



UTPL

La Universidad Católica de Loja

Modalidad Abierta y a Distancia

Aplicación de Matemáticas y Estadística en Tecnologías de la Información

Guía didáctica



Facultad de Ingenierías y Arquitectura

Departamento de Ciencias de la Computación y Electrónica

Aplicación de Matemáticas y Estadística en Tecnologías de la Información

Guía didáctica

Carrera	PAO Nivel
▪ Tecnologías de la Información	VIII

Autora:

Valdiviezo Díaz Priscila Marisela



ESTA_4009

Asesoría virtual
www.utpl.edu.ec

Universidad Técnica Particular de Loja

Aplicación de Matemáticas y Estadística en Tecnologías de la Información

Guía didáctica

Valdiviezo Díaz Priscila Marisela

Diagramación y diseño digital:

Ediloja Cía. Ltda.

Telefax: 593-7-2611418.

San Cayetano Alto s/n.

www.ediloja.com.ec

edilojacialtda@ediloja.com.ec

Loja-Ecuador

ISBN digital - 978-9942-39-215-2



Los contenidos de este trabajo están sujetos a una licencia internacional Creative Commons **Reconocimiento-NoComercial-CompartirIgual 4.0 (CC BY-NC-SA 4.0)**. Usted es libre de **Compartir** — copiar y redistribuir el material en cualquier medio o formato. **Adaptar** — remezclar, transformar y construir a partir del material citando la fuente, bajo los siguientes términos: **Reconocimiento-** debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante. **No Comercial-** no puede hacer uso del material con propósitos comerciales. **Compartir igual-** Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original. No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

20 de septiembre, 2021

Índice

1. Datos de información.....	7
1.1. Presentación de la asignatura.....	7
1.2. Competencias genéricas de la UTPL.....	7
1.3. Competencias específicas de la carrera.....	7
1.4. Problemática que aborda la asignatura.....	7
2. Metodología de aprendizaje.....	8
3. Orientaciones didácticas por resultados de aprendizaje.....	9
Primer bimestre	9
Resultado de aprendizaje 1	9
Contenidos, recursos y actividades de aprendizaje	9
Semana 1	9
Unidad 1. Modelos determinísticos vs. estocásticos.....	9
1.1. Modelo determinístico	10
1.2. Modelo estocástico.....	11
Semana 2	12
1.3. Probabilidades.....	12
1.4. Probabilidad condicional	13
Actividad de aprendizaje recomendada	14
Autoevaluación 1	15
Semana 3	17
Unidad 2. Distribución de probabilidad Discreta.....	17
2.1. Variable aleatoria y distribución de probabilidad	17
2.2. Distribuciones de probabilidad de Bernoulli	18
Semana 4	19
2.3. Distribuciones de probabilidad Binomial.....	19
Actividad de aprendizaje recomendada	21
2.4. Distribuciones de probabilidad de Poisson	21
Actividad de aprendizaje recomendada	23

Autoevaluación 2	24
Resultado de aprendizaje 2	26
Contenidos, recursos y actividades de aprendizaje	26
Semana 5	26
Unidad 3. Pruebas de Hipótesis	26
3.1. Formulación de Hipótesis	26
Actividades de aprendizaje recomendadas	29
3.2. Prueba de hipótesis respecto a una proporción	30
Actividad de aprendizaje recomendada	31
Resultado de aprendizaje 2	31
Contenidos, recursos y actividades de aprendizaje	31
Semana 6	32
3.3. Prueba de hipótesis respecto a una media y a una desviación estándar.	32
Autoevaluación 3	34
Actividades finales del bimestre	37
Semana 7	37
Semana 8	37
Segundo bimestre	38
Resultado de aprendizaje 3	38
Contenidos, recursos y actividades de aprendizaje	38
Semana 9	38
Unidad 4. Correlación y regresión	38
4.1. Correlación	38
4.2. Coeficiente de correlación lineal	41
Semana 10	43
4.3. Regresión	43

Actividad de aprendizaje recomendada 47

Autoevaluación 4 48

Resultado de aprendizaje 4 50

Contenidos, recursos y actividades de aprendizaje 50

Semana 11 50

Unidad 5. Análisis de Varianza..... 50

5.1. ANOVA de un factor 50

Actividad de aprendizaje recomendada 53

Semana 12 53

5.2. ANOVA de dos factores 53

Actividad de aprendizaje recomendada 55

Autoevaluación 5 57

Resultado de aprendizaje 5 59

Contenidos, recursos y actividades de aprendizaje 59

Semana 13 59

Unidad 6. Pruebas no paramétricas 59

6.1. Conceptos básicos de las pruebas no paramétricas..... 59

6.2. Pruebas del signo..... 60

Actividad de aprendizaje recomendada 62

Semana 14 62

6.3. Prueba de rangos con signo de Wilcoxon 62

Autoevaluación 6 64

Actividades finales del bimestre..... 66

Semana 15 66

Semana 16 66

4. Solucionario 67

5. Referencias bibliográficas 73



1. Datos de información

1.1. Presentación de la asignatura



1.2. Competencias genéricas de la UTPL

- Orientación a la investigación e innovación

1.3. Competencias específicas de la carrera

- Construir modelos específicos de ciencias de la computación mediante esquemas matemáticos y estadísticos, para propiciar el uso y explotación eficiente de datos e información.

1.4. Problemática que aborda la asignatura

El medio social en el cual se desenvolverá el profesional en TI, maneja una gran cantidad de información que debe ser analizada, modelada y transmitida. Por ello se requiere soluciones donde el estudiante pueda aplicar técnicas adecuadas que puedan llevarlo a diferentes conclusiones.

Así como transmitir los resultados de la aplicación de estas técnicas usando un lenguaje estadístico.

Por otro lado, en la actualidad la estadística es la herramienta más usada en todo tipo de investigación. Todo trabajo de nivel universitario o profesional de carácter investigativo debe estar basado en un diseño experimental y uso de técnicas de contraste de hipótesis. Por ello la importancia del estudio de esta asignatura.



2. Metodología de aprendizaje

Con el objeto de contribuir al logro de los resultados de aprendizaje, durante el desarrollo de esta asignatura se aplicará el proceso metodológico de **aprendizaje basado en problemas** que conlleva la investigación y análisis que debe realizar el profesional en formación para obtener una solución frente a un problema planteado en la asignatura.

La metodología de aprendizaje basado en problemas le ayudará en la adquisición de habilidades y conocimientos que pueden ser aplicados en la solución de problemas orientados a las tecnologías de la información y comunicación, así como innovar en la aplicación de soluciones tecnológicas para el desarrollo de los contenidos.

Con base a esto, el estudiante obtendrá los conocimientos y habilidades a través de problemas y situaciones reales. Será él mismo quien busque e investigue la forma de solucionar un problema tal cual como lo haría durante su actividad profesional.

Para mayor información sobre el aprendizaje basado en problemas, revise el texto "[Aprendizaje basado en problemas](#)".



3. Orientaciones didácticas por resultados de aprendizaje



Primer bimestre

Resultado de aprendizaje 1

- Describe la diferencia entre el análisis estocástico y el análisis determinista.

Con base en el resultado de aprendizaje, usted podrá distinguir entre un modelo determinístico y un modelo estocástico, identificará las características de cada uno de ellos y los elementos necesarios para realizar un análisis estocástico y determinístico.

Contenidos, recursos y actividades de aprendizaje



Semana 1

Iniciaremos con el estudio de los contenidos de la unidad 1.

Unidad 1. Modelos determinísticos vs. estocásticos

Es hora de revisar los temas 1.1 Modelo determinístico y 1.2 Modelo estocástico, que serán cubiertos en esta primera semana, usted revisará aspectos generales relacionados con estos dos modelos y conocerá algunas de sus características. Estos temas serán complementados con la lectura del [Recurso web](#), el cual le permitirá profundizar en cada tipo de modelo y conocer algunos ejemplos en los cuales pueden ser aplicados.

1.1. Modelo determinístico

Es importante recordar qué es un modelo. De acuerdo a Rustom (2012), “Todo modelo es una representación aproximada de la realidad” (p. 31). Esto significa que un modelo no necesariamente va a representar de forma exacta la realidad, pero si debe ser apropiado y simple.

Ahora ¿Qué es un modelo determinístico?, un modelo determinístico se entiende como un modelo matemático donde los datos se suponen son conocidos con certeza, es decir, se tiene disponible toda la información para el modelado. En estos modelos las entradas iniciales producirán las mismas salidas o resultados, es decir, los resultados esperados son predecibles. Además, no se contempla la existencia del azar, o el principio de incertidumbre en el modelado. De acuerdo al diccionario de [Geotecnia](#), un modelo determinista está relacionado con la creación de entornos basados en simuladores para el estudio de situaciones hipotéticas. Estos modelos no pueden predecir correctamente la mayor parte de sus características.

Podemos decir que, un modelo determinístico se expresa en forma de ecuaciones que relacionan las variables y constantes asociadas con un problema. Veamos el siguiente ejemplo:

Ejemplo:

La ecuación de una recta $y = a + bx$ es un modelo matemático determinístico, que puede ser utilizada para el cálculo de predicciones. En este modelo las variables satisfacen exactamente las ecuaciones.

De esta forma, cuando las variables que forman parte de la ecuación no son aleatorias, podemos aplicar un modelo determinista.

Veamos otro caso: $y = 2x$, al igual que en el ejemplo anterior las variables se relacionan de forma determinística, ya que dado un valor para una variable el valor de la otra se determina de manera exacta sin errores.

Los modelos determinísticos son muy aplicados en problemas de administración, para tomar decisiones óptimas en base al análisis de los resultados obtenidos a partir de estos modelos.

Ahora lo invito a revisar otros ejemplos de problemas que pueden ser resueltos con modelos determinísticos:

Modelo determinístico: Ejemplos

Con relación a su campo de estudio, es posible aplicar este tipo de modelos si hay una relación determinista entre las entradas de un sistema informático, el tiempo de ejecución y las salidas o resultados finales asociados al sistema. Donde además no se contemple la existencia al azar.

Luego de la revisión de los ejemplos propuestos acerca de los modelos determinísticos es momento de conocer otro tipo de modelo matemático.

1.2. Modelo estocástico

A diferencia de los modelos anteriores, un modelo estocástico es aquel modelo probabilístico donde algún elemento no se conoce con anticipación, las acciones están dadas por azar, es decir que en estos modelos se incorpora el principio de incertidumbre. En los modelos probabilísticos (o estocásticos) el comportamiento de un sistema es no determinista, donde sus estados siguientes están dados por elementos aleatorios. De ahí el concepto de proceso estocástico donde los datos que se manejan varían con el tiempo, es decir están formados por variables dadas aleatoriamente que dependen de argumentos o parámetros.

Es momento de revisar, algunos ejemplos de problemas que se modelan con procesos estocásticos:

Modelos estocásticos- ejemplos

Ahora veamos en el ámbito informático que se podría realizar con este tipo de modelos. Bien, se puede aplicar un modelo estocástico para predecir el rendimiento de un sistema informático, para predecir el tráfico de una red, para el análisis de redes sociales, etc.

Es decir que, los modelos estocásticos se utilizan cuando tenemos la presencia de variables aleatorias, cuando hay procesos que varían con el tiempo, o si existe incertidumbre en el proceso modelado.

Después de la lectura del tema 1.1 y 1.2 usted está en la capacidad de identificar a un modelo determinístico de uno estocástico (probabilístico). Para ello, le invito a revisar el recurso web [Economía de la Empresa](#), específicamente el punto 3, modelos de gestión de stock.

¿Cómo le fue con la revisión de este recurso?, me imagino pudo analizar la diferencia entre los modelos determinísticos y estocásticos, así como sus aplicaciones. Ahora usted conoce que las variables que utilizan los modelos determinísticos son no aleatorias, por el contrario, en los modelos estocásticos una característica de ellos, es que alguna de las variables que utilizan son aleatorias y por ende en estos modelos se pueden realizar estimaciones en términos de probabilidad.



Semana 2

En esta semana se revisan algunos fundamentos de probabilidades y ejemplos de problemas que se pueden trabajar con modelos estocásticos también conocidos como modelos probabilísticos.

Es momento de leer comprensivamente de que tratan estos temas recurriendo al capítulo 4 de su texto básico. Con la lectura de este capítulo usted recordará algunos conceptos de probabilidades, propiedades o reglas, así como el tema de probabilidad condicional empleada en el área de la estadística.

1.3. Probabilidades

La probabilidad de un evento aleatorio se puede definir como el grado de posibilidad de que dicho evento ocurra. Este grado de ocurrencia es expresado comúnmente con valores entre 0 y 1, o bien en porcentajes entre el 0% y 100%, respectivamente.

Como parte del estudio de este tema, es necesario que profundice en dos componentes desarrollados en el capítulo 4 de su texto básico, estos son: 1) Conceptos básicos de probabilidad, 2) Regla de la suma y regla de la multiplicación.

Una vez realizada la lectura de este tema lo invito a revisar el siguiente ejercicio de probabilidad, que le permitirá poner en práctica los conceptos de probabilidad.

Ejercicio:

En una empresa de telefonía móvil existen 87 celulares almacenados en bodega, y 68 son marca Samsung, si se escoge uno al azar, ¿cuál es la probabilidad de que este sea marca Samsung?

Solución:

Dividir la cantidad de celulares que son marca Samsung (68) por la cantidad total de celulares almacenados en bodega (87).

$$P(\text{celular marca Samsung}) = \frac{\# \text{ Celulares marca Samsung}}{\# \text{ Total celulares}} = \frac{68}{87} = 0.7816$$

Como usted habrá observado hay una probabilidad del 78,16%, de que el celular escogido al azar sea marca Samsung.

Ahora lo invito a completar el ejercicio identificando cuál de los tres métodos para calcular la probabilidad de un evento mencionados en el texto básico es utilizado en este ejercicio. Para ello, le sugiero que en su libreta de apuntes anote la respuesta a este planteamiento de tal forma que posteriormente le sirva para recordar lo estudiado en este apartado.

1.4. Probabilidad condicional

La probabilidad condicional se emplea para determinar la ocurrencia de un evento dada la ocurrencia de otro evento. Para conocer como calcular este tipo de probabilidad lo invito a desarrollar la lectura del capítulo 4, relacionado con la probabilidad condicional.

Como pudo darse cuenta en su texto básico para que se trate de una probabilidad condicionada necesariamente se debe suponer la ocurrencia de otro evento, siendo de esta manera su notación $P(B|A)$, la cual se lee la

probabilidad de B dado A, es decir, la probabilidad de que el evento B ocurra dada la ocurrencia del evento A.

Luego de las lecturas realizadas sobre los temas tratados en esta unidad, usted está en la capacidad de dar respuesta a las siguientes interrogantes:

- ¿Cuándo decimos que un evento es muy poco probable?
- ¿Qué es un evento simple y un espacio muestral?
- ¿Cuáles son los métodos comunes para encontrar la probabilidad de un evento?
- ¿Cuál es la diferencia entre evento independiente y dependiente?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

¿Cómo le fue con las respuestas a las interrogantes?, estoy segura de que estas preguntas le permitirán fortalecer lo aprendido en esta unidad.



Actividad de aprendizaje recomendada

En su texto básico capítulo 4 se resuelven algunos ejercicios de probabilidades usando la regla de la suma y de la multiplicación, así como de probabilidad condicional, lo invito a revisarlos, y así dar solución al ejercicio 25 “Redundancia en discos duros de computadora”, propuesto en la sección regla de la suma y regla de la multiplicación: 4-2 Habilidades y conceptos básicos.

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Seguramente tiene dificultades para desarrollar este ejercicio, tomar en cuenta que para dar solución al ejercicio indicado debe revisar el principio de redundancia: aplicación importante de la regla de la multiplicación mencionado en su texto básico.

Para dar por concluido con el estudio de la primera unidad lo invito a desarrollar la autoevaluación 1.



Autoevaluación 1

Seleccione la alternativa correcta en cada uno de los siguientes enunciados.

1. Un modelo determinístico es un modelo matemático en donde:
 - a. Existe incertidumbre en el proceso de modelado.
 - b. Las mismas condiciones iniciales (entradas) producirán el mismo resultado (salida).
 - c. La salida es aleatoria.
2. En un modelo estocástico:
 - a. Al menos una de las variables es tomada al azar.
 - b. Los datos se conocen con certeza.
 - c. La salida del modelo siempre será la misma.
3. Un evento es simple cuando:
 - a. Se puede descomponer en más eventos.
 - b. Es un espacio muestral.
 - c. No se puede descomponer en componentes más simples.
4. Una probabilidad pequeña como 0.001, se puede interpretar como:
 - a. Un evento que es bastante probable que ocurra.
 - b. Un evento que ocurre en pocas ocasiones.
 - c. El evento tiene la misma probabilidad de ocurrir o no ocurrir.
5. La notación $P(A \text{ o } B)$ indica:
 - a. La probabilidad de que ocurra el evento A u ocurra el evento B.
 - b. La probabilidad de que ocurra el evento A y el evento B.
 - c. La regla de la multiplicación.

6. Cuando la probabilidad de un evento es afectada por el conocimiento de que algún otro evento ha ocurrido, entonces se habla de:
- Probabilidad complementaria.
 - Probabilidad condicional.
 - Probabilidad disjunta.
7. La probabilidad condicional de que ocurra el evento B dado que el evento A ya ha ocurrido se expresa con la notación:
- $P(B \text{ y } A)$
 - $P(B) \cdot P(A)$
 - $P(A)$
8. El método formal para encontrar $P(A)$ consiste en:
- Dividir la probabilidad de que el evento A y B ocurra por la probabilidad del evento A.
 - Asumir que la probabilidad A ha ocurrido y luego calcular la probabilidad de B.
 - Dividir la probabilidad de que el evento B ocurra por la probabilidad del evento A y B.
9. Si tengo la siguiente notación $P(A|M)$, ésta se lee como:
- Probabilidad de M dado A.
 - Probabilidad de A dado M.
 - Probabilidad de A y M.
10. El método intuitivo para encontrar $P(B|A)$:
- Es más probable que resulte en errores.
 - Se basa en la manipulación de la fórmula.
 - Es mucho más fácil de usar.

[Ir al solucionario](#)



Unidad 2. Distribución de probabilidad Discreta

2.1. Variable aleatoria y distribución de probabilidad

Una vez que usted ha revisado los conceptos de probabilidad, lo invito a iniciar con el estudio de las distribuciones de probabilidad discretas, éstas a diferencia de las distribuciones de probabilidad normal o de variable continua que estudió en la materia de Estadística, utilizan valores contables. Es decir que, con este tipo de distribución se describe la probabilidad de ocurrencia de una variable aleatoria discreta. Pero ¿A qué nos referimos con variable aleatoria discreta?, para aclarar esta interrogante es necesario realizar una lectura del texto básico capítulo 5, sobre el tema “Distribuciones de probabilidad”, y revise los conceptos de variable aleatoria, distribución de probabilidad, y conozca la diferencia entre variable aleatoria discreta y continua.

Bien, como podrá haberse dado cuenta las variables aleatorias pueden ser discretas o continuas. Algunos ejemplos de ellas se ilustran en el siguiente recurso interactivo:

Variables continuas y discretas

¿Le pareció interesante los ejemplos?, supongo habrá notado que la diferencia entre cada tipo de variable aleatoria radica en el valor que puede tomar la misma.

Continuando con esta temática, es importante recordar que una distribución de probabilidad puede ser representada mediante una tabla o gráfica, y que la distribución de probabilidad tiene algunos parámetros por ejemplo: la media, varianza, desviación estándar, etc.

Por tanto, del estudio de este tema es conveniente que tenga claro cómo se representa una distribución de probabilidad, los parámetros de una distribución de probabilidad discreta, la definición de valor esperado y cuando el valor de una variable aleatoria es significativamente alto o bajo.

2.2. Distribuciones de probabilidad de Bernoulli

Cómo usted conoce a menudo en la vida real nos encontramos con eventos que solo tienen dos resultados, por ejemplo, en el caso de un nuevo emprendimiento tendríamos: éxito o fracaso, en el caso de un examen: aprueba o reprueba, operatividad de un sistema: funciona o falla, etc. En estas situaciones podemos aplicar el concepto de probabilidad de Bernoulli o ensayos de Bernoulli.

La distribución de Bernoulli toma valor 1 para la probabilidad de éxito (p) y valor 0 para la probabilidad de fracaso ($q=1-p$). Por ejemplo: Si X es una variable aleatoria que mide el “número de éxitos”, y se realiza un único experimento con dos posibles resultados (éxito o fracaso), se dice que la variable aleatoria X se distribuye como una distribución de Bernoulli de parámetro p . Su notación es: $X=Be(p)$

Su función de probabilidad viene dada por:

$$f(x) = P(x) = p^x \cdot q^{n-x} \quad (1)$$

Esto quiere decir que, si partimos de un experimento que tiene solo 2 resultados posibles, a uno de ellos lo llamaremos éxito y al otro fracaso. Esto se conoce como ensayo de Bernoulli. Por ejemplo, realicemos un experimento sencillo que consiste en preguntar a un estudiante de esta asignatura, que es seleccionado al azar, si le gusta la programación. Si dice que sí, lo consideramos un éxito, si dice que no, lo consideramos un fracaso. Por tanto, recuerde que:

Un ensayo de Bernoulli se conoce como a todo experimento aleatorio que tiene solo dos resultados posibles, que suelen llamarse éxito o fracaso

Par completar lo estudiado en este tema lo invito a revisar el video denominado [Distribución de Bernoulli](#)

Como usted podrá haberse dado cuenta en una distribución de Bernoulli el experimento se realiza en un ensayo y éste tiene dos resultados posibles:

éxito o fracaso. Además, tanto la probabilidad de éxito o fracaso no cambian de un ensayo a otro.



Semana 4

2.3. Distribuciones de probabilidad Binomial

Existe otro tipo de distribución discreta que es la distribución Binomial, en la mayoría de los casos, las personas se confunden con los dos términos 'Bernoulli' y 'Binomial'. Bien veamos en qué se diferencian.

Realice una lectura comprensiva del texto básico capítulo 5, sección "Distribuciones de probabilidad binomial". Durante la lectura, centre su atención en la definición y los ejemplos mencionados en esta sección.

Ahora que ya conoce a qué se refiere esta distribución de probabilidad, podrá observar que la diferencia entre una distribución Binomial y de Bernoulli radica principalmente que en la distribución binomial se realizan n ensayos idénticos, es decir n ensayos de Bernoulli. Por el contrario, en la distribución de Bernoulli el experimento solo se realiza en un ensayo. Esto lo podemos comprobar comparando las funciones de probabilidad dadas para cada tipo de distribución de probabilidad. Lo invito a revisar el siguiente ejemplo.

Ejemplo:

Considere que, para un experimento binomial, donde p es la probabilidad de "éxito" y $1-p$ la probabilidad de "fracaso" en un solo ensayo; entonces la probabilidad de obtener x éxitos en n ensayos, está dada por la función de probabilidad:

$$f(x) = P(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \quad (2)$$

Donde el coeficiente binomial $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

Si observa en la función de probabilidad de Bernoulli no se incluye $(n \times)$.

En su texto básico se especifica la función de probabilidad con el coeficiente binomial incorporado, por ello es necesario su revisión.

Usted habrá notado además que en el texto básico se mencionan tres métodos para calcular probabilidades binomiales: 1) Usando la fórmula de probabilidad binomial, 2) Usando tecnología y 3) Basados en la Tabla A1 del apéndice A del texto básico. Este último punto lo vamos a obviar en esta asignatura, ya que para esto se utilizará mejor la tecnología.

A continuación, detallo algunas herramientas tecnológicas que usted podría utilizar:

- **Statdisk:** paquete de análisis estadístico, incluye más de una serie de funciones y ensayos, con sus correspondientes datos y representaciones gráficas. Usted puede hacer uso de esta herramienta de manera online ingresando al siguiente enlace [Statdisk](#).
- **R:** contiene numerosas funciones para la realización de cálculos asociados a distintas distribuciones de probabilidad, y otro tipo de operaciones estadísticas.
- **Microsoft Excel:** aunque es un software muy conocido, este pone a disposición un conjunto de funciones para efectuar cálculos sobre los datos.

Un aspecto importante de esta sección también es el cálculo de los siguientes parámetros: media, varianza y desviación estándar para este tipo de distribución de probabilidad. En el texto básico en el apartado mencionado anteriormente, se muestran como calcular estos parámetros. Para profundizar en la lectura que previamente hizo, lo invito a revisar el siguiente **ejercicio**:

[Distribución de probabilidad binomial](#)

Bien, como se mencionó anteriormente las probabilidades binomiales se pueden calcular mediante tres métodos, en este caso el ejercicio mencionado en el recurso interactivo está resuelto usando la fórmula de probabilidad binomial (ecuación 2).

Tenga en cuenta además, que usted puede calcular la probabilidad binomial usando el método 2, por ejemplo Excel, seleccionando la función estadística “DISTR.BINOM.N”, al ingresar los parámetros correspondientes obtendrá el mismo resultado. Lo invito a hacer el ensayo.

Recuerde, antes de usar la fórmula de probabilidad binomial, verificar que se encuentre ante un experimento binomial.

Ahora que ya conoce a que hace referencia la probabilidad binomial, es momento de poner en práctica lo aprendido con el desarrollo de la siguiente actividad recomendada.



Actividad de aprendizaje recomendada

Con base al ejercicio de probabilidad binomial presentado en el recurso interactivo de la sección 2.3, sobre el examen de Computación con 10 preguntas de Verdadero o Falso, calcular:

- a. La probabilidad de obtener 5 aciertos en el examen
- b. La probabilidad de no obtener ningún acierto

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Resuelva estos ejercicios usando la fórmula de probabilidad binomial y compruebe los resultados mediante el uso de tecnología.

2.4. Distribuciones de probabilidad de Poisson

Otro tipo de distribución de probabilidad discreta es la distribución de Poisson, en la cual la ocurrencia de un evento sucede durante un intervalo dado. Lo invito a realizar una lectura de su texto básico capítulo 5, sección “Distribuciones de probabilidad de Poisson”, para conocer su definición, la fórmula utilizada para calcularla, los requisitos que se deben cumplir y los parámetros de esta distribución.

Luego de la lectura desarrollada, usted estará en la capacidad de distinguir una distribución de Poisson de una Binomial.

Algunos **ejemplos** en los que se puede emplear esta distribución a parte de los mencionados en su texto básico son:

- El número de llamadas telefónicas en una central telefónica que se da en un día (hora, minutos, etc.).
- El número de pacientes que llegan por hora (día, mes) al hospital de la UTPL.
- El número de clientes que llegan a una oficina por hora.
- El número de servidores web que se accede por minuto.

Bien, ahora que ya conoce de que trata esta distribución lo invito a revisar el siguiente ejercicio:

Ejercicio:

Los mensajes de correo que llegan a un servidor de correo electrónico lo hacen siguiendo una distribución de Poisson con una tasa promedio de 0.1 mensajes por minuto. Calcular la probabilidad de que lleguen dos mensajes en un minuto:

Siguiendo la fórmula especificada en su texto básico, los datos del problema serían:

Parámetro $\mu=0.1$

Variable $x=2$

La probabilidad de que lleguen dos mensajes estaría dada por:

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} = \frac{0.1^2 \cdot e^{-0.1}}{2} = 0.004524$$

Recuerde que puede comprobar los resultados con el uso de tecnología por ejemplo Excel mediante la función POISSON.

En este ejemplo hemos calculado la probabilidad de que lleguen dos mensajes dado el valor promedio. Pero ¿Qué sucede cuando tengo solo una tabla con el valor de probabilidad para diferentes valores aleatorios? En este caso recuerde el cálculo de los parámetros de una distribución de probabilidad: media, varianza y desviación estándar.

En resumen, tener presente que este tipo de distribución es aplicable cuando queremos modelar situaciones en las que nos interesa determinar la probabilidad de que ocurra un determinado número de eventos durante un intervalo de tiempo o espacio. Analicemos este planteamiento realizando la siguiente actividad recomendada.



Actividad de aprendizaje recomendada

Use la Distribución de Poisson para resolver el siguiente ejercicio: ***Una empresa que ensambla computadoras determina que el número de componentes que fallan antes de cumplir 80 horas de actividad es una variable aleatoria de Poisson. Considerando que el número promedio de fallos es 8, calcular la probabilidad de que falle un componente en 20 horas.***

Imagino se preguntará cómo resolver el ejercicio, en este caso tener en cuenta que en 80 horas se tiene un promedio de fallos de 8, por tanto, en 20 horas se tendría una tasa promedio de fallos de 2. Entonces los datos del problema a considerar serían $\lambda=2$ y $x=1$.

Para el desarrollo de esta actividad recuerde revisar los ejemplos resueltos en su texto básico.

Es momento de medir los conocimientos adquiridos desarrollando la segunda autoevaluación.



Autoevaluación 2

Seleccione la alternativa correcta en cada uno de los siguientes enunciados.

1. Una distribución de probabilidad es:
 - a. Una variable que tiene un único valor numérico.
 - b. Una descripción que da la probabilidad para cada valor de la variable aleatoria.
 - c. Colección de valores que es finito o contable.
2. Los valores de la media, la varianza y desviación estándar en una distribución de probabilidad se conocen como:
 - a. Parámetros.
 - b. Variables aleatorias discretas.
 - c. Significancia estadística.
3. El valor esperado de una variable aleatoria discreta es:
 - a. El valor medio de los resultados.
 - b. La varianza de los resultados.
 - c. Un valor significativo bajo.
4. La distribución de probabilidad binomial se utiliza cuando:
 - a. Los resultados pertenecen a más de dos categorías.
 - b. Se tienen circunstancias en las que hay una sola categoría.
 - c. Los resultados pertenecen a dos categorías.
5. En una distribución de Bernoulli el experimento se realiza en:
 - a. Varios ensayos.
 - b. Un solo ensayo.
 - c. Ningún ensayo.

6. Para encontrar las probabilidades binomiales, es posible:
- Usar la fórmula de probabilidad binomial o algún software.
 - Recurrir algún texto que contenga los resultados de las probabilidades.
 - Buscar en internet la solución.
7. ¿Cuál de las siguientes alternativas es una distribución de probabilidad discreta?
- Distribución de Gauss.
 - Distribución de Poisson.
 - Distribución Multinomial.
8. Si tengo el enunciado: "Número de usuarios de internet que ingresan a un sitio web en un día", se refiere a un ejemplo de:
- Distribución de Bernoulli.
 - Distribución Binomial.
 - Distribución de Poisson.
9. En una distribución de probabilidad de Poisson, las ocurrencias deben ser:
- Dependientes entre sí.
 - Medianamente dependientes entre sí.
 - Independientes entre sí.
10. Una distribución de Poisson está determinada por:
- La media y la varianza.
 - Únicamente la media.
 - La media, desviación estándar y varianza.

[Ir al solucionario](#)

Resultado de aprendizaje 2

- Dada una situación de ejemplo, formular una hipótesis nula apropiada para una pregunta simple y realizar una prueba apropiada para determinar su aceptabilidad.

Mediante el desarrollo de ejemplos prácticos, usted conocerá como formular una hipótesis nula y alterna, y estará en la capacidad de aplicar algunas pruebas de hipótesis para validar resultados.

Contenidos, recursos y actividades de aprendizaje



Semana 5

Hemos llegado a la última unidad del Primer bimestre, en la cual usted conocerá un aspecto importante de la inferencia estadística, la prueba de hipótesis.

Unidad 3. Pruebas de Hipótesis

3.1. Formulación de Hipótesis

Cuando se trabaja en un proyecto o investigación, muchas veces se requiere tomar una decisión entre aceptar o rechazar un supuesto sobre los parámetros. Este supuesto recibe el nombre de hipótesis. Gran parte de los problemas de toma de decisiones, experimentos, estudios de ingeniería, etc. pueden formularse como problemas de pruebas de hipótesis, por ello la importancia del estudio de esta unidad. Bajo este contexto lo invito a realizar una lectura de su texto básico capítulo 8. Pruebas de Hipótesis, sección “Fundamentos de las pruebas de hipótesis”, donde conocerá algunos conceptos básicos de las pruebas de hipótesis y los Errores Tipo I y Tipo II, que se suscitan al rechazar o no rechazar una hipótesis nula.

Realizada la lectura usted estará en capacidad de dar respuesta a lo siguiente:

- ¿Qué es una hipótesis y una prueba de hipótesis?
- ¿Cuál es la diferencia entre una hipótesis nula y una alterna?
- ¿Cuáles son los pasos que se siguen para realizar una prueba de hipótesis?
- ¿A qué hace referencia en nivel de significancia?
- ¿Cuál es la diferencia entre Error tipo I y tipo II?

Nota: conteste las interrogantes en un cuaderno de apuntes o en un documento Word.

En el siguiente cuadro se presentan dos conceptos importantes que se resaltan en su texto.

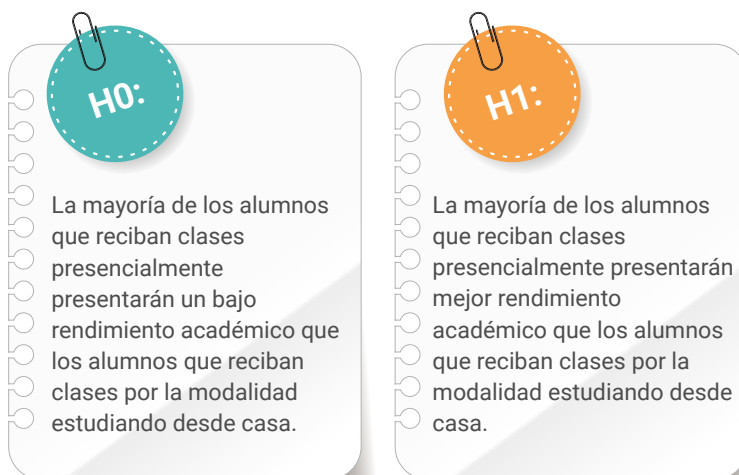
Hipótesis	Prueba de Hipótesis
Suposición que se realiza de manera previa al desarrollo de una determinada investigación.	Procedimiento que conduce a una decisión sobre una hipótesis en particular.

Recuerde que la conclusión de un experimento o proyecto siempre se refiere a la hipótesis nula (rechaza o acepta H_0 en lugar de la H_1).

En la siguiente figura se presenta un ejemplo de hipótesis nula y alterna:

Figura 1.

Ejemplo de hipótesis nula y alterna



Nota. Ejemplo de la formulación de una hipótesis nula y alterna.

Siguiendo los pasos señalados en su texto básico puede dar una representación simbólica de la afirmación representada en la figura anterior. Lo invito a realizarlo considerando el ejemplo 1 de los conceptos básicos de las pruebas de hipótesis.

Una vez que se tiene las expresiones simbólicas (paso 3 de su texto básico), es conveniente seleccionar el nivel de significancia e identificar el estadístico de prueba, para esto revise en qué consisten los pasos 4 y 5 explicados en su texto básico, podrá conocer las distribuciones muestrales que se utilizan para cada parámetro, así como su correspondiente estadístico de prueba.

Como usted observará los estadísticos de prueba se pueden utilizar para determinar si es posible rechazar la hipótesis nula.

Una vez que haya completado la lectura de los pasos para realizar una prueba de hipótesis, lo invito a conocer a qué se refiere el Error tipo I y tipo II.

Errores tipo I y tipo II

La decisión de rechazar o no rechazar una hipótesis nula a veces es correcta o incorrecta, para esto se utilizan dos tipos de errores. En su texto básico se describen en qué consiste cada uno, por ello es necesario que realice una lectura comprensiva de los Errores tipo I y tipo II de la sección “Fundamentos de pruebas de hipótesis”.

Veamos ahora el siguiente ejemplo donde se muestra la aplicación de estos dos tipos de errores:

Ejemplo:

Con base en la siguiente afirmación: Un tratamiento médico diseñado para aumentar la probabilidad de recuperación de un paciente de coronavirus leve es efectivo, de esta manera la probabilidad de recuperación de un paciente es $p > 0.5$, considerar la siguiente hipótesis nula y alternativa para determinar el Error tipo I y tipo II.

$$H_0: p = 0.5$$

$$H_1: p > 0.5$$

La afirmación que describe el Error tipo I y tipo II sería:

Error tipo I: error de rechazar la hipótesis nula verdadera.

En este caso se concluye que el tratamiento médico es efectivo cuando en realidad no tiene ningún efecto. Esto significa que estamos dando por aceptado que $p > 0.5$ cuando en realidad $p = 0.5$

Error tipo II: no rechazar la hipótesis nula cuando ésta es falsa.

Concluir que el procedimiento médico no tiene ningún efecto, cuando en realidad es efectivo para aumentar la probabilidad de recuperación de un paciente de coronavirus leve. Es decir, realmente $p > 0.5$, pero no se concluye esto.

Bien, ahora es momento de poner en práctica lo estudiado en esta sección realizando la siguiente actividad:



Actividades de aprendizaje recomendadas

Recurra a su texto básico y resuelva el ejercicio 5: Afirmación sobre datos en línea, de la sección de 8-1 habilidades y conceptos básicos, de los Fundamentos de las pruebas de hipótesis.

Para dar solución a este ejercicio le recomiendo considerar los siguientes literales:

- Expresar la afirmación original en forma simbólica
- Identificar la hipótesis H_0 y H_1
- Realizar estimaciones subjetivas, para decidir si los resultados son significativamente bajos o significativamente altos.
- Proporcionar afirmaciones que identifiquen el Error tipo I y tipo II, que corresponden a la afirmación dada en el ejercicio.

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

3.2. Prueba de hipótesis respecto a una proporción

Una vez que conoce los pasos para realizar una prueba de hipótesis, es necesario revisar como se prueba una hipótesis sobre una proporción poblacional p . Para tratar este tema lo invito a realizar una lectura de su texto básico, sección “Prueba de una hipótesis respecto a una proporción”.

De la lectura realizada, usted podrá darse cuenta de que hay ciertos elementos que son claves para probar este tipo de hipótesis, es posible usar una distribución normal como una aproximación a una distribución binomial, este método es conocido como aproximación normal.

En el texto básico también se mencionan algunos otros métodos equivalentes para probar una hipótesis respecto a una proporción, estos son: método del valor P , método del valor crítico y método del intervalo de confianza, estudiados en la sección anterior.

Ahora revisemos el ejercicio propuesto a continuación, el cual muestra los pasos que son necesarios seguir utilizando el método del valor crítico.

Ejercicio:

Para facilitar la comprensión del tema estudiado, le invito a revisar el siguiente recurso:

Pasos para realizar una prueba de hipótesis

Como usted habrá observado, para dar solución al ejercicio es conveniente primero identificar los datos del problema y luego utilizar el método más apropiado para probar la hipótesis. Es necesario considerar que en el paso 6 especificado en la solución del ejercicio, para el cálculo del valor crítico correspondiente al nivel de confianza dado, pueden basarse también en la tabla A-2 del capítulo 6 de su texto base, sección “Distribución normal estándar” y observar cómo se calcula este valor. En internet también pueden consultar los valores críticos para un nivel de confianza dado.

Bien, a diferencia del ejercicio descrito anteriormente, también se puede utilizar las pruebas de hipótesis en la medición del desempeño de diferentes tipos de software; para probar la utilidad de programas de ordenador para el desarrollo de un proyecto; para analizar encuestas aplicadas a una proporción de consumidores de productos tecnológicos antes de lanzar al mercado un nuevo producto, etc.

Es momento de poner en práctica lo aprendido realizando la siguiente actividad.



Actividad de aprendizaje recomendada

Usando los resultados de la encuesta *USA Today* que menciona el enunciado de los ejercicios 1-4 planteado en su texto básico, sección 8-2 Habilidades y conceptos básicos, desarrolle los literales 1 al 3.

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Para resolver esta actividad puede basarse en los elementos claves para una prueba de hipótesis.

Resultado de aprendizaje 3

- Dada una muestra de dos variables aleatorias, calcula la prueba t , z -test, y Chi-cuadrado pruebas estadísticas y determina si existe significación estadística.

A través del presente resultado de aprendizaje usted conocerá que existen diferentes estadísticos para realizar una prueba de hipótesis e identificará cuál aplicar dependiendo de la afirmación que se tenga.

Contenidos, recursos y actividades de aprendizaje



Ahora que entendemos como formular hipótesis es importante conocer cómo realizar una prueba de hipótesis de una afirmación acerca de una media poblacional y de una desviación estándar. Recuerde que existen ciertos estadísticos que son necesarios utilizar para realizar las pruebas de hipótesis.

3.3. Prueba de hipótesis respecto a una media y a una desviación estándar.

En la prueba de hipótesis respecto a la media se utilizará el estadístico de prueba t , el cual sigue una distribución llamada distribución t de *Student*. Es necesario que revise en que consiste esta prueba de hipótesis desarrollando una lectura comprensiva de la sección “Prueba de hipótesis respecto a una media”.

Con la lectura de esta sección usted conocerá como realizar una prueba de hipótesis acerca de la media poblacional con desviación estándar desconocida y conocida. Revise los ejemplos desarrollados y los métodos equivalentes para la prueba de t -student que conducen a las mismas conclusiones.

Como usted observará al igual que la prueba de hipótesis anterior se pueden aplicar similares pasos para los métodos: del valor P , de valor crítico, e intervalo de confianza.

Una vez comprendido el tema, realice otra lectura de la sección 8-4 Prueba de una hipótesis respecto a una desviación estándar o varianza. ¿Cuál es el dato estadístico de prueba utilizado en este caso? Supongo que pudo observar que el estadístico de prueba utilizado es X^2 (ji-cuadrada o chi-cuadrada) y que este posee algunas propiedades. Bien, ahora veamos como se aplica analizando los ejemplos presentados en la sección correspondiente a este tema en su texto básico.

Además de los ejemplos del texto base, usted puede aplicar las prueba de hipótesis respecto a una media poblacional, por ejemplo para: probar la incidencia de las redes sociales en el estilo de vida de las personas; probar

afirmaciones relacionadas con la velocidad de transferencia de datos, uso de una red social mediante la aplicación de encuestas a usuarios de Internet, etc.

En el caso de pruebas de hipótesis respecto a una desviación estándar se puede utilizar para analizar la variación de los resultados de exámenes de una determinada asignatura, en los tiempos de respuesta de aplicaciones web, etc.

La siguiente tabla resume las pruebas de hipótesis más comunes y sus respectivos estadísticos de prueba revisados en esta unidad.

Tabla 1.
Estadísticos de prueba para diferentes pruebas de hipótesis

Prueba de hipótesis para:	Estadístico de prueba
Un proporción	$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$
La media de una población con desviación estándar desconocida	$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$
La media de una población con desviación estándar conocida	$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
Una desviación estándar o varianza	$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$

Nota. Adaptado de la tabla 8-2 presentada en Triola (2018).

Como usted se habrá dado cuenta existen algunos requisitos que se deben cumplir para aplicar cada tipo de prueba de hipótesis. En la tabla 8-2 de su texto básico puede revisar un resumen de los requisitos que debe cumplir cada tipo de prueba a realizar.

Es momento de medir el nivel de conocimiento adquirido del estudio de esta unidad, desarrollando la autoevaluación 3.



Autoevaluación 3

Seleccione la alternativa correcta en cada uno de los siguientes enunciados.

1. En estadística, una hipótesis se define como:
 - a. Un procedimiento para probar una hipótesis.
 - b. Una afirmación o declaración sobre una propiedad de una población.
 - c. El valor de un parámetro poblacional.
2. Las pruebas de hipótesis también se conocen como:
 - a. Pruebas de significación.
 - b. Hipótesis nula.
 - c. Intervalo de confianza.
3. La hipótesis alternativa es:
 - a. Una afirmación de que el valor de un parámetro poblacional es igual a algún valor declarado.
 - b. Una afirmación de que el parámetro tiene un valor que difiere en alguna forma de la hipótesis nula.
 - c. Expresión simbólica de que el parámetro es igual al valor fijo bajo consideración.
4. El nivel de significancia se entiende como:
 - a. La probabilidad de rechazar la hipótesis nula cuando ésta es falsa.
 - b. El valor de probabilidad utilizado para aceptar la hipótesis nula.
 - c. La probabilidad de rechazar erróneamente la hipótesis nula cuando es verdadera.

5. El método que permite probar una hipótesis, tomando una decisión al comparar el estadístico de prueba con el (los) valor(es) crítico(s), se conoce como:
- Método del valor P.
 - Método del intervalo de confianza.
 - Método del valor crítico.
6. El error de rechazar la hipótesis nula cuando en realidad es verdadera se conoce como:
- Error tipo I.
 - Error tipo II.
 - Error tipo III.
7. En la prueba de hipótesis respecto a una media, para la prueba t , el método del valor P, del valor crítico, y del intervalo de confianza, se consideran:
- Equivalentes, porque todos conducen a las mismas conclusiones.
 - No equivalentes, porque no todos conducen a las mismas conclusiones.
 - Disyuntos, porque no llevan a los mismos resultados.
8. Los métodos para una prueba de hipótesis respecto a una desviación estándar o varianza utilizan la distribución:
- χ^2 cuadrada.
 - t Student.
 - Normal.
9. El dato estadístico de prueba para probar una hipótesis sobre una media es:
- Prueba t .
 - Prueba z .
 - Prueba X^2 .

10. El dato estadístico de prueba para probar una hipótesis sobre una proporción es:

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

a.

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

b.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

c.

[Ir al solucionario](#)



Actividades finales del bimestre



Semana 7

Una vez completado el estudio de las unidades del primer bimestre lo invito a desarrollar la siguiente actividad, la cual le permitirá poner en práctica los conocimientos aprendidos y entrenarse para la evaluación presencial.

Por lo expuesto, le solicito revise el siguiente recurso:

[Actividad Final del Primer Bimestre](#)



Semana 8

En la presente semana debe prepararse para la evaluación presencial del Primer bimestre, para ello le recomiendo realizar un repaso de las unidades que se ha presentado, revisar las actividades recomendadas y las autoevaluaciones propuestas en cada unidad, a fin de comprender mejor los contenidos que debe estudiar. Si tiene alguna inquietud, recuerde que puede consultar a través de la plataforma EVA.



Segundo bimestre

Resultado de aprendizaje 4

- Dado un conjunto de datos, determina el mejor modelo de regresión.

Usted tendrá una visión clara de cómo determinar la correlación entre variables, y cómo aplicar la teoría de regresión, para determinar el mejor modelo a utilizar en diferentes casos de estudio.

Contenidos, recursos y actividades de aprendizaje



Semana 9

Es momento de iniciar con el estudio de los contenidos del segundo bimestre, empecemos revisando la Unidad 4.

Unidad 4. Correlación y regresión

Analicemos el tema de correlación entre variables y cómo desarrollar modelos de regresión lineal y no lineal.

4.1. Correlación

Cuando hablamos de correlación nos referimos a la proporcionalidad entre dos variables, o a la relación recíproca que se da entre dos o más variables. Para comprender mejor esta definición es necesario que realice una lectura comprensiva del capítulo 10. Correlación y Regresión, de su texto básico. Ahí encontrará los conceptos básicos relacionados a la correlación. Además, mediante el análisis de gráficas podrá identificar cuándo existe una correlación lineal.

De la lectura realizada usted podrá darse cuenta de que una herramienta importante para determinar la relación entre dos variables cuantitativas es el diagrama de dispersión.

Veamos el siguiente **ejemplo**:

Ejemplo

En un estudio sobre la utilización de una impresora en una oficina del Departamento de Informática, se midió en un día los minutos transcurridos entre los sucesivos usos y el número de páginas impresas, produciéndose los siguientes resultados:

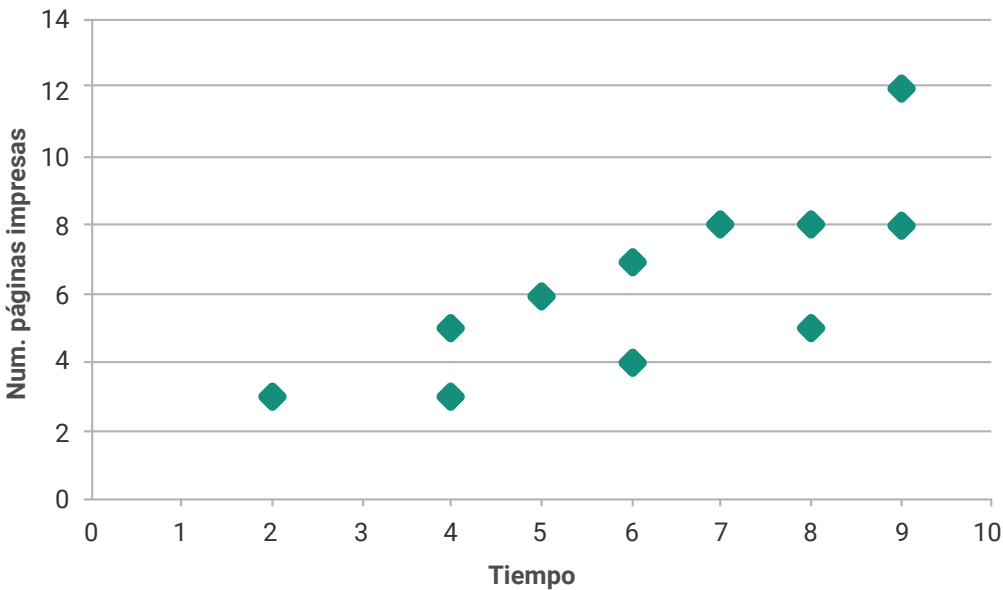
Tabla 2.
Tiempo en minutos del número de páginas impresas

Tiempo (X)	Núm. Páginas Impresas (Y)
9	8
5	6
4	3
6	4
8	5
4	5
7	8
6	7
8	8
2	3
9	12
9	12

Nota: Tabla que muestra los resultados del uso de una impresora en una oficina de informática.

Usando la Herramienta R, dibujaremos el diagrama de dispersión para observar la relación entre la variable independiente “X” que hace referencia al tiempo transcurrido y la variable dependiente “Y” que representa el número de páginas impresas. En estos casos decimos que Y depende de X.

Figura 2.
Diagrama de dispersión del uso del tiempo transcurrido de sucesivos usos de la impresora y el número de páginas impresas. Correlación positiva $r=0.802$



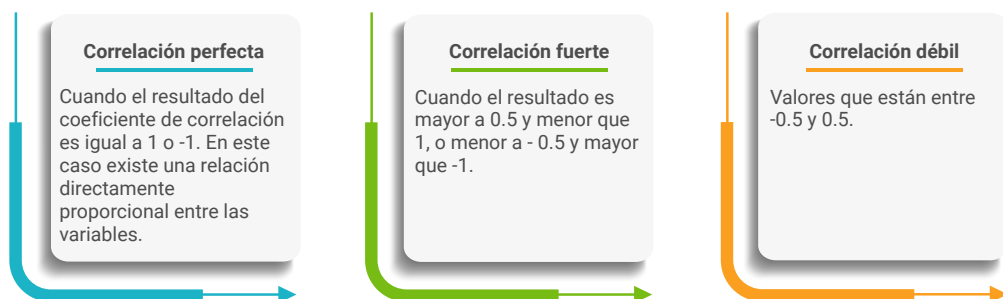
Nota: Imagen obtenida usando la herramienta R, con la función plot().

Al analizar la gráfica se puede observar que hay una relación entre las variables. Para calcular el valor de correlación se puede utilizar algunas de las herramientas tecnológicas mencionadas en el texto básico e ingresar los valores que se muestran en la tabla anterior. Además, en la siguiente sección se presenta cómo calcular el valor del coeficiente de correlación r mediante el uso de una fórmula.

Adicional a lo señalado en su texto básico, a continuación, se presentan los tipos de correlaciones que puede existir de acuerdo a la fuerza de la relación que se da entre las variables:

Figura 3.

Fuerza de la correlación



Nota: La fuerza de la correlación se mide en base al resultado del valor del coeficiente de correlación.

Esto significa que mientras más cercano es el valor de correlación entre las variables más fuerte es la relación entre ellas. Es decir a mayor correlación, los valores de las variables tienden a formar más una línea recta en el diagrama de dispersión (figura 2).

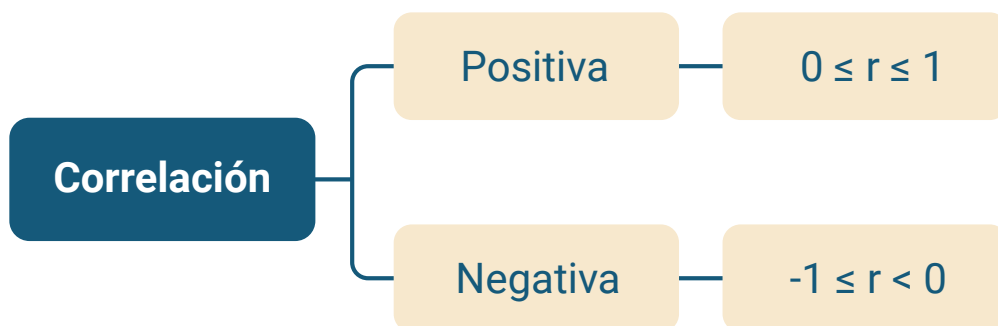
4.2. Coeficiente de correlación lineal

En esta sección usted conocerá cómo obtener e interpretar el coeficiente de correlación lineal r realizando una lectura de su texto básico, concerniente a los conceptos relacionados al coeficiente de correlación lineal y el cálculo del mismo usando tecnología.

De la lectura realizada, usted identificará que una correlación puede ser positiva o negativa de acuerdo a los cambios de las variables, medida por el coeficiente de correlación lineal r . En la siguiente figura se muestra la escala considerada para determinar el tipo de correlación de acuerdo a los resultados del valor de r .

Figura 4.

Tipo de correlación con base en el valor del coeficiente de correlación



Nota: En base al valor del coeficiente de correlación se conoce si ésta es positiva o negativa.

Además, la correlación es lineal cuando en el diagrama de dispersión se observa un patrón lineal entre las variables. Por ejemplo, en la sección anterior, figura 2, se puede observar que existe una correlación positiva y además una relación lineal entre la variable tiempo transcurrido en minutos con respecto al número de páginas impresas.

Recuerde que si usted desea conocer la relación entre dos variables, entonces debe calcular el coeficiente de correlación, y dependiendo del valor obtenido en este coeficiente: positivo o negativo, se puede observar si los cambios en una variable corresponden a cambios en la otra. Estos pueden ser en el mismo sentido, es decir, cuando una variable aumenta, la otra también aumenta: correlación positiva, o en dirección opuesta, cuando una variable aumenta, la otra disminuye: correlación negativa. Veamos un ejemplo: si tenemos un software con miles de líneas de código la complejidad aumenta (correlación positiva), por el contrario si tenemos pocas líneas de código probablemente la complejidad disminuya (correlación negativa).

Ahora lo invito a realizar el cálculo del coeficiente de correlación lineal del ejercicio planteado anteriormente sobre el tiempo transcurrido de los sucesivos usos de la impresora, aplicando la fórmula que se menciona en su texto básico para obtener el coeficiente r , de esta manera usted podrá comprobar si el resultado obtenido en el apartado anterior es el correcto.



Semana 10

En esta semana se revisarán temas relacionados al modelado de problemas mediante métodos de regresión lineal y no lineal. Lo invito a continuar con el estudio.

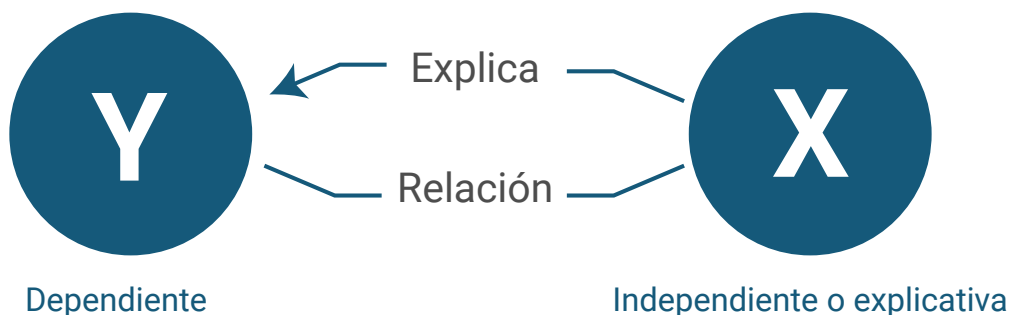
4.3. Regresión

Ahora que ya conoce cuándo existe una correlación lineal, vamos a determinar la ecuación de la línea recta que mejor se ajusta a los puntos del diagrama de dispersión utilizado para representar los datos. A la línea recta se la conoce como línea de regresión. Para revisar los métodos que se utilizan para encontrar la ecuación de la línea recta, es necesario realizar una lectura comprensiva del Capítulo 10 de su texto básico, el tema relacionado a Regresión donde usted podrá observar algunos conceptos básicos de regresión, la representación de una ecuación de regresión y las diferentes variables que intervienen en ella. Asimismo, es importante que revise la notación utilizada para la ecuación de una línea de regresión, e identifique sus elementos.

De la lectura realizada usted podrá determinar que la regresión consiste en generar una ecuación (modelo) que, basándose en la relación existente entre dos variables, permita predecir el valor de una a partir de la otra. En este caso estamos frente a un modelo de regresión lineal simple, tal como se representa en la siguiente imagen.

Figura 5.

Variables del modelo de regresión simple



Nota. Rodríguez (2007).

Además, usted podrá haberse dado cuenta que la forma de obtener estas rectas es mediante el método de los mínimos cuadrados. Para enfatizar en este procedimiento matemático voy a resaltar lo expresado por Rodríguez (2007), quién menciona que para construir una recta de regresión “se trata de colocar una recta entre los puntos dados, de la forma mejor balanceada con el criterio de hacer que la suma de las distancias de la recta a los puntos sea la menor posible” (p. 273). De esta forma a esta recta también se la conoce como recta de mínimos cuadrados.

Lo invito ahora a complementar lo estudiado en su texto básico, revisando el siguiente recurso web [Modelo de regresión lineal](#) de forma específica el apartado de conceptos básicos del análisis de regresión lineal.

Luego de la lectura que usted realizó podrá concluir que:

La regresión lineal simple es un método utilizado para predecir la variable dependiente (Y) en función de los valores de la variable independiente (X). Por ejemplo, predecir el consumo de energía de una aplicación web en función del tamaño, predecir el tiempo de computación de un programa en función de la velocidad del procesador, estimar el sueldo de una persona en función de los años de experiencia, etc.

Pero ¿qué son las variables dependientes y las variables independientes?, en el texto básico podrá identificar la diferencia entre las dos variables, sin embargo, complementaremos la explicación dada en su texto, señalando que la variable independiente (X) o variable explicativa es la característica que se usará para predecir algún valor dado de Y, y la variable dependiente (Y) conocida también como variable respuesta, es la característica a predecir que depende de X.

Bien, imagino que usted se preguntará ¿Cuándo utilizar la regresión lineal?, para dar respuesta a esta interrogante revise algunas ideas claves que se presentan a continuación:

Figura 6.

Claves para identificar una regresión lineal.

Cuando se evidencie una tendencia en los datos (creciente o decreciente), para esto es necesario dibujar los datos, puede ser mediante un diagrama de dispersión.

Determinar la relación entre la variable. Si la correlación es alta, más fuerte será la tendencia y más apropiado será aplicar un modelo de regresión lineal.

Nota: Diagrama de dispersión y coeficiente de correlación factores importante para conocer si hay regresión lineal.

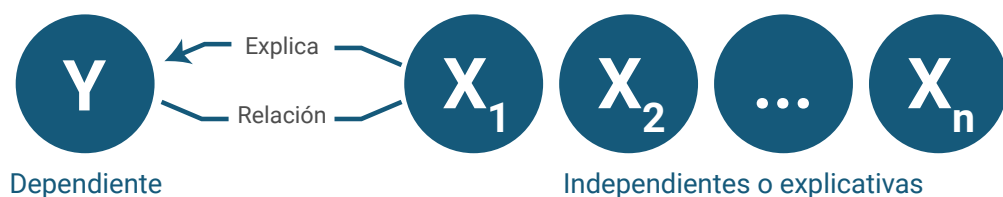
Vamos ahora a conocer los tipos de regresión que suelen darse dependiendo del número de variables explicativas que se tenga. En su texto básico, se explica cómo encontrar las ecuaciones de regresión lineal simple mediante la estimación de los coeficientes que hacen referencia a la pendiente e intersección con la recta, y como representar una ecuación de **regresión lineal múltiple** teniendo en cuenta más de una variable explicativa. Durante la lectura de su texto básico es importante que ponga énfasis en:

- Estrategias para predecir los valores de la variable dependiente del modelo de regresión simple.
- ¿Qué son los valores atípicos y valores influyentes, y cómo identificarlos?
- ¿Qué es el residuo, cómo se calcula?
- Conceptos básicos de una ecuación de regresión lineal múltiple.
- Procedimiento a seguir para encontrar la ecuación de regresión lineal múltiple.
- Coeficiente de determinación y coeficiente de determinación ajustado.

Una vez realizada la lectura, asumo que usted ya tiene clara la diferencia entre el modelo de regresión simple y múltiple, ya que en el primero se predice el valor de la variable dependiente a partir de una variable explicativa, mientras que, en el **modelo de regresión lineal múltiple** se genera una ecuación (modelo) que basándose en la relación existente entre variables, permita predecir el valor de la variable dependiente a partir de dos o más variables explicativas, como se representa a continuación:

Figura 7.

Variables del modelo de regresión lineal múltiple



Nota: El modelo de regresión lineal múltiple está formado de una variable dependiente y dos o más variables independientes.

Algunos ejemplos en los que podemos usar un **modelo regresión lineal múltiple** son:

- Predecir el sueldo de una persona, en función de los años de experiencia y el cargo que desempeña.
- Predecir el tiempo de cómputo de un programa, en función de la velocidad del procesador, y el tamaño del programa.
- Estimar el consumo de energía de un CPU al momento de ejecutar una aplicación considerando líneas de código y complejidad computacional.

Así mismo, es importante que conozca que existen relaciones entre variables que no necesariamente son lineales, en su texto básico también se explica el tema de **Regresión no lineal**, y los métodos para encontrar funciones no lineales mediante el uso de tecnología.

Puede complementar lo desarrollado en su texto básico revisando el siguiente recurso didáctico, donde encontrará una explicación más detallada de los tipos de regresión, así como algunos ejemplos de aplicación.

Regresión

Como pudo darse cuenta los modelos de regresión lineal son aplicables cuando todos los datos son continuos o cuantitativos, pero en el caso de que se tenga variables categóricas o cualitativas lo ideal es utilizar **regresión logística**.

Por ejemplo, si deseamos predecir el rendimiento de un sistema y tenemos como variable respuesta (dependiente) una variable cualitativa, llamada rendimiento con valores: ALTO o BAJO, en este caso lo más aplicable es un modelo de regresión logística.

Usted habrá notado también que la estimación de los parámetros de los modelos de regresión se puede realizar mediante el uso de tecnología. A continuación, se resume los pasos a seguir para hacer el cálculo del coeficiente de correlación y coeficientes del modelo de regresión usando Microsoft Excel.

- Para obtener el diagrama de puntos. Ir a la opción Gráficos. Seleccionar “Dispersión”.
- Dar *click* derecho sobre la serie de datos graficados y agregar línea de tendencia.
- Ir a Función: Estadística.
 - Para el coeficiente de correlación seleccionar la función “COEF.DE.CORREL”.
 - Para los coeficientes del modelo de regresión lineal, seleccionar función “ESTIMACION.LINEAL”, tanto para regresión simple o múltiple.

Lo invito a revisar también el apartado de “**Centro de tecnología**” en su texto básico donde se muestra paso a paso como calcular el coeficiente de correlación usando [Statdisk](#).

En este momento es necesario desarrollar la siguiente actividad para reforzar más los conocimientos aprendidos.



Actividad de aprendizaje recomendada

Realice los ejercicios del 1-4, de la sección de Regresión: 10-4 Habilidades y conceptos básicos de su texto básico. Los ejercicios correspondientes a: Notación, línea de mejor ajuste y pendiente.

Para el desarrollo de la actividad recomendada usted deberá aplicar los conocimientos estadísticos adquiridos durante el estudio de esta sección.

Es hora de poner en práctica lo estudiado desarrollando la autoevaluación 4.



Autoevaluación 4

Seleccione la alternativa correcta en cada uno de los siguientes enunciados.

1. Existe una correlación lineal cuando:
 - a. La gráfica de dispersión presenta una distante separación entre los valores de las variables.
 - b. Existe correlación y los puntos graficados de los datos pareados dan como resultado un patrón que se puede aproximar mediante una línea recta.
 - c. No hay relación entre las variables.
2. Una correlación es negativa cuando el valor de r está en el rango de:
 - a. $-1 \leq r < 0$
 - b. $0 \leq r \leq 1$
 - c. $r = 0$
3. El modelo de regresión lineal simple explica la relación entre:
 - a. La variable dependiente y una o más variables independientes.
 - b. La variable dependiente y una única variable independiente.
 - c. Variables cualitativas.
4. Se conoce como línea de regresión a:
 - a. La línea recta que mejor se ajusta al diagrama de dispersión de los datos.
 - b. La relación entre los datos muestrales.
 - c. La gráfica que representa los datos pareados.
5. La ecuación de regresión expresa una relación entre:
 - a. La variable explicativa y la variable independiente.
 - b. La variable independiente y la variable predictora.
 - c. La variable explicativa y la variable respuesta.

6. En la ecuación de la recta el valor del estadístico b_0 , hace referencia a:
- La intersección "y" de la ecuación de regresión.
 - Pendiente de la ecuación de regresión.
 - El valor atípico.
7. El coeficiente de determinación R^2 representa:
- Una medida de qué tan bien se ajusta la ecuación de regresión múltiple a los datos muestrales.
 - Una relación lineal entre más de una variable.
 - Proporción de la variación en Y.
8. El coeficiente de determinación R^2 es muy bueno si:
- Es cercano a 0.
 - Es cercano a 1.
 - Es igual a 1.
9. La regresión logística es utilizada en problemas de:
- Predicción de una variable cuantitativa.
 - Clasificación binaria.
 - Ajuste de una recta.
10. Si tenemos un conjunto de datos muestrales con las estaturas y peso de hombres y mujeres, donde la variable respuesta representa el género: 1=Hombre y 0= Mujer. En este caso es conveniente utilizar:
- Regresión lineal múltiple .
 - Regresión lineal simple.
 - Regresión logística.

[Ir al solucionario](#)

Resultado de aprendizaje 5

- Realiza un análisis estadístico del rendimiento de un sistema.

A través del presente resultado de aprendizaje, usted determinará que el análisis de varianza puede ser utilizado para probar el efecto de las variables en problemas de TI.

Contenidos, recursos y actividades de aprendizaje



Semana 11

Unidad 5. Análisis de Varianza

Veremos ahora otro interesante tema que es el análisis de varianza (ANOVA), un tipo de prueba estadística paramétrica que estudia el efecto de uno o más factores sobre la media de una variable continua. En este caso se denominan factores a las variables independientes. Por tanto, esta prueba se aplica cuando se desea comparar las medias de dos o más grupos. Es decir, si queremos analizar si el lenguaje de programación es sinónimo de buen rendimiento de un software, podríamos plantear un problema de esta forma:

Tenemos tres grupos de software según el lenguaje de programación en el que estén desarrollados. ¿Cómo sé si estos grupos tienen un buen rendimiento de cómputo? Para esto necesitamos una herramienta que nos compare estos tres grupos de software y nos diga si estadísticamente éstos son distintos o no. Esta herramienta puede ser el ANOVA. Si el resultado del ANOVA es significativo entonces habrá diferencias en el rendimiento según el lenguaje de programación en el que han sido desarrollados.

5.1. ANOVA de un factor

Existen diferentes tipos de ANOVA, en esta unidad se estudia una clasificación dependiendo del número de factores. Veamos esto con más detenimiento realizando una lectura comprensiva de su texto básico:

Capítulo 12: Análisis de varianza, sección “ANOVA de un factor”, y revise los conceptos básicos de este tipo de prueba, la distribución que se requiere para realizar una prueba ANOVA, los requisitos a cumplir, y los cálculos e identificación de medias que son diferentes.

¿Cómo estuvo la lectura? es interesante ¿verdad?. Al realizar la lectura de esta sección se observó como verificar los requisitos de esta prueba y los criterios a considerar para dar por aceptada o no una afirmación de igualdad de medias. Además, se observó que el ANOVA de un factor se utiliza con datos categóricos de un factor, y que se requiere calcular el estadístico de prueba F para aceptar o rechazar la igualdad de medias.

Resumiendo, podemos decir que la aplicación del ANOVA de un factor, se basa en un contraste de hipótesis, tal como se muestra a continuación:

Figura 8.
Contraste de hipótesis. ANOVA de un factor



De la imagen anterior, podemos ver que la hipótesis nula que se contrasta en el ANOVA de un factor consiste en que las medias poblacionales son iguales. Tener en cuenta que, si damos por aceptada la hipótesis nula, significa que los grupos no difieren en el valor medio de la variable dependiente (respuesta) y que, por tanto, dicho valor medio se podrá considerar independiente del factor.

Haciendo referencia al ejemplo inicial de los grupos de software la hipótesis nula sería que no hay diferencias entre las medias del rendimiento de los grupos de software, es decir las medias son iguales.

Es momento de poner en práctica lo estudiado, revisando el siguiente ejemplo a fin de comprender mejor el ANOVA de un factor.

Ejemplo:

Se desea comprobar si el uso de tres computadoras produce resultados diferentes que son estadísticamente significativos en el tiempo de ejecución de un programa. Sean dos grupos de programas seleccionados al azar, con

las mediciones obtenidas en tiempo de ejecución después de que cada uno de ellos fue ejecutado en diferente computador.

En la tabla 3, se muestran los resultados obtenidos del tiempo de ejecución (segundos) de los programas para cada computador:

Tabla 3.

Tiempo de ejecución de cada programa

Programa	Computador A	Computador B	Computador C
1	50	100	150
2	100	150	120
Suma	150	250	270
Promedio	75	125	135

Nota. Esta tabla contiene el tiempo de ejecución de programas de computadora.

Si observamos los valores promedios, todo parece indicar que existen diferencias en el tiempo de ejecución entre los computadores. Ahora bien, ¿son dichas diferencias significativas? La prueba ANOVA permite responder a esta interrogante.

El objetivo del ANOVA es comparar los diversos valores medios para determinar si alguno de ellos difiere significativamente del resto.

En base al ejemplo propuesto lo invito a realizar el siguiente ejercicio:

Ejercicio:

Verifique la hipótesis de que las medias del tiempo de ejecución son iguales utilizando el nivel de significación 0.05. Puede seguir los siguientes pasos:

1. Formule la hipótesis nula y alternativa.
2. Encuentre la varianza entre las muestras.
3. Determine la varianza dentro de las muestras.
4. Calcule el estadístico de prueba.
5. Si desea puede elaborar una tabla ANOVA
6. Interpretación: Indicar la decisión respecto a la hipótesis nula.

Recuerde también que puede utilizar la herramienta Statdisk o cualquier otra herramienta de las mencionadas en su texto básico para obtener los resultados ANOVA.

Es momento de poner en práctica lo estudiado, realizando la siguiente actividad recomendada:



Actividad de aprendizaje recomendada

Desarrollar los ejercicios 1-4 de la sección 12-1 Habilidades y conceptos básicos del tema de ANOVA de un factor en su texto básico.

Un punto importante para desarrollar los ejercicios, es hacer una analogía con el ejemplo analizado anteriormente, donde las columnas de la tabla (Computadoras), se corresponden con los vuelos (Vuelo 1, Vuelo 19 y Vuelo 21) del ejercicio a resolver.



Semana 12

5.2. ANOVA de dos factores

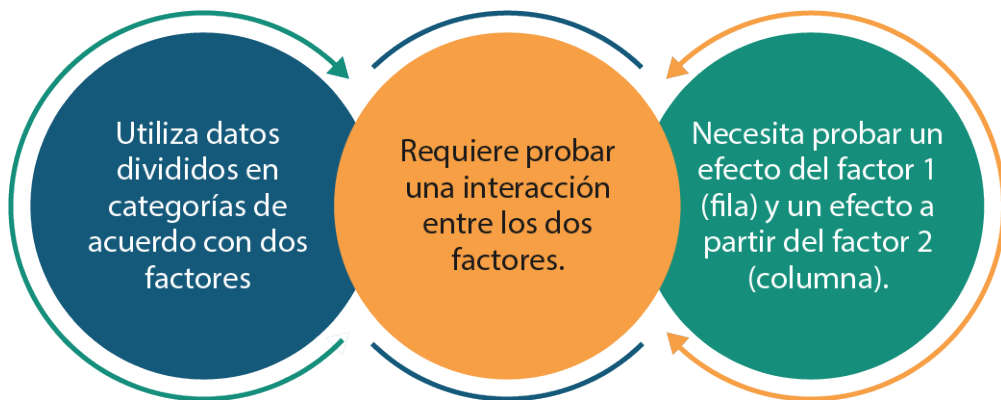
¿Animado para continuar adelante con el siguiente tema? Seguro que sí. Ahora revisaremos el método de análisis de varianza con dos factores. Para ello, es necesario revisar los conceptos claves que se requieren para aplicar este método, desarrollando una lectura comprensiva de la sección “ANOVA de dos factores” de su texto básico. En esta sección usted encontrará algunas definiciones importantes a tener en cuenta de este método, el procedimiento a seguir para el ANOVA de dos factores, y algunos ejemplos en los que se detalla paso a paso como realizar esta prueba estadística.

¿Fue comprensible la lectura? Espero que sí y que le haya sido de mucha ayuda.

Luego de la lectura del texto básico, se puede resumir que:

Figura 9.

Anova de dos factores



Recuerde que se puede identificar posibles interacciones de los dos factores de forma gráfica utilizando “gráficos de interacción”. Si los segmentos de líneas son aproximadamente paralelas no hay interacción.

Algunos ejemplos de aplicación de este tipo de análisis son:

- Eficacia de distintos tratamientos con un determinado medicamento. Factores: tratamiento, edad del paciente.
- Estudio de diversos tratamientos para conocer la evolución de la hipertensión arterial según el sexo. Factores: género y fármaco.
- Estudio de diferentes aplicaciones web para conocer su rendimiento según el tiempo de ejecución. Factores: Lenguaje de programación y sistema operativo.
- Efectos sobre el retardo de los mensajes transmitidos por una red de computadoras. Factores: Algoritmo de encadenamiento de mensajes y nivel de carga de la red.

Muy bien, para entender mejor lo que acabamos de mencionar, lo invito a revisar el siguiente recurso con un ejemplo práctico del ANOVA de dos factores.

ANOVA

¡Estamos de vuelta!, ¿terminamos la revisión del video? Espero que le haya sido de utilidad.

Ahora considerando su campo de estudio completemos el siguiente ejercicio:

Si contamos con una tabla que registra los tiempos de ejecución de algunos algoritmos, donde el tiempo se ve afectado por el tipo de lenguaje de programación y el sistema operativo en el que se ejecutan, ¿cuáles serían los factores a considerar? Puede completar este ejercicio colocando valores al azar o utilizar algoritmos desarrollados por usted mismo y ejecutarlos en dos sistemas operativos diferentes y así determinar los tiempos de ejecución. Luego seguir el procedimiento indicado en su texto básico para el análisis de varianza de dos factores.

Bien, como usted habrá observado el ANOVA de dos factores es aplicable cuando tenemos dos variables independientes, es decir ésta prueba el efecto de dos variables independientes sobre una variable dependiente. De acuerdo al ejemplo anterior, estas dos variables independientes son: lenguaje de programación y sistema operativo. Por tanto la diferencia del ANOVA de un factor y dos factores radica en el número de variables independientes a utilizar para el análisis de una prueba de varianza.

Recuerde que cuando se utiliza el ANOVA de dos factores (una variable de fila y una de columna) es necesario realizar las siguientes pruebas (formuladas como hipótesis):

- Hipótesis acerca del efecto de la interacción entre los factores.
- Hipótesis acerca del efecto del factor fila.
- Hipótesis acerca del efecto del factor columna.

Es momento de poner en práctica lo estudiado, realizando la siguiente actividad recomendada a fin de comprender de mejor manera la aplicación de esta prueba estadística.



Actividad de aprendizaje recomendada

- Con base en el ejercicio 6. Pesos y 7. Estaturas, del análisis de varianza de dos factores propuesto en su texto básico, sección 12-2 habilidades y conceptos básicos, analice los resultados y determine qué se puede concluir.

Con el desarrollo de esta actividad usted podrá recordar los conceptos analizados en esta sección dando respuesta a las interrogantes que se plantean en cada uno de estos ejercicios.

¡Muy bien!, hemos culminado con éxito el desarrollo de esta unidad. En este momento lo invito a que resuelva la siguiente autoevaluación para comprobar lo aprendido.



Autoevaluación 5

Para cada enunciado identifique la alternativa correcta.

1. El ANOVA de un factor se utiliza para realizar pruebas de hipótesis de que:
 - a. Tres o más poblaciones tienen medias que no son iguales.
 - b. Tres o más poblaciones tienen medias que son todas iguales.
 - c. Dos o más poblaciones tienen medias que no son todas iguales.
2. Uno de los requisitos para aplicar el ANOVA de un factor es que:
 - a. Las muestras sean independientes entre sí.
 - b. Las poblaciones tienen diferente varianza.
 - c. Las distribuciones de las poblaciones son binomiales.
3. Los valores más grandes del dato estadístico de prueba producen:
 - a. Valores P más grandes.
 - b. Varianzas distintas.
 - c. Valores P más pequeños.
4. El ANOVA de un factor requiere calcular el estadístico de prueba F para:
 - a. Aceptar la diferencia entre las poblaciones.
 - b. Rechazar la diferencia entre las medias poblacionales.
 - c. Aceptar o rechazar la igualdad de las medias.
5. En el ANOVA de un factor el dar aceptada la hipótesis nula significa:
 - a. Que los grupos no difieren en el valor medio de la variable respuesta.
 - b. Que el valor medio se podrá considerar dependiente del factor.
 - c. Que los grupos difieren en el valor medio de la variable dependiente.

6. En el ANOVA de dos factores al utilizar “gráficos de interacción”, si los segmentos de líneas son aproximadamente paralelas significa que:
- No hay interacción de los dos factores.
 - Hay interacción entre los factores.
 - No es posible identificar la interacción entre los factores.
7. Una interacción entre dos factores existe cuando:
- El efecto de uno de los factores no cambia para diferentes categorías.
 - El efecto de uno de los factores cambia para diferentes categorías del otro factor.
 - El efecto de los dos factores cambia para diferentes categorías de ambos factores.
8. Cuando se realiza la prueba del efecto del factor fila, si el valor P es pequeño ejemplo (menor a 0.05), se concluye:
- No hay efecto del factor de fila.
 - Hay un efecto del factor de fila.
 - Hay un efecto del factor de columna.
9. En el ANOVA de dos factores, uno de los requisitos a cumplir es que para cada celda los valores muestrales deben provenir de una población con una distribución que es:
- Aproximadamente normal.
 - Binomial.
 - Categórica.
10. Si tengo el consumo de CPU de una aplicación web categorizados por el intervalo de tiempo de ejecución y el tipo de sistema operativo: Windows y MACOS, ¿cuál método sería más recomendable utilizar para este caso?
- ANOVA de un factor.
 - ANOVA de dos factores.
 - ANOVA de tres factores.

[Ir al solucionario](#)

Resultado de aprendizaje 6

- Determina si una prueba paramétrica o no paramétrica es apropiada.

A través del presente resultado de aprendizaje usted determinará las principales pruebas no paramétricas que pueden ser aplicadas a problemas reales e identificará cuándo éstas son apropiadas utilizarlas.

Contenidos, recursos y actividades de aprendizaje



Semana 13

Unidad 6. Pruebas no paramétricas

Luego de haber revisado acerca de las pruebas paramétricas, en esta unidad nos enfocaremos al estudio de otro tipo de pruebas, las no paramétricas que son parte del conjunto de pruebas estadísticas que se pueden utilizar para el análisis de los datos.

6.1. Conceptos básicos de las pruebas no paramétricas

Para conocer en qué consisten las pruebas no paramétricas, sus ventajas y desventajas, vamos a ir al texto básico y revisar el capítulo 13: Pruebas no paramétricas, a la sección correspondiente a conceptos básicos, prestemos atención entre otras cosas a las definiciones que se presentan y a la comparación que se realiza entre pruebas paramétricas y no paramétricas.

Luego de revisar el texto básico es preciso señalar que las pruebas no paramétricas son aquellas que no requieren que los datos tengan una distribución particular, es decir, los datos no están organizados de forma normal como en una prueba paramétrica.

Bien, usted se preguntará ¿Cuándo aplicar una prueba no paramétrica o una prueba paramétrica? Para enfatizar en el uso de este tipo de pruebas, y cuándo utilizarlas a continuación se complementa la comparación realizada en su texto básico sobre estas pruebas y se resaltan algunos criterios

importantes a tener en cuenta al momento de elegir una prueba paramétrica o no paramétrica.

Tabla 4.

Cuadro comparativo de las pruebas paramétricas y no paramétricas

Criterio de comparación	Prueba paramétrica	Prueba no paramétrica
Nivel de medición	Por intervalos	Catégoricos: nominales u ordinales.
Tamaño de la muestra	Grande	Pequeña
Tipo de muestra	Muestra aleatoria	Muestra no aleatoria
Distribución	Normal	Libre

En la tabla anterior usted podrá observar que antes de aplicar una prueba paramétrica o no paramétrica, es importante conocer el tamaño de la muestra poblacional y la escala en la que están medidos los datos. Recuerde que las pruebas no paramétricas pueden ser utilizadas con datos catégoricos y que no presentan una distribución normal. Por tanto, este tipo de pruebas pueden ser aplicadas a una variedad de situaciones donde sea necesario contrastar una hipótesis y se requiera analizar datos en escala nominal (catégoricos). Por ejemplo: En una muestra de usuarios de telefonía móvil donde es necesario emplear el género para probar una hipótesis de que existe diferencia entre el uso del celular por parte de los hombres y el uso del celular en las mujeres.

Lo invito ahora a profundizar en el estudio de las pruebas no paramétricas que se listan en la tabla 13-2 de su texto básico, revisando las siguientes secciones.

6.2. Pruebas del signo

Luego de revisar su texto básico sobre la prueba del signo, es conveniente resaltar que este tipo de prueba no paramétrica permite el contraste de hipótesis respecto a la mediana, más no de una media. Además, se basa en signos positivos y negativos para evaluar diferentes afirmaciones, de esta forma se analizan las frecuencias de estos signos para determinar si son diferentes significativamente.

Para enfatizar en el uso de la prueba del signo, veamos el siguiente ejemplo propuesto el recurso interactivo que se muestra a continuación:

Ejemplo prueba del signo

¿Cómo les fue con el ejercicio propuesto en el recurso anterior?, ¿Cuál es el valor crítico para $n=11$? Espero que haya podido identificar que el valor crítico para $n=11$ con un valor de significancia de 0.05 es igual a 1 en dos colas, por tanto, el estadístico de prueba $x=4$ no es menor o igual al valor crítico, esto quiere decir que, no es posible rechazar la hipótesis nula. Como pudo observar en este ejemplo, la idea es probar si hay diferencia en las ventas de diferentes tipos de software, así mismo usted podría aplicar la prueba del signo a otras situaciones relacionadas en su campo profesional.

Bien, hasta ahora hemos visto la aplicación de la prueba del signo para datos numéricos, pero ¿qué sucede con aquellas afirmaciones en las cuales tenemos datos nominales o categóricos? Revisemos el texto básico sección “Afirmaciones que implican datos nominales con dos categorías” y veamos cómo usar la prueba del signo.

Cómo usted habrá observado en estos casos se debe identificar la proporción de datos nominales que pertenecen a una categoría específica, por **ejemplo**, para el caso de los productos en software de Amazon mostrado en el recurso interactivo anterior, se tienen dos categorías: software para dispositivos móviles y software en la nube. En este caso se representa al software de dispositivos móviles con signo positivo (+) y al software en la nube con signo negativo (-).

A fin de complementar lo estudiado en su texto básico, puede realizar la lectura del siguiente recurso web [Pruebas no paramétricas](#), especialmente lo relacionado a la prueba de los signos para el contraste de hipótesis sobre una variable y dos variables.

Otro importante tema a resaltar es la prueba del signo para “Afirmaciones sobre la mediana de una sola población”. En su texto básico usted encontrará el procedimiento a seguir para probar este tipo de afirmaciones, además podrá observar que en este caso los signos positivos o negativos se basan en el valor declarado de la mediana.



Actividad de aprendizaje recomendada

Antes de continuar con el siguiente tema, es importante que realice el ejercicio 6, para el uso de la prueba del signo para datos de pares relacionados, especificado en la sección 13-2 habilidades y conceptos básicos.

Para este ejercicio es conveniente que siga el procedimiento de la prueba del signo especificado en el diagrama de flujo de su texto básico.



Semana 14

6.3. Prueba de rangos con signo de Wilcoxon

Bien, ahora revisaremos otra prueba no paramétrica para datos pareados. Para conocer en qué consiste este tipo de prueba es conveniente realizar la lectura de la sección “Prueba de rangos con signo de Wilcoxon para datos pareados” de su texto básico”.

¿Cómo le fue con la lectura? ¿Tenemos claro en qué consiste la prueba de **Wilcoxon**? Entonces para resumir lo mencionado en su texto básico, podemos decir que la prueba de Wilcoxon utiliza rangos para probar las siguientes afirmaciones:

Tabla 5.

Usos de la prueba de rangos con signo de Wilcoxon

Afirmaciones que involucran datos pareados	Una población de datos pareados posee la propiedad de que los pares relacionados tienen diferencias con una mediana igual a cero.
Afirmaciones sobre la mediana de una sola población	Una sola población de valores individuales tiene algún valor declarado de la mediana.

Es oportuno que ponga atención en el procedimiento a seguir para la prueba de rangos con signo de Wilcoxon, usted observará en su texto básico el

desarrollo de ejemplos, en los cuales se aplican algunos pasos a seguir hasta llegar a la conclusión de aceptar o no la hipótesis nula.

Como se habrá dado cuenta la diferencia de la prueba del signo con la prueba de rangos de Wilcoxon radica en que en esta última las hipótesis se basan en rangos y la mediana. Por ejemplo se puede utilizar este tipo de prueba no paramétrica para probar alguna hipótesis relacionada con la evaluación de un sistema informático en la que se considere resultados de evaluaciones de usuarios acerca del funcionamiento y usabilidad del sistema. La hipótesis a probar podría ser de que no hay diferencia entre las evaluaciones realizadas por los usuarios en términos de funcionamiento y usabilidad.

Es momento de verificar los temas estudiados en esta unidad, realizando la siguiente autoevaluación.



Autoevaluación 6

Para cada enunciado identifique la alternativa correcta.

1. Las pruebas no paramétricas a diferencia de las paramétricas:
 - a. Requieren que las muestras provengan de distribuciones normales.
 - b. No requieren que las muestras provengan de poblaciones con distribuciones normales o cualquier otra distribución.
 - c. Necesitan que las muestras sigan una distribución binomial.
2. La prueba del signo utiliza signos positivos y negativos para evaluar afirmaciones:
 - a. Afirmaciones sobre la moda de una población.
 - b. Que involucra numéricos con tres categorías.
 - c. Afirmaciones sobre la mediana de una población.
3. Una de las ventajas de las pruebas no paramétricas es que éstas:
 - a. Se pueden aplicar a más tipos de datos que las pruebas paramétricas.
 - b. Solo pueden ser aplicadas a ciertos tipos de datos.
 - c. Se puede aplicar a limitadas situaciones.
4. Para evaluar afirmaciones que involucra pares relacionados de datos muestrales se usa la prueba no paramétrica:
 - a. Prueba del signo o prueba de rangos con signo de Wilcoxon.
 - b. Prueba de la suma de rangos de Wilcoxon.
 - c. Prueba de correlación lineal.
5. En las pruebas basadas en rangos, un rango se entiende como:
 - a. Un número asignado a un elemento muestral individual de acuerdo con su lugar en la lista ordenada.
 - b. Colocar un elemento muestral en una lista ordenada.
 - c. Un elemento muestral asignado en orden a la posición en la lista.

6. En afirmaciones sobre pares relacionados es importante tener en cuenta que:
 - a. Se debe incluir los empates del par relacionado cuando ambos valores son iguales.
 - b. Se excluyen los empates borrando cualquier par relacionado en el que ambos valores sean iguales.
 - c. La cantidad de signos positivos debería ser mayor a la cantidad de signos negativos.
7. El requisito que se debe cumplir en este tipo de pruebas es:
 - a. Los datos muestrales sean una muestra no aleatoria.
 - b. La muestra sea una muestra estratificada.
 - c. La muestra sea una muestra aleatoria simple.
8. La prueba de rangos con signo de Wilcoxon para datos pareados es un tipo de prueba:
 - a. Paramétrica.
 - b. No paramétricas.
 - c. Paramétrica y no paramétrica.
9. Con la prueba de rangos con signo de Wilcoxon para datos pareados, se puede probar una afirmación de que:
 - a. Una sola población de valores individuales tiene una mediana igual a algún valor declarado.
 - b. Una población de valores individuales tiene una media igual a algún valor declarado.
 - c. Una población de datos pareados tiene diferencias con una media igual a cero.
10. En la prueba del signo con datos pareados se usa:
 - a. Solo los signos de las diferencias.
 - b. Las magnitudes de las diferencias
 - c. Datos muestrales en rangos

[Ir al solucionario](#)



Actividades finales del bimestre



Semana 15

Una vez completado el estudio de las unidades del segundo bimestre lo invito a desarrollar la siguiente actividad, que le permitirá poner en práctica los conocimientos aprendidos y entrenarse para la evaluación presencial.

Actividades finales del segundo bimestre



Semana 16

En la presente semana debe prepararse para la evaluación presencial del Segundo Bimestre, para ello le recomiendo realizar un repaso de las unidades que se ha presentado, revisar las actividades recomendadas y las autoevaluaciones propuestas en cada unidad, a fin de comprender mejor los contenidos que debe estudiar. Si tiene alguna inquietud, recuerde que puede consultar a través de la plataforma EVA.



4. Solucionario

Autoevaluación 1		
Pregunta	Respuesta	Retroalimentación
1	b	Un modelo determinista implica que, dadas algunas entradas, la salida siempre será la misma.
2	a	En un modelo estocástico algún elemento no se conoce con anticipación, alguna variable del modelo es tomada como un dato al azar.
3	c	Es un resultado o un evento que no se puede descomponer en otros más simples.
4	b	Corresponden a eventos que son muy poco probables de que ocurran.
5	a	Es la probabilidad de que ocurra el evento A u ocurra el evento B (o que ambos ocurran).
6	b	La probabilidad condicional es una probabilidad obtenida con información adicional de que algún otro evento ya ha ocurrido.
7	c	Es la notación utilizada para indicar la probabilidad condicional de que ocurra el evento B dado que el evento A ya ha ocurrido.
8	a	Con el método formal la se calcula dividiendo la probabilidad de que el evento A y B ocurra por la probabilidad del evento A.
9	b	Representa la notación de una probabilidad condicional, por tanto se lee como probabilidad de A dado M.
10	c	El método intuitivo es mucho más fácil de usar y menos probable que resulte en errores.

Ir a la
autoevaluación

Autoevaluación 2

Pregunta	Respuesta	Retroalimentación
1	b	La distribución de probabilidad puede entenderse como una lista que proporciona los resultados de los valores de las variables aleatorias junto con la probabilidad de ocurrencia asociada a estos valores.
2	a	Los valores de la media, la varianza y desviación estándar, son parámetros, no valores estadísticos.
3	a	Es el valor medio de los resultados.
4	c	Se usa en circunstancias en las que se tiene solo dos resultados posibles, es decir dos categorías.
5	b	La probabilidad de Bernoulli el experimento se realiza en un solo ensayo.
6	a	Dos de los métodos para calcular la distribución binomial son: uso de la fórmula y uso de software (tecnología).
7	b	La distribución de Poisson es otra categoría de distribución discreta.
8	c	Es un ejemplo o aplicación de distribución de Poisson.
9	c	Uno de los requisitos para una distribución de Poisson es que las ocurrencias deben ser independientes entre sí.
10	c	Una distribución de Poisson está determinada únicamente por la media.

[Ir a la autoevaluación](#)

Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1	b	Una afirmación o declaración sobre una propiedad de una población.
2	a	Las pruebas de hipótesis también se llaman pruebas de significancia.
3	b	La hipótesis alterna es lo que se espera probar que es cierta. Es decir, indica que un parámetro de población es diferente del valor hipotético de la hipótesis nula.
4	c	El nivel de significancia para una prueba de hipótesis es la probabilidad de rechazar erróneamente la hipótesis nula cuando es verdadera.
5	c	Los valores críticos separan la región crítica de los valores del estadístico de prueba que no conducen al rechazo de la hipótesis nula.
6	a	El error tipo I se produce cuando rechazamos una hipótesis nula cuando ésta es verdadera.
7	a	Para la prueba t , el método de valor P , el de valor crítico y el método de intervalo de confianza, son todos equivalentes en el sentido de que todos conducen a las mismas conclusiones.
8	a	Los métodos para llevar a cabo una prueba de hipótesis de una afirmación hecha sobre una desviación estándar o varianza usan la distribución ji cuadrada.
9	a	Para probar una prueba de hipótesis acerca de una media poblacional el dato estadístico utilizado se conoce como <i>prueba t</i> .
10	c	El dato estadístico de prueba para probar una hipótesis sobre una proporción es la prueba z .

Ir a la
autoevaluación

Autoevaluación 4

Pregunta	Respuesta	Retroalimentación
1	b	Hay correlación lineal cuando existe una correlación y los puntos graficados de los datos pareados dan como resultado un patrón que se puede aproximar mediante una línea recta.
2	a	Existe correlación negativa puesto que a medida que los valores x aumentan, los valores correspondientes de y disminuyen, en este caso el valor de r debe estar en el rango de $-1 \leq r < 0$.
3	b	El modelo de regresión simple explica la relación entre la variable Y y una única variable dependiente X .
4	a	Es la línea de mejor ajuste.
5	c	Explica la relación entre la variable independiente (explicativa) y la variable dependiente (respuesta).
6	a	El estadístico muestral b_0 es la intersección "y"
7	a	Es una medida de qué tan bien se ajusta la ecuación de regresión múltiple a los datos muestrales.
8	b	Un ajuste es muy bueno si es cercano a 1.
9	b	La regresión logística permite generar un modelo para predecir la clase de cada observación.
10	c	Se podría utilizar regresión logística ya que estamos frente a un problema de clasificación, donde la variable respuesta es de tipo categórica.

[Ir a la autoevaluación](#)

Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	b	Se utiliza para probar de que tres o más poblaciones tiene medias que son todas iguales como: $H_0: \mu_1 = \mu_2 = \mu_3$.
2	a	Las muestras no están relacionadas o pareadas de ninguna forma, es decir, que son independientes entre sí.
3	c	Los valores más grandes del dato estadístico de prueba dan como resultado valores P más pequeños.
4	c	El ANOVA de un factor requiere calcular el estadístico de prueba F para aceptar o rechazar la igualdad de las medias.
5	a	Si damos por aceptada la hipótesis nula, significa que los grupos no difieren en el valor medio de la variable respuesta y que, dicho valor medio se podrá considerar independiente del factor.
6	a	Si los segmentos de líneas son aproximadamente paralelas no hay interacción de los dos factores.
7	b	Existe una interacción entre dos factores si el efecto de uno de ellos cambia para diferentes categorías del otro factor.
8	b	En la prueba del efecto del factor fila, usando el valor P, si éste es menor a 0.05, entonces hay un efecto del factor de fila.
9	a	Para cada celda, los valores muestrales provienen de una población con una distribución que es aproximadamente normal.
10	b	Al analizar el ejemplo se ve que se tiene dos factores: intervalo de tiempo y tipo de sistema operativo. Por tanto, se debe aplicar el ANOVA de dos factores.

[Ir a la autoevaluación](#)

Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	b	Las pruebas no paramétricas no requieren que las muestras provengan de poblaciones con distribuciones normales o cualquier otra distribución particular.
2	c	La prueba del signo utiliza signos positivos y negativos para evaluar diferentes afirmaciones, entre ellas las afirmaciones sobre la mediana de una población.
3	a	Las pruebas no paramétricas se pueden aplicar a más tipos de datos que las pruebas paramétricas.
4	a	Para relaciones de datos muestrales se usa la prueba del signo o prueba de rangos con signo de Wilcoxon.
5	a	Un número asignado a un elemento muestral individual de acuerdo con su lugar en la lista ordenada.
6	b	Se registra solamente el signo de la diferencia encontrada al restar el valor de la segunda variable de la primera. Los empates se excluyen.
7	c	Los datos muestrales son una muestra aleatoria simple.
8	b	Prueba de rangos con signo de Wilcoxon para datos pareados, es una prueba no paramétrica que utiliza rangos para probar afirmaciones.
9	a	Prueba de rangos con signo de Wilcoxon para datos pareados, permite probar una afirmación de que una sola población de valores individuales tiene una mediana igual a algún valor declarado.
10	a	La prueba del signo se puede usar con datos pareados, pero ésta solo usa los signos de las diferencias.

[Ir a la autoevaluación](#)



5. Referencias bibliográficas

- Rustom, A. (2012). *Estadística descriptiva, Probabilidad e Inferencia*. Universidad de Chile. <https://cutt.ly/ihq9VcR>
- Diccionario de Geotecnia (14 de mayo de 2021). *Análisis Determinístico*. <https://www.diccionario.geotecnia.online/palabra/analisis-deterministico/>
- Guzmán, V. (2009). Economía de la Empresa I. Open Course Ware Universidad de Málaga. https://ocw.uma.es/pluginfile.php/1520/mod_resource/content/0/GuzmanParraOcwT10.pdf
- Triola, M. (2018). *Estadística*. Editorial: Pearson. Edición: 12.
- Rodríguez, L. (2007). *Probabilidad y Estadística Básica para Ingenieros*. Escuela Superior Politécnica del Litoral, ESPOL. https://archuto.files.wordpress.com/2011/02/probabilidad_y_estadistica_basica.pdf
- Molina, G, Rodrigo, M. (2010). *El Modelo de Regresión lineal*. Open Course Ware Universidad de Valencia. http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_09nuevo.pdf
- Molina, G, Rodrigo, M. (2014). *Pruebas no paramétricas*. Open Course Ware Universidad de Valencia. http://ocw.uv.es/ciencias-de-la-salud/estadistica-ii/est2_t5.pdf