−1.695	1.695
0.10	90%	−1.285	1.285	−1.645	1.645

Ilustración 25 Comparativo del valor crítico de Z

Ubicando los valores de Z calculada y Z de tablas calculada con la distribución normal, queda de la siguiente manera

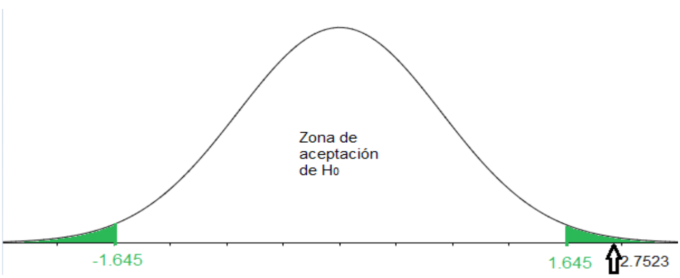


Ilustración 26 Regiones de Hipótesis

El valor calculado se encuentra en la región de rechazo de H_0.Así H_1:$\mu \neq 73.58$ es la hipótesis aceptada.

Existe evidencia estadísticamente significativa para afirmar que la esperanza de vida en el estado de Jalisco es diferente a años.

Ejercicio:

1. El director de la biblioteca del centro universitario considera que el número de estudiantes que asisten solo a consulta de libros son 250 diarios. Para verificar esta información se toma una muestra de 30 días de los cuales se obtiene un promedio de 241 estudiantes con una desviación de 8. Realizar una prueba de hipótesis para la veracidad de lo dicho por el director considerando un nivel de significancia de 0.02

2. Con muestras pequeñas (n<30)

Al igual que en la sección anterior es factible trabajar con una prueba de hipótesis en la que se conoce la desviación estándar poblacional (σ), sin embargo en la práctica común con muestras (s) pequeñas se realiza utilizando la desviación estándar muestral , de modo que la fórmula para calcular el valor estadístico cambia así:

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

La distribución fue desarrollada por William Sealy Gosset bajo el seudónimo Student debido a la necesidad de estimar medias de poblaciones cuando se tienen muestras pequeñas, debido a esto se dice que es efectiva en n<30 pues a medida que crece los valores se van aproximando a los de la distribución normal Z.

La distribución t de Student es una distribución continua con una gráfica acampanada, simétrica, depende de los grados de libertad. Estos grados de libertad son $gl=n-1$, esto es, el número de muestras menos uno.

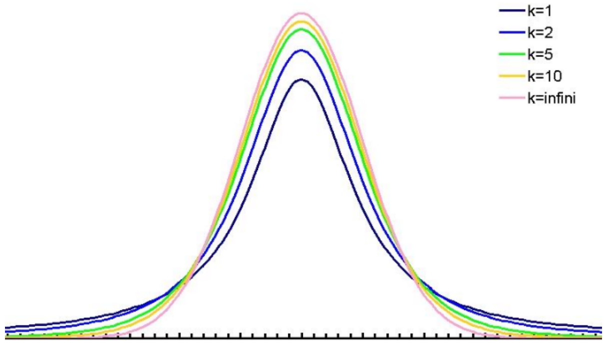


Ilustración 27 Familia de campanas de la distribución t de Student

Existe una gran familia de distribuciones t, cada grado de libertad genera una distribución nueva.

Ejemplos

1. Una máquina expendedora de llenado de garrafones de agua está programado para llenarlos con 19 litros, sin embargo, algunos clientes dicen que despacha menos. El dueño decide realizar un muestreo aleatorio a 6 garrafones en los que obtiene los siguientes resultados

18.68	18.90	19.31
19.08	18.61	19.40

Realizar una prueba de hipótesis con un nivel de significancia de para verificar si es necesario revisar la máquina de llenado.

Solución

Se comienza planteando las hipótesis según lo que dice el problema. En este caso, los clientes dicen que despacha *menos* de lo que debería, así las hipótesis son:

$$H_0: \mu \leq 19$$
$$H_1: \mu > 19$$

El estadístico a utilizar es

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

Para resolver este problema será necesario encontrar la media de la muestra y la desviación estándar muestral, así que se mostrará cómo obtenerlas con fórmulas preestablecidas en Excel.

La media se calcula la función =MEDIA.GEOM (la cual calcula la media o promedio de las celdas seleccionadas.

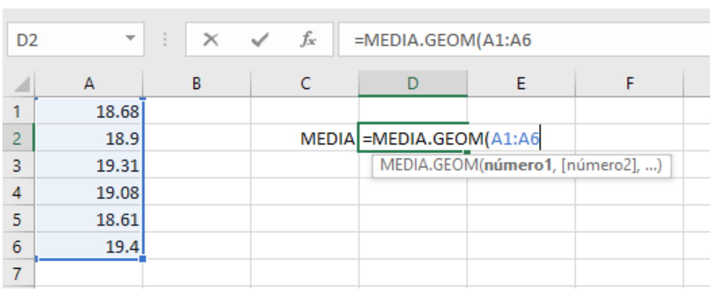
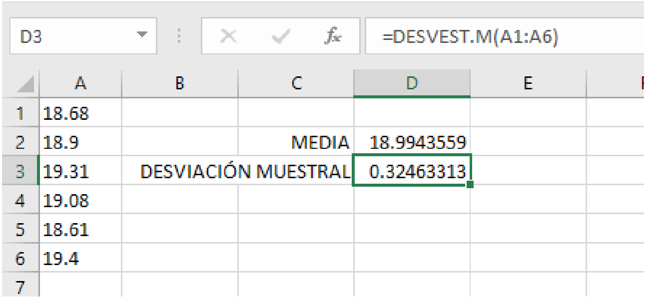


Ilustración 28 Calculo de la Media

De forma similar se obtiene la desviación estándar muestral, que como se mencionó al inicio de la sección es la conveniente al tratarse de muestras pequeñas. La fórmula correspondiente es =DESVEST.M(



D3 | =DESVEST.M(A1:A6)

	A	B	C	D	E
1	18.68				
2	18.9		MEDIA	18.9943559	
3	19.31	DESVIACIÓN MUESTRAL		0.32463313	
4	19.08				
5	18.61				
6	19.4				
7					

Ilustración 29 Desviación estándar muestral

Ahora ya tenemos los datos necesarios para nuestro estadístico.

X=18.99 al truncar el número a dos decimales

s=0.32 desviación estándar muestral

μ=19 es lo que queremos probar

n=6 cantidad de garrafones muestreados

$$t = \frac{18.99 - 19}{\frac{0.32}{\sqrt{6}}}$$

$$t = -0.0765$$

Para tomar la decisión de aceptar o rechazar la hipótesis nula es necesario encontrar el valor crítico de t para esto usamos una función de Excel en la que es necesario conocer los grados de libertad (n-1), el nivel de confianza establecido en el problema y si la prueba es unilateral o bilateral.

En este caso al observar la hipótesis alternativa H_1:μ>19 es visible que la prueba es unilateral, de echo es de una cola derecha por lo que será necesario utilizar la formula =DISTR.T.INV(2*

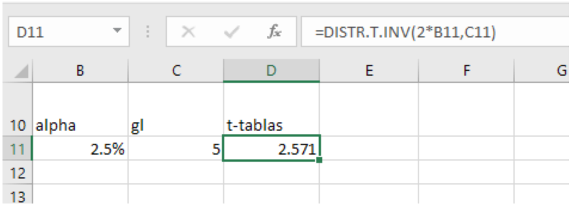


Ilustración 30 Valor crítico de t

Como el problema dice que el nivel de confianza es de 97.5% la probabilidad α es 0.025 y los grados de libertad son 5 pues 6-1=5 de esta forma el valor crítico de t es 2.571

En la siguiente grafica se ilustra a la derecha una región sombreada que representa el área de rechazo de la hipótesis nula. Al ubicar el valor calculado de t está en la región de aceptación de la hipótesis nula H_0:$\mu \leq 19$

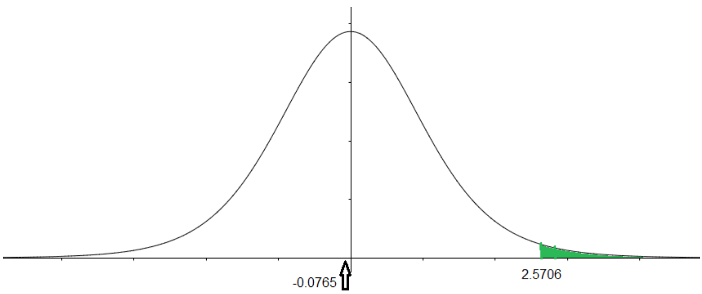


Ilustración 31 Regiones delimitantes de la hipótesis

Con estos resultados se concluye que existe evidencia estadísticamente significativa de que la maquina expende 19 litros o menos por lo según el dueño de la máquina expendedora, se deberá revisar para calibrar nuevamente.

1. El supervisor de una recaudadora gubernamental asegura que el tiempo promedio de espera en la fila es de 10 minutos. Se tomaron al azar una muestra de 16 personas con las que se obtuvo un promedio de 12.5 minutos y una desviación estándar de 4.1 minutos. Realice una prueba de hipótesis con el 90% de confianza.

Solución

Se comienza planteando las hipótesis según lo que plantea el problema

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

Ahora se extraen los datos dados en el problema

$X=12.5 \quad \mu=10 \quad s=4.1 \quad n=16 \quad \alpha=0.1$

El estadístico a usar

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

Al sustituir los valores en el estadístico se obtiene

$$t = \frac{12.5 - 10}{\frac{4.1}{\sqrt{16}}}$$

$t=2.44$ al redondear a dos decimales.

Para tomar una decisión es necesario hacer el comparativo con una gráfica de t de student, para este caso la hipótesis alternativa tiene el signo de ≠ por lo que la prueba es de dos colas. El valor critico se obtendrá con la misma función solo que ahora la probabilidad se deja sencilla, esto es, NO se multiplica por 2 (algo que si se hizo en la prueba de una cola.

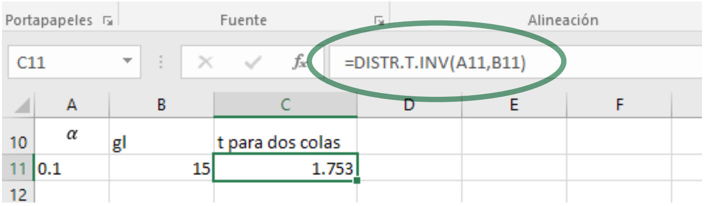


Ilustración 32 Distribución t para dos colas

Este valor crítico se coloca en los dos extremos de la campana para marcar las regiones de rechazo de la hipótesis nula, el izquierdo con signo negativo y la cola derecha en positivo.

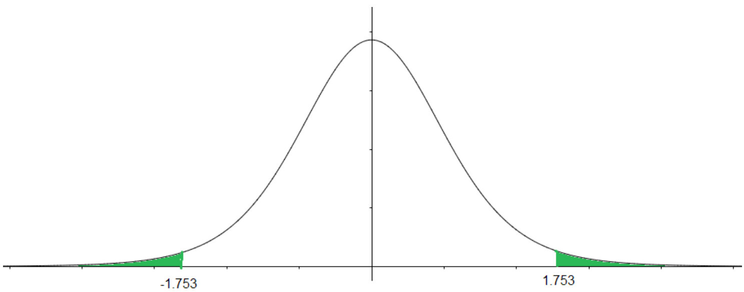


Ilustración 33 Región de rechazo de hipótesis nula

El valor de t calculado es 2.44 y se ubicará en la cola derecha por lo que se rechaza la hipótesis nula, de esta manera: existe evidencia estadísticamente significativa para afirmar que el tiempo medio en la fila en la recaudadora es diferente de 10 minutos.

Ejercicios:

1. La operadora del estacionamiento del aeropuerto internacional de Guadalajara estima que el tiempo medio que le toma a un cliente encontrar lugar para estacionarse es de 4 minutos al ver que esto molesta a muchos usuarios se realiza un muestreo de un auto en cada

hora del día del cual se obtienen los siguientes resultados expresados en minutos: X=3.7 s=0.7 Realizar una prueba de hipótesis con un 99% de confianza.

2. Sean las hipótesis $H_0: \mu \leq 100$

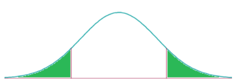
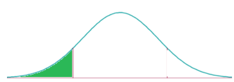
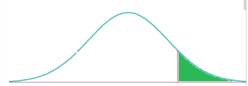
   $H_1: \mu > 100$

   a. Encuentre la media y desviación estándar para los valores
   b. 81 92 85 119 100 87 87 108 95
   c. ¿Cuál es el valor del estadístico de prueba?
   d. Con un nivel de significancia de 0.05, ¿Se puede aceptar la hipótesis nula?

Capítulo 3. Prueba de hipótesis para igualdad de dos medias

En ocasiones es preciso realizar comparaciones entre dos poblaciones por lo que es necesario realizar dos muestreos, es cuando se requiere utilizar este tipo de pruebas. La metodología utilizada es muy parecida a la de hipótesis para una muestra solo que ahora se parte del supuesto que las medias poblacionales son iguales.

Las hipótesis se plantean de la misma forma que en las anteriores, unilateral derecha, unilateral izquierda o bilateral según el signo de la hipótesis nula.

Bilateral	Unilateral izquierda	Unilateral derecha
$H_0: \mu_1=\mu_2$ $H_1: \mu_1\neq\mu_2$	$H_0: \mu_1-\mu_2\geq 0$ $H_1: \mu_1-\mu_2<0$	$H_0: \mu_1-\mu_2\leq 0$ $H_1: \mu_1-\mu_2>0$
 $-\alpha/2$ $\quad \alpha/2$	 $-\alpha$	 α

1. Con varianzas conocidas y muestras grandes ($n_1+n_2\geq 60$)

Supóngase que se toma una muestra aleatoria de una primera población de tamaño n_1 y otra muestra aleatoria de tamaño n_2, también se supone que estas dos muestras provienen de una población normalmente distribuidas de manera independiente con medias μ^1, μ^2 y varianzas σ_1^2, σ_2^2 que provienen de poblaciones independientes.

Se tiene entonces el siguiente estadístico comparando con valores de prueba Z

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$$

Para facilitar el trabajo del estudiante se presentan nuevamente los valores críticos de Z, sin embargo, solo aparecen los más habituales: 90% usado cuestiones de carácter político, 95% usado en investigaciones educativas y generales, 99% usado en investigaciones relacionadas con ciencias de la salud.

Confianza	Significancia	2 colas	Cola izquierda	Cola derecha
90%	0.10	±1.65	-1.28	1.28
95%	0.05	±1.96	-1.65	1.65
99%	0.01	±2.58	-2.33	2.33

Ejemplos:

1. El centro universitario desea comparar el conocimiento en el idioma ingles de los alumnos que provienen de bachilleratos de la zona metropolitana y de los bachilleratos foráneos. Se recaba una muestra aleatoria de 40 alumnos de cada grupo a los que se les aplico un examen. En la siguiente tabla se muestran los resultados obtenidos.

Escuelas foráneas	Escuelas de zona metropolitana
n_1=40 X^1=82.5 σ_1=4.3	n_2=40 X^2=85.6 σ_2=3.9

Determine si existe evidencia significativa de $\mu1$ respecto a $\mu2$ utilizando α=0.05

Solución

Se comienza planteando las hipótesis respecto a lo que se desea conocer

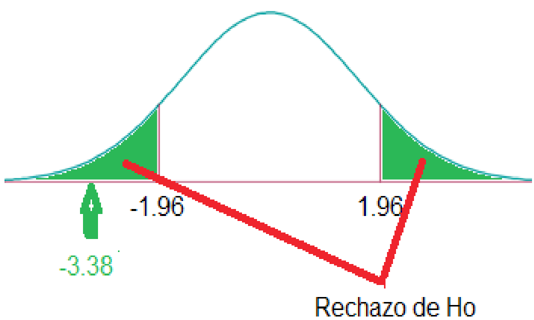
$$H_0{:}\mu_1=\mu_2$$
$$H_1{:}\mu_1\neq\mu_2$$

Así que la prueba será de dos colas. El estadístico a usar será el de la diferencia entre dos medias muéstrales. Al suponer que $H_0{:}\mu_1=\mu_2$ el valor de μ_1-μ_2=0 de esta forma al sustituir los datos en la formula queda:

$$Z=\frac{(82.5-85.6)-(0)}{\sqrt{\frac{(4.3)^2}{40}+\frac{(3.9)^2}{40}}}$$

$$Z=-3.38$$

Al ser una prueba de dos colas con α=0.05 los valores críticos para Z según la tabla corresponde a $Z_{\alpha/2}=\pm1.96$ Como -3.38<-1.96, se rechaza la hipótesis nula.



2. El departamento de recursos humanos de una empresa recibe una queja de que los empleados hombres ganan más que las mujeres en los mismos puestos. La empresa asegura que la queja es falsa así que se decide realizar un muestreo aleatorio al sueldo quincenal de 100

mujeres y a 75 hombres cuyos resultados aparecen en la siguiente tabla

Hombres	Mujeres
n_1=75 X^1=191.20 σ_1=17	n_2=100 X^2=189.50 σ_2=14

Los analistas necesitan realizar una prueba de hipótesis para analizar la diferencia salarial con un nivel de significancia de 0.05 ¿Deberán poner una denuncia en la secretaria de trabajo y previsión social sobre la brecha salarial?

Solución

Se comienza planteando las hipótesis respecto a lo que se desea conocer

$$H_0{:}\mu_1\text{-}\mu_2\leq 0$$
$$H_1{:}\mu_1>\mu_2$$

La hipótesis nula plantea que el salario de los hombres es mayor que el de las mujeres por lo que se utilizara una prueba unilateral derecha. Recuerde que en la hipótesis nula se supone que las medias son iguales por lo que nuevamente el lado derecho del numerador en la formula se sustituye por cero.

Al sustituir los datos se obtiene

$$Z=\frac{(191.20-189.50)-(0)}{\sqrt{\frac{(17)^2}{75}+\frac{(14)^2}{100}}}$$
$$Z=0.71$$

La regla de decisión a establecer busca comparar el valor estadístico de prueba de Z calculado y comprarlo con el valor de tablas, esto es, $0.71<1.645$ por lo que se acepta la hipótesis nula, esto quiere decir que los salarios son estadísticamente equivalentes así que no es necesario realizar la denuncia.

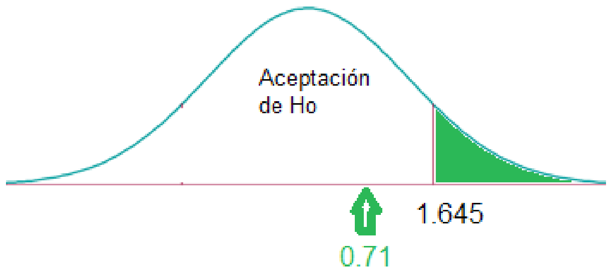


Ilustración 34 Decisión respecto a la hipótesis nula

Ejercicios

1. Se busca conocer los efectos publicitarios de una campaña contra el uso del tabaco en dos grupos de personas para ver si hay disminución en su consumo. Se seleccionaron al azar 90 de ellos que pertenecen al grupo de menores de 40 años y otros 90 que son mayores de 40 años. Realizar una prueba de hipótesis con un 90% de confianza si se sabe que los resultados son los siguientes:

Característica	Media disminución de cigarros consumidos diarios	Desviación estándar
Menores de 40	7.95	1.97
Mayores de 40	6.44	2.15

2. Se realizó un estudio para comparar los años promedios de servicio entre hombres y mujeres que se jubilaron el año pasado de una empresa de cobertura

nacional. Con un nivel de significancia de se puede concluir que los hombres que se retiraron trabajan más años que las mujeres?

Característica	Desviación estándar
X=30.4	X=27.1
σ=3.3	σ=2.6
n=40	n=45

2. Con varianzas desconocidas y muestras pequeñas ($n_1+n_2 \leq 60$)

En la sección anterior se trabajó con un estadístico usando la distribución normal Z y dos muestras que sumaban más de 60, ahora se presenta el estadístico para dos muestras que suman a lo más 60, sin embargo, el estadístico requiere modificaciones con la finalidad de hacer más preciso el estadístico. Un supuesto general para este método es suponer que las desviaciones estándar de la población son estadísticamente iguales, al tiempo que las muestras siguen una distribución normal.

A pesar que se parte del supuesto que las muestras provienen de una población normal, este estadístico se utiliza la distribución *t de student.* La significancia es de la misma manera que para una muestra, los grados de libertad será n_1+n_2-2

El estadístico se mostrará en dos partes, una es la estimación conjunta (formará parte del denominador de todo el estadístico) y el estadístico *t calculada.* La estimación conjunta se calcula como:

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

Donde s_1^2 es la varianza (o desviación estándar al cuadrado) de la muestra 1, s_2^2 es la varianza de la muestra 2.

La prueba de dos medias con σ desconocida y muestras pequeñas se calcula con el estadístico:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2{}_c\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

El criterio de decisión es similar al del estadístico de Z pues al plantear la hipótesis nula se puede obtener una criterio unilateral derecho, unilateral izquierdo o bilateral tomando en cuenta que cuando sea bilateral se debe considerar el valor de $\pm\alpha/2$

Ejemplos:

1. Se desea probar la duración de dos pequeños jabones antibacteriales en los baños de la escuela. A continuación, se muestran los resultados del tiempo en horas que duraron en uso estos jabones.

Jabón 1	4.34	2.70	3.79	3.07	5.84
Jabón 2	4.60	3.23	4.75	4.71	

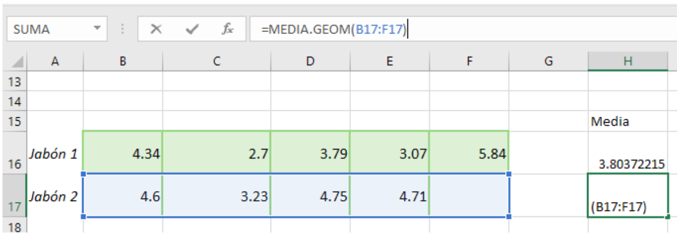
¿Se puede decir que la duración de los dos jabones es equivalente al usar un nivel de significancia de $\alpha=0.05$?

Solución

Lo primero para resolver cualquier problema de prueba de hipótesis es plantear las hipótesis nula y alternativa.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Ahora se realizan los cálculos utilizando las funciones básicas de Excel. La media de los datos con la función =MEDIA.GEOM(



SUMA | =MEDIA.GEOM(B17:F17)

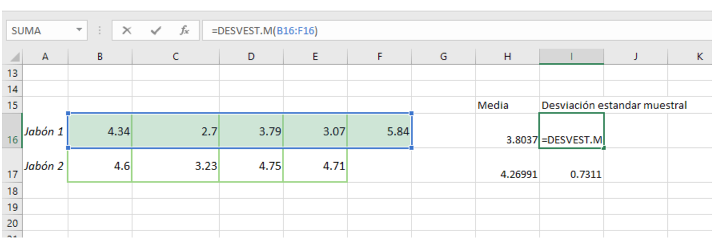
	A	B	C	D	E	F	G	H
13								
14								
15								Media
16	Jabón 1	4.34	2.7	3.79	3.07	5.84		3.80372215
17	Jabón 2	4.6	3.23	4.75	4.71			(B17:F17)
18								

Ilustración 35 Media obtenida con los datos dados

Es necesario recordar que la desviación estándar muestral y desviación estándar muestral son estadísticos muy parecidos sin embargo no dan el mismo resultado.

Desviación estándar poblacional	Desviación estándar muestral
$\sigma = \sqrt{\dfrac{\Sigma(X - \bar{X})^2}{N}}$	$S = \sqrt{\dfrac{\Sigma(X - \bar{X})^2}{n - 1}}$

La desviación estándar de la muestra se calcula usando la función =DESVEST.M(y seleccionando el rengo de los valores de los cuales se quiera calcular la desviación estándar.



SUMA | =DESVEST.M(B16:F16)

	A	B	C	D	E	F	G	H	I	J	K
13											
14											
15								Media	Desviación estandar muestral		
16	Jabón 1	4.34	2.7	3.79	3.07	5.84		3.8037	=DESVEST.M		
17	Jabón 2	4.6	3.23	4.75	4.71			4.26991	0.7311		
18											
19											
20											

Ilustración 36 Desviación estándar muestral

¿También se usará la función =CONTAR(para encontrar el número de muestras de cada tipo.

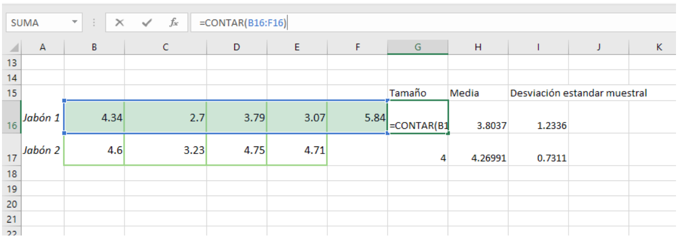


Ilustración 37 Contar número de datos

Al sustituir los valores en la fórmula de estimación conjunta

$$S^2{}_c = \frac{(5-1)(1.2336)^2 + (4-1)(0.7311)^2}{5+4-2}$$

$$S^2{}_c = 1.10$$

Ahora se calcula el estadístico tomando en cuenta que se parte del supuesto que las medias poblacionales son iguales.

$$t = \frac{(3.80 - 4.27) - (0)}{\sqrt{1.10\left(\frac{1}{5} + \frac{1}{4}\right)}}$$

$$t = -0.67$$

Ahora se realiza el comparativo con los valores críticos que se obtiene a partir de la tabla de la distribución *t de Student* con una prueba de dos colas. Tal y como se hizo en la prueba de una muestra, se utiliza la función =DISTR.T.INV(

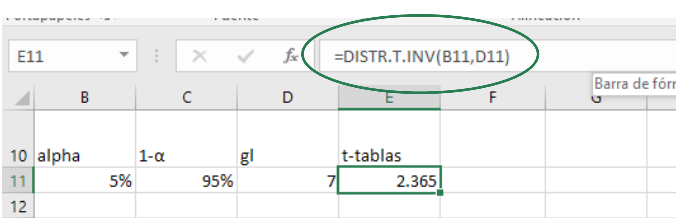


Ilustración 38 Valor crítico con distribución t para dos colas

Este valor es el que delimita las dos colas que marcan la zona de rechazo de H_0 sin embargo, en este caso el valor calculado de t se encuentra en la zona de aceptación de H_0.

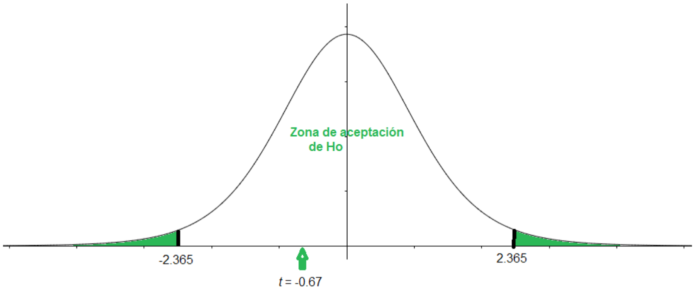


Ilustración 39 Decisión sobre Ho

Al aceptar la hipótesis nula, se puede afirmar que los dos jabones son estadísticamente equivalentes.

2. De dos poblaciones independientes se toman muestras aleatorias obteniendo los siguientes datos:

Muestra 1	Muestra 2
$n_1=15$ $X^1=641$ $\sigma_1=6.3$	$n_2=25$ $X^2=629$ $\sigma_2=7.6$

Tomando en cuenta las hipótesis

$$H_0: \mu_1 \leq \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Poner a prueba con $\alpha=0.05$ la hipótesis.

Solución

Se comienza sustituyendo los datos en la estimación conjunta

$$S^2{}_c = \frac{(15-1)(6.3)^2 + (25-1)(7.6)^2}{15+25-2}$$

$$S^2{}_c = 51.10$$

Ahora sustituir en la fórmula del estadístico de prueba tomando en cuenta que en la hipótesis nula se supone que en algún momento las medias poblacionales son iguales.

$$t = \frac{(641-629)-(0)}{\sqrt{51.10\left(\frac{1}{15}+\frac{1}{25}\right)}}$$

$$t = 5.14$$

Se establece la regla de decisión ubicando los valores críticos de t en la gráfica para delimitar las regiones de aceptación y rechazo $t_{(0.05,15+25-2)}$

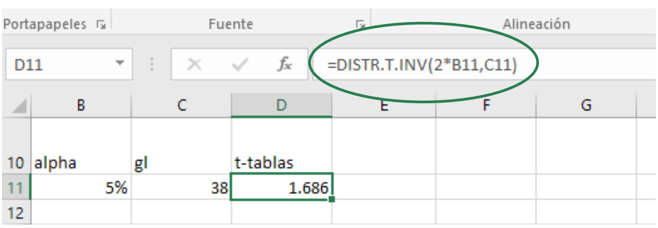
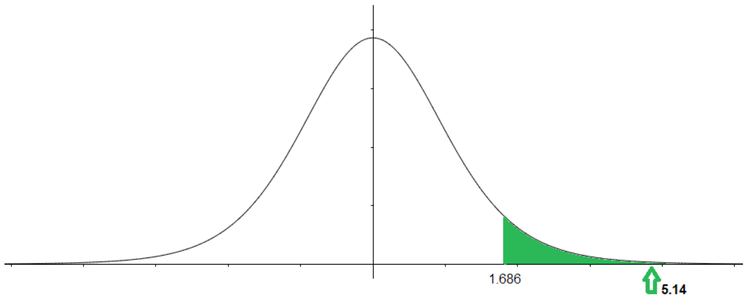


Ilustración 40 Valor crítico con distribución t unilateral

Se utiliza la función =DISTR.T.INV(pero al tratarse de una prueba de una cola (la derecha por cierto) la probabilidad se reúne en un solo lugar por lo que es necesario en la formula multiplicarla por dos ya que Excel está programado para dar el valor de $\alpha/2$.



Cuando se ubican los valores en la campana es visible que el valor de t calculado en el estadístico es mayor que el valor que se obtiene de $t_{(0.05,38)}$ por lo que está ubicado en la región de rechazo de H_0, esto quiere decir que existe evidencia estadísticamente significativa de que $\mu_1>\mu_2$ con una significancia de $\alpha=0.05$.

Ejemplos

1. Una marca de ropa desea determinar si los hombres de entre 18 y 60 años ven más tiempo la televisión que las mujeres del mismo rango de edad, se realizaron 50 llamadas telefónicas elegidas al azar registrando el número de horas que dijeron ver televisión durante una semana. El resultado obtenido se muestra en la tabla

Hombres	Mujeres
$n_1=29$ $X^1=10.79$ $\sigma_1=7.2$	$n_2=21$ $X^2=12.07$ $\sigma_2=5.1$

Realizar una prueba de hipótesis con el 95% de confianza para la H_0: $\mu_1-\mu_2\leq 0$

1. Un médico desea probar si un nuevo tratamiento para una enfermedad difiere en el tiempo en que los pacientes tienen síntomas fuertes así que decide administrar

el nuevo tratamiento a 20 personas y el tratamiento tradicional a otros 20 pacientes. Los resultados del nuevo tratamiento son una media de 20.15 horas con una desviación estándar muestral de 2.76 horas. El tratamiento tradicional dio una media de 20.87 horas con una desviación estándar de 2.44 horas. Realizar una prueba de hipótesis con α=0.01 para probar la hipótesis del médico.

Capítulo 4. Prueba de hipótesis para proporciones

Existen muchas situaciones en las que al investigador le interesa saber datos a cerca de una variable aleatoria que siga una distribución binomial, por ejemplo: en el proceso de fabricación de un artículo se clasifican como aceptables y defectuosos. En otras palabras, se utiliza en experimentos donde solo se pueden tomar dos valores posibles y son mutuamente excluyentes.

En este tipo de experimentos los éxitos se representan con la letra p y los fracasos son $1-p$. El estimador de este parámetro poblacional puede calcularse con: $\hat{p}=x/n$, estos estimadores son una parte de un todo por lo que su valor siempre se encuentra $0 \leq p \leq 1$. Los porcentajes también entran en este tipo de pruebas de hipótesis pues se representa en valores de decimales $35\%=35/100=.35$:

1 Con una proporción

En este método estadístico de pruebas de hipótesis se considera probar las siguientes

$H_0: p_p = p_x$ $H_1: p_p \neq p_x$	$H_0: p_p \leq p_x$ $H_1: p_p > p_x$	$H_0: p_p \geq p_x$ $H_1: p_p < p_x$
Prueba bilateral	Unilateral derecha	Unilateral izquierda

Donde p_x es el valor supuesto de la población y p_p la verdadera proporción poblacional.

El estadístico para esta prueba es:

$$Z = \frac{p_m - p_x}{\sqrt{\frac{p_x(1-p_x)}{n}}}$$

Como se observa en el estadístico, esté se trabajará con una distribución normal estándar por lo que, para tomar decisiones respecto a las hipótesis se usará la tabla de distribución Z con los valores que en el capítulo anterior se calcularon.

Ejemplos:

1. La gerencia de un restaurante desea incluir en sus bebidas cervezas artesanales pues se cree que a la mitad de la población le gusta ésta, para esto realizan un estudio con llamadas a 80 personas elegidas aleatoriamente preguntando sobre la aceptación de esta nueva bebida. El resultado fue que el 62% de la muestra probaría esta bebida. ¿Es conveniente que el restaurant incluya cervezas artesanales en su menú? Realizar una prueba de hipótesis con un 90% de confianza.

Solución

Primero se plantean la hipótesis acorde a lo que plantea la problemática a investigar

$$H_0: p = 0.5$$
$$H_1: p \neq 0.5$$

Ahora los datos obtenidos de la investigación y se sustituyen en la formula

$$n = 80 \qquad p_m = .62 \qquad p_x = .5$$

$$Z = \frac{.62 - .5}{\sqrt{\frac{.5(1-.5)}{80}}}$$

$$Z = 2.15$$

Como la hipótesis alternativa tiene el signo se trata de una prueba bilateral así que se usa una gráfica de dos colas

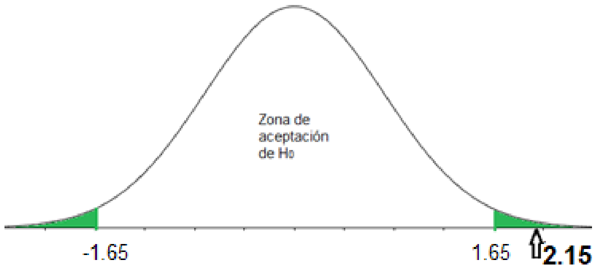


Ilustración 41 Comparativo entre Z crítica y Z calculada

Como el valor de Z calculada (2.15) es mayor que el valor obtenido en la tabla de Z crítica (±1.65) se rechaza H_0 para afirmar que existe evidencia estadísticamente significativa de que la proporción de clientes que consumirían cerveza artesanal es diferente al 50% de manera que será buena opción incluir este tipo de bebidas en el menú.

2. Realice una prueba de hipótesis con α =0.01 si

$$H_0{:}p \geq 60\%$$
$$H_1{:}p < 60\%$$
$$p = .54 \quad n = 350)$$

Solución

En este problema ya están definidas las hipótesis y el ellas p_x=60%=.6 así que los datos a utilizar son

$$n = 350 \qquad p_m = .54 \qquad p_x = .6$$

$$Z = \frac{.54 - .6}{\sqrt{\frac{.6(1 - .6)}{350}}}$$

Z=-2.29 es el valor de Z calculada que se ubicará en la gráfica de manera que sea posible derminar si se acepta o rechaza la hipótesis nula.

El valor de Z calculada es menor que el de Z crítica de manera que se ubica en la región de rechazo de H_0.

Ejercicios

1. El departamento de recursos humanos de una empresa sospecha que al menos el 75% de sus trabajadores no están en su lugar de trabajo en la primera hora que llegan al trabajo, por tal motivo, se revisaran las cámaras de vigilancia de 110 empleados elegidos al azar, en el video observan que 65 están fuera de su estación de trabajo. ¿Tiene razón el departamento o están exagerando? Realizar una prueba de hipótesis con una significancia de α=0.05
2. Realice una prueba de hipótesis con α=0.01 si

$$H_0: p=0.66$$
$$H_0: p\neq 0.66$$
$$p=.24 \quad n=250)$$

3. Un candidato a diputado de un distrito piensa que si el 57% de la población vota por él ganará la elección. Su equipo de campaña realiza una consulta en las principales calles del distrito tomando una muestra de 500 votantes de los cuales 305 afirman van votar por ese candidato ¿Indica esto que el candidato será el ganador de las elecciones? Realizar una prueba de hipótesis con un 90% de confianza.

2 Con dos proporciones

Cuando se desea construir una hipótesis acerca de dos proporciones de poblaciones es necesario considerar las hipótesis a probar

Prueba bilateral	Unilateral derecha	Unilateral izquierda
$H_0{:}p_1=p_2$ $H_1{:}\ p_1\neq p_2$	$H_0{:}p_1\leq p_2$ $H_1{:}\ p_1>p_2$	$H_0{:}p_1\geq p_2$ $H_1{:}\ p_1<p_2$

Como en la hipótesis nula siempre considera la igualdad entre las proporciones poblacionales, es necesario calcular un estimador de la diferencia entre proporciones ($\hat{p}$) mediante un promedio ponderado con la formula siguiente

$$\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Así el estadístico a utilizar es

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

El procedimiento para realizar esta prueba de hipótesis es similar a las pruebas anteriores de hipótesis en donde intervienen dos poblaciones

Ejemplos:

1. Se desea comparar el porcentaje de alumnos aprobados dos grupos de alumnos de la materia de Estadística, unos llevan la materia a las 7:00 am y otros a las 8:00 pm. Los alumnos tienen la hipótesis que en el grupo de la mañana tiene mayor cantidad de aprobados. Se seleccionan al azar 40 alumnos de cada horario, en el grupo de las 7:00 am aprobaron 34 alumnos y del grupo de la tarde 31 alumnos. ¿La suposición de los alumnos es correcta? Realizar una prueba de hipótesis con una significancia de $\alpha=0.10$

Solución

Para este problema (al igual que todas las pruebas de hipótesis) primero se plantean las hipótesis nula y alternativa

$$H_0: p_1 \leq p_2$$
$$H_1: p_1 > p_2$$

Se extraen los datos del problema

$$n_1=40 \qquad p_1=34/40=.85 \qquad n_2=40 \qquad p_2=31/40=.78$$

Ahora se sustituyen para calcular el promedio ponderado

$$\bar{p} = \frac{40(.85) + 40(.78)}{40 + 40}$$

$$\bar{p} = 0.82$$

Este será utilizado para calcular el estadístico Z calculada

$$Z = \frac{(.85 - .78) - (0)}{\sqrt{0.82(1 - 0.82)\left(\frac{1}{40} + \frac{1}{40}\right)}}$$

$$Z = 0.81$$

Este valor se compara con el de Z crítico de tabas para tomar la decisión.

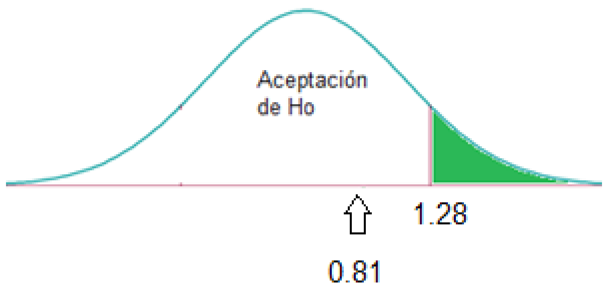


Ilustración 42 Delimitación de zonas de Ho

Como 0.81<1.28 se acepta la hipótesis nula, de manera que existe evidencia estadísticamente significativa de $p_1 \leq p_2$ esto quiere decir que el grupo de la mañana tiene menor o igual porcentaje de alumnos aprobados.

2. Una dependencia de salud pública selecciona una muestra aleatoria de 35 hombres de los cuales 30 son fumadores y en otra muestra de 30 mujeres 18 fumaban. ¿Podemos decir que la proporción de fumadores es igual en hombre y mujeres? Usar α=0.05

Lo primero es establecer las hipótesis

$$H_0: p_1 = p_2$$
$$H_0: p_1 \neq p_2$$

Ahora obtener los valores de p_1 y p_2

$$p_1 = {}^{30}\!/_{35} = 0.86 \quad n_1 = 35 \qquad p_2 = {}^{18}\!/_{30} = 0.60 \quad n_2 = 30$$

Ahora se sustituyen para calcular el promedio ponderado

$$\bar{p} = \frac{30 + 18}{35 + 30}$$

$$\hat{p} = 0.74$$

Este será utilizado para calcular el estadístico Z calculada recuerde que en la hipótesis nula se supone que las proporciones son iguales por eso la segunda parte del numerador el cero.

$$Z = \frac{(.86 - .60) - (0)}{\sqrt{0.74(1 - 0.74)\left(\frac{1}{35} + \frac{1}{30}\right)}}$$

$$Z = 2.38$$

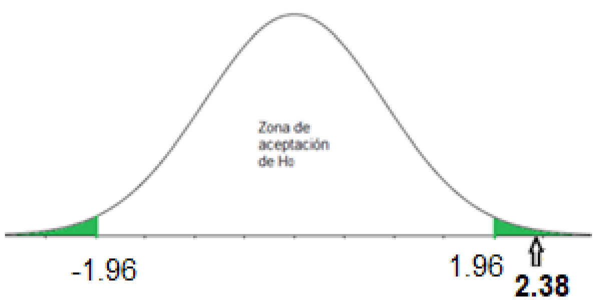


Ilustración 43 Toma de decisión con 95% de confianza

El valor de Z calculado se ubica en la zona de rechazo de la hipótesis nula, por lo que se puede afirmar que existe evidencia estadísticamente significativa para decir que la proporción de hombres fumadores es diferente a la proporción de mujeres fumadoras con un nivel de confianza del 95%.

Ejercicios

I. Considere los datos siguientes para realizar una prueba de hipótesis para dos proporciones

Muestra 1	Muestra 2
$p_1=34\%$ $n_1=250$	$p_2=43\%$ $n=350$

Ponga a prueba la hipótesis $H_0: p_1=p_2$

$$H_1: p_1 \neq p_2 \text{ con } \alpha=0.05$$

II. En una investigación para conocer el impacto del color en los anuncios del periódico, se realiza una investigación en 400 personas divididas en dos grupos de personas seleccionadas al azar (200 en cada uno). Al primer grupo se les dio el periódico con el anuncio de color, el 69% de ellos dice que lo noto. Al segundo grupo se les dio el periódico con el anuncio en blanco y negro 57% notó el anuncio. Realizar una prueba de hipótesis con un nivel de confianza del 90%

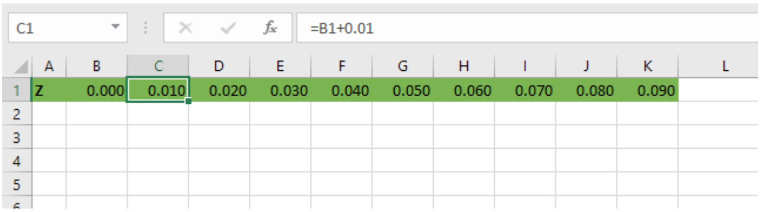
Referencias

Hipótesis. (2009). En el Diccionario Larousse (4ª ed.). Puebla, México: Ediciones Larousse.
Lind, D (2015) Estadística aplicada a los negocios y la economía. (16 ed) México D.F. México. Mc Graw Hill Education
Lopez, M. (Sin Fecha) Excel como una herramienta asequible en la enseñanza de la Estadística. Recuperado el 3 de junio de 2020 de https://campus.usal.es/~teoriaeducacion/rev_numero_07/n7_art_lopez_lagunes_herrera.htm
Rrodriguez, J. (2016) Estadistica para administración (2da. ed) Ciudad de México, México. Grupo Editorial Patria
ProfeIO2 (28 de noviembre de 2016) Tabla de distribución normal estándar con Excel (Archivo de video) Recuperado el 20 de julio de 2020 de https://www.youtube.com/watch?v=WpYptb59pbI

Apéndice

Generación de tabla Z con Excel

En la primera fila se determinan los centésimos de cada número en la primera columna.

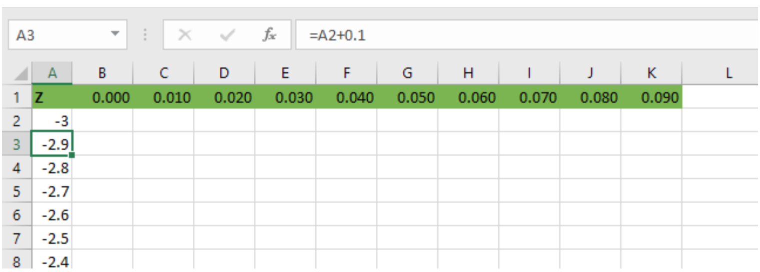


C1 | =B1+0.01

	A	B	C	D	E	F	G	H	I	J	K	L
1	Z	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090	
2												
3												
4												
5												

Ilustración 44 Rellenar columnas consecutivas

Ahora correremos los números correspondientes a las filas partiendo de la celda A2 con el número -3 hasta 0 disminuyendo de 0.1 en cada.



A3 | =A2+0.1

	A	B	C	D	E	F	G	H	I	J	K	L
1	Z	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090	
2	-3											
3	-2.9											
4	-2.8											
5	-2.7											
6	-2.6											
7	-2.5											
8	-2.4											

Ilustración 45 Rellenar filas consecutivas

Al llegar a 0 lo escribimos dos veces porque es necesario la parte positiva y negativa al hacer referencia a la gráfica (campana de Gauss) y proseguimos con la numeración sumando 0.1 hasta el 3

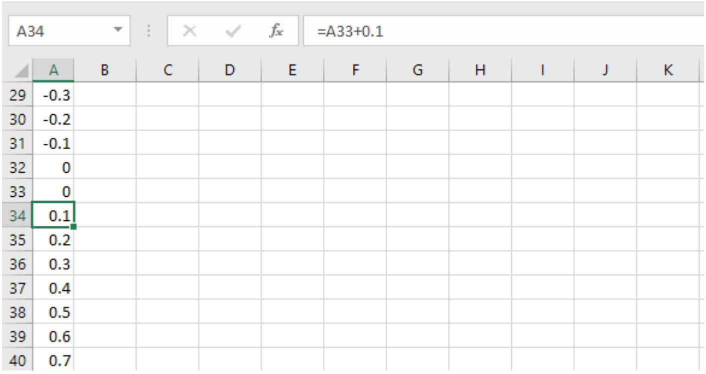


Ilustración 46 Valores necesarios para la distribución normal

Ahora será necesario utilizar la función de Distribución normal estándar DISTR.NORM.ESTAND.N(z,acumulado) fijando los valores de filas y columnas de esta manera al jalar la selección de celdas se llenará la tabla.

En la parte de valores de z negativo se usa l forma =DISTR.NORM.ESTAND.N($A2**-**B$1,1).

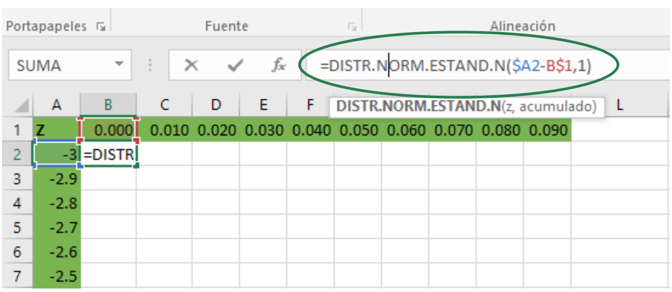
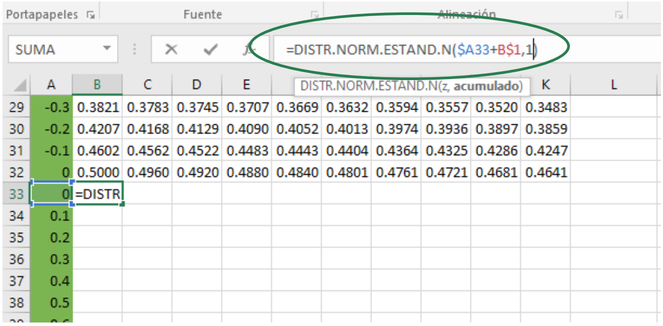


Ilustración 47 Generar parte negativa

En la parte de z positiva se utiliza la formula =DISTR.NORM.ESTAND.N($A43**+**B$1,1) que difieren en el signo entre la relación de fila y columna: restando en los negativos y sumando en los positivos.



Portapapeles | Fuente | Alineación

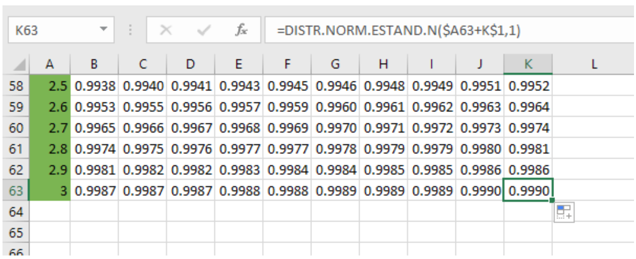
SUMA | =DISTR.NORM.ESTAND.N($A33+B$1,1)

DISTR.NORM.ESTAND.N(z, acumulado)

	A	B	C	D	E	F	G	H	I	J	K	L
29	-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483	
30	-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859	
31	-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247	
32	0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641	
33	0	=DISTR										
34	0.1											
35	0.2											
36	0.3											
37	0.4											
38	0.5											

Ilustración 48 Generar parte positiva

Al completar la tabla se observa que en el valor menor correspondiente a la probabilidad acumulada es cercana a 0 y en el valor contrario positivo -3.09 es cercana a 1, lo que concuerda con el área bajo la curva.



K63 | =DISTR.NORM.ESTAND.N($A63+K$1,1)

	A	B	C	D	E	F	G	H	I	J	K	L
58	2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	
59	2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	
60	2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	
61	2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	
62	2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	
63	3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	
64												
65												

Ilustración 49 Tabla completa con probabilidad de 0 a 1

La elaboración de esta tabla mediante la función DISTR.NORM.ESTAND.N(z,acumulado) es calculada al usar la formula:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

PRUEBA DE HIPÓTESIS ESTADÍSTICA CON EXCEL
se terminó de imprimir
en octubre de 2020
en los talleres gráficos
de Amateditorial, S.A. de C. V.
Prisciliano Sánchez 612, Colonia Centro
Guadalajara, Jalisco
Tel.: 36120751 / 36120068
amateditorial@gmail.com
www.amateditorial.com.mx
Edición al cuidado del autor