TECNOLÓGICO
UNIVERSITARIO
PICHINCHA

2
Nivel

Estadística

Guía general de estudios de la asignatura

Modalidad de Educación a Distancia
Tecnología Superior “Talento Humano”

Autor:
MSc. Jhonson Peralta

Periodo académico
octubre 2023 – marzo 2024

Modalidad
Distancia

Estadística
Guía general de estudios de la asignatura

© MSc. Jhonson Peralta

ISBN: 978-9942-672-22-3

Edición: Julio 2024

Texto digital proporcionado por el autor.

TALLPA Publicidad Impresa - 2540 662 - 09 9561 4887
Quito - Ecuador

PRÓLOGO

Ha sido y es objetivo fundamental del instituto utilizar herramientas esenciales para que nuestros estudiantes logren alcanzar una formación integral. Bajo esta consideración ponemos a disposición estas guías de estudio que posibilitarán, sin duda, puedan organizarse para comprender el contenido de las diferentes asignaturas.

Estas guías han sido creadas por un equipo de profesionales altamente capacitados en cada asignatura, con el objetivo de convertir su proceso de aprendizaje en una experiencia enriquecedora.

Nuestros docentes han recopilado información, han sintetizado temas, organizado conceptos y aspectos relevantes para que cada guía se presente cuidadosamente elaborada para responder a la realidad actual, con contenidos actualizados y a la vanguardia del conocimiento. La didáctica empleada facilitará la comprensión y aprendizaje de cada tema, permitiéndoles avanzar de manera efectiva en su formación profesional. En la elaboración de estas guías se denota el compromiso del instituto para lograr el éxito académico.

La diagramación de estas guías ha sido pensada para ser clara y atractiva, transmitiendo los conocimientos de manera amena y accesible. Queremos que nuestros estudiantes disfruten del proceso de aprendizaje encontrando en cada página una herramienta útil que les motive a salir adelante en su camino educativo.

Estimados estudiantes: Les deseamos éxito en su recorrido académico, que el Instituto Tecnológico Universitario Pichincha estará siempre pendiente por vuestro éxito educativo.

Dr. Edgar Espinosa. MSc.
RECTOR ISTP-U

ÍNDICE

Presentación de la asignatura

La estadística en general, es el campo del conocimiento; parte de la matemática aplicada, que se encarga de recopilar, ordenar y procesar información de los diversos contextos humanos, utilizando diversas técnicas, con la finalidad de obtener conclusiones sobre temas específicos que necesitan ser investigados para la posterior toma de decisiones.

En este sentido, la asignatura de estadística, como parte de carrera de administración, tiene particular importancia, puesto que considera los aspectos más relevantes del mundo empresarial para su análisis cuantitativo o cualitativo y permite conocer efectivamente las características de las poblaciones que intervienen en la administración de negocios y en la economía en general.

Siendo la estadística una asignatura que involucra la utilización constante de cálculos numéricos, necesita de la aplicación de los conocimientos de la matemática básica que un estudiante en este nivel debe tener, tales como las operaciones con expresiones algebraicas y desarrollo de fórmulas que procesen cuantitativamente las informaciones que el campo administrativo requiere.

Por todo lo expuesto, se determina que la estadística contribuye significativamente en el conocimiento de las características del entorno empresarial, permitiendo así, la oportuna toma de decisiones frente a las diversas situaciones o problemas que constantemente se presentan en este campo.

El presente texto contiene la información básica de cada uno de los temas y subtemas a tratar en el desarrollo de la asignatura, con ejemplos y problemas propuestos para que sean resueltos en el trascurso del semestre.

El uso de las TIC'S ha proporcionado paso a la eliminación de diversos trabajos rutinarios, además de permitir la reestructuración de campos laborales y los cambios de actitud y comportamiento de sus trabajadores.

En las organizaciones formadas por asociatividad, la tecnología es un factor clave en las tareas y responsabilidades de sus miembros, pues esta ayuda a determinar las tareas según el cargo que se posee.

Competencias específicas de la asignatura para la carrera

Aplicar con ética los conocimientos científicos y tecnológicos, en el campo de la investigación relacionada con las empresas; - Elaborar y asesorar estudios de planificación, ejecución y evaluación de proyectos de emprendimiento, de acuerdo con las dimensiones de sustentabilidad y principios de administración; - Asesorar en procesos para promocionar el procesamiento, conservación y comercialización de productos con alto componente de materia prima y valor agregado locales. - Promover y ejecutar la gestión administrativa de las empresas. -Elaborar diagnósticos y análisis de la realidad local. - Aplicar los conocimientos, la investigación y la vinculación con la sociedad, al desarrollo local integral, sistematizando los planteamientos de solución de acuerdo a la prioridad de los problemas que presenta el sistema desde el punto de vista social y productivo, a través de la creatividad e innovación.

Metodología de aprendizaje

El presente curso se basará en un aprendizaje activo centrado en todas las actividades que están planificadas y que usted estimado estudiante desarrollará con mucho compromiso, siendo el protagonista del proceso de aprendizaje. Esta metodología debe promover la reflexión y relacionar los nuevos aprendizajes con conocimientos previos que posea. El proceso de enseñanza aprendizaje debe hacer énfasis en la lectura, comprensión, cuestionamiento, discusión, aplicación de conceptos y solución de problemas.

La solución de problemas se presenta como aplicación de los conceptos y procedimientos vistos en las unidades didácticas; son problemas que se relacionan con su entorno y su futura profesión. Con esta estrategia usted aprende a analizar información y datos, a interpretarlos y se entrena en la toma de decisiones, considerando también que al uso de las tecnologías ya que facilita los aprendizajes autónomos. Es por eso que usted tendrá interacción permanente con las TICs y la plataforma Moodle.

UNIDAD 1. ESTADÍSTICA

Resultado de aprendizaje

Resume la información obtenida mediante cuadros de distribución de frecuencias con sus respectivas gráficas e interpreta el comportamiento de las variables involucradas.

Contextualización

El estudiante debe conocer y diferenciar los tipos de variables estadísticas, de manera que, sea capaz de elaborar tablas de distribución de frecuencias, con sus respectivas gráficas y realizar la interpretación correspondiente.

Contenidos:

1. Introducción a la estadística

1.1 Conceptos de Estadística

La estadística está presente en innumerables aspectos de la vida cotidiana, y la utilizamos frecuentemente, aunque de manera empírica y sin darnos cuenta, pues basta con que pensemos en la distancia diaria que recorremos a nuestro trabajo, cuánto gastamos semanalmente en alimentación, con qué frecuencia me lavo las manos, o cuántas horas veo televisión en la semana, etc. Son ejemplos sencillos de actividades cotidianas en las que está presente la estadística.

En sentido formal, diremos que la estadística es la ciencia, parte de la matemática, que se encarga de recopilar, organizar y procesar información obtenida de algún contexto de estudio, mediante una serie de métodos, técnicas, herramientas y parámetros, que ayudan a recolectar, organizar, clasificar y analizar datos, todo con la finalidad de obtener ciertos indicadores y conclusiones de la realidad observada para tomar las decisiones más oportunas según sea el caso.

1.1.1 Estadística descriptiva e inferencial

La estadística descriptiva se encarga del procesamiento y análisis de los datos con el fin de proporcionar las conclusiones acerca de un tema de investigación y con ello, tener elementos de juicio para tomar decisiones, pero siempre y cuando se trabaje con poblaciones enteras; es decir, no se deben realizar inferencias aplicables a la población a partir de una muestra de la misma.

En cambio, la estadística inferencial, se encarga de realizar estudios similares a los de la descriptiva pero una muestra representativa de la población y luego sus resultados pueden ser generalizados y obtener conclusiones acerca de la población entera. Un ejemplo muy claro es el trabajo de las encuestadoras en las elecciones sean seccionales o nacionales, pues a partir de muestras representativas, tomadas a boca de urna, se sabe quiénes son los candidatos vencedores antes de empezar a contar los votos de las urnas. Por supuesto, depende mucho de la seriedad de la encuestadora y de la no intervención de factores que puedan tergiversar la información.

También, mediante la estadística inferencial de pueden probar hipótesis y realizar investigaciones a nivel de post – grado.

1.1.2. Población y muestra

La población es la totalidad de los individuos, objeto del estudio estadístico, y no se refiere solamente a personas, sino, al número total de objetos o sujetos de quienes se va a tomar la información. Por ejemplo, en un estudio sobre la calidad de la comida en los restaurantes de un centro turístico, la población son los restaurantes que existen en ese lugar.

La muestra es un subconjunto de la población, pero con sus mismas características, es decir, que una conclusión a la que se llegue sobre una muestra, pueda ser generalizada a la población entera. Se utilizan cuando las poblaciones son muy extensas o se desconoce el tamaño real de la población.

1.1.3. Parámetros y estadísticos

Las medidas representativas de una población se conocen con el nombre de parámetros y esas mismas medidas aplicadas sobre una muestra, se llaman estadísticos. Un ejemplo de ello, es la mediana muestral y la mediana poblacional. Si la muestra es representativa de la población, los estadísticos pueden ser inferidos a la población; todo depende del nivel de confianza que exista y para determinar dicho nivel, también existen indicadores estadísticos que se pueden calcular.

1.2. Variables.

Las variables son las características observables y medibles de una población o muestra; en otras palabras, son los aspectos a observar, por ejemplo, si en un grupo de estudiantes del instituto Superior Tecnológico Pichincha, queremos investigar su “rendimiento académico”, esa es la variable, o si en un conjunto de bares de la ciudad, deseamos investigar la “calidad de la atención al cliente”, esa es la variable y las has de varios tipos.

1.2.1. Variables cualitativas.

Denotan ciertas categorías no numéricas, por ejemplo, el género (masculino – femenino), o también pueden denotar algún orden como nivel de satisfacción en la atención al cliente en una tienda (excelente, bueno, malo, pésimo).

1.2.2. Variables cuantitativas.

Estas denotan categorías numéricas, como la edad (20 años, 25 años, etc.), otro ejemplo el ingreso mensual ($450.76, $817.21)

Según el grado de influencia de una variable sobre otra también pueden considerarse las variables independientes y dependientes.

1.3. Escalas de medición

Para determinar la medida de cada categoría de la variable, sea esta cualitativa y cuantitativa, se hace necesario recurrir a las escalas de medición, que son niveles utilizados como referencia

1.3.1. Escala nominal.

Indican la característica no numérica (**cualitativa**) de un objeto o sujeto, pero sin ningún orden en particular, un ejemplo de ello es el género, que puede ser masculino o femenino y ninguno de ellos tienen mayor o menos jerarquía. No se pueden calcular parámetros o estadísticos con estas variables.

1.3.2. Escala ordinal.

Expresan características no numéricas, (**cualitativas**) pero denotan orden o jerarquía, por ejemplo, el nivel de satisfacción por la atención al cliente en un bar, cuya escala puede ser: muy satisfecho, satisfecho, poco satisfecho e insatisfecho.

1.3.3. Escala de intervalo.

Es de tipo numérico (cuantitativa) con la característica de que el cero no es absoluto, lo que significa que no indica ausencia de valor y solo es un punto de referencia, por ejemplo, la temperatura ambiental, en la cual, el CERO, no significa la ausencia de temperatura, sino que es muy frío pero que hay temperaturas mucho más frías menores a cero en lugares cercanos a los polos y en épocas invernales y también hay temperaturas muy cálidas como las de las costas en la zona ecuatorial sobretodo del África.

1.3.4. Escala de razón.

Esta escala considera al cero absoluto, es decir, el cero significa ausencia de algo, por ejemplo, el dinero que llevo en el bolsillo, que al no tener nada, significa ausencia de dinero.

Actividad De Aprendizaje recomendada n° 1

- Elabore un mapa conceptual sobre los conceptos básicos de estadística.

Actividad De Aprendizaje recomendada n° 2

- Elabore un cuadro comparativo de las variables estadísticas, con ejemplos de cada una.

1.3. El operador sumatoria.

Su símbolo es la letra griega sigma mayúscula (Σ); es un operador matemático que representa sumas con varios, muchos e incluso infinitos sumandos.

$$\sum_{i=m}^{n} x_i = x_{m+1} + x_{m+2} + \dots + x_n$$

Notación: n = límite superior

m = límite inferior

i = índice de la sumatoria

xi = valores a sumar

$$\sum_{i=1}^{6} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

Ejemplo:

La suma de los seis primeros dígito $\sum_{i=1}^{6} x_i = 1+2+3+4+5+6=15$

Actividad recomendad nº 3

Resuelva las siguientes sumatorias:

1) $\sum_{i=2}^{5} x_i$

2) $\sum_{i=3}^{7} x_i$

3) $\sum_{i=1}^{3} x_i + \sum_{i=4}^{8} x_i$

4) $\sum_{i=1}^{3} 2x_i$

5) $\sum_{i=1}^{5} i(i+1)$

Escriba en forma de sumatoria:

6) $x_3 + x_4 + x_5$

7) $\frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{1}{x_5} + \frac{1}{x_6}$

8) $3+6+9+12$

9) $(3+4+5)-(6+7+8)$

10) La suma de los dígitos impares

1.4. Parámetros estadísticos para datos sin agrupar

1.4.1. La media aritmética.

Es la suma de todos los datos numéricos dividido para el número de los mismos

Para poblaciones completas la fórmula es: $\mu = \frac{\sum x_i}{N}$

Fórmula para muestras: $\overline{X} = \frac{\sum x_i}{n}$

Ejemplo: obtener el promedio de la siguiente muestra da datos: 2, 7, 8, 11, 13, 17, 19

$$\overline{X} = \frac{2+7+8+11+13+17+19}{7} \qquad \overline{X} = \frac{77}{7} \qquad \overline{X} = 11$$

1.4.2. La media ponderada.

Mientras que en la media aritmética todos sus valores tienen la misma importancia, en la media ponderada, cada valor puede tener una importancia o "ponderación" diferente, por ejemplo, si las tres calificaciones de estadística se califican sobre 10, y un estudiantes obtuvo 6.5, 6 y 8, al sumar y dividir para 3, se está dando un peso de 1/3 a cada una de las calificaciones y su promedio sería **6,83** con lo que perdería la asignatura.

Ahora, si la primera calificación tiene un peso del 25%, la segunda, el 15% y la tercera el 60%, su promedio se calcularía de la siguiente manera:

$$\overline{X}_p = \frac{6.5 \times 0{,}25 + 6 \times 0{,}15 + 8 \times 0{,}6}{0{,}25 + 0{,}15 + 0{,}60} =$$

7,325 Esta calificación sí le permitiría aprobar la asignatura.

Fórmula para la media ponderada

$$\overline{X}_p = \frac{\sum_{i=1}^{n} x_i \cdot w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_n}{w_1 + w_2 + \ldots + w_n}$$

1.4.3. La mediana.

En realidad, la mediana es una medida de ubicación del dato que está en el centro. Para ello es necesario ordenarlos de menor a mayor o viceversa y si el número de datos es impar, aquel que se encuentre ubicado en el centro ya es la mediana, mientras que, si el número de datos es par, se seleccionan los dos datos que se encuentran en el centro y la mediana será el promedio simple de estos.

En los datos del primer ejemplo de media aritmética, se cuantifican 7 datos que se encuentran ordenados de menor a mayor, así, en 2, 7, 8, **11**, 13, 17, 19, el número 11 se encuentra en el cuarto puesto, ubicándose tres datos por debajo de ese valor y tres datos por encima del mismo, por lo tanto, le **11** es la mediana.

Si los datos son 2, 7, 8, **11**, **13**, 17, 19, 20, existen 8 valores y por lo tanto hay dos intermedios, que en este caso son 11 y el 13, con los que se procede a obtener el promedio entre los dos, (11 + 13) / 2 = 24/2 = **12**, entonces esa sería la mediana.

Actividad calificada n° 1

Resolver las siguientes situaciones

1. Francisco se dedicó a practicar videojuegos cada día de la semana durante las siguientes horas: 3, 2, 3, 3, 2, 6, 3, calcule el promedio de horas que jugó diariamente
2. Se presentan a continuación las veces que Mariana asiste a la iglesia en un mes: 1, 2, 3, 3, 4, 2, 1; calcular su promedio y su mediana.
3. Alicia obtuvo las siguientes notas en las diferentes asignaturas del colegio: 7, 5, 6, 8, 7, 8, 8, 9, 10, 10. Calcule la media aritmética y la mediana.
4. Seis compañeras de trabajo conversan sobre sus edades. Si ellas tienen 27, 32, 32, 29, 28 y 35 años. Calcule la media aritmética y la mediana.
5. Los jugadores de un equipo de baloncesto tienen las siguientes estaturas en metros: 1.92, 1.95, 1.83, 1.86 y 1.79. calcule la media aritmética y la mediana.
6. Las calificaciones de 10 estudiantes de estadística son: 7.0, 6.5, 4.2, 2.1, 9.0, 5.0, 8.0, 8.5, 2.0 y 5.5. Calcule la media aritmética y la mediana de las notas de la clase.
7. En la clase de estadística Pablo ha obtenido tres calificaciones con diferentes ponderaciones. Las 2 tareas integradoras tienen un valor de 20% y 20% respectivamente, y el examen de 60%; las calificaciones respectivas son de 6.4, 9.2 y 7.3. Calcule su media ponderada.
8. Una agencia de viajes tramita durante 5 semanas los siguientes viajes: 20, 73, 75, 80, 82. Calcule la media aritmética y la mediana.
9. Las ventas de aceite comestible de litro en un micro mercado durante la semana fueron las siguientes: 40, 55, 62, 43, 50, 60, 65. Calcule la media aritmética y la mediana
10. La ponderación de los exámenes de inglés en un curso internacional son 25%, 35% y 40%. Si un estudiante obtuvo 7, 5 y 9 respectivamente, calcular la media ponderada

1.4.4. La varianza

La varianza es una medida de dispersión; es un promedio de los cuadrados de las diferencias de cada uno de los datos respecto a la media aritmética. Sirve de base para el cálculo de otros parámetros estadísticos muy importantes en la investigación.

Su fórmula es $\vartheta^2 = \frac{\sum (X_i - \mu)^2}{N}$,si se va a calcular sobre la población entera,y $\vartheta^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$,si se calcula sobre una muestra.

Ejemplo: (za para la siguiente serie de datos: 12, 6, 7, 3, 15, 10, 18, 5.

Primeramente, se obtiene la media aritmética, que en este caso es 9,5

Ahora se resta cada uno de los datos con la media aritmética.

12	–	9,5	=	2,5
6	–	9,5	=	– 3,5
7	–	9,5	=	– 2,5
3	–	9,5	=	– 6,5
15	–	9,5	=	5,5
10	–	9,5	=	0,5
18	–	9,5	=	8,5
5	–	9,5	=	– 4,5

Se eleva al cuadrado cada una de las diferencias

$(2,5)^2$	=	6,25
$(-3,5)^2$	=	12,25
$(-2,5)^2$	=	6,25
$(-6,5)^2$	=	42,25
$(5,5)^2$	=	30,25
$(0,5)^2$	=	0,25
$(8,5)^2$	=	72,25
$(-4,5)^2$	=	20,25

Se obtiene la media aritmética de los cuadrados obtenidos y ese valor será la varianza.

$$\vartheta^2 = \frac{190}{8} \qquad \vartheta^2 = 23{,}75$$

1.4.5. La desviación estándar

Es la medida de dispersión más común, indica qué tan dispersos están los datos y qué tan alejados se encuentran estos datos de la media. Se representa con la letra sigma y se obtiene simplemente extrayendo la raíz cuadrad a la varianza.

En una distribución normal, dentro de las tres primeras deviaciones estándar, tanto para la derecha como para la izquierda de la media, se encuentran ubicados prácticamente la totalidad de los datos (99,7%), por lo tanto, no es confiable un promedio si la desviación estándar tiene una gran medida. Para entender mejor este concepto, calcularemos la desviación estándar con los datos del ejercicio anterior.

$$\vartheta^2 = 23{,}75$$

$$\vartheta = \sqrt{23{,}75}$$

$$\vartheta = 4{,}47$$

Cuestionario de repaso

1. La masa corporal en kg de los trabajadores de una microempresa es:

    a) Cuantitativa, continua, de intervalo

    b) Cualitativa, nominal

    c) Cuantitativa, discreta, de razón

    d) Cuantitativa, continua, de razón

2. Una variable estadística es:

a) Una muestra representativa de una población o muestra

b) Una característica observable y medible en una población o muestra

c) Un parámetro que se debe observar y medir

d) Un aspecto de la población que tiene variabilidad

3. Seleccione el resultado de la sumatoria $\sum_{i=5}^{16} i - \sum_{i=2}^{10} i$

a) 64

b) 72

c) 90

D) 81

4. Alicia obtuvo las siguientes calificaciones en Estadística 8.4, 9.1, 7.2, 6.8, 8.7, 7.8. Calcular la media aritmética simple:

a) 7,52

b) 8,00

c) 7,89

d) 8,05

5. Las temperaturas promedio en la ciudad de Quito durante todos los días del mes de junio fueron: 13º, 13º, 17º, 18º, 15º, 16º, 15º, 11º, 12º, 15º, 11º, 18º, 17º, 11º, 11º, 18º, 19º, 12º, 17º, 20º, 17º, 11º, 16º, 19º, 14º, 16º, 17º, 11º, 13º, 12º. ¿Cuál es su varianza?

a) 3,16º

b) 2,07º

c) 2,84º

d) 3,95º

6. Se midieron las estaturas en centímetros a una muestra de 35 estudiantes de la carrera de administración. Calcule la varianza muestral si las mediciones fueron las siguientes: 167, 161, 161, 169, 176, 158, 172, 164, 169, 171, 159, 173, 162, 163, 176, 174, 160, 161, 157, 159, 159, 175, 168, 176, 166, 173, 165, 166, 176, 178, 158, 168, 173, 176, 163. Determine la desviación típica.

   a) 6,58 cm.

   b) 5,49 cm.

   c) 7,12 cm.

   d) 4,92 cm.

7. Las edades en años de 25 trabajadores de una fábrica de calzado son: 29, 58, 51, 53, 40, 46, 5 8, 47, 24, 48, 52, 24, 44, 55, 43, 24, 28, 24, 33, 27, 49, 58, 23, 31, 57. ¿Cuál es la mediana de los datos?

   a) 44

   b) 47

   c) 39

   d) 42

8. Si las calificaciones promedio obtenidas por Ismael sobre 50 puntos fueron: 46, 35 y 27 pero tienen pesos del 20%, 30% y 50%. ¿Aprueba la signatura?; ¿Por qué?

   a) Sí porque su promedio ponderado es 35,33

   b) No porque su promedio ponderado es 33,50

   c) Sí porque su promedio ponderado es 71,20

   d) No porque su promedio ponderado es 32,40

9. ¿Cuál de las siguientes expresiones es incorrecta?

a) La media aritmética indica el punto medio de todos los datos cuantitativos recopilados

b) Un estadístico es un parámetro aplicado a la muestra

c) La muestra es un subconjunto representativo de la población

d) La varianza es la raíz cuadrada de la desviación estándar.

10. ¿Cuál es el procedimiento para determinar la mediana?

a) Eliminar los datos que estén por sobre o debajo del valor medio

b) Ubicar en orden alfabético y encontrar el punto medio

c) Ubicar de mayor a menor y seleccionar el dato que está en el medio

d) Sumar todos los valores y dividir para el número de valores.

UNIDAD 2. PARÁMETROS Y ESTADÍSITICOS PARA AGRUPACIÓN SIMPLE

Resultado de aprendizaje

Calcula e interpreta las medidas de tendencia central y dispersión tanto de datos simples como agrupados.

Contextualización

El estudiante debe ser capaz de procesar los datos estadísticos de variables cuantitativas tanto discretas como continuas, obteniendo los parámetros o estadísticos que le permitirán tomar decisiones en las situaciones que lo ameriten.

2. Medidas de posición y dispersión

2.1. Coeficiente de variación.

Considerando el ejercicio de la unidad anterior en el que la media aritmética es 9.5 y la desviación estándar es 4.47, resulta que, relacionando los dos valores, la desviación típica es más del 51% de la media; para ser más precisos, se dividirá la desviación estándar para la media aritmética; así:

$$CV = \frac{\vartheta}{\mu} \qquad CV = \frac{4{,}47}{9{,}5} \qquad CV = 0{,}51$$

Una desviación tan alta significa que existen datos muy alejados de la media y puede llegar a deslegitimar las conclusiones del estudio.

Al cociente obtenido al dividir la desviación estándar para la media aritmética, se lo conoce con el nombre de ***coeficiente de variación,*** que es una medida relativa

2.2. Sesgo

El sesgo en estadística muestra la forma en la que están distribuyen los datos respecto a la media; es decir, que puede ocurrir que el mato mayor se encuentre muy alejado de la media mientras el menor de los datos sea cercano a éste o viceversa, y en ese caso el valor promedio será menos representativo de la población.

El sesgo puede clasificarse en:

2.2.1. El sesgo de selección.

Cuando se elige la muestra sin criterios técnicos u objetivos y se consideren aspectos como la afinidad o intereses creados

2.2.2. Sesgo de información.

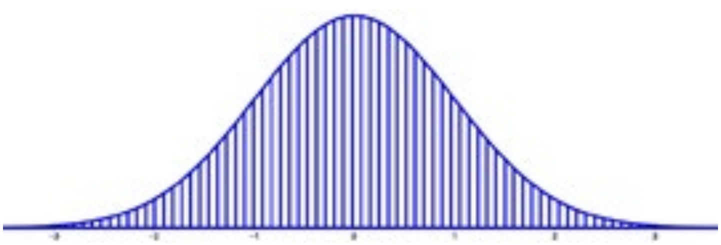
Cuando los datos que se van a utilizar en el análisis estadístico no sean los suficientes.

2.2.3. Sesgo de confusión.

En este caso existe una variable llamada así, de confusión, que es la que provoca el sesgo. Suele ser difícil encontrar dónde está el problema.

2.2.4. Coeficiente de asimetría de Fisher.

Si las frecuencias del conjunto de datos se distribuyen normalmente, forman una especie de campana simétrica, esto significa que existe igual distancia entre los datos centrales y sus extremos mayor y menor, tal como se muestra en la figura.



Sin embargo, en muchas ocasiones, la distancia entre el centro de los datos y su valor extremo mayor, puede ser distinta a la del mismo centro y su valor extremo menor, generándose curvas asimétricas, lo que demuestra que existe un sesgo sea positivo o negativo, tal como se muestra en las figuras.

Asimétrica a la derecha (sesgo positivo

Asimétrica a la izquierda (sesgo negativo)

Para determinar el sesgo en los datos, se utiliza el coeficiente de asimetría de Fisher, cuya fórmula se muestra a continuación.

$$g_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{ns^3}$$

Donde:

n = tamaño de la muestra

xi = valor da cada una de las observaciones

$\bar{x}$ =media aritmética

s = desviación estándar de los datos.

Un coeficiente positivo, indica que existe asimetría hacia la derecha. Si el coeficiente es negativo, existe asimetría hacia la izquierda. Si el coeficientre es exactamente cero, significa que no existe asimetría.

Ejemplo:

Calcular el coeficiente de asimetría de Fisher con los siguientes datos:

13404, 13443, 13445, 13447, 13449, 13450, 13453, 13455, 13457, 13460, 13465

La media aritmética es 13448

Su desviación estándar es: 15,26732

N°	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3$
1	13404	-44	-85184
2	13443	-5	-125
3	13445	-3	-27
4	13447	-1	-1
5	13449	1	1
6	13450	2	8
7	13453	5	125
8	13455	7	343

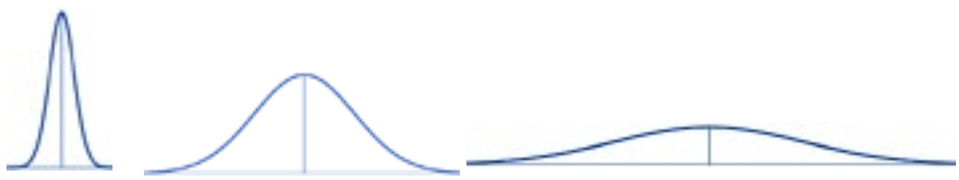
9	13457	9	729
10	13460	12	1728
11	13465	17	4913
TOTAL	11		-77490

$$g_1 = \frac{-77490}{11 \cdot 15{,}2673^3} = -1{,}979543$$

Puesto que el coeficiente es negativo, se concluye que la distribución es asimétrica hacia la izquierda.

2.3. La curtosis.

También conocida como medida de apuntamiento, determina si los datos tienen mayor o la mayor o menor concentración de frecuencias alrededor de la media, así, si los datos están concentrados muy cerca de la media, la curva es más apuntada y se llama leptocúrtica, si los datos se distribuyen normalmente alrededor de la media, la curva es mesocúrtica y si los datos se alejan mucho de la media, la curva es aplanada y se llama platicúrtica. Sus representaciones gráficas son las siguientes:



Leptocúrtica Mesocúrtica Platicúrtica

La fórmula para el cálculo de la curtosis es: $g_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4} - 3$

Ejemplo: En el ejercicio anterior, cuya media aritmética es 13448 y desviación estándar igual a 15,26732, la curtosis se calcularía de la siguiente manera:

N°	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^4$
1	13404	-44	3748096
2	13443	-5	625
3	13445	-3	81
4	13447	-1	1
5	13449	1	1
6	13450	2	16
7	13453	5	625
8	13455	7	2401
9	13457	9	6561
10	13460	12	20736
11	13465	17	83521
TOTAL	**11**		**3862664**

$$g_2 = \frac{3862664}{11 \times 15{,}26732^4} - 3$$

$$g_2 = 3{,}463140$$

La curtosis es positiva, lo que significa que la curva es leptocúrtica, es decir, los datos alrededor de la media están muy cercanos.

2.4. Los cuantiles

Los cuantiles que dividen los datos en intervalos iguales. Los más importantes son:

2.4.1. Mediana.

Es un valor intermedio y divide a los datos en dos partes iguales.

$$\begin{cases} \text{impar: la mediana está en el lugar } \dfrac{n+1}{2} \\ \text{par: la mediana se ubica entre los datos } \dfrac{n}{2} \text{ y } \dfrac{n}{2}+1 \end{cases}$$

Ejemplo: obtener la mediana de las calificaciones de estadística de 20 estudiantes.

5, 5, 8, 7, 9, 10, 7, 6, 8, 7, 8, 9, 10, 10, 8, 7, 6, 5, 9, 6

Como el número de valores es par (20), se ordenan de forma ascendente o descendente y se determinan las ubicaciones de los dos datos centrales, en este caso, 20/2 = 10 y el siguiente que sería el onceavo y se obtiene el promedio de ambos.

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, **8, 8,** 8, 8, 9, 9, 9, 10, 10, 10

Como los dos datos centrales son iguales (8), entonces, su promedio es el mismo y en consecuencia, la **mediana** del conjunto de datos es 8.

2.4.2. Cuartiles

Son tres valores intermedios que dividen a la población en cuatro intervalos iguales. Para su cálculo se utiliza $(N+1)/4$ y $3(N+1)/4.$. Si el resulta

do es un decimal, se utilizará para d terminar el

valor del cuartil. Así, en el eje $Q_1 = x_i + d(x_{i+1} - x_i)$ cede de la siguiente manera.

5, 5, 5, 6, **6, 6,** 7, 7, 7, 7, 8, 8, 8, 8, **9, 9,** 9, 10, 10, 10

Como hay 20 datos, para ubicar el cuartil 1 se aplica la fórmula:

$(20+1)/4 = 5{,}25$. Si resulta un decimal, aplicamos: $Q_1 = x_i + d(x_{i+1} - x_i)$

$$Q_1 = 6 + 0{,}25(6-6)$$

$$Q_1 = 6$$

$3(20+1)/4 = 15{,}75$ **que también es un decimal y entonces aplicamos:**

$Q_3 = 9 + 0{,}25(9-9)$

$Q_3 = 9$

2.4.3. Quintiles.

Son cuatro y dividen a la población en cinco intervalos de igual tamaño y la fórmula para el cálculo de los quintiles es $\frac{k(n+1)}{5}$ y al igual que los cuartiles, si resulta un decimal, se debe aplicar $K = x_i + d(x_{i+1} - x_i)$. . Así, con los datos planteados

vamos a determinar el valor del tercer quintil.

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, **8, 8,** 8, 9, 9, 9, 10, 10, 10

$$\frac{3(20+1)}{5} = 12{,}6$$

En vista de que la respuesta es decimal, aplicamos $K_3 = 8 + 0{,}6(8-8)$

Entonces, $K_3 = 8$

2.4.4. Deciles.

Son nueve y dividen a la población en 10 intervalos iguales. Sus fórmulas son $\frac{k(n+1)}{10}$ y $K = x_i + d(x_{i+1} - x_i)$ si resulta decimal.

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, **8, 9,** 9, 9, 10, 10, 10

Con los mismos datos, se determinará el decil siete.

$$\frac{7(20+1)}{10} = 14{,}7$$

Como se observa, está ubicado entre los valores 14 y 15, entonces procedemos con el cálculo. D_7*= 8+0,7 (9-8)*, de esta manera se determina que D_7 = 14,7

2.4.5. Percentiles.

Son 99 y dividen a la población en 100 intervalos de igual tamaño, para lo cual, su fórmula sería *k(n+1)*.

Ejemplo. Determinar el percentil 84 con los datos anteriores.

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, **9, 10,** 10, 10

$\frac{84\,(20+1)}{100} = 17{,}64$ 100

Luego se procede con C_{84} =9+00,64 (10-9)

C_{84} = 9,64

2.5. Datos agrupados

Cuando el número de datos es considerable, se dificulta analizarlos individualmente y entonces se los agrupa de acuerdo a su valor y se cuenta la frecuencia con la que se repite cada uno de ellos. Así, a los datos 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 10, 10, 10, se los agrupa de la siguiente manera:

N°	CATEGORÍAS	frecuencia absoluta (*f*)	frecuencia relativa (*fr*)	%
1	5	3	0,15	15%
2	6	3	0,15	15%
3	7	4	0,2	20%
4	8	4	0,2	20%
5	9	3	0,15	15%
6	10	3	0,15	15%
		20	1	100%

Como se observa, se formaron seis categorías de números y cada una se repite varias veces, a esta repetición de un mismo dato se le conoce como frecuencia absoluta y se representa con la letra f. Esta información se ubica en la denominada ***"tabla de distribución de frecuencias simples"***, en la que también se ubican la frecuencia relativa que es la fracción del grupo entero al que representa el dato y además se ubica también el porcentaje que es la proporción en relación al 100 al que representa cada dato; en otras palabras, partimos al grupo entero en cien partes iguales, y de ellas tomamos algunas.

2.6. Agrupación por intervalos.

Cuando el número de observaciones de una variable cuantitativa es considerable y además sus valores son muy diversos, se tornaría un trabajo tedioso y poco fructífero generar categorías simples y entonces se recurre a la ***tabla de distribución de frecuencias con intervalos de clase***, que no es más que la formación de grupos de datos en intervalos iguales. Para ello, necesitamos calcular el ***"rango"***, que es la diferencia entre los valores más alto y más bajo de los datos, luego se calcula la ***"amplitud de cada intervalo"*** que es la distancia que se genera entre los valores más alto y más bajo de cada intervalo, y para ello, se divide el rango para el número de intervalos (categorías) que haya decidido hacer el investigador y se redondea al inmediato superior, aunque el decimal sea menor que 5.

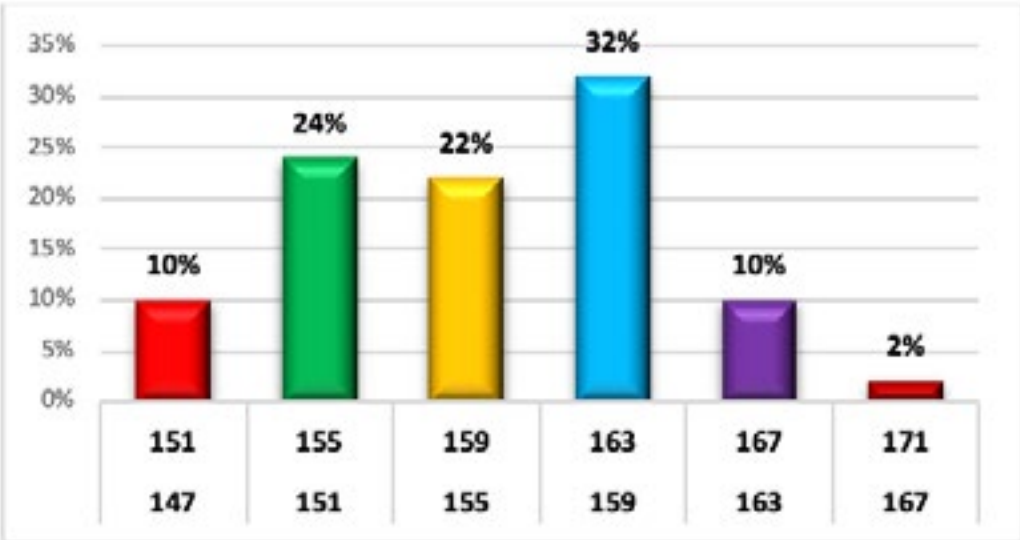
El número de intervalos (categorías) cuando se trata de variables cuantitativas, no es una camisa de fuerza y depende mucho del criterio del investigador, aunque hay algunas fórmulas que sugieren ese número como: 1,32xlog (N) o $\sqrt{N}$.

Ejemplo, con siguiente grupo de números, corresponde a las masas corporales de 50 empleados en una empresa. Elabore la tabla de distribución de frecuencias con intervalos.

156, 159, 163, 162, 165, 161, 159, 151, 162, 162, 153, 149, 157, 154, 153, 159, 158, 157, 147, 164, 155, 159, 153, 156, 153, 147, 157, 160, 154, 156, 150, 162, 159, 162, 154, 168, 152, 162, 162, 149, 165, 153, 159, 156, 154, 158, 152, 163, 156, 162

1. Rango (R) = Valor más alto es 168, el menor es 147 y su diferencia **(rango)** es 21
2. Se elaborarán 6 intervalos (nº elegido arbitrariamente por el investigador)
3. Amplitud de cada intervalo: $i = \frac{R}{m}$, $i = \frac{21}{6}$, $i = 3.5$, $i = 4$

N°	Intervalos	*f*	*fr*	*fa*	%
1	147 – 151	5	0,1	5	10%
2	151 – 155	12	0,24	17	24%
3	155 – 159	11	0,22	28	22%
4	159 – 163	16	0,32	44	32%
5	163 – 167	5	0,1	49	10%
6	167 – 171	1	0,02	50	2%
		50	**1**		**100%**



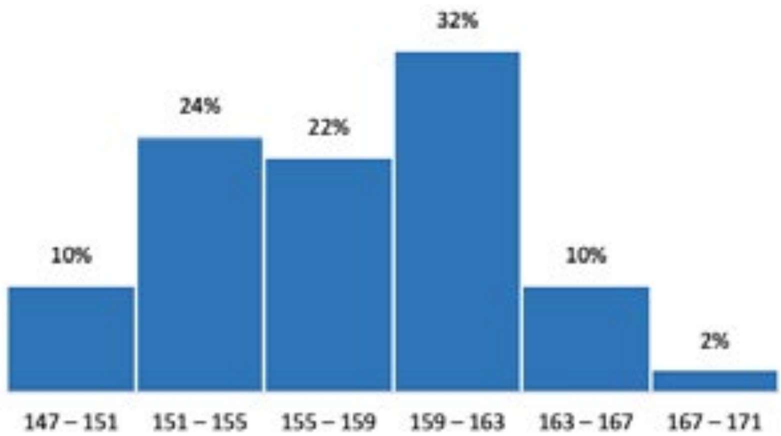
Interpretación: como se observa, la gran mayoría de los encuestados tiene una masa corporal que oscila entre las 151 y 163 libras, evidenciando que se encuentran todavía dentro de los límites saludables.

Las dimensiones de las barras del gráfico son proporcionales a sus frecuencias, así mismo, si se decide elaborar una gráfica de sectores circulares, las medidas angulares de cada sector, son proporcionales a su frecuencia o porcentaje.

Histogramas

Expresa el comportamiento de los datos continuos y poder comprobar si los datos se distribuyen de manera normal o no. A continuación se presenta el histograma de los datos del anterior ejercicio.

Polígonos de frecuencias



Pasteles.

La gráfica de sectores circulares es una de las más utilizadas para la representación de datos estadísticos, de manera especial cuando se trata de datos cualitativos. En él, cada sector representa a una de las categorías de la variable. Es muy fácil de interpretar

Actividad calificada n° 2

Diseñar una encuesta en línea con un mínimo de 10 ítems y aplicarla a 30 personas o más y elaborar las tablas de distribución de frecuencias con los datos obtenidos

Cuestionario de repaso

1. El coordinador de transporte del municipio, realiza un estudio sobre la velocidad a la que los conductores manejan en el centro de la ciudad. A continuación, se muestran las velocidades de 50 conductores, a los que se les aplicó aleatoriamente el lector de velocidad en km/h.

   15, 32, 45, 46, 42, 39, 68, 47, 18, 31, 48, 49, 56, 52, 39, 48, 69, 61, 44, 42, 38, 52, 55, 58, 62, 58, 48, 56, 58, 48, 47, 52, 37, 64, 29, 55, 38, 29, 62, 49, 69, 18, 61, 55, 49, 60, 52, 51, 61, 66

   Si se hace una tabla de distribución de frecuencias con 7 intervalos: ¿Qué porcentaje corresponde al cuarto intervalo?

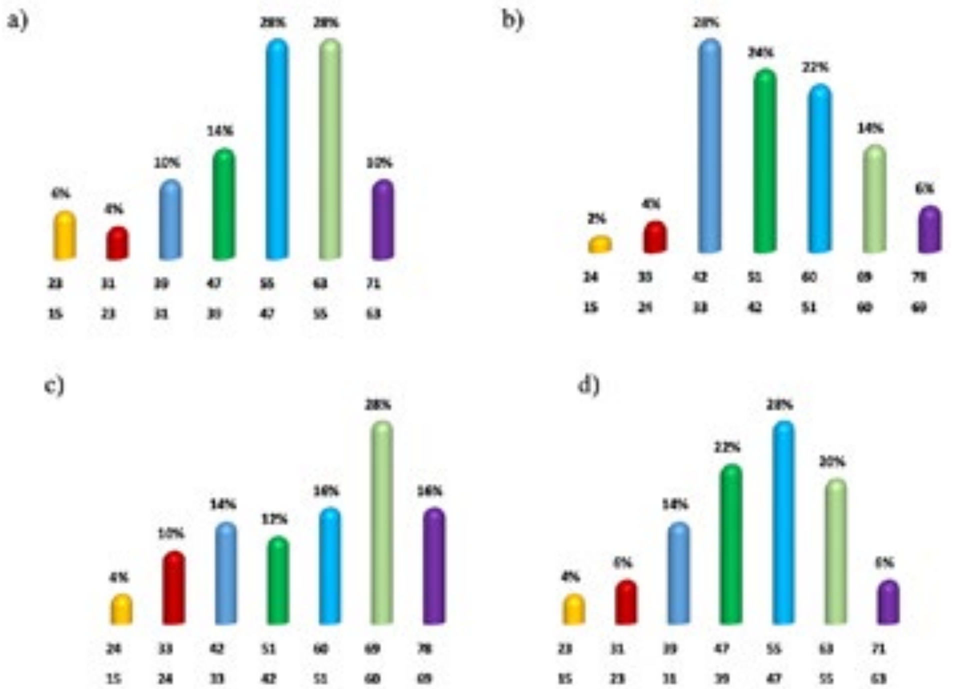
   a) 26%

   b) 14%

   c) 28%

   d) 10%

   Seleccione la gráfica que corresponde a los datos.



2. El GAD provincial realiza una encuesta a una muestra de 50 familias sobre el número de hijos que tienen con el fin de determinar la variación en la densidad demográfica, con respecto a años anteriores, obteniendo los siguientes resultados:

0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 6

Encuentre:

2.1. Cuartil 3

a) 3

b) 3,25

c) 3,75

d) 4

2.2. Decil 8

a) 3,9

b) 4

c) 4,9

d) 5

2.3. Percentil 84

a) 4,84

b) 4

c) 3,84

d) 3

2.4. Coeficiente de asimetría

a) 0,75

b) 0,37

c) 0,63

d) 0,42

2.4. La curtosis

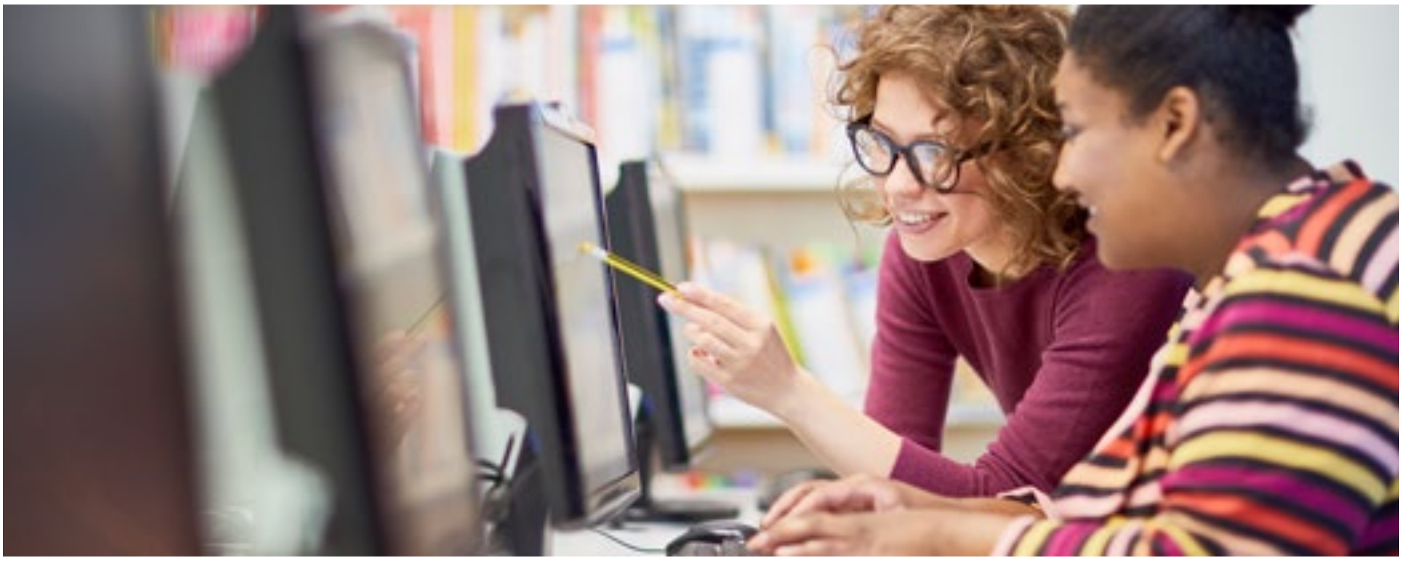
a) 1,62

b) 0,96

c) 1,07

d) 1,28

UNDAD 3. PARÁMETROS PARA DATOS AGRUPADOS EN INTERVALOS DE CLASE



Resultado de aprendizaje

Procesa e interpreta la información obtenida con técnicas de recolección de datos de una población o muestra y determina los parámetros estadísticos pertinentes

Contextualización

El desarrollo de la temática de la presente unidad extenderá al estudiante las herramientas matemáticas y tecnológicas para procesar información cuantitativa con gran variabilidad, agrupando en intervalos y obteniendo conclusiones válidas en el contexto de las investigaciones que se realicen dentro de su carrera.

3. Parámetros estadísticos para datos agrupados en intervalos de clase

3.1. Media aritmética en datos agrupados en intervalos

Para obtener la media aritmética para datos agrupado se multiplica la marca de clase (punto medio de cada intervalo) por su correspondiente frecuencia absoluta, se sumas esos productos y se divide para el número de datos, ejemplo.

Los datos que se presentan a continuación corresponden a las masas corporales medidas en Kg. de ochenta estudiantes del tecnológico Pichincha.

60, 66, 77, 70, 66, 68, 57, 70, 66, 52, 75, 65, 69, 71, 58, 66, 67, 74, 61, 63, 69, 80, 59, 66, 70, 67, 78, 75, 64, 71, 81, 62, 64, 69, 68, 72, 83, 56, 65, 74, 67, 54, 65, 65, 69, 61, 67, 73, 57, 62, 67, 68, 63, 67, 71, 68, 76, 61, 62, 63, 76, 61, 67, 67, 64, 72, 64, 73, 79, 58, 67, 71, 68, 59, 69, 70, 66, 62, 63, 66

Agrupar los datos en intervalos y encuentre la media aritmética.

N°	Masas	f	fr	%	X_m	fX_m
1	52 – 58	5	0,06	6%	55	275
2	58 – 64	17	0,21	21%	61	1037
3	64 – 70	34	0,43	43%	67	2278
4	70 – 76	16	0,20	20%	73	1168
5	76 – 82	7	0,09	9%	79	553
6	82 – 88	1	0,01	1%	85	85
		80	**1**	**100%**		**5396**

$$\overline{X} = \frac{\sum fX_m}{N}$$

$$\overline{X} = \frac{5396}{80}$$

$$\overline{X} = 67,45 \text{ kg}$$

3.2. La varianza para datos agrupados con intervalos

Se divide la sumatoria de los cuadrados de las diferencias entre la media aritmética y la marca de clase para el número de datos.

N°	Masas	f	fr	%	X_m	fX_m	$f(x_m-\bar{x})^2$
1	52 – 58	5	0,06	6%	55	275	775,0125
2	58 – 64	17	0,21	21%	61	1037	707,2425
3	64 – 70	34	0,43	43%	67	2278	6,885
4	70 – 76	16	0,20	20%	73	1168	492,84
5	76 – 82	7	0,09	9%	79	553	933,8175
6	82 – 88	1	0,01	1%	85	85	308,0025
		80	**1**	**100%**		**5396**	**3223,8**

$$\vartheta^2 = \frac{3223,8}{80}$$

$$\vartheta^2 = 40,30$$

3.3. La desviación estándar para datos agrupados con intervalos

De la misma manera que con los datos simples, la desviación estándar es la raíz cuadrada de la varianza

$$\vartheta = \sqrt{40,30}$$

$$\vartheta = 6,35$$

3.4. El coeficiente de variación para datos agrupados con intervalos

Es la razón entre la desviación estándar y la media aritmética.

$$CV = \frac{\vartheta}{X}$$

$$CV = \frac{6{,}35}{67{,}45}$$

$$CV = 0{,}094$$

3.5. Coeficiente de asimetría de Fisher para datos agrupados con intervalos.

La diferencia respecto al mismo parámetro calculado con frecuencias simples, es que se usa la marca de clase da cada intervalo (Xm) en lugar de la medición Xi. Así, con los datos del ejemplo anterior, tenemos:

$$g_1 = \frac{\sum f\left(X_m - \overline{X}\right)^3}{n\vartheta^3}$$

N°	Masas	*f*	*fr*	*%*	*Xm*	*fXm*	$f(x_m\text{-}x)^3$
1	52 – 58	5	0,06	6%	55	275	-9648,90563
2	58 – 64	17	0,21	21%	61	1037	-4561,71413
3	64 – 70	34	0,43	43%	67	2278	-3,09825
4	70 – 76	16	0,20	20%	73	1168	2735,262
5	76 – 82	7	0,09	9%	79	553	10785,5921
6	82 – 88	1	0,01	1%	85	85	5405,44388
		80	**1**	**100%**		**5396**	**4712,6**

$$g_1 = \frac{3223{,}8}{80 \times 6{,}35^3}$$

$$g_1 = 0{,}2303$$

Existe un ligero sesgo hacia la derecha.

3.6. La curtosis para datos agrupados en intervalos de clase

$$g_2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{ns^4} - 3$$

g2 = – 0,0428

En razón de que el valor de la curtosis en bajo y negativo, se concluye que la distribución de frecuencias es ligeramente platicúrtica.

3.7. La moda para datos agrupados con intervalos

En sentido estricto, la moda es el valor del dato que tiene mayor frecuencia absoluta, o en términos sencillos; es el valor que más se repite, pero al tratarse de tablas con intervalos, no se puede determinar un solo valor, puesto que cada uno de ellos comprende más de un valor distinto; entonces, se procede a elegir el intervalo que tiene la mayor frecuencia absoluta y a calcular la moda con la siguiente fórmula:

$$Mo = Li + \frac{f - f_{(-1)}}{\left(f - f_{(-1)}\right) + \left(f - f_{(+1)}\right)} \cdot i,$$ donde:

Li = límite inferior de la clase modal; en otras palabras, el límite inferior del intervalo que tenga mayor frecuencia absoluta.

f = frecuencia absoluta del intervalo seleccionado

$f(-1)$ = frecuencia absoluta anterior al de la clase modal

$f(+1)$ = frecuencia absoluta posterior al de la clase modal

i = amplitud del intervalo

Siguiendo con el ejemplo anterior, tenemos.

N°	Masas	*f*	*fr*	%	*Xm*	*fXm*	*fa*
1	52 – 58	5	0,06	6%	55	275	5
2	58 – 64	17	0,21	21%	61	1037	22
3	64 – 70	34	0,43	43%	67	2278	56
4	70 – 76	16	0,20	20%	73	1168	72
5	76 – 82	7	0,09	9%	79	553	79
6	82 – 88	1	0,01	1%	85	85	80
		80	**1**	**100%**		**5396**	

$$Mo = 64 + \frac{34-17}{(34-17)+(34-16)} \cdot 6$$

Mo = 66,91

3.8. Mediana para datos agrupados en intervalos de clase.

Se selecciona el intervalo en donde se ubica el dato intermedio; para ello se busca en la columna de las frecuencias acumuladas y se aplica la fórmula:

$Me = Li + \dfrac{\frac{n}{2} - f_{am}}{f} \cdot i$ donde:

Li = Límite inferior da la clase en donde se ubica el dato n/2

n = número de datos

fam = frecuencia acumulada menor, es decir, en la columna de las frecuencias acumuladas, se selecciona la anterior a la del intervalo que se seleccionó anteriormente.

f = Frecuencia absoluta del intervalo seleccionado

i = amplitud de cada intervalo

En el ejemplo anterior, tenemos:

N°	Masas	*f*	*fr*	%	*Xm*	*fXm*	*fa*
1	52 – 58	5	0,06	6%	55	275	5
2	58 – 64	17	0,21	21%	61	1037	22
3	64 – 70	34	0,43	43%	67	2278	56
4	70 – 76	16	0,20	20%	73	1168	72
5	76 – 82	7	0,09	9%	79	553	79
6	82 – 88	1	0,01	1%	85	85	80
		80	**1**	**100%**		**5396**	

$$Me = 64 + \frac{\frac{80}{2} - 22}{34} \cdot 6$$

$$Me = 67{,}18$$

3.9. Los cuantiles para datos agrupados en intervalos de clase.

El procedimiento para el cálculo de todos los cuantiles es similar al que se realizó con la mediana, utilizando las fórmulas, según el cuantil que se necesite. Así:

Cuartiles: $$Q_k = Li + \frac{\frac{kn}{4} - f_{am}}{f} \cdot i$$

Deciles: $$D_k = Li + \frac{\frac{kn}{10} - f_{am}}{f} \cdot i$$

Percentiles: $$P_k = Li + \frac{\frac{kn}{100} - f_{am}}{f} \cdot i$$

Con los datos anteriores, calcular el percentil 93

N°	Masas	f	fr	%	Xm	fXm	fa
1	52 – 58	5	0,06	6%	55	275	5
2	58 – 64	17	0,21	21%	61	1037	22
3	64 – 70	34	0,43	43%	67	2278	56
4	70 – 76	16	0,20	20%	73	1168	72
5	76 – 82	7	0,09	9%	79	553	79
6	82 – 88	1	0,01	1%	85	85	80
		80	**1**	**100%**		**5396**	

Primeramente calcular $\frac{kn}{100}$ para determinar el intervalo con el que se va a trabajar.

$$\frac{(93)(80)}{100} = 74,4$$

Ubicamos el intervalo en donde se ubica el dato n° 74

$$P_{93} = 76 + \frac{74,4 - 72}{7} \times 6$$

$$P_{93} = 78,06$$

3.10. Puntuaciones tipificadas.

La puntuación típica "z", determina el número de desviaciones estándar que algún punto ***Xi*** se desvía de la media.

Su fórmula de cálculo es $z = \frac{Xi - \overline{X}}{\vartheta}$

3.11. La distribución normal

Es una función que permite calcular la probabilidad de que un valor de la variable X, se encuentre dentro de un intervalo (a, b), tiene forma acampanada y se la conoce también como campana de Gauss. Se la usa para realizar inferencias estadísticas, una vez conocidas la media aritmética y la desviación estándar.

Las inferencias estadísticas realizadas con este medio, ayudan en la toma de decisiones bajo condiciones de incertidumbre y tiene aplicación en muchos escenarios y campos de la vida real.

3.11.1. Propiedades de la distribución normal

- La distribución normal tiene forma de campana.
- El área bajo la curva o la probabilidad desde menos infinito a más infinito vale 1.
- La distribución normal es simétrica, es decir cada mitad de curva tiene un área de 0,5.
- La escala horizontal de la curva se mide en desviaciones estándar.

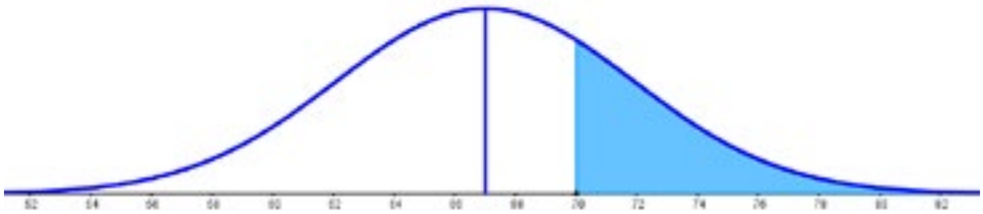
Fórmulas para la distribución normal:

$$z = \frac{Xi - \overline{X}}{\vartheta} \qquad \vartheta = \frac{X - \mu}{Z} \qquad X = \vartheta Z + \mu \qquad \mu = X - \vartheta Z$$

$$\mu = Np \qquad \vartheta = \sqrt{Npq}$$

Ejemplos de distribución normal resueltos

1. Los resultados de un examen de ingreso a la universidad, siguen una distribución normal con media 67 y varianza 24 ¿Cuál es la probabilidad de que una persona que se presenta el examen obtenga una calificación superior a 70?



$$z = \frac{X-\mu}{9} \qquad z = \frac{70-67}{\sqrt{24}} \qquad z = 0{,}6124 \qquad P = 0{,}2701$$

2. Un estudio ha mostrado que, en una institución de educación superior, el 40% de los estudiantes tienen automóvil. Se elige al azar una muestra de 60 estudiantes en la institución. ¿Cuál es la probabilidad de que entre 25 y 30 estudiantes tengan vehículo propio?

$$\mu = (60)(0{,}4) = 24$$

$$z = \frac{25-24}{\sqrt{60(0{,}4)(0{,}6)}} = 0{,}26 \qquad P = 0{,}9431$$

$$z = \frac{25-24}{\sqrt{60(0{,}4)(0{,}6)}} = 0{,}26 \qquad P = 0{,}6039$$

$$P = (0{,}9431 - 0{,}6039) = 0{,}3392$$

Problemas propuestos de distribución normal

1. El departamento de talento humano de una empresa requiere que los aspirantes a un puesto, en una prueba de aptitud alcancen una califi-

cación mínima de 70. Si las calificaciones de la prueba se distribuyen normalmente con media $\mu = 49$ y desviación estándar $\sigma = 3$ ¿Qué porcentaje de los solicitantes pasará la prueba?

2. Una fábrica de cerveza artesanal envasa botellas con su producto, utilizando instrumento de envasado manual. Si la cantidad de cerveza en cada botella sigue una distribución normal con media de 600 cm3 y una varianza de 25. ¿Qué porcentaje de las botellas se llenan con agua entre 592 y 608 cm3?

3. Los ingresos mensuales de los profesionales jóvenes se distribuyen de forma normal con un promedio de $1100 y una desviación estándar de $250. Calcular el porcentaje de profesionales que ganan:

a) menos de $500 al mes

b) entre $900 y $1400 al mes

c) más de $2000 al mes

4. Un examen tiene 200 ítems de verdadero o falso y se necesitan contestar correctamente 140 o más para aprobar. Si un estudiante contesta al azar, calcule la probabilidad de que apruebe el examen.

5. El 40 % de la población de una ciudad vive en conjuntos habitacionales privados. Si se pregunta a 1500 de sus habitantes, su lugar de vivienda, calcule la probabilidad de que menos de 550 vivan en conjuntos privados.

3.11.2. Áreas bajo la distribución normal en hoja de cálculo.

Existen tablas que nos dan los valores aproximados de una probabilidad, una vez que se conozca el valor de la puntuación “z”. A continuación, se muestra un ejemplo y su modo de uso.

Supongamos que deseamos saber la probabilidad correspondiente a una puntuación z = 0.53, se procede a buscar la décima 0,5 en la columna de la izquierda y la centésima 0,03 en la fila superior, lo que sumado será 0,53. La probabilidad correspondiente se ubica en la celda donde se cruzan los dos valores; así, la probabilidad correspondiente es 0,7019.

z	0	*0,01*	*0,02*	*0,03*	*0,04*	*0,05*
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596
0,2	0,5793	0,5832	0,5871	0,5910	67	2278
0,5948	0,5987	16	0,20	20%	73	1168
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088

Sin embargo, de que estas tablas son de fácil acceso, puesto que actualmente se encuentran en la web, en la actualidad, tenemos también a la mano las hojas de cálculo, que permiten obtener tanto los valores de las puntuaciones tipificadas "z", como las probabilidades de una manera más rápida y precisa. Se utilizará para ello la función:

=DISTR.NORM.ESTAND (0,53;VERDADERO)

De la misma manera, si se tiene el valor de la probabilidad y se desea obtener la puntuación tipificada "Z", se procede con el inverso de esta función.

=INV.NORM.ESTAND( )

El en paréntesis se ubica la probabilidad

Cuestionario de repaso

Las utilidades diarias en dólares de un establecimiento comercial en Quito, durante sesenta días consecutivos fueron las siguientes: 73, 82, 95, 92, 91, 64, 86, 63, 85, 96, 98, 87, 68, 99, 108, 97, 96, 71, 84, 93, 76, 62, 97, 76, 99, 91, 99, 97, 110, 101, 68, 95, 105, 99, 64, 97, 88, 106, 83, 66, 105, 102, 81, 79, 98, 108, 67, 103, 106, 95, 78, 85, 94, 85, 70, 87, 110, 104, 86, 79. Si se elaboró una tabla de frecuencias con 7 intervalos, encuentre:

1. La media aritmética para datos agrupados en intervalos

   a) 90,22

   b) 89,53

   c) 88,79

   d) 91,14

2. La desviación estándar

   a) 14,71

   b) 12,11

   c) 13,49

   d) 12,83

3. La mediana

   a) 92,69

   b) 90,31

   c) 89,75

   d) 91,24

4. La moda

    a) 89,92

    b) 91,45

    c) 92,00

    d) 90,00

5. El coeficiente de variación

    a) 12%

    b) 15%

    c) 14%

    d) 17%

6. El percentil 35

    a) 79,15

    b) 86,23

    c) 85,10

    d) 83,78

Los directivos de una empresa pública que tiene 1856 empleados, están preocupados por la salud de los mismos y han dispuesto que se mida la masa corporal de cada uno de ellos. Si estas medidas se pueden aproximar a una distribución normal, con una media de 73 kg y desviación estándar de 5,4 kg, determine:

7. El porcentaje de trabajadores cuya masa corporal está comprendida entre los 68 y 80 kg.

    a) 73,14%

    b) 72,53%

c) 69,80%

d) 71,63%

8. La probabilidad de que un empleado tenga una masa corporal mayor a 80 Kg.

    a) 9,74%

    b) 10,17%

    c) 8,96%

    d) 11,00%

El 40 % de la población de una ciudad vive en conjuntos habitacionales privados. Si se pregunta a 1500 de sus habitantes, su lugar de vivienda, calcule:

9. La probabilidad de que menos de 870 personas vivan en conjuntos privados.

    a) 4,95%

    b) 6,19%

    c) 3,49%

    d) 5,69%

10. La probabilidad de que vivan en conjuntos privados entre 885 y 925 familias

    a) 69,16%

    b) 48,71%

    c) 72,41%

    d) 57,21%

UNIDAD 4. MUESTREO

Resultado de aprendizaje

Determina modelos de regresión, relacionando variables y obteniendo de conclusiones

Contextualización

El estudiante habrá de conocer los principales métodos de muestreo y desarrollar modelos de regresión y correlación, a partir de las cuales podrá realizar inferencias estadísticas con mucha precisión.

4. Muestreo

Muchas veces se imposibilita el estudio individual de todos los elementos de una población, ya sea por la extrema dificultad de llegar a todos cuando esta es infinita o por sus elevados costos. En estos casos, se aplica la investigación a un subconjunto llamado muestra, el mismo que debe ser representativo; esto significa, que las características de la muestra son las mismas que las de la población con la finalidad de que las conclusiones puedan ser inferidas para todos.

4.1. Tipos de muestreo

Muestreo aleatorio simple

Este tipo de muestreo consiste en seleccionar elementos de la población al azar, de manera que todos tengan la misma oportunidad de ser parte de la muestra. El muy importante que los individuaos que forman parte de la muestra, tengan características similares a las de toda la población para que sea representativa de la misma, de manera que, las conclusiones obtenidas a través del estudio de la muestra, sean aplicables a toda la población.

Muestreo aleatorio estratificado

Un estrato es un subconjunto de la población con características específicas y exclusivas. Cuando se identifican estos grupos llamados estratos, deben considerarse para el estudio a un número de individuos proporcional a del tamaño de cada estrato, para garantizar la participación de todos los sectores de la población.

Un ejemplo de ello podría ser, determinar el número de hombres y mujeres que participarán en una encuesta, según la cantidad que se estime que existe en una población completa. Así, si la población está compuesta por 65% de mujeres y 35% de hombres, la muestra debe tener una proporción igual de cada sexo.

Muestreo aleatorio sistemático

Se ordenan los datos obtenidos con algún criterio (por ejemplo, antigüedad en una empresa), se divide el tamaño de la población para el tamaño de la muestra, se elige uno entre los primeros n elementos según resulta en la división, así, si la población es de 2000 y la muestra es de 50; 2000/50 = 40, entonces, de los primeros 40 se selecciona uno, por ejemplo, el que está ubicado en el puesto 15 y se va sumando 40 a partir de ahí, los siguientes seleccionados son los puestos 55, 95, 135, 175, ...

Muestreo por conglomerados

Cuando la población se encuentra dividida en sectores con todas sus características, pero de manera natural, es decir, son representativos de la población, pueden considerarse solamente algunos de estos grupos o conglomerados para la ejecución del estudio.

Un ejemplo de este tipo de muestreo sería si se quiere realizar una investigación sobre los hábitos alimenticios a nivel nacional, bastaría con tomar muestras en unas pocas provincias representativas y realizar el estudio en ellas.

Es importante no confundir estratos con conglomerados, pues en el primero cada subgrupo es homogéneo y en el segundo, los grupos tienen todas las características de la población entera.

4.2. Errores de muestreo

Una encuesta, realizada en distintos momentos y a la misma población, puede arrojar resultados distintos. Esto se debe a los errores que se generan en el proceso de la recopilación de los datos.

Se conoce como error muestral a la diferencia entre los promedios calculados sobre la muestra y sobre la población total, por ejemplo, si se requiere calcular la estatura media de las mujeres entre 20 y 30 años de una población en la que habitan trecientas mil mujeres aproximadamente, se toma una muestra al azar de 300 mujeres, la estatura promedio obtenida, no necesariamente en la misma de toda la población, aunque si la información no está sesgada, estos valores serán muy cercanos.

El margen de error se puede determinar con la fórmula $\varepsilon = \frac{Z\sigma}{\sqrt{n}}$

Ejemplo, determinar el error muestral si se tiene una desviación estándar de 15, un nivel de confianza del 95% y una muestra de 250.

$$\varepsilon = \frac{(1{,}96)(15)}{\sqrt{250}} = 1{,}86$$

Esto significa que el promedio se encuentra en el intervalo $\left(\overline{X} - \varepsilon,\ \overline{X} + \varepsilon\right)$

4.3. Cálculo del tamaño de la muestra

Para el cálculo del tamaño de la muestra se emplearán dos fórmulas en caso de la que la población se finita o infinita; así:

$$n = \frac{NpqZ^2}{(N-1)e^2 - pqZ^2}$$ para poblaciones finitas

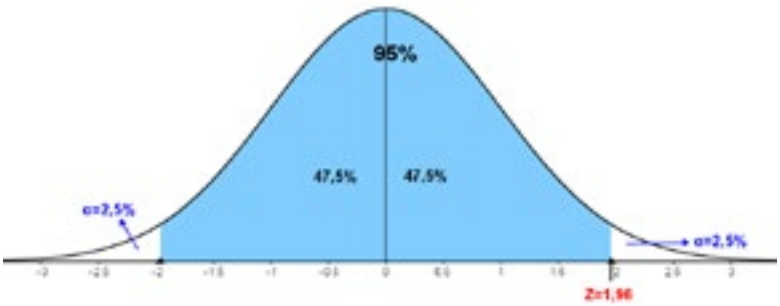
$$n = \frac{pqZ^2}{e^2}$$ para poblaciones infinitas

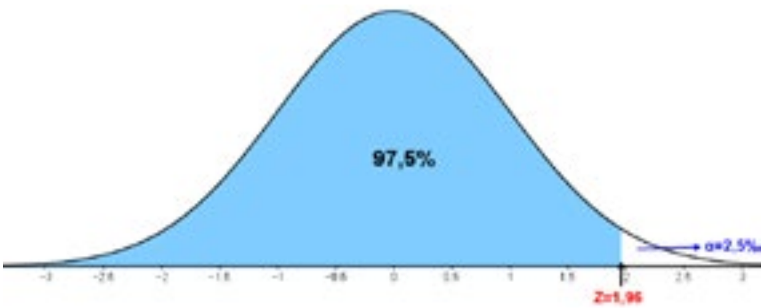
n es el tamaño de la muestra;

Z **nivel de confianza;** Es el porcentaje de seguridad con el que el investigador puede realizar las inferencias de los resultados hacia la población. Un nivel de confianza del 100% significaría realizar el estudio con toda la población. Generalmente el nivel de confianza pasa del 90% y el más comúnmente utilizado es el 95%, sin embargo, en su elección prima el criterio del investigador.

Para obtener el valor de la puntuación Z se debe considerar que el porcentaje de confianza se encuentra en el centro de la curva normal, distribuido simétricamente de la mitad hacia la izquierda y hacia la derecha y puesto que en Excel o en las tablas de área bajo la curva normal, la probabilidad se determina desde un punto hacia la izquierda, se debe tomar como referencia la mitad de porcentaje de confianza más el 50% de la izquierda. Ejemplo.

Para 95% de confianza, se suma la mitad de la derecha del mismo, 47.5%, más el 50% de toda la izquierda y obtenemos una puntuación Z = 1,96.





En Excel se obtiene con la función **=INV.NORM.ESTAND(0,95/2+0,5)**

p **es la variabilidad positiva;** Es la probabilidad de que ocurra el evento en que se está investigando, esto en el caso de que existan estudios anteriores o se realicen pruebas piloto.

q **es la variabilidad negativa;** Es el complemento de la variabilidad positiva, es decir, es la probabilidad de que el evento que se está investigando no ocurra.

N **es el tamaño de la población;** Es el número total de elementos que componen la población.

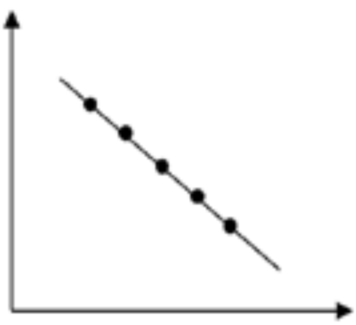
e **error máximo aceptado;** Es la probabilidad de aceptar una afirmación sea falsa como si fuera verdadera, o viceversa. Un error del 0% implica realizar la investigación a la población entera. Este error debe ser muy bajo para garantizar que los resultados se puedan generalizar con confianza.

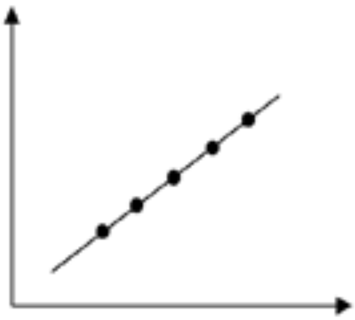
Es importante considerar que el nivel de confianza y el margen de error no son complementarios; es decir, puede trabajarse por ejemplo con un nivel de confianza del 95% y un margen de error del 3%.

5. Modelos de regresión correlación.

La correlación mide el grado de relación que existe entre dos variables cuantitativas, mientras que la regresión lineal determina la ecuación de la recta, entorno a la que se ubican los pares de valores relacionados. Con base en estos cálculos, se puede predecir el valor de una variable a partir de otra.

El coeficiente de correlación se designa con la letra r y puede tomar cualquier valor entre –1.00 y +1.00. Un coeficiente de correlación de – 1.00 o de +1.00 indica correlación perfecta.





Si no existe ninguna relación entre los dos conjuntos de variables, el coeficiente r será cero. Una correlación cercana a cero significa la relación es mínima, mientras que un coeficiente cercano a uno como 0,92 o –0,92 demuestra que hay una relación muy fuerte entre los dos conjuntos de variables.

El coeficiente de correlación se calcula con la siguiente fórmula.

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left(n\left(\Sigma X^2\right) - \left(\Sigma X\right)^2\right)\left(n\left(\Sigma Y^2\right) - \left(\Sigma Y\right)^2\right)}}$$

Donde:

n	es el número de pares de observaciones.
ΣX	es la suma de las variables X.
ΣY	es la suma de las variables Y.
(ΣX^2)	es la suma de los cuadrados de la variable X.
$(\Sigma X)^2$	es la suma de las variables X elevada al cuadrado.
(ΣX^2)	es la suma de los cuadrados de la variable Y.
$(\Sigma X)^2$	es la suma de las variables elevada al cuadrado.
ΣXY	es la suma de los productos de X y Y.

Ejemplo:

Un gran almacén vende televisores de todos tamaños. La gerencia desea saber si las llamadas telefónicas influyen significativamente en la venta de televisores. Se reúne la información sobre la relación entre el número de llamadas de venta y el número de televisores vendidos. Se realizó el estudio con una muestra aleatoria de 10 vendedores, obteniendo información que consta en la tabla sobre el número de llamadas de venta que hicieron el mes pasado y el número de televisores vendidos. Calcular el coeficiente de correlación.

Nº	Representante de ventas	Llamadas de venta (X)	Televisores vendidos (Y)	X2	Y2	XY
1	José Armijos	20	30	400	900	600
2	Marco Álvarez	40	60	1600	3600	2400
3	Silvia Carranza	20	40	400	1600	800
4	Martha Bayas	30	60	900	3600	1800
5	Cindy Cobos	10	30	100	900	300
6	María Jiménez	10	40	100	1600	400
7	Juan Pérez	20	40	400	1600	800
8	Pablo Paucar	20	50	400	2500	1000
9	Melisa Zambrano	20	30	400	900	600
10	Darío Zurita	30	70	900	4900	2100
	TOTAL	220	450	5600	22100	10800

$$r=\frac{10(10800)-(220)(450)}{\sqrt{\left(10(5600)-(220)^2\right)\left(10(22100)-(450)^2\right)}}=0{,}759$$

La correlación de 0.759 es positiva, por lo que se observa que hay una relación directa fuerte entre el número de llamadas de venta y el número de televisores vendidos. Se podría afirmar entonces que un el aumento de un 40 por ciento en las llamadas llevaría, posiblemente, a un aumento de un 40 por ciento en las ventas.

5.1. Regresión lineal simple

Es un modelo matemático para expresar la relación entre dos variables y estimar el valor de la variable dependiente "Y" basándonos en el valor de la variable independiente "X". A la técnica que se usa para desarrollar la ecuación de la línea y hacer estas predicciones se le llama análisis de regresión.

Para el ejemplo del número de llamadas de venta y el número de televisores vendidos con una muestra de 10 vendedores, se desarrollará un modelo matemático que exprese la relación entre el número de llamadas de venta y el número de unidades vendidas. A la ecuación de la línea que se usa para estimar Y basándose en X se le llama ecuación de regresión.

Ecuación de regresión, Es una ecuación que define la relación entre dos variables.

Principio de los mínimos cuadrados

Este método proporciona lo que se conoce comúnmente como recta de “mejor ajuste”. Determina una ecuación de regresión minimizando la suma de los cuadrados de la distancia vertical entre el valor real de Y, y el valor predictorio de Y.

La forma general de la ecuación de regresión es $Y' = mx + b$

Donde:

Y’ se lee Y prima, es el valor predictorio de la variable Y para un valor de X seleccionado.

b es la intersección con el eje Y. Es el valor estimado de Y cuando X = 0. Otra manera de expresar esto es: m es valor estimado de Y en donde la línea de regresión cruza el eje Y cuando X es cero.

m es la pendiente de la línea, o el cambio promedio en Y’ por cada cambio en una unidad (ya sea aumentando o disminuyendo) de la variable independiente X.

X es el valor que se escoge para la variable independiente.

Las fórmulas para m y b son:

Pendiente de la línea de regresión $$m = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$$

Intersección con el eje y $b = \frac{\Sigma Y}{n} - a\frac{\Sigma x}{n}$

Donde:

X es un valor de la variable independiente.

Y es un valor de la variable dependiente.

n es el número de elementos en la muestra.

Retomando el problema de los televisores, se usará el método de mínimos cuadrados para expresar la relación entre las dos variables. ¿Cuál es el número esperado de venta de televisores para un empleado que hace 20 llamadas?

La ecuación de regresión es Y' = 1,1842X+18,9476. Si un vendedor hace 20 llamadas, puede esperar vender 42,6326 televisores, lo cual se obtiene de Y = 1,1842 (20)+18,9476.

5.2. Regresión potencial

La relación entre las variables se realiza mediante la función y = ax^b, la misma que se procede a linealizar, aplicando las propiedades de los logaritmos de la siguiente manera.

$$\log y = \log a + b \log x$$

Si se reemplazan los logaritmos por nuevas variables, tendremos, $\log y = Y'$, $\log a = A'$,y $\log x = X'$, , con los que tendríamos la ecuación lineal $Y' = BX' + A'$

$$B = \frac{n(\Sigma X'Y') - (\Sigma X')(\Sigma Y')}{n(\Sigma X'^2) - (\Sigma X')^2}$$

5.3. Regresión exponencial

La función de ajuste es $y = ae^{bx}$; $a \neq 0$. Para linealizar esta función nuevamente aplicamos logaritmos de la siguiente manera:

$$\ln y = \ln a + bx$$

Y luego, reemplazando variables será $\ln y = Y'$, $\ln a = A$, y la ecuación linealizada quedará

$$Y' = A + bx$$

5.4. Regresión múltiple

Se presenta cuando más de una variable independiente influyen sobre la variable dependiente y entonces su función de correspondencia es **y = f(x, w, z, …)**

Ejemplo.

La gerencia de una importadora de electrodomésticos quiere determinar la relación entre los ingresos en cierto periodo y el número de artículos vendidos de cada línea. Línea blanca, línea marrón, motocicletas y equipos informáticos.

A continuación, se presentan los datos de las ventas del último periodo, siendo, Y = ingresos en miles de dólares, X = artículos vendidos en la línea blanca, W = artículos de la línea marrón y Z = equipos tecnológicos.

Y	X	W	Z
450	50	105	75
465	40	140	68
470	35	110	70
515	45	130	64
508	51	125	67
475	55	115	72
450	53	100	70
510	48	103	73
480	38	118	69
460	44	98	74

$$y = m_1x_1 + m_2x_2 + m_3x_3 + \ldots\ldots + m_nx_n + b$$

$$\begin{cases} bn + m_1\sum X_1 + m_2\sum X_2 + m_3\sum X_3 = \sum Y \\ b\sum X_1 + m_1\sum X_1^2 + m_2\sum X_1X_2 + m_3\sum X_1X_3 = \sum X_1Y \\ b\sum X_2 + m_1\sum X_1X_2 + m_2\sum X_2^2 + m_3\sum X_2X_3 = \sum X_2Y \end{cases}$$

Problemas propuestos:

1. Se cree que los gastos en alimentación de una familia son dependientes de los ingresos mensuales y del número de integrantes familiares. Se realiza la investigación a una muestra de 15 familias. Determine la recta de regresión con los datos de la tabla.

Gastos	Ingresos	Integrantes
430	2100	3
310	1100	4
320	900	5
460	1600	4
1250	6200	4
440	2300	3
520	1800	6
290	1000	5
1290	8900	3
350	2400	2
350	1200	4
780	4700	3
430	3500	2
470	2900	3
380	1400	4

2. Una desea estimar los gastos en alimentación de una familia Y en base a la información que proporcionan las variables regresora **X1= "ingresos mensuales" y X2= "número de miembros de la familia"**.Para ello se recoge una muestra aleatoria simple de 15 familias cuyos resultados son los de la tabla adjunta (El gasto e ingreso está dado en miles de dólares)

Cuestionario de repaso

1. El gerente de una importadora recibe un embarque de 4200 piezas, necesarias para la fabricación de un artículo y se quiere evaluar la calidad de las mismas. Según informes previos, 180 piezas suelen ser defectuosas ¿Cuántas piezas se deben examinar si se considera un nivel de confianza del 96% y un error permitido del 3%?

    a) 215

    b) 184

    c) 178

    d) 190

La tabla siguiente muestra las estaturas en centímetros de una muestra de 20 padres y sus hijos mayores.

Estatura padre	178	162	182	172	177	174	179	185	173	165	167	173	171	170	168	169
Estatura hijo	180	172	177	169	185	174	177	182	172	171	175	178	167	167	173	176

Determine:

2. La estatura que tendrá un hijo mayor si su padre mide 172 cm

    a) 180 cm

    b) 172 cm

    c) 174 cm

    d) 176 cm

3. El coeficiente de correlación r

    a) 0,56

    b) 0,67

c) 0,59

d) 0,62

4. La recta de regresión

a) y = 0,508x + 86,90

b) y = 0,492x + 23,62

c) y = 0,539x + 91,27

d) y = 0,515x + 73,43

5. En una fábrica de insumos médicos generalmente el 10% de los artículos presenta algún defecto de fabricación. Se requiere realizar un control de calidad a los artículos que están listos para la venta y para ello se necesita una muestra significativa de los mismos. ¿Cuántos artículos deben ser elegidos si se desea una confianza de 97% y un error de 4%?

a) 168

b) 265

c) 324

d) 159

Examen de Fin de módulo

Referencias Bibliográficas

Libro Base:

- Triola, Mario F (2013). Estadística, México DF, Pearson Educación.

Complementaria:

- Díaz, A (2013). Estadística aplicada a la administración y la economía, México DF.
- Lind, D & Wathen, S & Marchal, W (2012). Estadística aplicada a los negocios y la economía, México DF , McGraw-Hill
- Santamaría, María Benigna (2012). La enseñanza de la Estadística y el profesor de matemática, Saarbrücken, Académica Española

FORMATO DE REVISIÓN DE GUÍAS GENERAL DE ESTUDIOS POR PARES ACADÉMICOS (MODALIDAD A DISTANCIA)

IDENTIFICACIÓN DE LA GUÍA GENERAL DE ESTUDIOS		
TÍTULO DE LA GUÍA GENERAL DE ESTUDIOS DE LA ASIGNATURA: ESTADÍSTICA		
FECHA DE ENTREGA DE LA GUÍA GENERAL DE ESTUDIOS DE LA ASIGNATURA: 31/8/2023	FECHA DE ENTREGA DE LA REVISIÓN REALIZADA: 17/10/2023	
2. DATOS DEL PAR ACADÉMICO (Los siguientes datos deben ser suministrados por el para académico y son de carácter obligatorio)		
NOMBRE Y APELLIDOS: Diego Enrique Polanco Calvachi	DIRECCIÓN: Av. Buenos Aires OE1-16 y Av. 10 de agosto	TELÉFONOS; 0999216079
CORREO ELETRÓNICO: dpolanco@tecnologicopichincha.edu.ec	CIUDAD: Quito	PAÍS: Ecuador
CARGO: Docente	INSTITUCIÓN: Instituto Universitario Pichincha	ÁREAS DE INTERÉS: Tecnologías Innovación Fuente De Energías Renovables
ÚLTIMO TÍTULO ACADÉMICO OBTENIDO: Cuarto Nivel: Magister en Pedagogía y Docencia en Innovación Educativa	N°. DE IDENTIFICACIÓN/ PASAPORTE: 1720749892	

I. INSTRUCCIONES

1. Por favor responda **todas** las preguntas de este formulario.
2. Diligencie el formulario en computador.
3. **No modifique o altere las preguntas u opciones de este formulario**. La estructura de esta evaluación está planificada y responde a las políticas de publicación de las Guías General de Estudios de la MED.
4. Una vez finalice su diligenciamiento, debe devolverlo firmado vía e-mail a la persona que lo contactó.
5. Sea claro y preciso en sus respuestas.

6. Las respuestas del aparte de la fundamentación científica deben ser detalladas.
7. **En caso de no poder cumplir con el plazo** establecido, por favor **informar oportunamente al equipo editorial de la MED.**
8. En caso de detectar plagio, citación indebida o cualquier mala práctica, por favor comuníquelo al equipo editorial.

II. La guía de aprendizaje contiene:

ASPECTOS DE ESTILO A REVISAR	SI CUMPLE	NO CUMPLE
Márgenes	OK	
Numeración de páginas	OK	
Jerarquización de títulos	OK	
Tipo de letra	OK	
No existencia de encabezados o pies de páginas	OK	
Viñetas estandarizadas	OK	
Referencias de cuadros / Gráficos	OK	
Portada en acuerdo a Manual de estilo	OK	
Índice	OK	
Estructura de la guía		
4 unidades	OK	
Resultados de aprendizaje	OK	
Autoevaluación por cada unidad	OK	
Recursos de la guía	OK	
Redacción	OK	
Ortografía	OK	
Referencia Bibliográfica Norma APA séptima edición	OK	
Informe anti-plagio	OK	

III. Fundamentación científica

ASPECTOS DE ESTILO A REVISAR	SI CUMPLE	NO CUMPLE
¿Los objetivos del texto están claramente enunciados y sustentados?	OK	
¿Utiliza una metodología adecuada para el desarrollo de los objetivos?	OK	
¿La presentación y argumentación de las ideas es coherente?	OK	
¿El manejo de conceptos, teorías y datos es preciso?	OK	
¿Existe relación entre el título, el problema, los objetivos, el marco teórico o metodológico y las conclusiones?	OK	
¿El tema es pertinente y brinda aportes a su área de conocimiento?	OK	

IV. Presentación de la información

ASPECTOS DE ESTILO A REVISAR	SI CUMPLE	NO CUMPLE
¿El autor utiliza un lenguaje claro y conciso?	OK	
¿Hay coherencia en la presentación y desarrollo de las ideas?	OK	
¿Las partes del trabajo se articulan entre sí y responden a los objetivos planteados?	OK	
¿Utiliza fuentes bibliográficas actualizadas (últimos tres años)?	OK	

¿Es adecuado el manejo del idioma por parte el autor (ortografía, redacción, sintaxis, puntuación)?	OK	
¿El texto se puede considerar original?	OK	

V. Recomendaciones

- Publicar sin modificaciones: [X]
- Publicar con modificaciones: [ ]
- No publicar: [ ]

V. Comentarios adicionales

El trabajo es coherente y reúne los requisitos para su publicación: [X]

FIRMA DEL EVALUADOR
Nombre: Msc. Diego Enrique Polanco Calvachi
ID: 1720749892

CERTIFICADO DE ANÁLISIS
magister

Guía Estadística GTH

4% Textos sospechosos

4% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas
0% Idiomas no reconocidos

Nombre del documento: Guía Estadística GTH.docx
ID del documento: 7cb649aa2efabcba7d859ca630950e5fe39651e9
Tamaño del documento original: 653,88 kB

Depositante: PABLO FABIAN CARRERA TOAPANTA
Fecha de depósito: 8/3/2024
Tipo de carga: interface
fecha de fin de análisis: 8/3/2024

Número de palabras: 12.839
Número de caracteres: 74.050

Ubicación de las similitudes en el documento:

Fuentes principales detectadas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	**estadisticasoctys.freetzi.com** http://estadisticasoctys.freetzi.com/Estadistica1/Estadisticadescriptiva/unidadiii.pdf **1 fuente similar**	3%		Palabras idénticas: **3%** (400 palabras)
2	**repositorio.utn.edu.ec** http://repositorio.utn.edu.ec/bitstream/123456789/2341/1/1_Interaprendizaje de Estadística Básica.... **2 fuentes similares**	< 1%		Palabras idénticas: **< 1%** (37 palabras)
3	**Documento de otro usuario** #b64928 El documento proviene de otro grupo	< 1%		Palabras idénticas: **< 1%** (22 palabras)

Fuentes con similitudes fortuitas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	**es.plusmaths.com** | Ejercicios de media aritmética | Problemas de media aritmética https://es.plusmaths.com/ejercicios/media-aritmetica	< 1%		Palabras idénticas: **< 1%** (19 palabras)
2	**es.slideshare.net** | Regresión lineal múltiple | PPT https://es.slideshare.net/alksa/regresin-39901424	< 1%		Palabras idénticas: **< 1%** (12 palabras)
3	**preguntame.es** | ¿Cómo interpretar el coeficiente de variación? - Preguntame.es https://preguntame.es/como-interpretar-el-coeficiente-de-variacion/	< 1%		Palabras idénticas: **< 1%** (10 palabras)
4	**www.asf.gob.mx** https://www.asf.gob.mx/Trans/Informes/IR2018c/Documentos/Auditorias/2018_1590_a.pdf	< 1%		Palabras idénticas: **< 1%** (13 palabras)
5	**repositorio.unjfsc.edu.pe** https://repositorio.unjfsc.edu.pe/bitstream/handle/20.500.14067/2040/TEXTO REGRESION Y CORREL...	< 1%		Palabras idénticas: **< 1%** (10 palabras)

TECNOLÓGICO UNIVERSITARIO PICHINCHA

Buenos Aires OE1-16 y Av. 10 de Agosto
099 516 2499
(02) 2 238 291
www.tecnologicopichincha.edu.ec

Modalidad Distancia

ISBN: 978-9942-672-22-3
9789942672223