TEXTOS POR COMPETENCIAS

ESTADÍSTICA APLICADA A LA GASTRONOMÍA

LICENCIATURA EN GASTRONOMÍA

Fecha: Agosto 2016

MARÍA DEL PILAR ROSAS DEL BARRIO

INDICE

INTRODUCCIÓN

La unidad de aprendizaje de Estadística Aplicada a la Gastronomía se presenta dentro del núcleo sustantivo en el plan de estudios de la Licenciatura en Gastronomía. Mediante esta unidad de aprendizaje se mostrará un conjunto de métodos estadísticos que se pueden utilizar para diseñar estudios estadísticos y analizar e interpretar sus resultados; estos métodos permiten la obtención de información útil para una mejor toma de decisiones en la administración de las empresas gastronómicas, así como para el análisis y la investigación del fenómeno gastronómico en general.

PROPÓSITO DE LA UNIDAD DE APRENDIZAJE:

El estudiante será capaz de seleccionar el método estadístico más apropiado para analizar un conjunto de datos surgidos de un problema que se desea resolver, así como interpretar los resultados obtenidos; esto le permitirá obtener un mayor conocimiento en el área de la gastronomía. Lo que apoyará en el diseño, formulación e instrumentación de proyectos, planes y procesos administrativos e investigativos tendientes a generar alternativas de solución frente al conocimiento disciplinar de la gastronomía.

Para lograr lo anterior, esta unidad de aprendizaje se conforma de cuatro unidades descritas a continuación:

UNIDAD DE COMPETENCIA I: Inferencia estadística para una o dos poblaciones.

En esta unidad conocerás y aplicarás tanto la estimación por intervalos de confianza como la prueba de hipótesis tanto para una como para dos poblaciones con la finalidad de efectuar una mejor toma de decisiones.

UNIDAD DE COMPETENCIA II: Correlación y análisis de regresión. En esta unidad podrás predecir el comportamiento de una variable en función de otras, en situaciones relacionadas con la gastronomía.

UNIDAD DE COMPETENCIA III: Diseño y análisis de experimentos. Mediante esta unidad adquirirás los conocimientos para planear, ejecutar, analizar e interpretar y evaluar resultados de experimentos relacionados con la gastronomía.

UNIDAD DE COMPETENCIA IV: Muestreo. Mediante esta unidad podrás aplicar las técnicas de muestreo probabilístico y no probabilístico en diferentes situaciones.

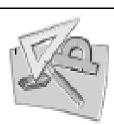
FORMA DE UTILIZAR EL TEXTO

Querido alumno:

El presente texto está elaborado pensando en ti y en que te sea más fácil la comprensión y aplicación de la estadística inferencial. Está basado en el programa de estudio por competencias de la licenciatura en gastronomía vigente a la fecha.

El texto está constituido por teoría así como ejercicios para resolver.
Se ha tratado de emplear lenguaje de fácil comprensión así como manejar ejemplos que pudieras encontrar en tu labor cotidiana dentro del ámbito de la gastronomía así como en la administración de empresas de alimentos y bebidas. Puedes ampliar los conocimientos o profundizar en algún tema de interés acudiendo a las recomendaciones documentales indicadas en el cuerpo de texto.
Te sugiero que para obtener un máximo aprovechamiento primero trates de comprender los conceptos generales expuestos al inicio de cada una de las cuatro unidades que constituyen esta unidad de aprendizaje, luego te esfuerces en desarrollar la habilidad de aplicar los procedimientos matemáticos relacionados con la unidad en estudio, después interpretar correctamente los resultaos obtenidos y por último leer una vez más el texto para comprender toda la teoría. Tendrás más éxito si comprendes lo que se está haciendo en lugar de aplicar mecánicamente los pasos necesarios para llegar a los resultados que sin interpretación no tienen sentido alguno.
Espero que este texto te sea de utilidad y sea el inicio o la base para futuros aprendizajes relacionados con el campo de la estadística.

Las actividades están señaladas por diferentes capsulas las cuales se explican a continuación:

	**Lo aplico**: Aquí se proponen ejercicios que te permiten aplicar los conocimientos teóricos y procedimientos estadísticos aprendidos en clase.
	**Para saber más:** Se incluye referencias bibliográficas que te ayuden a complementar el tema así como más ejercicios de los temas estudiados.

	**Para mi evaluación**-. Ejercicios que te permiten verificar lo aprendido y afirmar los conocimientos adquiridos. Estos se integrarán al portafolio de evidencias.
	**Mi Proyecto**: Trabajo de diseño de experimentos que se elaborará para aplicar lo aprendido en relación al tema.

UNIDAD I

INFERENCIA ESTADÍSTICA

Estimación de azúcar en las naranjas

En Florida, los miembros de la industria de los cítricos usan profusamente métodos estadísticos.

Una aplicación específica tiene que ver con la forma en que se paga a los agricultores por las naranjas que se usan para elaborar jugo de naranja. Cuando llega un camión cargado con naranjas, primero se pesa la carga en la planta receptora, luego se elige al azar una muestra de una docena de naranjas. La muestra se pesa, se exprime y se mide la cantidad de azúcar que contiene el jugo. Con base en los resultados de la muestra, se estima la cantidad total de azúcar contenida en toda la carga del camión. El pago por la carga de naranjas se basa en la estimación de la cantidad de azúcar, ya que las naranjas más dulces son más valiosas que las menos dulces, aunque las cantidades de jugo sean iguales. (Triola, 2009).

PROPÓSITO DE LA UNIDAD DE COMPETENCIA: Propiciar una mejor toma de decisiones en la solución de problemas relacionados con la gastronomía, mediante la aplicación de las técnicas de inferencia estadística para una o dos poblaciones.

En vista del propósito anterior, en esta unidad estudiaremos primeramente que es la estadística inferencial, conceptos generales que se utilizan en esta unidad así como en las siguientes por lo que es muy importante su comprensión. También conoceremos los dos métodos que existen para efectuar inferencias (Estimación y prueba de hipótesis) tanto para una como para dos poblaciones.

La **estadística inferencial** es la rama de la estadística que comprende los métodos y procedimientos para obtener conclusiones generales para toda la población, a partir de la información que proporciona una muestra proveniente de la de la misma y el grado de fiabilidad o significación de los resultados obtenidos.

- Existen dos procedimientos generales para hacer inferencias sobre parámetros poblacionales desconocidos:
  1. ***Estimar*** *un parámetro poblacional y*
  2. ***Probar*** una hipótesis o afirmación con respecto a un parámetro poblacional

La estimación y la prueba de hipótesis comparten elementos de la teoría de probabilidades.

PARA MI EVALUACIÓN:
Investigar los conceptos de Estadística, estadística descriptiva, probabilidad, distribuciones de probabilidad.
Criterios de evaluación: Entregar trabajo en tiempo y forma, conceptos correctos, con letra legible.

I.1 ESTIMACIÓN

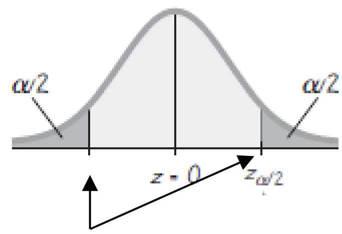
En la estimación mediante en el estudio de una muestra de una población se quiere generalizar las conclusiones al total de la misma. Existen dos tipos de estimaciones para parámetros; ***puntuales y por intervalo de confianza.***

Un **estimado puntual** es un valor individual (o punto) que se usa para aproximar un parámetro de población.
Un **intervalo de confianza** (o **estimado del intervalo**) es un rango (o un intervalo) de valores que se usa para estimar el valor real de un parámetro de población. El intervalo de confianza suele abreviarse como IC.

Aunque un estimado puntual es el *mejor* valor individual para estimar un parámetro poblacional, no nos da ninguna indicación precisa de qué tan *bueno* es este mejor estimado. Sin embargo, un intervalo de confianza nos ofrece información que nos permita comprender mejor la exactitud del estimado. El intervalo de confianza se asocia con un nivel de confianza, como 0.95 (o 95%) por ejemplo. (Levine, D.M., Krehbiel, T.C., Berenson, M.L. 2006).

El **nivel de confianza** es la probabilidad 1-α (a menudo expresada como el valor de porcentaje equivalente), que es la proporción de veces que el intervalo de confianza realmente contiene el parámetro de población, suponiendo que el proceso de estimación se repite un gran número de veces. El nivel de confianza también se llama **grado de confianza** o **coeficiente de confianza**.

Un **valor crítico** es el número en la línea limítrofe que separa estadísticos muestrales que tienen mayor probabilidad de ocurrir de aquellos que no tienen probabilidad de ocurrir. El número Z α/2 es un valor crítico, una puntuación Z con la propiedad de que separa una área de α/2 en la cola derecha de la distribución normal estándar.



Valores críticos

Figura 1.1

Los métodos de esta sección y muchos de los demás métodos estadísticos que se encuentran en los capítulos siguientes incluyen el uso de una puntuación Z estándar. Una puntuación Z de este tipo se llama *valor crítico* (definido anteriormente). Los valores críticos se basan en las siguientes observaciones.

1. Sabemos, que en ciertas condiciones, la distribución muestral de las proporciones muestrales puede aproximarse a una distribución normal.

2. Las proporciones muestrales tienen una probabilidad relativamente pequeña (denotada por α) de caer en una de las colas sombreadas de la figura 1.1

3. Denotando el área de cada cola sombreada como α/2, vemos que existe una probabilidad total de a de que una proporción muestral caiga en cualquiera delas dos colas sombreadas.

4. Por la regla de los complementos, existe una probabilidad de 1 – α de que una proporción muestral caiga dentro de la región interior sombreada de la figura 1.1.

5. La puntuación *z* que separa la región de la cola derecha generalmente se denota por Z α/2 y se conoce como *valor crítico*, puesto que está en la frontera que separa las proporciones muestrales que tienen probabilidad de ocurrir de aquellas que no tienen probabilidad de ocurrir.

EJERCICIO: Calcule el valor crítico Z *α/*2 que corresponde a un nivel de confianza del 95%. Anexo 1)

SOLUCIÓN**:** Para calcular el valor crítico *z* de un nivel de confianza del 95% *no* busque 0.95 en el cuerpo de la tabla del anexo 1. Un nivel de confianza del 95% corresponde a α=0.05. El área en cada una de las colas sombreadas es α/2=0.025. Buscamos este número en la tabla de la distribución Normal estandarizada (Anexo 1)y encontramos que *Z α/*2 =1.96 señalando que toda el área a su izquierda debe ser 1 - 0.025, o 0.975.

LO APLICO:

De manera similar calcular el valor crítico para 90 y 99% de confianza y registrar en la tabla siguiente:

Nivel de confianza	α	Z α/2
90%		
95%		
99%		

I.2 ESTIMACIÓN DE INTERVALO DE CONFIANZA DE UNA POBLACIÓN

FÓRMULAS PARA EL CÁLCULO DE INTERVALOS DE CONFIANZA PARA UNA POBLACIÓN:

INTERVALO DE CONFIANZA PARA PROPORCIONES

$$p - Z\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + Z\sqrt{\frac{p(1-p)}{n}}$$

$$p = \frac{X}{n}$$

Donde:
π = proporción de la población
p = proporción de la muestra
= número de elementos con característica de interés / n
n= tamaño de muestra
Z = Valor crítico de la distribución normal estandarizada dependiente del nivel de confianza deseado

INTERVALO DE CONFIANZA PARA LA MEDIA CON σ CONOCIDA**:**

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

Donde:
μ= media poblacional
σ= desviación estándar poblacional
$\bar{\chi}$ = media muestral
N= tamaño de muestra
Z = Valor crítico de la distribución normal estandarizada dependiente del intervalo de confianza deseado.

INTERVALO DE CONFIANZA PARA LA MEDIA CON σ DESCONOCIDA:

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

μ= media poblacional
s= desviación estándar muestral
$\bar{\chi}$ = media muestral
n= tamaño de muestra
t = valor crítico de la distribución t de student con n – 1 grados de libertad.

Las opciones más comunes para el nivel de confianza son 90% (con α=0.10), 95% (con α=0.05) y 99% (con α= 0.01). La opción del 95% es la más común puesto que provee un buen equilibrio entre precisión (reflejada en el ancho del intervalo de confianza) confiabilidad (expresada por el nivel de confianza).

EJEMPLO:

A un restaurante nuevo regresan una segunda vez a comer el 44% de los clientes. Calcule el intervalo de confianza de la proporción de clientes que vuelven a consumir al restaurante para el 95% de confianza cuando la muestra es de 280 clientes.

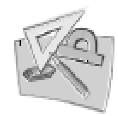
Solución. La proporción poblacional π es $0.381 < \pi < 0.497$

Interpretación de un intervalo de confianza:

Se debe ser cuidadoso para interpretar los intervalos de confianza correctamente.

Correcta:" Se tiene una confianza del 95% de que el intervalo de 0.381 a0.497 realmente contiene el valor verdadero de π". Esto significa que si seleccionamos muchas muestras diferentes de tamaño 280 y construimos los intervalos de confianza correspondientes, el 95% de ellos incluirían realmente el valor de la proporción poblacional π.

Errónea: "Existe un 95% de probabilidades de que el valor real de π esté entre 0.381 y 0.497".



LO APLICO:

1.1 Al cuestionar a los gerentes de 500 restaurantes sobre si contratarían los servicios de una empresa de capacitación externa sobre manejo higiénico de alimentos 135 gerentes contestaron que si la contratarían.

Construya una estimación del intervalo de confianza del 99% de la proporción poblacional de gerentes que contratarían los servicios de la empresa de capacitación en manejo higiénico de los alimentos.

1.2 El Administrador de una pastelería desea estimar la media del valor al menudeo de pasteles de cumpleaños que la pastelería elaboró este mes. Una muestra aleatoria de 20 pasteles indica una media de valor de $ 230 y una desviación estándar muestral de $50.

Suponiendo una distribución normal construya una estimación del intervalo de confianza del 95% para la media de valor de todos los pasteles elaborados en este mes.

1.3 Para calcular los salarios de graduados universitarios que tomaron un curso de estadística en la universidad se tomó una muestra de n = 41 universitarios. El sueldo promedio resultó de $67,200 y se sabe que σ es $18,277.

Calcule el intervalo de confianza para la media poblacional del salario de los universitarios que tomaron un curso de estadística para un nivel de confianza del 95%.

PARA MI EVALUACIÓN:

1.4 Una de las principales medidas en el servicio de calidad proporcionado por cualquier organización es la velocidad con la que se responde a las quejas al cliente. Una gran empresa de servicio de alimentos y bebidas ha tomado el tiempo en que el personal de contacto con el cliente ha tardado en solucionar sus quejas. Una muestra de 26 tiempos (en minutos) de solución de las quejas del mes en curso arrojó los siguientes datos:

12	8	5	5	6	6	10	10	9	7	10	7	7
5	0	10	6	9	12	1	5	10	8	5	5	9

a)Construya un intervalo de confianza del 95% para la media poblacional del tiempo de solución a las quejas del cliente.

b)¿Cómo podrían ser útiles los resultados obtenidos en el inciso anterior para la empresa de Alimentos y bebidas?

PARA MI EVALUACIÓN:

1.5 Una encuesta realizada a 600 trabajadores de empresas de Alimentos y bebidas y que abandonaron sus trabajos por razones de salud, indicó que el 66% de ellos deseaban regresar al mismo ramo.

Construya un intervalo de confianza del 95% para la proporción poblacional de extrabajadores de empresas de alimentos y bebidas que desean regresar a trabajar.

1.6 El pulso de las personas es sumamente importante. Se Tomó el pulso a 40 mujeres seleccionadas al azar; los estadísticos son los siguientes: media=76.3 latidos por minuto y s =12.5 latidos por minuto

a) Utilice esta muestra para calcular el mejor estimado puntual de la media poblacional μ de los pulsos cardiacos de todas las mujeres.

b) Utilice esta muestra para calcular el intervalo de confianza del 90% de la media poblacional de los pulsos cardiacos de todas las mujeres.

1.7 Un estudiante encuesta a 100 compañeros de clase y les pregunta si tienen deudas pendientes. Después de calcular la proporción muestral de esta muestra de n =100 sujetos, ¿se pueden utilizar los métodos de esta sección para estimar la proporción todos los adultos que tienen deudas pendientes? ¿Por qué?

1.8 En una zona con límite de velocidad de 90 Km/h se selecciona una muestra de n =90 conductores, el promedio de velocidad resulta ser 110 Km/h, y se sabe que σ es de 6 Km/h.
Calcular un intervalo de confianza del 95% para la media poblacional de conductores multados en una zona con límite de velocidad de 90 Km/h.

Criterios de evaluación: Entregar los ejercicios resueltos en tiempo y forma, con letra legible y con resultados correctos

PARA SABER MÁS:

1.9 Calcule un intervalo de confianza para la media poblacional del peso perdido por una dieta de Weight Watchers para una muestra de 40 individuos, la media de peso perdido es de 3.0 kg y se distribuye normalmente, la desviación estándar muestral de 4.9 kg y el nivel de confianza es del 95%.
1.10 Mediante un estudio efectuado a una muestra aleatoria de 21 computadoras de escritorio se calculó un periodo de vida media de 6.8 años y una desviación estándar de 2.4 años. Calcule un intervalo de confianza del 99% para la media poblacional del periodo de vida de las computadoras de escritorio suponiendo que la población tiene una distribución normal.
1.11 Las puntuaciones de CI de estudiantes de estadística efectuada a una muestra de 25 estudiantes es en promedio 118.0 y la desviación estándar de la muestra es igual a 10.7.Calcule un intervalo de confianza del 95% para la media poblacional del coeficiente intelectual de los estudiantes de estadística. Además escriba un enunciado que interprete el intervalo de confianza.
1.12 El periodo de vida de los teléfonos celulares calculado por el fabricante en una muestra aleatoria de n= 27 celulares tiene una media de 4.6 años y una desviación estándar s= 1.9 años.
Calcule un intervalo de confianza del 95% para la media poblacional del periodo de vida de los teléfonos celulares y redacte un enunciado que interprete dicho intervalo.

I.3 PRUEBA DE HIPÓTESIS

En estadística, una **hipótesis** es una aseveración o afirmación acerca de una propiedad de una población.

Una **prueba de hipótesis** (o **prueba de significancia**) es un procedimiento estándar para probar una aseveración acerca de una propiedad de una población.
Por medio de la prueba de hipótesis se determina si las proposiciones efectuadas con relación a la población son compatibles o no con los datos disponibles.

COMPONENTES DE LA PRUEBA DE HIPÓTESIS:
_ La **hipótesis nula** (denotada por H_0) es la afirmación de que el valor de un parámetro de población (como una proporción, media o desviación estándar) es *igual a* un valor aseverado.
Es la hipótesis que debe probarse.
La hipótesis nula se prueba en forma directa, en el sentido de que suponemos que es verdadera, y llegamos a una conclusión para rechazar H_0 o no rechazar H_0.
Se rechaza cuando a partir de la información de la muestra existe suficiente evidencia de que es falsa.

_ La **hipótesis alternativa** (denotada por H_1 o Ha o H_A) es la afirmación de que el parámetro tiene un valor que, de alguna manera, difiere de la hipótesis nula.
Es la hipótesis de la investigación.
Se creerá cierta si los datos de la muestra llevan al rechazo de H_0.
Representa la conclusión obtenida al rechazar H_0.

_ La **región crítica** (o **región de rechazo**) es el conjunto de todos los valores del estadístico de prueba que pueden provocar que rechacemos la hipótesis nula.

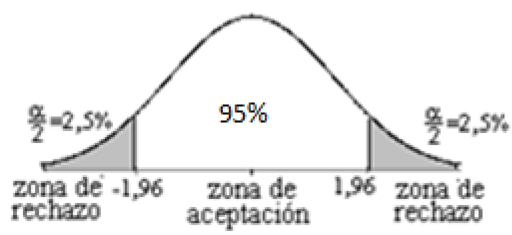
_ El **nivel de significancia** (denotado por α) es la probabilidad de que el estadístico de prueba caiga en la región crítica, cuando la hipótesis nula es verdadera.

Si el estadístico de prueba cae en la región crítica, rechazamos la hipótesis nula, de manera que α es la probabilidad de cometer el error de rechazar la hipótesis nula cuando es verdadera. Se trata de la misma α, donde definimos el nivel de confianza para un intervalo de confianza como la probabilidad 1 - α. Las opciones comunes para α son 0.05, 0.01 y 0.10, aunque la más común es 0.05. (Levine, D.M., Etal. 2006).

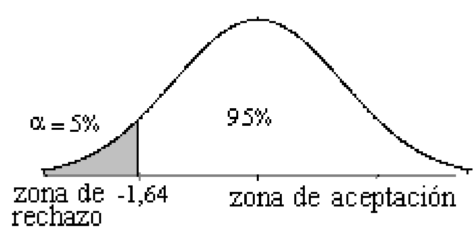
_ Un **valor crítico** es cualquier valor que separa la región crítica (donde rechazamos la hipótesis nula) de los valores del estadístico de prueba que no conducen al rechazo de la hipótesis nula. Los valores críticos dependen de la naturaleza de la hipótesis nula, de la distribución muestral que se aplique y del nivel de significancia α.

_Las *colas* en una distribución son las regiones extremas limitadas por los valores críticos. Algunas pruebas de hipótesis incluyen dos colas, otras la cola derecha y otras la cola izquierda.

_ **Prueba de dos colas:** La región crítica se encuentra en las dos regiones extremas (colas) bajo la curva.



_ **Prueba de cola izquierda:** La región crítica se encuentra en la región extrema izquierda (cola) bajo la curva.



_ **Prueba de cola derecha:** La región crítica se encuentra en la región extrema derecha (cola) bajo la curva.

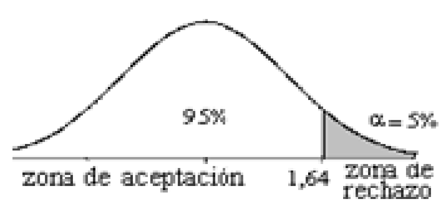


FIGURA 1.2

La hipótesis nula no se rechaza cuando la evidencia muestral sugiere que es más probable que ésta sea cierta que la hipótesis alternativa. Sin embargo, el no poder rechazar la H_0 no comprueba que sea cierta. (Levine, D.M., Etal. 2006).

No podemos demostrar que H_0 es correcta porque la decisión se basa sólo en información de la muestra, no de toda la población. Por lo tanto, si no se rechaza la H_0, solo podemos concluir que no existe evidencia suficiente para garantizar su rechazo.
Al rechazar H_0 se tiene una prueba estadística de que H_1 es correcta.
Si no se rechaza H_0, no se ha podido demostrar H_1.

_ Error tipo I y tipo II

Los errores tipo I y tipo II son los riesgos en la toma de decisiones al utilizar la metodología de la prueba de hipótesis. Cuando probamos una hipótesis nula, llegamos a la conclusión de rechazarla o no rechazarla. Tales conclusiones pueden ser correctas o incorrectas (incluso cuando hacemos todo correctamente). Existen dos tipos de errores denominándolos errores tipo I y tipo II.

<u>Error tipo I</u>: Se presenta cuando se rechaza la hipótesis nula H_0 siendo cierta y no debería rechazarse. La probabilidad de que se presente el error tipo I es α.
<u>Error tipo II</u>: Se presenta cuando no se rechaza la hipótesis nula H_0 siendo falsa y debería rechazarse. La probabilidad de que se presente el error tipo II es β.

Decisión estadística	H_0 verdadera	H_0 falsa
No rechazar H_0	Decisión correcta Coeficiente de confianza = $1-\alpha$	ERROR TIPO II P (error tipo II) = β
Rechazar H_0	ERROR TIPO I P (error tipo I) = α	Decisión correcta Potencia = $1- \beta$

Fuente: Estadística para administración. Lvine, krehbiel.

El **valor *P*** (o **valor de probabilidad**) es la probabilidad de obtener un valor del estadístico de prueba que sea *al menos tan extremo* como el que representa a los datos muestrales, suponiendo que la hipótesis nula es verdadera. La hipótesis nula se rechaza si el valor *p* es muy pequeño, tanto como 0.05 o menos.

_Decisiones y conclusiones:

El procedimiento convencional de prueba de hipótesis requiere que siempre probemos la hipótesis nula, de manera que nuestra conclusión inicial siempre será una de las siguientes:

1. Rechazo de la hipótesis nula.

2. No rechazo de la hipótesis nula.

Criterio de decisión: La decisión de rechazar o no rechazar la hipótesis nula suele realizarse por medio del método tradicional (o método clásico) de prueba de valor crítico o el método del valor-p.

En años recientes, el uso del método del valor-p ha aumentado, junto con la inclusión de valores *P* en los resultados de programas de cómputo.

Método tradicional:

Rechace H_0 si el estadístico de prueba cae dentro de la región crítica.

No rechace H_0 si el estadístico de prueba no cae dentro de la región crítica.

Método del valor *P*:

Rechace H_0 si el valor de $p < \alpha$ (donde α es el nivel de significancia, tal como 0.05).

No rechace H_0 si el valor $p > \alpha$.

Redacción de la conclusión final La conclusión debe redactarse en términos sencillos y sin tecnicismos al establecer el verdadero significado de la conclusión.

Aceptación=no rechazo Algunos libros de texto dicen “aceptar la hipótesis nula” en vez de “no rechazar la hipótesis nula”. Ya sea que usemos el término *aceptar* o *no rechazar*, debemos reconocer que *no estamos probando la hipótesis nula*; únicamente estamos diciendo que la evidencia muestral no es lo suficientemente fuerte como para justificar el rechazo de la hipótesis nula. (Cuando un jurado no encuentra evidencia suficiente para sentenciar a un sospechoso, emite un veredicto de no culpabilidad y no un veredicto de inocencia). El término *aceptar* es un poco confuso.

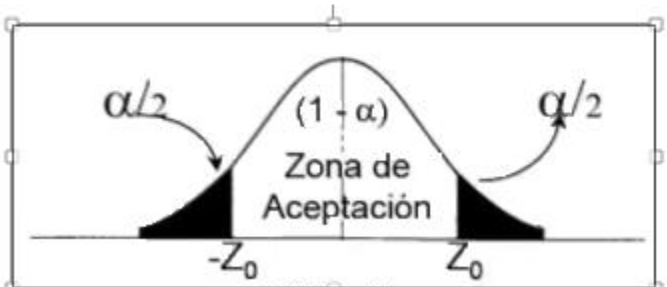


FIGURA 1.3

I.4 PRUEBA DE HIPÓTESIS PARA UNA POBLACIÓN

PRUEBA DE HIPÓTESIS PARA LA MEDIA CON σ CONOCIDA:

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

Donde:
μ= media poblacional
σ= desviación estándar poblacional
$\bar{\chi}$ = media muestral
n = tamaño de la muestra
Z = Estadístico de prueba

PRUEBA DE HIPÓTESIS PARA LA MEDIA CON σ DESCONOCIDA:

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Donde:
μ= media poblacional
s= desviación estándar muestral
$\bar{\chi}$ = media muestral
n= tamaño de muestra

El número de **grados de libertad** para un conjunto de datos muestrales recolectados es el número de valores muestrales que pueden variar después de haber impuesto ciertas restricciones a todos los valores de los datos.

PRUEBA DE HIPÓTESIS PARA LA PROPORCIÓN:

$$Z = \frac{P - \pi}{\sqrt{\frac{P\,(1-P)}{n}}}$$

Donde:
p =proporción de la muestra
n= tamaño de muestra
q= 1-p
π = Proporción poblacional
Z = Estadístico de prueba

LO APLICO:

1.13 Una empresa eléctrica fabrica focos que tienen una duración que se distribuye de forma aproximadamente normal con una media de 800 horas y una desviación estándar de 40 horas. Si una muestra aleatoria de 30 focos tiene una duración promedio de 788 horas, ¿muestran los datos suficiente evidencia para decir que la duración media ha cambiado? Utilice un nivel de significancia del 0.05.

1.14 Las leyendas en la etiqueta de frituras de una marca que ha tenido quejas de sus clientes indican un peso de llenado de 5.5 onzas. Se toma una muestra aleatoria de 64 bolsas de frituras y se determina que pesan en promedio 5.23 onzas con una desviación estándar de 0.24 onzas. Pruebe la hipótesis alternativa, $\mu < 5.5$ onzas en el nivel de significancia de 0.05.

1.15 El fabricante de un medicamento sostiene que el mismo tiene un 90 % de efectividad en el alivio de una alergia, por un periodo de 12 h. En una muestra de 200 individuos que tenían la alergia, la medicina suministrada alivió los síntomas de la alergia durante 12 horas a 160 personas. Determinar si la aseveración del fabricante es cincorrecta. Emplee un nivel de significancia de 0.01

1.16 Una muestra aleatoria de 100 casos de intolerancia a la lactosa registradas en enfermos de un hospital geriátrico este año muestra una edad promedio de 71.8 años. Suponga una desviación estándar poblacional de 8.9 años, ¿esto parece indicar que la edad de personas de la tercera edad intolerantes a la lactosa en dicho hospital es ahora mayor de 70 años? Utilice un nivel de significancia de 0.05

PARA MI EVALUACIÓN:

1.17 El peso teórico de llenado de un cereal de arroz es de 500 g por paquete; la desviación estándar del proceso es igual a 15 g. Se toma una muestra de 25 paquetes del cereal obteniéndose una media de 506g. Efectúe una prueba de hipótesis para determinar si el proceso es diferente de 500 g en cuyo caso debe suspenderse. ($\alpha = 0.05$)

1.18 Usted es gerente de un restaurante de comida rápida. Durante el mes pasado el tiempo de espera en la ventanilla de servicio medido a partir del momento en que el cliente realizó su pedido hasta que lo recibió fue de 3.7 min. Después de efectuar algunas mejoras al proceso con la finalidad de reducir el tiempo de espera usted toma una muestra aleatoria de 64 pedidos de los cuales la media y la desviación estándar muestrales resultaron 3.57 y 0.8 minutos respectivamente. Utilizando un nivel de significancia de 0.05 ¿Existe evidencia de que la media poblacional del tiempo de espera es ahora menor que 3.7 min?

1.19 El gerente administrativo de un salón de eventos supone que el munú preferido por los clientes es de un costo mayor de $300 por persona. Seleccionó una muestra de 100 eventos y registró los costos por platillo contratado para cada uno de ellos. Luego calculó un costo medio muestral de $315.40 y una desviación estándar de $ 43.2. Utilizando un nivel de confianza de 0.1 ¿Existe evidencia de que la media poblacional de los platillos seleccionados para su evento por los clientes esté por arriba de los $ 300?

1.20 Un fabricante de chocolate para repostería utiliza máquinas para empacar el chocolate. A pesar de que los paquetes están etiquetados con un contenido de 16 onzas, la empresa quiere que tengan 16.34 onzas, de tal manera que virtualmente ninguno de sus paquetes tenga menos de 16 onzas. De forma periódica se selecciona una muestra de 50 paquetes y si se encuentra evidencia de que la cantidad media suministrada es distinta de 16.34 onzas, se detiene el proceso de empacado. Suponga que la cantidad media abastecida en una muestra en particular de 50 paquetes posee 16.318 onzas con una desviación estándar muestral de 0.102 onzas. ¿Existe evidencia de que la media poblacional es diferente de 16.34 onzas? Emplee α = 0.05

1.21 El auditor de una empresa de A y B debe de evaluar los reembolsos por gastos médicos a los empleados. La auditoría se realiza sobre una muestra de 75 reembolsos con los siguientes resultados:

- En 12 casos se reembolsó una cantidad incorrecta.
- El monto de los reembolsos fue media muestral = $1405.5, S = $ 518.25

a) Utilizando un nivel de significancia de 0.05, ¿existe evidencia de que el reembolso es menor de $ 1500.00 pesos?

B) Utilizando un nivel de significancia de 0.05 ¿existe evidencia de que la proporción de reembolsos incorrectos en la población es mayor de 0.1?

1.22 La división de inspectores del departamento de pesos y medidas está interesada en determinar si en las botellas de 2 litros procesadas en una embotelladora se ha colocado la cantidad adecuada de bebida gaseosa. La empresa embotelladora informó a los inspectores que la desviación estándar de las botellas de 2 litros es de 0.05. De una muestra conformada por 100 botellas de dos litros tomada de la planta embotelladora se obtiene una media de 1.99 litros.

Con un nivel de significancia de 0.05 ¿existe evidencia de que la cantidad media en las botellas en distinta de 2 litros?

1.23 En un artículo de "Supermarket guru has a simple mantra" se afirmó que la visita típica al supermercado es de 22 minutos. Suponiendo que se quiere probar dicha afirmación, se selecciona una muestra de 50 compradores en el supermercado local. El tiempo de compras medio para la muestra de 50 compradores fue de 25,36 minutos, con una desviación estándar de 7.24 minutos. Utilizando un nivel de significancia de 0.1 ¿existe evidencia de que el tiempo de compras medio en el supermercado local es distinto al valor de 22 minutos que se afirma?

1.24 El propietario de una cafetería desea estudiar el consumo de leche en su negocio. Selecciona una muestra aleatoria de 60 días del año con los siguientes resultados.

- Cantidad empleada promedio= 11.3 litros, s = 31 litros
- 11 días se emplearon más de 5 litros de leche light.

a) Con un nivel de significancia de 0.05 ¿existe evidencia de que la compra media es distinta de 10 litros?

b) Con un nivel de significancia de 0.05 ¿existe evidencia de que menos del 20% de los días se emplearon más de 5 litros de leche light?

c) Cual sería su respuesta al inciso a) si la media muestral fuera de 10.3 litros?

I.5 INTERVALOS DE CONFIANZA PARA DOS POBLACIONES:

I. DE C. PARA LA DIFERENCIA DE MEDIAS CON σ CONOCIDA:

$$\bar{x}_1 - \bar{x}_2 - Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Donde:

$\bar{x}_1$ y $\bar{x}_2$= medias de las muestras tomadas de las Poblaciones 1 y 2 respectivamente

σ_1^2 y σ_2^2= Variazas de las poblaciones 1 y 2 respectivamente

n_1 y n_2= Tamaño de las muestras 1 y 2 respectivamente.

μ_1 y μ_2= Proporciones de las poblaciones 1 y 2 respectivamente.

Z = Valor crítico de la distribución normal estandarizada.

I. DE C. PARA LA DIFERENCIA DE MEDIAS CON σ DESCONOCIDA:

$$\bar{x}_1 - \bar{x}_2 - t\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

Donde:

$\bar{x}_1$ y $\bar{x}_2$= medias de las muestras tomadas de las poblaciones 1 y 2 respectivamente.

s_1^2 y s_2^2= Variazas de las muestras 1 y 2 respectivamente

n_1 y n_2= Tamaño de las muestras 1 y 2 respectivamente.

μ_1 y μ_2= Proporciones de las poblaciones 1 y 2 respectivamente.

t= Valor crítico de la distribución T de Student.

INTERVALO DE CONFIANZA PARA PROPORCIONES

$$(p_1 - p_2) \pm Z\sqrt{\frac{p_1\,(1\text{-}p_1)}{n_1} + \frac{p_2\,(1\text{-}p_2)}{n_2}}$$

Donde:

p_1 y p_2= proporciones de las muestra 1 y 2 respectivamente

n_1 y n_2= Tamaño de las muestras 1 y 2 respectivamente

Z = Valor crítico de la distribución normal estandarizada.

I.6 PRUEBAS DE HIPÓTESISPARA DOS POBLACIONES

PRUEBA DE HIPÓTESIS PARA DOS PROPORCIONES

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Donde:
p_1 y p_2= proporciones de las muestra 1 y 2 respectivamente
n_1 y n_2= Tamaño de las muestras 1 y 2 respectivamente
π_1 y π_2= Proporciones de las poblaciones 1 y 2 respectivamente.
Z = Estadístico de prueba para la diferencia entre dos Proporciones.

PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA DE LAS MEDIAS CON σ CONOCIDA

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

Donde:
$\bar{x}_1$ y $\bar{x}_2$= medias de las muestras tomadas de las poblaciones 1 y 2 respectivamente
σ_1^2 y σ_2^2= Variazas de las poblaciones 1 y 2 respectivamente
n_1 y n_2 = Tamaño de las muestras 1 y 2 respectivamente.
μ_1 y μ_2 = Proporciones de las poblaciones 1 y 2 respectivamente.
Z = Estadístico de prueba, sigue una distribución normal estandarizada.

PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA DE LAS MEDIAS CON σ DESCONOCIDA:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

Donde:

$\bar{x}_1$ y $\bar{x}_2$= medias de las muestras tomadas de las poblaciones 1 y 2 respectivamente.

s_1^2 y s_2^2= Variazas de las muestras 1 y 2 respectivamente

n_1 y n_2= Tamaño de las muestras 1 y 2 respectivamente.

μ_1 y μ_2= Proporciones de las poblaciones 1 y 2 respectivamente.

T = Estadístico de prueba, sigue una distribución de T de Student con $n_1 + n_2 - 2$ grados de libertad.

PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA DE LAS MEDIAS POBLACIONES RELACIONADAS:

$$t = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}}$$

$\bar{D}$ = Promedio de las diferencias

S_D = Desviación estandar de las diferencias

n = tamaño de muestra (pares de datos)

T = Estadístico de prueba, sigue una distribución de T de Student con $n_1 - 1$ grados de libertad.

LO APLICO: Resuelva Los siguientes ejercicios:

1.21 Se utilizan dos máquinas para llenar botellas de plástico con un volumen neto de16.0 onzas. Las distribuciones de los volúmenes de llenado pueden suponerse normales, con desviaciones estándar de 0.020 y 0.025 onzas. Un miembro del grupo de ingeniería de calidad sospecha que el volumen neto de llenado de ambas máquinas no es el mismo, sin importar si este es o no de 16 onzas. De cada máquina se toma una muestra aleatoria de 10 botellas. ¿Se encuentra el ingeniero en lo correcto? Utilice un nivel de significancia de 0.05.

MAQUINA 1	16.03	16.04	16.05	16.05	16.02	16.01	15.96	15.98	16.02	15.99
MAQUINA 2	16.02	16.03	16.04	16.02	16.01	16.00	15.97	15.96	16.01	15.99

1.22 La gerente de planta de una fábrica enlatadora de jugo de naranja está interesada en comparar el rendimiento de dos diferentes líneas de producción. Como la línea número 1 es relativamente nueva, sospecha que el número de cajas que se producen al día es mayor que el correspondiente a la vieja línea 2. Se toman datos al azar durante diez días para cada línea, encontrándose que la máquina 1 produce en promedio 824.9 cajas por día y la máquina 2 818.6 cajas por día. De la experiencia con la operación de este tipo de equipo se sabe qué $\sigma^2_1 = 40$ y $\sigma^2_2=50$. Con un nivel de significancia de 0.05 determine si la nueva máquina es más rápida que la antigua.

MI PROYECTO:

- Selecciona con quienes de tus compañeros formarás equipo.
- Redacta porque trabajarás con ellos y entrega ésta redacción junto con los nombres de los integrantes de tu equipo.

PARA MI EVALUACIÓN:

Buscar un artículo en una revista científica, que incluya una prueba de hipótesis como las que analizamos en este capítulo. Redacte un breve informe que describa la prueba de hipótesis y su papel en el contexto del artículo.

LO APLICO:

1.23 Un ingeniero de tráfico realizó un estudio de velocidades vehiculares en un segmento de calle en la cual se cambió varias veces el límite de velocidad señalizado. Cuando el límite era de 30 millas por hora, el ingeniero vigiló las velocidades de 100 vehículos elegidos al azar que transitaron por la calle y observo 49 violaciones del límite de velocidad. Después de que el límite se elevó a 35 millas por hora, el ingeniero volvió a vigilar las velocidades de 100 vehículos elegidos aleatoriamente y observo 19 que violaron el límite. Establezca un intervalo de confianza de 99% para la diferencia de proporciones donde p_1 es la proporción de vehículos que excedieron el límite de velocidad menor (30 millas por hora) y p_2 es la proporción de vehículos que (en condiciones de circulación similares) excedieron el límite de velocidad mayor (35 millas por hora).

1.24 Se efectúa una prueba de degustación a dos panecillos nuevos con beneficios para la salud. El primero posee una mayor cantidad de fibra que el producto de línea, el segundo menor cantidad de grasas adicionadas.120 de 200 jueces que degustaron los panecillos con mayor cantidad de fibra dijeron que los preferían a los normales y 240 de los 500 que degustaron los bajos en grasa dijeron que los preferían a los normales. Emplee una prueba de hipótesis para determinar si es mayor la proporción de jueces que prefieren los panecillos con más cantidad de fibra que los elaborados con menos cantidad de materia grasa. Emplee un nivel de significancia de 0.05

<u>Criterios de evaluación</u>: Entregar los ejercicios resueltos en tiempo y forma, con letra Legible y con resultados correctos.

PARA MI EVALUACIÓN:

1.25 Un grupo de estudiantes de ingeniería decide estudiar si los automóviles que supuestamente no necesitan gasolina de alto octanaje rinden más millas por galón utilizando gasolina regular o de alto octanaje. Prueban varios automóviles (en condiciones similares de carretera, altitud, clima y manejo entre otras) usando ambos tipos de gasolina en cada automóvil en tiempos diferentes. El kilometraje en millas de cada uno de los autos para cada tipo de gasolina es:

Automóvil

Gasolina	1	2	3	4	5	6	7	8	9	10
Regular	15	23	21	35	42	28	19	32	31	24
Alto octanaje	18	21	25	34	47	30	19	27	34	20

¿Existe evidencia de una diferencia en el kilometraje promedio entre las gasolinas regular y de alto octanaje? Emplee $\alpha = 0.05$

1.26 Un estudio comparó la pérdida de peso en pacientes obesos con una dieta baja en grasas y pacientes obesos con una dieta baja en carbohidratos seguida por estos durante 6 meses. Una muestra de 100 pacientes que siguieron la dieta baja en grasas perdieron en promedio 7.6 lbs con desviación estándar de 3.2 lbs. Mientras que la muestra de 100 pacientes que siguieron la dieta baja en carbohidratos perdieron en promedio 6.7 lbs con desviación estándar de 3.9
Con un nivel de significancia de 0.05 ¿existe evidencia de diferencia en la media de pérdida de peso en ambos grupos de pacientes?

1.27 La gerente de desarrollo de una empresa en la que se preparan alimentos empleando diferentes métodos de conservación desea determinar si existe alguna diferencia en el tiempo de vida útil de un alimento nuevo preparado con dos diferentes fórmulas. Mediante la fórmula I se preparó un alimento cuyo periodo de vida útil tiene una desviación estándar poblacional de 110 días y de la fórmula II de 125 días. Una muestra aleatoria de 25 productos de la fórmula I indica una vida media de 375 días y una muestra similar de la fórmula II indica una media muestral de 362 días. Usando un nivel de significancia de 0.05 ¿Existe alguna evidencia de la diferencia en el periodo de vida útil de los productos elaborados empleando las dos fórmulas?

1.29 Se encuestó a 500 dueños de empresas de servicio de alimentos y bebidas **sin** estudios concluidos en el ramo de la gastronomía y a 500 dueños de empresas de servicio de alimentos y bebidas **con** estudios concluidos en el ramo de la gastronomía. Los resultados indicaron que el 74% de los dueños sin estudios concluidos poseen ganancias superiores a $500,000 anuales y el 84% de los dueños con estudios concluidos poseen ganancias superiores a $500,000 anuales.

a) ¿Existe evidencia significativa de la diferencia de las ganancias entre los dueños de empresas de servicio de alimentos y bebidas **con** y **sin** estudios concluidos en el ramo de la gastronomía? (Use $\alpha = 0.1$).

b) Construya e interprete una estimación del intervalo de confianza para la proporción de la población de dueños de empresas de servicio de alimentos y bebidas con y sin estudios concluidos en el ramo de la gastronomía que poseen ganancias superiores a $500,000 anuales.

UNIDAD II

ANÁLISIS DE REGRESIÓN Y CORRELACIÓN

PROPÓSITO DE LA UNIDAD DE COMPETENCIA: A partir del análisis de las relaciones entre una variable dependiente y una o más variables independientes identificar la aplicabilidad de las técnicas de regresión en la vida cotidiana y en situaciones relacionadas a la gastronomía, previo conocimiento de los conceptos básicos de inferencia estadística.

II.1 CORRELACIÓN

II.1.1 COEFICIENTE DE CORRELACIÓN6 DE PEARSON

En esta unidad se recordará el significado de correlación(o relación) entre dos variables, también se explicará cómo describir este tipo de relación con una ecuación que pueda emplearse para predecir el valor de una variable.

Una correlación existe entre dos variables cuando una de ellas está relacionada con la otra de alguna manera.

El coeficiente de correlación lineal *r* mide la fuerza de la relación lineal entre los valores cuantitativos apareados *x* y *y* en una *muestra*. El coeficiente de correlación lineal también se conoce como coeficiente de correlación producto momento de Pearson, en honor de Karl Pearson (1857-1936), quien lo desarrolló originalmente.

Puesto que el coeficiente de correlación lineal *r* se calcula utilizando datos muestrales, se trata de un estadístico muestral empleado para medir la fuerza de la correlación lineal entre *x* y *y.* (Triola, 2009).

FORMULA PARA EL CÁLCULO DEL COEFICIENTE DE CORRELACIÓN DE PEARSON.

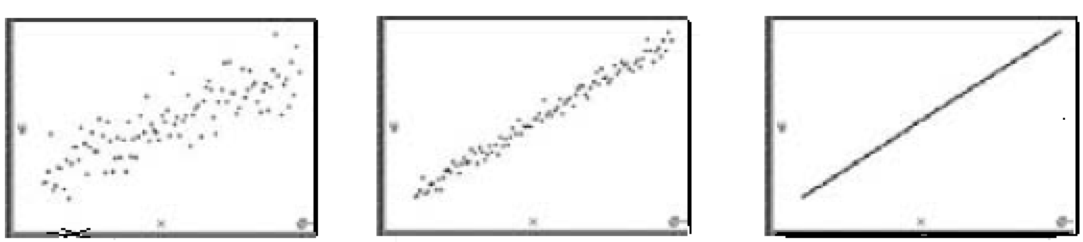
$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

n	representa el número de pares de datos presentes.
Σ	denota la suma de los elementos indicados.
Σx	denota la suma de todos los valores de x.
Σx^2	indica que cada valor de x debe elevarse al cuadrado y después deben sumarse esos cuadrados.
$(\Sigma x)^2$	indica que los valores de x deben sumarse y el total elevarse al cuadrado. Es sumamente importante evitar confundirse entre Σx^2 y $(\Sigma x)^2$
Σxy	indica que cada valor de x debe multiplicarse primero por su valor y correspondiente. Después de obtener todos estos productos, se calcula su suma.
r	representa el coeficiente de correlación lineal de una *muestra*.

Interpretación del coeficiente de correlación lineal:

El valor de *r* siempre debe estar entre +1 y -1. Si *r* se acerca a 0, concluimos que no existe una correlación lineal entre *x* y *y*, pero si *r* se acerca +1 o -1,concluimos que hay una fuerte correlación lineal entre *x* y *y.*

GRAFICAS DE DIFERENTES TIPOS DE RELACIONES LINEALES Y SUS CORRESPONDIENTES VALORES DE COEFICIENTE DE CORRELACIÓN DE PEARSON:

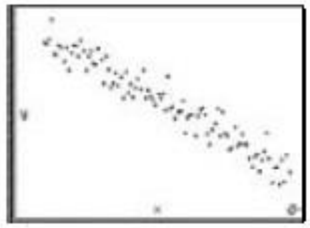


***a)* Correlación positiva:** $r = 0.851$

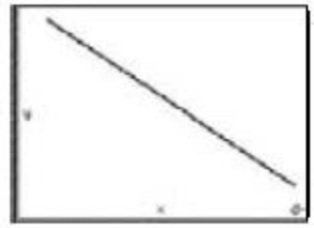
***b)* Correlación positiva:** $r = 0.991$

***c)* Correlación positiva perfecta:** $r = 1$

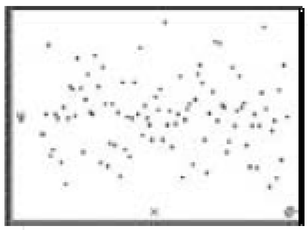
***d)* Correlación negativa:**
***r* = −0.702**



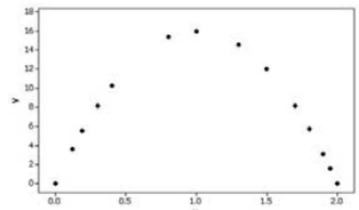
***e)* Correlación negativa:**
***r* = −0.965**



***f)* Correlación negativa perfecta:**
***r* = −1**



***g)* Sin correlación: *r* = 0**



***h)* Relación no lineal: *r* = −0.087**

FIGURA 2.1

LO APLICO:

2.1 Utilice la siguiente muestra aleatoria simple de datos para calcular el valor del coeficiente de correlación lineal *r*.

X	3	1	3	5
Y	5	8	6	4

2.2 Se realizó un estudio para determinar los efectos de no dormir en la capacidad de las personas para resolver problemas sencillos. En el estudio participaron 10 personas, dos para cada nivel de horas sin dormir. Después de un periodo específico de horas sin dormir se dio a cada persona un conjunto de problemas sencillos de sumas y se registró el número de errores. Se obtuvieron los siguientes resultados: (X= no. de horas sin dormir, Y= no. de errores)

X	8	8	12	12	16	16	20	20	24	24
Y	8	6	6	10	8	14	14	12	16	12

PARA MI EVALUACIÓN:

2.2 Los datos que se presentan a continuación son las estaturas (en pulgadas) pesos (en libras) de la supermodelos Michelle Alves, Nadia Avermann, Paris ,Kelly Dyer, Christy Turlington, Bridget Hall, Naomi Campbell, Valerie Mazza y Kristy Hume. ¿Existe una correlación entre la estatura y el peso? Si existe una correlación, ¿eso significa que hay una correlación entre la estatura y el peso de todas las mujeres adultas?

Estatura (pulgadas)	70	70.5	68	65	70	70	70	70	71
Peso (libras)	117	119	105	115	119	127	113	123	115

II.1.2 PRUEBA DE INDEPENDENCIA Ji CUADRADA

Para una tabla de contingencias que tiene *r* filas y *c* columnas, es posible generalizar la prueba de x^2 como una prueba de independencia para dos variables categóricas. Las hipótesis nula y alternativa son las siguientes:

H_0: Las dos variables categóricas son independientes (no hay relación entre ellas)
H_1: Las dos variables categóricas son dependientes (hay una relación entre ellas)

FÓRMULA PARA CALCULAR EL ESTADÍSTICO DE PRUEBA PARA LA INDEPENDENCIA:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left[f_{o_i} - f_{e_i}\right]^2}{f_{e_i}}$$

O_i = Valor observado en la i-ésimo celda.
E_i = Valor esperado en la i-ésimo celda.
K = Categorías o celdas

Se rechaza la hipótesis nula en el nivel de significancia α si el valor calculado del estadístico de prueba X^2es mayor que X^2_U, el valor crítico de la cola superior de una distribución de chi cuadrada con (r – 1)(c – 1) grados de libertad. (Levine, D.M., Etal. 2006).

CÁLCULO DE LAS FRECUENCIAS ESPERADAS:

La frecuencia esperada de cada celda es el producto del total de su fila y su columna dividido por el tamaño total de las muestras.

$$f_{e_{ij}} = \frac{(Total\ del\ renglón\ i)(total\ de\ la\ columna\ j)}{tamaño\ de\ la\ muestra}$$

Total del renglón= Suma de todas las frecuencias de la fila.
Total de la columna=Suma de todas las frecuencias de la columna.

LO APLICO:

2.3 Se ha tomado una muestra aleatoria simple de 150 consumidores de cerveza. Después de saborear cerveza ligera, clara y oscura, se les pide expresar cuál es su preferida. La siguiente tabla de contingencias resume las respuestas obtenidas.

	Cerveza preferida			
	Ligera	Clara	Oscura	Total
hombre	20	40	20	80
mujer	30	30	10	70

Determine si el tipo de cerveza preferida es independiente del sexo del consumidor. α = 0.05

II.2 REGRESIÓN LINEAL SIMPLE:

II.2.1 MODELO DE REGRESIÓN LINEAL SIMPLE

Al concluir que existe una correlación lineal entre *x* y *y*, podemos obtener una ecuación lineal que exprese *y* en términos de *x*, y la ecuación puede emplearse para predecir valores de *y* a partir de valores dados de *x*.

Esta recta se conoce como *recta de regresión* y su ecuación como *ecuación de regresión*. A partir de datos muestrales apareados, se calculan valores estimados de b_0, que es la intersección en *y*, y la pendiente b_1, de manera que se identifique una línea recta con la ecuación. En condiciones adecuadas, esa ecuación resulta útil para hacer predicciones.

La ecuación de regresión expresa una relación entre *x* (llamada variable explicativa, variable de predicción o variable independiente) y (llamada variable de respuesta o variable dependiente). La ecuación típica de una línea recta $y = mx + b$ está expresada en la forma , $y = b_0 + b_1X$ donde b_0 es el intercepto *y* (o la intersección en y) y b_1 es la pendiente. La notación dada muestra que b_0 y b_1 son estadísticos muestrales utilizados para estimar los parámetros poblacionalesβ_0 y β_1.

ECUACIÓN DE LA RECTA:

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Donde:

$\hat{y}$ = Y predicha

b_1 = Pendiente de la recta

b_0 = Ordenada al origen

n = Número de pares de datos

II.2.2 PREDICCIÓN

Las ecuaciones de regresión a menudo se utilizan para *predecir* el valor de una variable, dado algún valor particular de la otra variable. Si la recta de regresión se ajusta bastante bien a los datos, (*es un buen modelo para los datos),* entonces es sensato utilizar su ecuación para hacer predicciones, siempre y cuando no se efectúen con base en valores que rebasen las fronteras delos datos muestrales conocidos.

Además de reflexionar sobre la presencia o ausencia de una correlación lineal, también debemos considerar lo siguiente: las ecuaciones de regresión obtenidas de datos muestrales apareados con *r* muy cercana a -1 o +1 (porque los puntos en el diagrama de dispersión se acercan mucho a la recta de regresión) tienen mayores probabilidades de ser mejores modelos para los datos poblacionales que las ecuaciones de regresión de conjuntos de datos con valores de *r* que no son tan cercanos a -1 o +1 (aún cuando *r* resulte significativa). Con algunas muestras muy grandes de datos apareados, podríamos encontrar que *r* es significativa, aun cuando tenga un valor relativamente pequeño como 0.2. En este caso, la ecuación de regresión podría ser un modelo aceptable, pero las predicciones no serían tan exactas porque *r* no se acerca tanto a -1 o +1.
Si *r* no es significativa, entonces la recta de regresión no se ajusta bien a los datos, y la ecuación de regresión no debe utilizarse para hacer predicciones. (Triola, 2009).

LO APLICO:

2.4 Emplear los datos muestrales del ejemplo 2.1

Realice el diagrama de dispersión y efectúe una predicción de y para x = 4.

Lineamientos para el uso de la ecuación de regresión

1. Si no existe una correlación lineal, no utilice la ecuación de regresión para hacer predicciones.

2. Cuando utilice la ecuación de regresión para hacer predicciones, permanezca en el ámbito de los datos muestrales disponibles.

3. Una ecuación de regresión que está basada en datos antiguos no necesariamente es válida ahora.

4. No haga predicciones acerca de una población distinta de la población de donde se obtuvieron los datos muestrales.

II.2.3 VERIFICACIÓN DE SUPUESTOS

Residuales: La ecuación de regresión representa la recta que se ajusta "mejor" a los datos. Este criterio se basa en las distancias verticales entre los puntos de datos originales y la recta de regresión. Tales distancias se denominan *residuales*.

Definición de residual: Para una muestra de datos apareados (*x, y*), un residual es la diferencia entre un valor *y* muestral observado y el valor de que es el valor de *y* predicho por medio de la ecuación de regresión.

FÓRMULA DE RESIDUOS:

$$e_i = Y_i - \hat{Y}_i$$

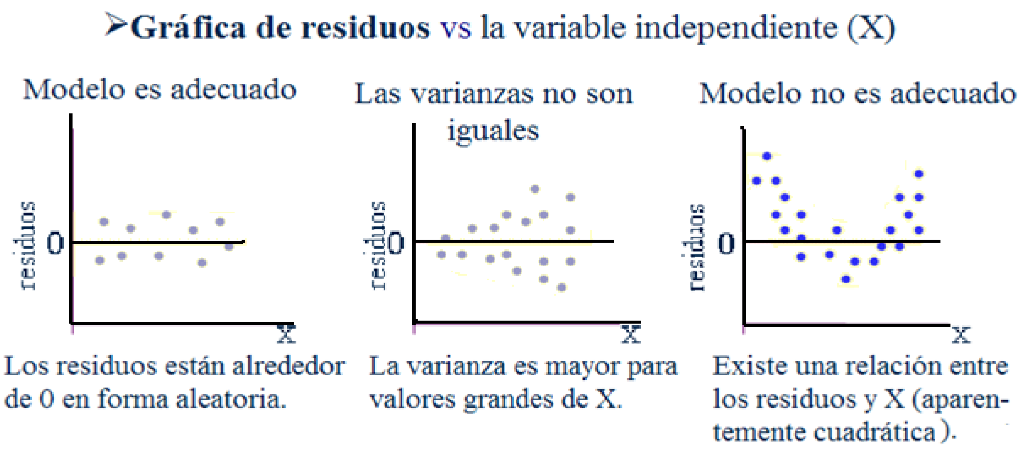
Donde:
e*i* = residuo o error observado
Y*i* = Valores de Y
$\hat{Y}_i$ = Y predicha

Gráficas residuales

En los análisis de regresión es indispensable realizar un diagrama de dispersión para verificar que el patrón de puntos se aproxime a una línea recta. También es necesario tomar en cuenta los valores extremos así como analizar la *gráfica residual* puede ser otra herramienta útil para analizar resultados de correlación y regresión, así como para verificar los requisitos o también llamados supuestos necesarios para hacer inferencias sobre una correlación y una regresión.

Para construir una gráfica residual, se utiliza el mismo eje *x* que en el diagrama de dispersión, pero en el eje de las *y* se grafican los valores residuales. Se recomienda dibujar una línea horizontal de referencia a partir del valor residual de 0 y luego se grafican los valores apareados de (*x*, *y* – y predicha). Se recomienda el uso de un programa de cómputo para facilitar el trabajo de análisis. Cuando se analiza una gráfica residual se busca un patrón en la configuración de los puntos y se utilizan los siguientes criterios:

- Si una gráfica residual no revela ningún patrón, la ecuación de regresión es una buena representación de la asociación entre las dos variables.
- Si una gráfica residual revela algún patrón sistemático, la ecuación de regresión no es una buena representación de la asociación entre las dos variable



Los supuestos de deben verificarse son:

- **Linearidad:** Para evaluar la linearidad, grafique los residuos en el eje vertical contra los valores correspondientes de X de la variable independiente en el eje horizontal. Si el modelo lineal es apropiado para los datos, no habrá un patrón aparente en el gráfico.
- **Normalidad:** Los gráficos empleados para analizar este supuesto son: el histograma y el gráfico de probabilidad normal entre otros. Si los puntos parecen ajustarse a una línea recta, puede decirse que parece indicar que los datos provienen de una distribución normal.
- **Independencia de errores:** Se grafican los residuos contra el orden en que se obtuvieron los datos. Para que se cumpla el supuesto de independencia, los datos no deben seguir ningún patrón bien definido en este gráfico.

- **Igualdad de varianza:** Evaluar la suposición de igualdad de varianza a partir de una gráfica de los residuos con X Para que se cumpla este supuesto no debe de existir ningún patrón aparente. (Levine, D.M., Etal. 2006).

PARA MI EVALUACIÓN:

2.6 Emplee los datos del problema 2.2 para efectuar una predicción del peso predicho de una supermodelo que mide 72 pulgadas.

II.3 MODELO DE REGRESIÓN LINEAL MÚLTIPLE:

En la sección anterior se explicó una relación lineal entre *dos* variables, en esta sección se presenta un método para analizar una relación lineal que incluye *más de dos* variables. Debido a la naturaleza tan compleja de las operaciones requeridas, los cálculos manuales son poco prácticos; así que en esta sección se recomienda efectuar los cálculos con un programa estadístico de cómputo.

Ecuación de regresión múltiple

Una ecuación de regresión múltiple expresa una relación lineal entre una variable de respuesta ***y*** (Variable endógena, explicada o dependiente), y dos o más variables de predicción(explicativas, exógenas o independientes). X_1 , X_2 , ..., X_n
la variable dependiente debe ser cuantitativa.

El modelo lineal viene dado por la ecuación lineal:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots\ b_k X_k + u$$

Los coeficientes (parámetros) b_1 , b_2 , ... , b_k denotan la magnitud del efecto de las variables explicativas (exógenas o independientes), esto es, representan los pesos de la regresión o de la combinación lineal de las predictoras X_1 , X_2 , ... X_k sobre la variable explicada (endógena o dependiente) **Y**. El coeficiente **b_0** se denomina término constante (o independiente) del modelo. Y al término **u** se le llama término de error del modelo o componente de **Y** no explicada por las variables predictoras.

UNIDAD III

DISEÑO Y ANÁLISIS DE EXPERIMENTOS

PROPÓSITO DE LA UNIDAD DE COMPETENCIA:

Con base en el conocimiento de la estadística descriptiva y de las técnicas de inferencia estadística, el estudiante de gastronomía podrá planear y ejecutar los experimentos más comunes en su área, así como analizar e interpretar los resultados de los mismos, para contribuir al enriquecimiento de la gastronomía.

III.1 PLANEACIÓN DE EXPERIMENTOS

Los experimentos se efectúan en muchos campos del conocimiento con la finalidad de descubrir algo acerca de un fenómeno, proceso o sistema.

El diseño de experimentos es un conjunto de técnicas activas que manipulan el proceso para inducirlo a proporcionar la información que se requiere para mejorarlo.(Gutiérrez, H.G., de la Vara, R., 2004)

A lo largo de esta unidad se estudiará el diseño de experimentos, mismo que conlleva necesariamente implicaciones estadísticas.

Para ello es necesario comprender primeramente los conceptos básicos, posteriormente se estudiarán algunos tipos de diseños experimentales como son el caso de diseño de experimentos completamente al azar y el diseño factorial.

Como trabajo final se efectuará un diseño de experimentos en el que se aplicarán los conocimientos adquiridos durante la presente unidad y algunos conocimientos de las anteriores.

Un Experimento es un cambio en las condiciones de operación de un sistema o proceso, que se hace con el objetivo de medir el efecto del cambio sobre una o varias propiedades del producto. Dicho experimento permite aumentar el conocimiento acerca del sistema. (Gutiérrez, H.G., de la Vara, R., 2004)

Existen dos formas de efectuar experimentos con la finalidad de efectuar mejoras en los productos o procesos:

-Estrategia pasiva: Se efectúa al observar o monitorear vía herramientas estadísticas hasta obtener señales útiles que permitan mejorar el proceso.

- Diseño de experimentos: Consiste en aplicar un conjunto de técnicas activas, efectuando cambios estratégicos y deliberados al proceso para provocar señales útiles que lo induzcan a proporcionar la información.

El Diseño de experimentos sigue el método científico, en el se plantean un conjunto de pruebas o corridas experimentales de manera que los datos así generados puedan ser analizados estadísticamente para obtener conclusiones válidas y objetivas. El diseño de experimentos debe ser:

1) Bien planeado
2) Conducido cuidadosamente
3) Analizado vía herramientas estadísticas
4) Interpretado correctamente

Algunos problemas que se pueden resolver en la industria con el diseño y análisis de experimentos son los siguientes:

- a dos o más proveedores del mismo material con la finalidad de elegir al que mejor cumple los requerimientos.
- Comparar varios instrumentos de medición para verificar si trabajan con la misma precisión y exactitud.
- Proponer una nueva manera de operar el proceso, variar sus condiciones y hacer cambios con el objetivo de reducir el número de defectos.
- Determinar los factores o fuentes de variabilidad que tienen impacto en la capacidad del proceso para cumplir con sus requerimientos más importantes.
- Localizar las condiciones de operación donde el proceso logra su desempeño óptimo.

- Proponer un nuevo método de muestreo igual de efectivo, pero más económico que el actual.
- Reducir el tiempo de ciclo del proceso.
- Hacer el proceso insensible o robusto a oscilaciones de variables ambientales.
- Apoyar en el diseño o rediseño del producto o proceso para mejorar su desempeño. (Gutiérrez, H.G., de la Vara, R., 2004)

En la investigación científica también es necesario emplear procesos bien planeados que proporcionen conocimientos objetivos y reales, por ello también se emplea el diseño y análisis de experimentos cuando es necesario.

III.1 CONCEPTOS BÁSICOS:

Variable de Respuesta: También llamada variable de salida, es la característica del producto cuyo valor interesa mejorar mediante el diseño de experimentos.
Factores controlables: Son variables del proceso que se pueden fijar en un punto o nivel de operación. Es la variable que manipula el investigador, para estudiar sus efectos sobre la variable dependiente.
Factores no controlables: Son variables que no se pueden controlar durante la operación normal del proceso como la luz, temperatura, humedad ambientales.
Factores estudiados: Son variables que se investigan en el experimento para observar cómo afectan o influyen en la variable de respuesta.
Nivel del Factor: Son los diferentes valores que se asignan a cada factor estudiado en el diseño experimental.
Factor Cualitativo: sus niveles se clasifican por atributos cualitativos.
Factor Cuantitativo: sus niveles son una cantidad numérica en una escala.
Tratamientos: Combinación de niveles de todos los factores.
Unidad Experimental: es la muestra de artículos que es necesario producir en una condición para obtener una medición o dato representativo.
Error Experimental: Componente del error aleatorio que refleja los errores del experimentador en la planeación y en la ejecución del experimento.
Error aleatorio: Es la variabilidad observada que no se puede explicar por los Factores estudiados, y resulta del pequeño efecto de los factores no estudiados y del error experimental.

Mediciones: Son los valores de la variable dependiente, obtenidos de las unidades experimentales luego de la aplicación de tratamientos.(Gutiérrez, H.G., de la Vara, R., 2004)

III.2 PRINCIPIOS BÁSICOS:

Se debe ser muy cuidadoso en la planeación y análisis de un experimento y el punto de partida para ello es la aplicación de los principios básicos del diseño de experimentos que son aleatorización, repetición y bloqueo, que aseguran que los datos sean útiles para lograr la validez de los resultados:

1) Aleatorización: Consiste en aplicar en forma aleatoria los tratamientos a las unidades experimentales. Este principio aumenta la posibilidad del que el supuesto de independencia de errores se cumpla. Para efectuar una aleatorización de las corridas experimentales ver el apartado VI.1.1 (Muestreo simple aleatorio)
2) Repetición: Es correr más de una vez un tratamiento o combinación de factores, permitiendo distinguir que parte de la variabilidad total de los datos se debe al error aleatorio y cual los factores.
3) Bloqueo: Es nulificar o tomar en cuenta en forma adecuada todos los factores que pueden afectar la respuesta observada. (Gutiérrez, H.G., de la Vara, R., 2004)

III.3 ETAPAS DEL DISEÑO DE EXPERIMENTOS

Las etapas del diseño de experimentos son una serie de pasos o etapas a realizar para llegar a los objetivos deseados. De forma breve estas etapas son las siguientes:

PLANEACIÓN: Formada por las siguientes partes:

1.- Detectar el problema de calidad que causa pérdidas a la compañía o que es de interés para el investigador.

2.-Determinar cuáles factores deben estudiarse o investigarse de acuerdo a la influencia que tienen sobre la respuesta.

3.-Elegir las variables de respuesta que serán medidas en cada punto del diseño y verificar que se mide de manera confiable.

4.- Seleccionar el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento.

5.- Planear y organizar el trabajo experimental.

6.- Realizar el experimento.

ANÁLISIS: Determinar el Análisis de varianza o técnica estadística que mejor describa el comportamiento de los datos.

INTERPRETACIÓN: analizar lo que ha pasado en el experimento, contrastar con las conjeturas iniciales, verificar los supuestos y elegir el tratamiento ganador.

CONCLUSIONES FINALES: Conclusiones del trabajo, toma de decisiones, medidas a implementar. (Gutiérrez, H.G., de la Vara, R., 2004)

III.3 ANÁLISIS ESTADÍSTICO: ANOVA

El Análisis estadístico generalmente se realiza mediante un análisis de varianza, por lo cual se realizará una revisión de esta prueba.

El Análisis de varianza o ANOVA (su acrónimo de **AN**alisys **O**f **Va**rianza) es un procedimiento estadístico que tiene como objetivo Analizar las diferencias entre las medias del grupo, no las varianzas.

Por lo tanto en el ANOVA es una prueba de hipótesis para probar la igualdad de medias de varias poblaciones para un factor, por lo tanto en ella se trata de probar si el efecto de un factor o tratamiento en la respuesta de un proceso o sistema es significativo, al realizar experimentos variando los niveles de ese factor

Las hipótesis para esta prueba de hipótesis son las siguientes:

$$\boxed{\begin{array}{l} Ho: \mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots = \mu_a \\ Ha: A\lg unas.\mu' s.son.diferentes \end{array}}$$

FORMULAS DEL ANOVA:

ANOVA

Source of Variation	SS	df	MS	F
Between	$SS_{between} = \sum_{j=1}^{j} n_j \left(\bar{Y}_j - \bar{Y}\right)^2$	$j-1$	$MS_{between} = \dfrac{SS_{between}}{j-1}$	$\dfrac{MS_{between}}{MS_{within}}$
Within	$SS_{within} = \sum_{j=1}^{j} \sum_{i=1}^{n_j} \left(Y_{ij} - \bar{Y}_j\right)^2$	$n-j$	$MS_{within} = \dfrac{SS_{within}}{n-j}$	
Total	$SS_{total} = \sum_{j=1}^{j} \sum_{i=1}^{n_j} \left(Y_{ij} - \bar{Y}\right)^2$	$n-1$		

Mediante la serie de cálculos se obtiene la Fcalculada y si esta resulta ser mayor que Ftablas (ver anexo 4) se rechaza H_0, aceptando Ha o H_1 donde las medias son diferentes.

Si se emplea el método del valor de *p* correspondiente a Fcalculada es menor de α se rechaza la H_0 obteniéndose la misma conclusión estadística.

- <u>Los Supuestos de Validez del modelo ANOVA son:</u>
  - Independencia de errores.
  - Normalidad.
  - Igualdad de Varianza.

PARA SABER MÁS:

Leer el artículo INTRODUCCIÓN AL DISEÑO ESTADÍSTICO DE EXPERIMENTOS en la dirección electrónica: http://ocw.univalle.edu.co/ocw/ingeniería-electronica-telecomunicaciones y afines/investigación-i/diseñoestadisticoexperimentos.pdf.

III.4 TIPOS DE DISEÑOS DE EXPERIMENTOS:

Un diseño experimental es una regla que determina la asignación de las unidades experimentales a los tratamientos. Existen muchos tipos de diseños para poder estudiar una gran diversidad de problemática. Aunque los experimentos difieren unos de otros en muchos aspectos, existen diseños estándar que se utilizan con mucha frecuencia.
La selección de uno u otro diseño depende de:
1.- El objetivo del experimento
2.- El número de factores a controlar
3.- En número de niveles de cada factor
4.- Los efectos que interesa investigar
5.- El consto del experimento, tiempo y precisión deseada

Algunos de los más utilizados son los siguientes:

III.4.1 Diseño completamente al azar:

Este es el más simple de todos los diseños y recibe ese nombre porque todas las corridas experimentales se realizan en orden aleatorio y en él se requiere comprar varios niveles de un solo factor con la finalidad de seleccionar la mejor opción de entre varias que existen. Este tipo de diseño se utiliza en experimentos que no requieren factores de bloqueo.
El diseño completamente al azar pertenece a la familia de los diseños para comparar tratamientos. Así podrían compararse por ejemplo varios materiales, procesos o máquinas.

El modelo matemático de este diseño tiene la forma:
Respuesta = Constante + Efecto Tratamiento + Error ó

$$\boxed{Y_{ij} = \mu_{..} + \tau_{.j} + \varepsilon_{ij}}$$

III.4.2 Diseño en bloques o con un factor bloque.

En este diseño el experimentador agrupa las unidades experimentales en bloques, a continuación determina la distribución de los tratamientos en cada bloque y, por último, asigna al azar las unidades experimentales a los tratamientos dentro de cada bloque.

En el análisis estadístico de un diseño en bloques, éstos se tratan como los niveles de un único factor de bloqueo, aunque en realidad puedan venir definidos por la combinación de niveles de más de un factor de ruido.

El modelo matemático de este diseño es:
Respuesta = Constante + Efecto Bloque + Efecto Tratamiento + Error

El diseño en bloques más simple es el denominado diseño en bloques completos, en el que cada tratamiento se observa el mismo número de veces en cada bloque.
El diseño en bloques completos con una única observación por cada tratamiento se denomina diseño en bloques completamente aleatorizado o, simplemente, diseño en bloques aleatorizado.
Cuando el tamaño del bloque es inferior al número de tratamientos no es posible observar la totalidad de tratamientos en cada bloque y se habla entonces de diseño en bloques incompletos.

.

III.4.3 Diseños factoriales:

En algunas ocasiones se está interesado en estudiar la influencia de dos (o más) factores en un proceso, para ello se hace un diseño de filas por columnas. En este modelo es importante estudiar la posible interacción entre los dos factores. Si en cada casilla se tiene una única observación no es posible estudiar la interacción entre los dos factores, para hacerlo hay que replicar el modelo, esto es, obtener k observaciones en cada casilla, donde k es el número de réplicas.
Los diseños factoriales producen experimentos más eficientes, pues cada observación proporciona información sobre todos los factores, y es factible ver las respuestas de un factor en diferentes niveles de otro factor en el mismo experimento. La respuesta a cualquier factor observado en diferentes condiciones indica si los factores actúan en las unidades experimentales de manera independiente.
La interacción entre factores ocurre cuando su actuación no es independiente

El modelo matemático de este diseño es:
Respuesta = Constante + Efecto Factor Fila + Efecto Factor Columna + Efecto Interacción + Error

Generalizar los diseños completos a más de dos factores es relativamente sencillo desde el punto de vista matemático, pero en su aspecto práctico tiene el inconveniente de que al aumentar el número de factores aumenta muy rápidamente el número de observaciones necesario para estimar el modelo (número elevado de corridas experimentales). (Kuehl, R. O. 2001)

El efecto de un factor es un cambio en la respuesta medida ocasionado por un cambio en el nivel de ese factor, los tres efectos de interés en un experimento factorial son los *simples,* los *principales y* los *de interacción;* Estos efectos se ilustran con las medias de población para los tratamientos factoriales de la figura 3.1

Los **efectos simples** de un factor son las comparaciones entre los niveles de un factor a un solo nivel de otro.
Los **efectos principales** de un factor son comparaciones entre los niveles de un factor promediados para todos los niveles de otro factor.
Los efectos de interacción miden las diferencias entre los efectos simples de un factor a diferentes niveles de otro. (Kuehl, R. O. 2001)

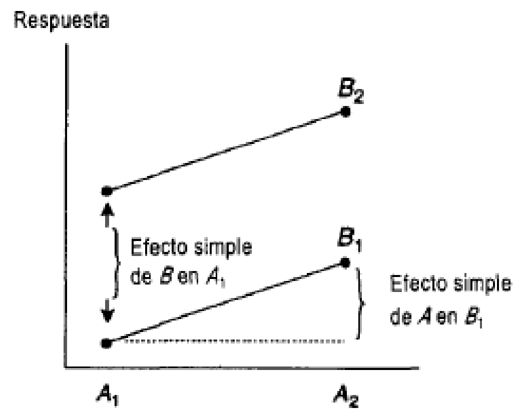


Ilustración de no interacción en un arreglo factorial

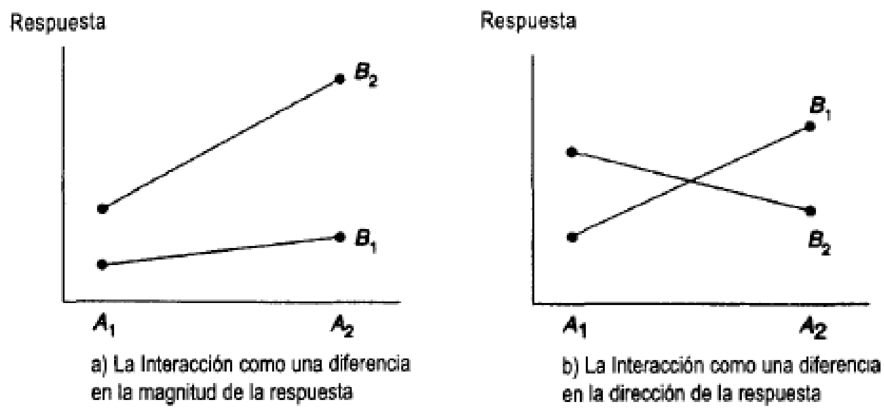


Ilustración de interacción en un arreglo factorial

FIGURA 3.1
DISEÑO DE EXPERIMENTOS (Kuehl, 2001)

Un camino alternativo a los diseños factoriales es utilizar fracciones factoriales que son diseños en los que se supone que muchas de las interacciones son nulas, esto permite estudiar el efecto de un número elevado de factores con un número relativamente pequeño de pruebas.

PARA MI EVALUACIÓN:

Elija un artículo de una publicación en su campo de estudio que proporcione los resultados de un experimento
comparativo o un estudio por observación. Identifique y describa brevemente (una o dos oraciones)
lo siguiente:
a) Hipótesis de investigación
b.)Tratamientos
c) Unidades experimentales (de observación)
d) Tipo de diseño de experimentos
e) Criterios para agrupamiento, bloqueo o compatibilidad en el estudio
f) Proporcione la cita del artículo

PARA SABER MAS:

http://viref.udea.edu.co/contenido/menu_alterno/apuntes/ac37diseno_experiment.pdf

PARA MI PROYECTO:

Entregar por escrito y engargolado el proyecto final (diseño de experimentos), trabajo que será realizado por equipo y cuya parte experimental se llevará a cabo en horas clase y en las instalaciones de la institución (talleres). Las partes del trabajo se enlistan a continuación con su ponderación.

<u>Criterios de evaluación:</u>

Evaluación del diseño de experimentos:

Planeación	40%
Realización del experimento	25%
Análisis	5%
Interpretación	5%
Conclusiones finales	5%
Evaluación del experimento	20%
Total	100%

UNIDAD IV

MUESTREO

PROPÓSITO DE LA UNIDAD DE COMPETENCIA: Con ayuda de conceptos básicos de encuestas y muestreo, se identificarán y aplicarán los esquemas de muestreo más comunes en la gastronomía con el fin de recopilar la información que sobre el fenómeno gastronómico sea de interés para el estudiante.

En esta unidad se conocerán algunos de los diferentes métodos de muestreo, también se estudiarán los métodos para calcular el tamaño muestral necesario para estimar una proporción poblacional así como para la media.
Muestra: Una muestra es un parte representativa de la población de la cual proviene.

IV.1 TIPOS DE MUESTREO:

El proceso de muestreo se inicia con la definición del marco. El marco es una lista de los elementos que constituyen la población. Los marcos son fuentes de datos como listas, directorios o mapas de la población. Las muestras se extraen de éstos marcos. Si los marcos excluyen algunos elementos de la población, los resultados serán inexactos o sesgados. (Levine, D.M., Etal. 2006).

IV.1.1 MUESTREO NO PROBABILÍSTICO:

En el muestreo no probabilístico se seleccionan los elementos o individuos sin conocer sus probabilidades de selección. Tienen ventajas como conveniencia rapidez y menor costo. Sin embargo su falta de exactitud por el sesgo de la selección y su falta de capacidad de generalización hacen que se empleen en estudios que preceden a investigaciones más rigurosas. Existen varias formas de realizar muestreo no probabilístico, algunas de ellas son las siguientes:

Muestreo de conveniencia: En el los elementos de la muestra se seleccionan solo con base en el hecho de que son fáciles, económicos o convenientes de muestrear. En algunos casos los participantes se autoeligen.

Muestreo por juicio: En este caso se recopilar las opiniones de expertos en el tema seleccionados previamente. (Levine, D.M., Etal. 2006).

IV.1.2 MUESTREO PROBABILÍSTICO:

El muestreo probabilístico comprende varios métodos en los cuales se seleccionan los elementos con base en probabilidades conocidas. Siempre que sea posible se debe utilizar métodos de muestreo probabilísticos. Las muestras basadas en estos métodos permiten elaborar inferencias sin sesgo acerca de la población de interés. En la práctica resulta complicado obtener una muestra probabilística, los cuatro tipos de muestras probabilísticas más comunes son: la aleatoria simple, la sistemática, la estratificada y la de conglomerados.

Muestra aleatoria simple: En este tipo de muestreo todos los elementos dentro del marco tienen las mismas posibilidades de selección que cualquier otro. Además, cada muestra de un tamaño fijo tiene las mismas posibilidades de selección que cualquier otra muestra del mismo tamaño. El muestreo aleatorio simple es la técnica de muestreo más básica y conforma la base de todas las demás técnicas de muestreo. (Levine, D.M., Etal. 2006).

Las muestras se seleccionan con reemplazo o sin reemplazo.

En el muestreo con reemplazo, después de seleccionar un elemento, este se devuelve al marco, donde tiene la misma posibilidad de ser seleccionado de nuevo.

En el muestreo sin reemplazo, una vez seleccionado un elemento, no podrá seleccionarse de nuevo.

Una forma de obtener una muestra aleatoria simple es empleando una tabla de números aleatorios.

<u>Pasos para el empleo de la tabla de números aleatorios:</u>

- 1º paso: localizar el punto de comienzo en una tabla de números al azar.
- 2^0 paso: Numerar los elementos que conforman la muestra, por ejemplo si n=30. numerar los elementos del 01 al 30.
- 3^{er} paso: Utilizando el punto de comienzo obtenido en el 1º paso, seleccionar n números de dígitos siguiendo siempre la misma dirección en la tabla de números

aleatorios localizando los números que constituyen su muestra. (numerados del 01 al 30 en este caso)

Muestreo sistemático: En este tipo de muestreo se elige un individuo al azar y a partir de el, a intervalos constantes se eligen los demás hasta completar la muestra. Los intervalos se calculan de la siguiente forma:

$$K = N/n$$

Donde:
N= Tamaño de la población
n= tamaño de la muestra
K= numero de elementos después de los cuales se Seleccionará un elemento de la muestra.

Muestreo por conglomerados: Se divide el área de la población en secciones (o conglomerados) representativas de la población, luego se eligen aleatoriamente algunos de estos conglomerados, y después se elige a todos los miembros de los conglomerados seleccionados. Los conglomerados son designaciones de suceso natural, como países, distritos electorales, manzanas de la ciudad o territorios de venta.

LO APLICO:

Se tiene una población de N = 100 elementos y se desea extraer una muestra n = 20. Obtenga una muestra aleatoria por:

a)Muestreo simple aleatorio.

b)Muestreo sistemático.

Muestreo estratificado: En él se divide a la población en clases o estratos que por naturaleza tienen Características similares y se selecciona aletoriamente un número de individuos de cada estrato. Esta cantidad de individuos es proporcional al número de elementos que componen el estrato.

LO APLICO:

En una empresa turística en la cual trabajan 600 empleados. 200 de ellos trabajan en el área de alimentos y bebidas, 150 en limpieza y mantenimiento y 150 en el área de servicios generales. Se desea tomar una muestra de 40 empleados. Calcule cuantos empleados de cada estrato se muestrearían.

IV.2 CÁLCULO DEL TAMAÑO DE MUESTRA:

Existen dos tipos de fórmulas para el cálculo del tamaño de muestra según el propósito con el que esa muestra es tomada:

- Cuando con esa muestra será calculada una media.
- Cuando se calcularán proporciones.

IV.2.1 CÁLCULO DE TAMAÑO DE MUESTRA PARA LA MEDIA:

FÓRMULA PARA EL TAMAÑO DE MUESTRA PARA LA MEDIA (POBLACIÓN INFINITA)

$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

Donde:

n = Tamaño de la muestra
σ = Desviación estándar poblacional
E = Margen de error de muestreo deseado
Z α/2 = Puntuación Z crítica basad en el nivel de confianza deseado

FÓRMULA PARA EL TAMAÑO DE MUESTRA PARA LA MEDIA (POBLACIÓN FINITA):

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

Donde:
N = Tamaño de la población
n = Tamaño de la muestra
σ = Desviación estándar poblacional
E = Margen de error de muestreo deseado
Z α/2 = Puntuación Z crítica basada en el nivel de confianza deseado

IV.2.2 CÁLCULO DE TAMAÑO DE MUESTRA PARA LA PROPORCIÓN:

FÓRMULA PARA EL TAMAÑO DE MUESTRA PARA PROPORCIONES (POBLACIÓN FINITA)

$$n = \frac{[z_{\alpha/2}]^2\hat{p}\hat{q}}{E^2}$$

Donde:
n = tamaño de muestra
p =proporción de la muestra
= x/n= proporción muestral de x éxitos
q = 1-p proporción muestral de x fracasos
E = Margen de error de muestreo deseado
Z α/2 = Puntuación Z crítica basada en el nivel de confianza deseado

FÓRMULA PARA EL TAMAÑO DE MUESTRA PARA PROPORCIONES (POBLACIÓN INFINITA)

$$n = \frac{N\hat{p}\hat{q}[z_{\alpha/2}]^2}{\hat{p}\hat{q}[z_{\alpha/2}]^2 + (N - 1)E^2}$$

Donde:

n = tamaño de muestra
N = tamaño de la población
p =proporción de la muestra
= x/n= proporción muestral de x éxitos
q = 1-p proporción muestral de x fracasos
E = Margen de error de muestreo deseado
Z α/2 = Puntuación Z crítica basada en el nivel de confianza deseado

LO APLICO:

En los ejercicios 4.1 al 4.4 , use el margen de error, el nivel de confianza y la desviación estándar poblacional σ indicados para calcular el tamaño de muestra mínimo requerido para estimar una media poblacional μ desconocida.

4.1 Margen de error: 0.5 pulgadas, nivel de confianza: 95%, σ= 2.5 pulgadas.

4.2 Margen de error: 0.25 segundos, nivel de confianza: 99%, σ=5.4 segundos.

4.3 Margen de error: $1, nivel de confianza: 90%, σ= $12.

4.4 Margen de error: 1.5 mm, nivel de confianza: 95%, σ= 8.7mm.

4.5 Suponga que queremos estimar la puntuación media del CI de la población de profesores de estadística. ¿Cuántos profesores de estadística deben seleccionarse al azar para efectuar pruebas de CI, si queremos tener una confianza del 95% de que la media muestral estará dentro de 2 puntos de CI de la media poblacional y desviación estándar= 10?

Para saber más:

4.6 Un estudio en la Secretaría de educación pública requiere estimar la media de la puntuación de coeficiente intelectual para la población de profesores. ¿Cuantos profesores deben seleccionarse al azar para efectuar pruebas de coeficiente intelectual si se quiere tener un nivel de confianza del 95% de que la media muestral estará dentro de 2 puntos de I.Q. de la media poblacional? Se conoce que la prueba está diseñada para que $\sigma = 15$

4.7 En el departamento de vinculación de una universidad se quiere calcular el salario promedio ganado actualmente por sus egresados de la carrera de gastronomía. A cuantos egresados se requiere encuestar si se conoce que la desviación estándar poblacional es de $500 para los siguientes casos:

a) Nivel de confianza del 95% y error de ± $50

b) Nivel de confianza del 95% y error de ± $100

c) Nivel de confianza del 95%, error de ± $50 Si la población de egresados de la carrera de gastronomía en esa universidad es de 2,300

4.8 El médico encargado de la salud de los alumnos de un campus universitario quiere estimar el porcentaje de estudiantes que ha sufrido una Enfermedad trasmitida por alimentos al consumir alimentos en la cafetería del mismo campus.

¿Cuántos estudiantes que se seleccionen al azar deben encuestarse para tener una confianza del 95% de que el porcentaje de la muestra tiene un margen de error de 3 puntos porcentuales?

a)Suponga que no existe información disponible que pueda utilizarse para estimar el porcentaje de estudiantes que han sufrido ETA's al acudir a la cafetería.

b)Suponga que otro estudio indicó que el 7% de los estudiantes universitarios han sufrido ETA's al acudir a la cafetería del Campus.

Para saber más:

4.9 En un estudio sociológico se requiere determinar el porcentaje actual de hogares en que se utiliza el correo electrónico. ¿Cuántos hogares deben encuestarse para tener un nivel de confianza del 90% de que el porcentaje muestral es erróneo por no más de 4 puntos porcentuales?

a) Utilice el siguiente resultado de un estudio pionero: en el 76% de hogares se empleaba el correo electrónico.

b) Suponga que no tenemos información previa que sugiera un posible valor de p.

4.10 En una encuesta de Gallup del año pasado el 29% de adultos dijeron que usaban internet para comprar al menos 5 veces al año. Calcule el tamaño de muestra de adultos que es necesario encuestar para poder calcular la proporción de adultos que en este año comprarán al menos en 5 ocasiones para 99% de confianza y error de muestreo de ±0.05 si la población en estudio es de 5000 personas.

4.11 Un director de personal quiere comparar la efectividad de dos métodos de entrenamiento para trabajadores industriales a fin de efectuar cierta operación de montaje. Se divide un número de operarios en dos grupos iguales: el primero recibe el método de entrenamiento 1, y el segundo, el método 2. Cada uno realizará la operación de montaje y se registrará el tiempo de trabajo. Se espera que las mediciones para ambos grupos tengan una desviación estándar aproximadamente de 2 minutos y error de muestreo de 0.5. Si se desea que la estimación dela diferencia en tiempo medio de montaje sea correcta hasta por un minuto, con una probabilidad igual a 0.95,

¿Cuántos trabajadores se tienen que incluir en cada grupo de entrenamiento?

4.12 Una campaña reciente se diseñó para convencer a los propietarios de automóviles de que debían llenar sus neumáticos con nitrógeno en lugar de aire. Aun costo de aproximadamente $5 por neumático, se supone que el nitrógeno tiene la ventaja de escaparse con una rapidez mucho menor que el aire, por lo que la presión ideal se puede mantener de manera más consistente. Antes de gastar grandes sumas en la publicidad del nitrógeno, sería pertinente realizar una encuesta para determinar el porcentaje de propietarios de automóviles que estarían dispuestos a pagar por el nitrógeno. ¿Cuántos propietarios de automóviles elegidos al azar se deben encuestar? Suponga que deseamos tener una confianza del 98% de que el porcentaje de la muestra está dentro de tres puntos porcentuales del porcentaje real de todos los dueños de automóviles que estarían dispuestos a pagar por el nitrógeno.

4.13 Se planea realizar un estudio de tiempos para estimar el tiempo medio de un trabajo, exacto dentro de 4 segundos y con una probabilidad de 0.90, para terminar un trabajo de montaje. Si la experiencia previa sugiere que σ= 16 seg. mide la variación en el tiempo de montaje entre un trabajador y otro al realizar una sola operación de montaje, ¿cuántos operarios habrá que incluir en la muestra?

BIBLIOGRAFÍA

1.- Anderson, D. R., Sweeney, D. J. y Williams, T. A. (2008). Estadística para Administración y Economía. México: International Thomson Editores.

2.- Gutiérrez, H.G., de la Vara, R. (2004) Análisis y diseño de experimentos. México: McGraw-Hill Interamericana.

3.- Kuehl, R. O. (2001) Diseño de experimentos. Principios estadísticos de diseño y análisis de investigación. México: Thomson Learning.

4.- Levine, D.M., Krehbiel, T.C., Berenson, M.L. (2006). Estadística para administración. México: Pearson Educación.

5.- Montgomery, D. (2002) Diseño y Análisis de experimentos. México: Ed. Limusa Wiley.

6.-. Scheaffer, R. L., Ott, L. R. y Mendenhall, W. (2006). Elementos de muestreo. (2006). Ed. Paraninfo Thomson

7.- Triola, M.F. (2009) Estadística. México: Pearson Educación

ANEXOS

El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

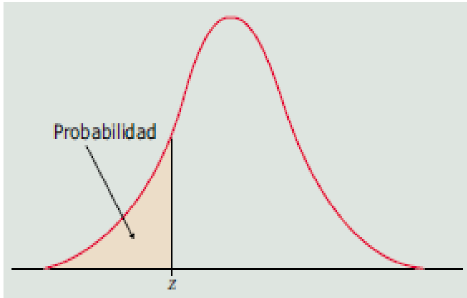


TABLA A: Probabilidades de la normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

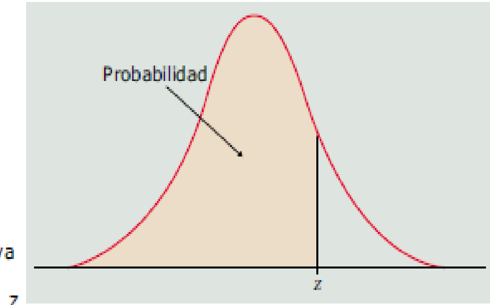
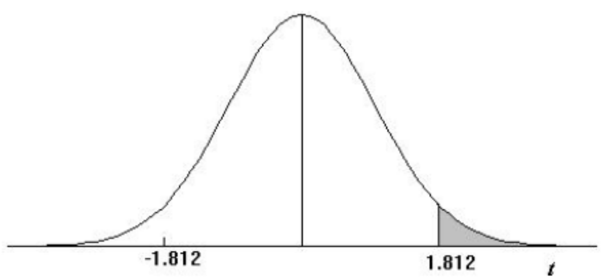


TABLA A: Probabilidades de la normal estándar (cont.)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TABLA 2: DISTRIBUCIÓN t DE STUDENT



Ejemplo

Para r = 10 grados de libertad:

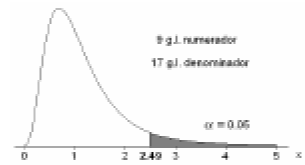
P[t > 1.812] = 0.05
P[t < -1.812] = 0.05

r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

TABLA 3: DE DISTRIBUCIÓN DE F

$\alpha = 0.05$

VALORES CRÍTICOS DE LA DISTRIBUCIÓN F



Grados de libertad del denominador	Grados de libertad del numerador																			
	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	240	1000
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.90	245.95	248.02	249.26	250.10	251.14	252.20	253.25	253.79	254.19
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.49	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57	8.55	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.64	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43	4.40	4.38	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74	3.70	3.69	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30	3.27	3.25	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01	2.97	2.95	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79	2.75	2.73	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62	2.58	2.56	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49	2.45	2.43	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38	2.34	2.32	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30	2.25	2.23	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22	2.18	2.15	2.14
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16	2.11	2.09	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11	2.06	2.03	2.02
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06	2.01	1.99	1.97
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02	1.97	1.94	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.90	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95	1.90	1.87	1.85
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.84	1.82
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89	1.84	1.81	1.79
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86	1.81	1.79	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84	1.79	1.76	1.74
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.74	1.72
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80	1.75	1.72	1.70
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79	1.73	1.70	1.68
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.68	1.66
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.75	1.70	1.67	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74	1.68	1.65	1.63
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64	1.58	1.54	1.52
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78	1.73	1.69	1.63	1.58	1.51	1.48	1.45
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53	1.47	1.43	1.40
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.79	1.70	1.64	1.60	1.54	1.48	1.41	1.37	1.34
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.77	1.68	1.62	1.57	1.52	1.45	1.38	1.33	1.30
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43	1.35	1.31	1.27
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.68	1.58	1.52	1.47	1.41	1.33	1.24	1.18	1.11