**Datos de identificación del programa de formación**

|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Gestión analítica de datos |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 210601026 | RESULTADOS DE APRENDIZAJE | 210601026-01. Recopilar datos de acuerdo con metodología estadística. |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | CF3 |
| NOMBRE DEL COMPONENTE FORMATIVO | Probabilidad y prueba de hipótesis |
| BREVE DESCRIPCIÓN | Este capítulo maneja bases de probabilidad. Se fundamenta en conceptos como el promedio o media de un conjunto de datos para comprender la distribución normal muy utilizada en análisis de datos; también plantea la hipótesis de medias, donde se efectúa una comparación de variables que permiten entender términos como: hipótesis nula, hipótesis alternativa, error tipo 1 y de intérvalo de confianza. |
| PALABRAS CLAVE | Estadística inferencial, probabilidades, prueba de hipótesis, estadística aplicada, distribución normal. |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | 4 – Ciencias sociales, educación, servicios gubernamentales y religión |
| IDIOMA | Español |

# **Tabla de contenidos**

**Introducción**

**1. Introducción a la probabilidad.**

**2. Introducción a la distribución normal.**

**3. Introducción a la prueba de hipótesis para la media.**

**Introducción**

|  |
| --- |
| Cuadro de texto |
| Estimado aprendiz, sea bienvenido a esta experiencia de aprendizaje orientado a la probabilidad y pruebas de hipótesis. Se desarrollarán los contenidos de forma gradual iniciando por las bases estadísticas; aprenderá a distinguir la probabilidad de ocurrencia de un evento, el manejo la distribución normal, entre otros. Este contenido le ayudará a distinguir los conceptos y estructuras más relevantes en la estadística inferencial para la analítica de datos como herramienta fundamental para su aplicación de manera efectiva en la solución de problemas. Antes de empezar, se sugiere revisar el siguiente video para identificar el contexto de aprendizaje:  ¡Le deseamos una experiencia de aprendizaje significativa y memorable! |

**Guion de video introductorio**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipo de recurso** | Video 2D | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración (voz en off)** | **Texto** |
| **1** | Imagen: 228131\_i300  <https://www.freepik.es/fotos-premium/tecnologia-omni-channel-negocios-minoristas-linea_7005465.htm#query=persona%20datos&position=20&from_view=search> | NA | En la actualidad las personas generan datos diaria y continuamente, | Introducción a la probabilidad |
| **2** | Imagen: 228131\_i301  <https://www.freepik.es/foto-gratis/mano-tarjeta-credito-portatil_978447.htm#query=transacciones%20bancarias%20y%20compras%20en%20l%C3%ADnea&position=0&from_view=search> | NA | en transacciones bancarias, compras en línea | Datos Web |
| **3** | Imagen: 228131\_i302  <https://www.freepik.es/fotos-premium/mujer-sosteniendo-telefono-inteligente-iconos-redes-sociales-pantalla-cafeteria_7704839.htm#query=redes%20sociales&position=21&from_view=search> | NA | interacciones en las redes sociales | Salen los letreros: Facebook, Twitter, Instagram… (*en general nombres de redes*) |
| **4** | Imagen: 228131\_i303  <https://www.freepik.es/fotos-premium/dispositivos-e-iconos-tecnologicos-conectados-al-planeta-tierra-digital_5074433.htm#query=informaci%C3%B3n%20web&position=38&from_view=search> | NA | al consultar información sobre diferentes temas: clima, videos, audio, imágenes, páginas web, bolsas de valores, salud, entre otros. | Salen los letreros respectivos: clima, videos, audio, imágenes, páginas web, bolsas de valores, salud, transporte, servicios… |
| **5** | Imagen: 228131\_i300  <https://www.freepik.es/vector-gratis/diseno-marco-rayos-brillantes-tecnologia-digital_9508843.htm#page=2&query=inteligencia%20artificial%20datos&position=17&from_view=search> | NA | Todo ese manejo de información es la fuente primaria para la inteligencia artificial. | AI (Inteligencia Artificial) |
| **6** | Imagen: 228131\_i305  <https://www.freepik.es/fotos-premium/manejar-archivos-exito-personal-ti-utiliza-sistema-gestion-documentos-dms-base-datos-documentacion-linea-automatizacion-procesos-asi-como-otras-tecnologias-comerciales-corporativas_22393425.htm#query=empresa%20manejo%20de%20datos&position=25&from_view=search> | NA | Las empresas en este mundo globalizado y competitivo giran alrededor de los datos y en lo que puede diferenciarlos de la competencia, | Datos en el mundo empresarial |
| **7** | Imagen: 228131\_i306  <https://www.freepik.es/foto-gratis/estadisticas-linea_5402510.htm#query=expertos%20en%20datos&position=0&from_view=search> | NA | por lo tanto, buscan personas que puedan comprender esos datos, analizarlos, visualizarlos en forma dinámica, | Expertos en el manejo de datos |
| **8** | Imagen: 228131\_i307  <https://www.freepik.es/foto-gratis/empresarios-que-trabajan-finanzas-contabilidad-analizan-financi_16068554.htm#query=estad%C3%ADsticas&position=3&from_view=search> | NA | utilizando las diferentes herramientas para la futura toma de decisiones, como el análisis predictivo, tratando de manejar las diferentes técnicas estadísticas existentes para a partir del historial de los datos o relaciones entre ellos, hacer una predicción lo más acertada posible con un menor margen de error o incertidumbre. | Herramientas estadísticas |
| **9** | Imagen: 228131\_i308  <https://www.freepik.es/vector-gratis/plantilla-presentacion-estrategia-empresarial-personas-reunion_16396383.htm#query=consejo%20empresarial&position=29&from_view=search> | NA |  | Toma de decisiones |
| **10** | magen: 228131\_i309  <https://www.freepik.es/foto-gratis/hermosa-chica-hacker-que-trabaja-otros-ciberdelincuentes-peligrosos-centro-hackers_19463732.htm#query=gobierno%20de%20los%20datos&position=5&from_view=search> | NA | Es tal la cantidad de datos que actualmente hay políticas gubernamentales en muchos países del mundo, incluido Colombia, para poner a disposición del público esa información, es el llamado “gobierno de datos”, el cual se enfoca en tener datos limpios, no repetidos, estandarizados, con formatos, reproducibles... | Gobernanza de datos |
| **11** | magen: 228131\_i310  <https://www.freepik.es/fotos-premium/politica-gobernanza-social-medioambiental-esg-empresas-moda_30541649.htm#query=gobernanza%20de%20datos&position=12&from_view=search> | NA | Con la meta de resolver problemas del sector público a fin de mejorar la vida del ciudadano, pensando en el futuro de las ciudades inteligentes. |
| **12** | magen: 228131\_i311  <https://www.freepik.es/foto-gratis/fondo-almacenamiento-nube-remezclado-dominio-publico-nasa_15606925.htm#query=data&from_query=data%20lake&position=1&from_view=search> | NA | Las empresas deben tomar esta gran cantidad de información y organizarla en bases de datos, mediante procesos de *ETL* (siglas en inglés de extraer, transformar y cargar datos) en una base de datos llamada *data lake*, | Data lake |
| **13** | Imagen: 228131\_i312  <https://www.freepik.es/fotos-premium/reunion-negocios-resolucion-problemas-gestion-problemas-solucion-problemas-comunicacion-intercambio-informacion-ejecutivos-discutiendo-resultados-empresa-leyendo-documentos-datos-graficos_23125543.htm#query=resoluci%C3%B3n%20de%20problemas%20de%20datos&position=43&from_view=search> | NA | para a partir de esta información hacer análisis correspondientes y resolver problemas de las empresas en diferentes áreas como fuga de clientes, competencia, entre otras. |
| **14** | magen: 228131\_i313  <https://www.freepik.es/foto-gratis/negocios-reunion-plan-analisis-grafico-empresa-estrategia-financiera_15873165.htm#query=anal%C3%ADtica%20de%20datos&position=0&from_view=search> | NA | Este componente le ayudará a distinguir los conceptos y estructuras más relevantes en la estadística inferencial para la analítica de datos como herramienta fundamental para su aplicación de manera efectiva en la solución de problemas con datos | En este componente aprenderá |
| **15** | Imagen: 228131\_i314  <https://www.freepik.es/foto-gratis/grafico-barras-creciente-digital-superposicion-manos-empresario_13312397.htm#query=estad%C3%ADstica&position=3&from_view=search> | NA | El componente se desarrollará de forma gradual iniciando por las bases estadísticas; |
| **16** | magen: 228131\_i315  <https://www.freepik.es/fotos-premium/comerciante-estudia-bolsas-valores-comprar-acciones-comprando-acciones-mercados-futuros-condiciones-favorables_21182438.htm#page=2&query=promedio%20de%20datos&position=12&from_view=search> | NA | aprenderá a distinguir la probabilidad de ocurrencia de un evento, el manejo de la distribución más utilizada en estadística, la distribución normal, sus parámetros, la media o promedio y la desviación estándar. |
| **17** | Imagen: 228131\_i316  <https://www.freepik.es/foto-gratis/hombre-tableta-manos_961257.htm#query=datos%20y%20toma%20de%20decisiones&position=2&from_view=search> | NA | Manejara el área bajo la curva centrada en la media y desplazamiento hacia la izquierda y la derecha el valor de la desviación. La utilización de esta información para la introducción a la prueba de hipótesis para la media, de acuerdo con los datos arrojados para la toma de decisiones. |
| **18** | Imagen: 228131\_i317  <https://www.freepik.es/foto-gratis/empresaria-usando-tableta-analizar-concepto-exito-estadisticas-estrategia-financiera-empresa-planificacion-futuro-sala-oficina_16068533.htm#query=informaci%C3%B3n%20estad%C3%ADstica&position=13&from_view=search> | NA | Se utilizarán términos como región de rechazo, región de aceptación, error tipo I, intervalo de confianza entre otros. |
| **19** | Imagen: 228131\_i318  <https://www.freepik.es/foto-gratis/concepto-estudio-inteligencia-aprendizaje-ideas-educacion_16437913.htm#query=estudio%20online&position=7&from_view=search> | NA | La metodología incluye diferentes herramientas audiovisuales con sus respectivas explicaciones teóricas, promoviendo el trabajo individual para maximizar la curva de aprendizaje. | Metodología y aprendizaje |
| **20** | Imagen: 228131\_i319  <https://www.freepik.es/fotos-premium/concepto-mercado-divisas-apreton-manos-dos-mujeres-holograma-grafico-financiero-multiexposicion-negocios-internacionales_29440516.htm#query=bienvenida%20online&position=29&from_view=search> | NA | Bienvenido al mundo de la probabilidad y las pruebas de hipótesis | ¡Bienvenido! |
| **Nombre del archivo** | **1:228131\_V1** | | | |

**Desarrollo de contenido**

# **Introducción a la probabilidad**

|  |
| --- |
| Cuadro de texto |
| El objetivo de este módulo es tener buenas bases estadísticas para el análisis de datos.  En este apartado se explicarán conceptos de la probabilidad, iniciando con la teoría, siguiendo con ejercicios prácticos referente a la probabilidad de ocurrencia de un evento. |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider pasos | |
| **Introducción** | La probabilidad se define como número de casos favorables dividido sobre número de casos posibles. | |
| **Slide 1** | Por lo tanto, la probabilidad es una división de dos términos y la fórmula sería: | Dónde: Número de casos posibles se simboliza con la letra **n**.  [se representan las ecuaciones como imágenes llamativas] |
| **Slide 2** | **Ejemplo:** En una bolsa oscura se encuentran 3 bolas negras y 1 blanca.  **¿Cuál es la probabilidad de sacar una bola negra?** |  |
| **Slide 3** | De acuerdo con la pregunta se tiene:   * Número de casos favorables = 3 bolas negras están dentro de la bolsa * Número de casos posibles es el total de valores que toma el experimento en este caso es 4, al sumar las 3 bolas negras y la blanca. | Imagen: 228131\_i320  <https://www.freepik.es/fotos-premium/embalaje-bolsa-cordon-negro-sobre-fondo-blanco_14092047.htm#page=4&query=balotas%20en%20bolsa&position=11&from_view=search&track=ais>  Armar composición de bolsa negra con bolas blancas y negras. |
| **Slide 4** | **¿Cuál es la probabilidad de sacar una bola blanca?** |
| **Slide 5** | De acuerdo con la pregunta se tiene:   * Número de casos favorables = 1 bola blanca está dentro de la bolsa. * Número de casos posibles = 4 bolas en total * La probabilidad siempre es un valor que está entre 0 y 1. * Se habla de un intervalo continuo cualquier valor real en el intervalo 0 y 1. * Se simboliza con la letra P abreviación de probabilidad | [se representan las ecuaciones como imágenes llamativas] |
| **Slide 6** | Se puede utilizar el símbolo de pertenecer a un conjunto o intervalo es este caso al intervalo  P puede tomar cualquier valor real en el intervalo  Es decir, P puede tomar el valor de cero (0), 1, cualquier valor real en el intervalo | [se representan las ecuaciones como imágenes llamativas] |
| **Slide 7** | **Ejemplo**   * P no puede ser negativo * P no puede ser mayor a 1 | P= 0.23; P= 0.45; P=0.69  [se representan como imágenes llamativas] |
| **Slide 8** | La probabilidad es 0 cuando no hay ningún caso posible, por ejemplo, la probabilidad de sacar una bola roja de la bolsa. |  |
| **Slide 9** | La probabilidad es 1 cuando los casos posibles son en número igual que los casos favorables. | Imagen de bolsa con bolas blancas, negras y rojas. |
| **Slide 10** | Ejemplo:  En una bolsa se tienen 4 bolas blancas  **¿Cuál es la probabilidad de sacar una bola blanca?**  Sería 1 debido a que los casos favorables son 4 y los casos posibles son 4 | Imagen de bolsa con bolas blancas |
| [se representa fórmula como imágenes llamativas] |

|  |
| --- |
| Cuadro de texto |
| En el ejemplo o experimento anterior podemos hablar de un término con reemplazo o sin reemplazo.   * Con reemplazo significa que saco la bola de la bolsa y para un siguiente evento la regreso. * Sin reemplazo significa que saco la bola de la bolsa y no la regreso a la bolsa. * A los casos posibles se les llama espacio muestral. |

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| Un espacio muestral sería el conjunto de valores que toma un experimento, en este caso serían el total de bolas negras y blancas , por lo tanto, el espacio muestral tendría un valor de 4 sumando las 3 bolas negras con la bola blanca.  **Imagen:** Bolsa negra con 3 bolas de color negro y una blanca | |
| **Cardinal o longitud**  El cardinal o longitud de ese espacio muestral es 4 que es el tamaño o la suma de todos los posibles valores del experimento.  **Imagen: Las 4 bolas (3 negras - 1 blanca)** | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | A continuación, se trabajará con probabilidad de conjuntos (Martínez, 2019) |
| Diferencia de conjuntos - Wikipedia, la enciclopedia libre  **Imagen que represente conjuntos** | |
| **Probabilidad de conjuntos:**  Probabilidad de la unión de dos conjuntos:    A y B sucesos independientes | |
| **Ejercicio 1**  En una universidad el 60% de los estudiantes aprobó el examen de matemáticas, el 70% el de inglés, el 40% aprobó ambos exámenes.  ¿Cuál es la probabilidad de elegir un estudiante al azar que haya aprobado inglés o matemáticas? | |
| **Análisis 1:**  En conjuntos si un elemento se encuentra en está en A o en B  En conjuntos si un elemento se encuentra en está en ambos A y B  P (aprobó inglés) = P(A) = 0.7  P (aprobó matemáticas) = 0.6  P (aprobó inglés aprobó matemáticas) = 0.4 | |
| **Solución 1:**  = P (inglés o matemáticas) = 0.7+ 0.6 - 0.4 = 0.9   * La probabilidad de que un estudiante elegido al azar haya aprobado inglés o matemáticas, es del 90% | |
| **Ejercicio 2**  En una universidad el 50% de los estudiantes aprobó el examen de programación, el 60% el de redes, el 70% aprobaron ambos.  ¿Cuál es la probabilidad de que un estudiante elegido al azar haya aprobado programación y redes? | |
| **Análisis 2**  P (programación) = 0.5  P (redes) = 0.6  P (programación redes) = 0.7 | |
| **Solución 2**  0.7 = 0.5 + 0.6 -  0.7 = 0.11-  La probabilidad es del 59% de tener aprobadas ambas materias: programación y redes. | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Tarjetas Animadas |
| **Introducción** | Ahora se trabajará con probabilidad condicional |
| Imagen: 228131\_i322  **Imagen:** Imagen de tienda online con camisas y zapatos | **Ejercicio 1**  En una tienda online el 40% de los usuarios compraron una marca de camisa y el 30% compraron una marca de zapato, el 20% ambos.  Si se elige un usuario al azar ¿Cuál es la probabilidad que compre una camisa, dado que compró unos zapatos? |
| Imagen: 228131\_i323  **Imagen:** representación de compra online de camisas y zapatos | **Fórmula de la probabilidad condicional**  La fórmula de la probabilidad condicional es: **P(B/A) =**  Significa que la probabilidad de que ocurra B una vez que ocurrió A, son sucesos dependientes.  P (compró camisa /compró zapatos) =  Significa que, si una persona compró zapatos, la probabilidad de que compre una camisa es del 66%. |
| Imagen: 228131\_i324  **Imagen:** representación de compra online de camisa | **Ejercicio 2**  En la misma tienda online se selecciona un usuario al azar  ¿Cuál es la probabilidad de que compre zapatos, teniendo en cuenta que ya compró una camisa? |
| Imagen: 228131\_i325  **Imagen:** representación de compra online de camisas y zapatos | **Fórmula de la probabilidad condicional**  P (compró zapatos/compró camisa) =  Significa que para una persona que compró una camisa, la probabilidad de que compre zapatos es del 50%  Se observa que en la segunda compra de la camisa, hay un 16% de probabilidades de que esta sea de zapatos. |
| Imagen: 228131\_i326  <https://es.wikipedia.org/wiki/Intersecci%C3%B3n_de_conjuntos>  **Imagen:** representación de intersección de conjuntos, ojalá con ejemplo de camisas y zapatos de diferente tipo. | **Intersección de conjuntos:**  El símbolo en conjuntos es un elemento que está en la intersección de los dos conjuntos, es decir en ambos. |

# **Introducción a la distribución normal**

|  |
| --- |
| Cuadro de texto |
| En esta parte aprenderá sobre la distribución normal, iniciando con los conceptos teóricos-prácticos sobre cálculo del promedio y desviación de un conjunto de datos. |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Pestañas o tabs horizontales | |
| **Introducción** | Explicación de los conceptos de promedio, desviación Lind estándar y estadística diferencial antes de entrar en el tema de la distribución normal o campana de Gauss ( (Lind, Marchal, & Wathen, 2019) ) | |
| **Promedio** | El promedio o media  se calcula sumando todos los datos y dividiendo entre el número total de datos.  El promedio se simboliza con la letra X y una barra lineal encima de esa letra:  = promedio = media | Imagen: 228131\_i327  Imagen que represente un promedio de datos |
| **Ejemplo de p** **romedio** | Calcular el promedio de los siguientes valores: 3, 4, 5, 6. | Representar la solución en imagen |
| **Desviación estándar** | También conocida como **desviación típica o desvío típico,** mide qué tanto se alejan los datos del promedio, es decir, si hay mucha variación entre ellos.  Si los datos se alejan, el valor de la desviación va a crecer, si se acercan, el valor de la desviación va a disminuir. | Imagen: 228131\_i328  **Poner una imagen de ejemplo de desviación estándar, similar a ésta, tomada de Wikipedia:** [**https://es.wikipedia.org/wiki/Desviaci%C3%B3n\_t%C3%ADpica**](https://es.wikipedia.org/wiki/Desviaci%C3%B3n_t%C3%ADpica) |
| **Representación de la desviación estándar** | La desviación se simboliza con la letra griega sigma:  La desviación es la raíz cuadrada de la varianza:  La varianza se simboliza con  c | Representar como imagen |
| **Estadística inferencial** | Consiste en estimar ciertas características de la población en estudio a partir de una muestra y tomar decisiones sobre la población, utilizando los resultados obtenidos, siempre que la muestra sea efectivamente representativa de la población en estudio. | Imagen: 228131\_i328  Representar en imagen población y muestra |

**Distribución Normal:**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | | Pestañas o tabs Verticales |
| **Introducción** | | La distribución normal o campana de Gauss, es una distribución teórica y en su fórmula involucra la media o promedio de la población y la desviación estándar ( de la población a estudiar. |
| Imagen: 228131\_i330  **Utilizar una representación de la Campana de Gauss, por ejemplo ésta tomada de:** [**https://calculo.cc/temas/temas\_estadistica/binomial\_normal/teoria/normal.html**](https://calculo.cc/temas/temas_estadistica/binomial_normal/teoria/normal.html) | | |
| **Media de la población** | se simboliza con la letra mu ( del alfabeto griego.  En la distribución normal la media es 0 y la desviación es 1, ; =1.  Los datos se encuentran agrupados alrededor de la media de la población , podemos hablar de porcentajes de datos agrupados en los intervalos (,); (,); (,) | |
| **Distribución normal estandarizada** | En la distribución normal se utiliza un estadístico llamado z, para que dado un conjunto de datos con media ydesviación mediante una traslación el conjunto de datos quede con ; =1.  Esa traslación tiene una fórmula y se simboliza con la letra se llama **distribución normal estandarizada**.  Después de utilizar esta traslación se obtendrá:   * Los valores z positivos corresponden a los valores x que son mayores que la media * Los valores z negativos corresponden a los valores x que son menores que la media | |
| **Ejemplo**  **nombre del botón** | **Hallando la media:**  En el ejemplo anterior calculamos la media o promedio para estos datos3, 4, 5, 6  **La media o promedio es**  = 4.5 la desviación =1.11 si utilizamos  **¿Cuál es el valor de z para cada uno de los datos?**   * Sustituyendo en z * Sustituyendo en z =-0.45 * Sustituyendo en z * Sustituyendo en z =1.35   Se observa que valores , que están por debajo de la media o promedio, los z correspondiente son negativos  Los valores que están por encima de la media o promedio los z correspondiente son positivos | |
| **Hallando la distribución normal estandarizada:**  En la **distribución normal estandarizada** se tiene ; =1, la notación es N( , ) = N(0,1)  Es la **distribución normal estandarizada con media 0 y desviación estándar 1.**  Sustituyendo en los intervalos agrupados de arriba el valor y =1, resulta (,); (,); (,) en el primer intervalo de una **distribución normal estandarizada** estaría el **68.2 %** de los datos, en el segundo intervalo el **95.4%** de los datos y en el último intervalo el **99.6%** de los datos | |
| **Gráfico 1** | **Representación de la media**  **Figura 1** *Representación de la media*    Gráfico realizado con Google Colab – Fuente propia  Imagen: 228131\_i331 | |
| **Gráfico 2** | **Representación de la distribución normal**  **Figura 2** *Representación de la distribución normal*    Gráfico realizado con MS Excel – Fuente propia  Imagen: 228131\_i332 | |
| **Gráfico 3** | **Figura 3** *Distribución normal con porcentajes de área bajo la curva*    Gráfico realizado con MS Excel – Fuente propia  Imagen: 228131\_i333 | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva p unto caliente | |
| **Texto introductorio** | **Distribución normal estandarizada** | |
| **Figura 4** *Distribución normal estandarizada*    Imagen: 228131\_i334  Gráfico realizado con MS Excel – Fuente propia  228131\_3 | | |
| **Código de la imagen** | Gráfico 228131\_3 | |
| **Punto caliente 1** | En el intervalo z en (-1, 1) se tiene el 68.2% de los datos | P1 en amarillo |
| **Punto caliente 2** | En el intervalo z en (-2, 2) se tiene el 95.4% de los datos | P2 en verde |
| **Punto caliente 3** | En el intervalo z en (-3, 3) se tiene el 99.6% de los datos | P3 en azul |

# **Introducción a la prueba de hipótesis para la media**

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| Ahora abordará la prueba de hipótesis con dos medias o promedios, para esto debe manejar términos como hipótesis nula, hipótesis alternativa, región de rechazo error tipo1.    Imagen: 228131\_i335  **Imagen sugerida:** [**https://www.iartificial.net/contraste-de-hipotesis-1-como-no-aceptar-lo-falso/**](https://www.iartificial.net/contraste-de-hipotesis-1-como-no-aceptar-lo-falso/) **o imagen similar** | |
| **Hipótesis nula e hipótesis alternativa:**  **La hipótesis nula y la hipótesis alternativa** **son opuestas**.  Si se trabaja con dos colas hablamos de en cada una de las colas de la gráfica de la distribución normal y la diferencia de medias , estaría en el centro de la gráfica de la distribución normal, sería la hipótesis nula, es decir, las dos medias son iguales y la hipótesis alternativa, es que las dos medias sean diferentes, es decir:  **Figura 5** *prueba de hipótesis para media*    **Prueba de hipótesis para medias.**  Gráfico realizado con MS Excel – Fuente propia  Imagen: 228131\_i336 | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | **Ejemplo de prueba de hipótesis para la media** | |
| Ahora comparará dos medias:  El jefe de una empresa dice que en promedio las mujeres que trabajan allí, se enferman igual que los hombres. | | **I**  Imagen: 228131\_i337  **Imagen de muestra, sugerida:** <https://www.freepik.es/foto-gratis/silueta-ejecutivos-comerciales-seguros_860252.htm#query=hombres%20y%20mujeres%20empresa&position=8&from_view=search> |
| **Solución:**  Se toma una muestra aleatoria de cada uno:  De 20 hombres arrojó un promedio de 8.4 y una desviación de 1.5  De 25 mujeres arrojó un promedio de 12.2 y una desviación de 2.0.  Se toma como hipótesis que se distribuyen normalmente.  Se quiere validar la hipótesis acerca de las dos poblaciones.  Con un error del 5%, este error es llamado error tipo 1, se simboliza con la letra .  El promedio de las mujeres que se enferman y el promedio de los hombres que se enferman  Para ello se propone una prueba de hipótesis:  Las dos medias son iguales , pasando a la izquierda resulta  Las dos medias son diferentes , pasando a la izquierda resulta  **Resumen:**  Ho = hipótesis nula =  H1 = hipótesis alternativa = | | Imagen: 228131\_i338  **Imagen:** Simular un gráfico de # porcentajes entre hombres y mujeres |
| **Error tipo 1**  En estadística cuando se plantea una prueba de hipótesis se habla de un error que se puede cometer, se llama **error tipo 1 ( el cual ocurre cuando se rechaza la hipótesis nula siendo esta verdadera.**  El error tipo 1( se conoce también como **error de tipo alfa (α)** o falso positivo. | | Imagen: 228131\_i339  **Imagen: sugerida para representar error tipo 1**  Fuente: <https://calculo.cc/temas/temas_estadistica/hipotesis/teoria/errores.html> |
| Con un intervalo de confianza del 95%, el resto sería el error, es decir, el error tipo 1 es = 5%= 0.05, como se está diciendo que son diferentes, no mayor ni menor, se debe trabajar con 2 colas el error se divide entre las dos colas  **.**  El valor **z** en la **distribución normal** **estandarizada** correspondiente a ese Intervalo de Confianzadel **95% es Z=1.96** | | **Imagen:** 228131\_4  Imagen: 228131\_i340  **Nota: se utiliza la misma imagen 4 para ambos textos.** |
| Siguiendo con el ejemplo, calculará el **estadístico**, si ese estadístico cae en el intervalo de z **(-1.96, 1.96)** **significa que cae en la región de aceptación de Ho,** por lo tanto, se puede decir que con un 95% de confianza no hay diferencia significativa entre los hombres y las mujeres que se enferman.  Pero si el **estadístico** cae fuera del intervalo de z (-1.96, 1.96), es decir, valores mayores a z = 1.96 o valores menores a z = -1.96 **significa que cae en la región de rechazo de Ho,** por lo tanto, aceptación de la hipótesis alternativa H1, se puede afirmar que con un 95% de confianza si hay diferencia significativa entre los hombres y las mujeres que se enferman. | |
| Ahora bien, el equivalente para las dos poblaciones de hombres y mujeres sería , siendo los hombres y las mujeres.  La desviación para las dos poblaciones de hombres y mujeres seria , siendo los hombres y las mujeres. | | magen: 228131\_i341  **Imagen: que simula el tema** |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva p unto caliente | |
| **Texto introductorio** | Aquí podrá apreciar el gráfico que representa la hipótesis nula y la hipótesis alternativa: | |
| **Figura 6** *Hipótesis nula y alternativa*    Imagen: 228131\_i342 | | |
| **Código de la imagen** | Imagen: 228131\_i342 | |
| **Punto modal 1** | Si el estadístico cae en el intervalo de z (-1.96, 1.96) significa que cae en la región de aceptación de Ho con un 95 % de confianza y un 5% de error tipo 1, por lo tanto, se acepta la hipótesis nula (Ho) | Punto 1 en amarillo |
| **Punto modal 2** | Si el estadístico no cae en el intervalo de z (-1.96, 1.96) significa que cae en la región de rechazo de Ho con un 95 % de confianza y un 5% de error tipo 1, por lo tanto, se rechaza la hipótesis nula (Ho) y se acepta la hipótesis alternativa(H1) | Punto 2 en verde |

|  |
| --- |
| Cuadro de texto |
| Continuando con el ejemplo de los hombres y mujeres en la empresa, se tiene que:  El **estadístico de prueba** para la diferencia de medias seria:    Donde:  =promedio de la muestra de hombres   * = desviación de la muestra de hombres * = tamaño de la muestra de hombres * = promedio de la población de hombres     promedio de la muestra de mujeres   * = desviación de la muestra de mujeres * = tamaño de la muestra de hombres * = promedio de la población de mujeres   Los valores del problema serían = 8.4, = 1.5,=30, =12.2, =2.0, =35, suponiendo que = 0  Sustituyendo el estadístico de prueba resulta:    El estadístico z = -8.74 cae por fuera del intervalo de z (-1.96, 1.96), **significa que cae en la región de rechazo de Ho,** por lo tanto, aceptación de la hipótesis alternativa H1, se puede afirmar que con un 95% de confianza si hay diferencia significativa entre los hombres y mujeres que se enferman (Vladimirovna, 2016). |

**Ahora se analizará con mayor detalle la gráfica de la distribución normal:**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Carrusel de tarjetas | |
| **Introducción** | La distribución normal, representada en la Campana de Gauss, tiene una función asociada que depende de la desviación, varianza, el promedio y el valor de x, tiene como nombre la función de densidad su fórmula es:  Media **=** desviación = , varianza = | |
| **Figura 7** *Distribución normal con diferente media*    Imagen: 228131\_i343  **Distribución normal con diferente media con la misma desviación y varianza. Fuente propia realizada con Maple** | | |
| Se obtendrán dos gráficas cambiando el valor de la media, y dejando la desviación y la varianza igual | | **Imagen: representar fórmula como imagen** |
| Dados los valores lo sustituye en la función para graficarla se reemplazan valores para x | | **Imagen: representar fórmula como imagen** |
| En la gráfica se observa en color rojo que tiene media positiva está hacia la derecha y en color azul que tiene media negativa está hacia la izquierda. | | Imagen: 228131\_i344 |

**Introducción a Google Colab**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | En Google Colab se tienen máquinas virtuales con tarjetas gráficas que Google facilita. Permite ejecutar código en el lenguaje Python dentro de la nube con el formato de Jupyter Notebook extensión ipynb.  Jupyter Notebook se puede utilizar localmente o en la nube con Google Colab. Para usarlo localmente debe instalarse en el computador.  Google Colab se encuentra en la dirección: <https://colab.research.google.com/?hl=es> | |
| Dentro de *Google Colab* se está en un entorno interactivo denominado *notebook* de *Colab*, que permite escribir y ejecutar código celda por celda o todo de una vez, se obtiene en la pestaña c ódigo.  Escribe en un lenguaje etiquetado especial llamado *Markdown*, se obtiene en la pestaña de nombre Texto. | | **Figura 8** *Pantalla de inicio de Google colab*    Pantalla de inicio de Google Colab  Imagen: 228131\_i345  Imagen propia |
| Instalar una librería o paquete  Ejemplo:  pip install pandas  Instalaría la librería de nombre pandas si ya está instalada lo dirá.  Importar librería con un alias(as)  Ejemplo:  import pandas as pd  Cargar la librería panda con el alias o abreviación pd  Ejemplo:  import math  Cargar la librería de nombre math referente a matemática.  print(dir(math))  Muestra las funciones que están dentro de la librería math  pow(2,3) una de las funciones que están en la librería math, es la función potencia en inglés (Power), pow(2,3) significa que eleva el número 2 a la 3 y da como resultado 8. | | **Figura 9** *Importar libreria math y utilizar la función potencia*    Importar la librería math y utiliza la función potencia: pow()  Imagen: 228131\_i346  Imagen propia |
| Un Dataframe es una estructura de datos con dos dimensiones (filas y columnas), se puede guardar datos de distintos tipos (caracteres, enteros, valores de punto flotante, etc.) en columnas. Es similar a una hoja de cálculo o una tabla de SQL. | | Dataframe  Imagen de muestra. Reproducir una imagen similar.  Imagen: 228131\_i347 |
| Un archivo CSV (valores separados por comas) es un archivo de texto que tiene un formato específico, que permite guardar los datos en un formato de tabla estructurada.  Un archivo CSV puede tener diferentes separadores como:   |  |  | | --- | --- | |  | Coma | |  | Punto y coma | |  | Barra espaciadora | |  | Guion medio | |  | Guion bajo | | | CSV  Imagen de muestra, representar un archivo CSV  Imagen: 228131\_i348 |
| Para leer un archivo en formato CSV se utiliza el comando pd.read\_csv('nombre del archivo.csv'), donde pd es la abreviación de la librería pandas.  En la parte nombre del archivo se debe escribir la ruta o dirección donde está el archivo al final de esa ruta el nombre con extensión.csv  Se puede utilizar el comando pwd para saber la ruta donde se está ubicado.  Para cambiar el directorio o ruta se utiliza el comando cd  Ejemplo: cd /content/sample\_data/ | | magen: 228131\_i349  Librería pandas  Imagen sugerida  Fuente: <https://www.analyticslane.com/2020/11/30/pandas-leer-archivos-csv-con-diferentes-delimitadores-en-pandas/> |
| Ejemplo:   * Leer un archivo csv, utilizando el alias de pandas pd * Nombrarlo data\_frame * Visualizarlo con el nombre que se le colocó en este caso data\_frame.   import pandas as pd  data\_frame = pd.read\_csv('california\_housing\_test.csv')  data\_frame | | **Figura 10** *Lectura y contenido del data frame*    Lectura y contenido del data frame utilizando el alias pd.read\_csv()  Imagen: 228131\_i350  Imagen propia |
| Para obtener las primeras 5 filas del data frame se escribe:   * data\_frame.head(5)   Para obtener las últimas 5 filas del data frame se escribe:   * data\_frame.tail() | | **Figura 11** *Contenido de las primeras y últimas cinco filas del data frame*    Contenido de las primeras y últimas 5 filas del data frame  Imagen: 228131\_i351  Imagen propia |
| Para ver la dimensión (largo y ancho), filas y columnas también el nombre de las columnas, en este caso 3000 filas, 9 columnas se utilizan los comandos:   * print("Dimensión: ", data\_frame.shape) * print("Columnas: ", data\_frame.columns) * Shape: (3000, 9) | | **Figura 12** *Dimensión y nombre de las columnas del data frame*    Dimensión y el nombre de las columnas del data frame  Imagen: 228131\_i352  Imagen propia |
| **Para ver i**nformación sobre el tipo de dato y si tiene datos únicos no repetidos   * print(dataframe.info()) * print("\Conteo de valores únicos:\n") * print(data\_frame.nunique()) | | **Figura 13** *Tipos de datos*    Tipo de dato y si existencia de datos únicos no repetidos  Imagen: 228131\_i353  Imagen propia |
| Información sobre los estadísticos descriptivos de la columna con nombre 'total\_rooms', se escribe el nombre entre comillas sencillas o dobles.  data\_frame['total\_rooms'].describe()  También se puede: data\_frame["total\_rooms"].describe()  Redondear la salida de los valores a 2 decimales.  round(data\_frame['total\_rooms'].describe(),2) | | **Figura 14** *Información estadísticos descriptivos de la columna "total\_rooms"*    Información sobre los estadísticos descriptivos de la columna con nombre 'total rooms'  Imagen: 228131\_i354  Imagen propia |
| Los estadísticos descriptivos son: (min), valor máximo (max), media o promedio (mean), desviación estándar (std), total de filas(count) y cuartiles()   |  |  | | --- | --- | | **min** | Valor mínimo | | **max** | Valor máximo | | **mean** | Media o promedio | | **std** | Desviación estándar | | **count** | Total de filas | | **25%,50%,75%** | Cuartiles | | | Estadística descriptiva  Imagen de muestra  Imagen: 228131\_i355  Link: <http://3.bp.blogspot.com/-6K1ymY5aO1g/VfMxVXz4HZI/AAAAAAAAEZ0/hw-j-Mc370E/s1600/technology-report-752x483.jpg> |
| Para mayor información ver Toomey, 2017; Toomey, 2018 | | Imagen: 228131\_i356  Imagen opcional – mostrar parte de la carátula del libro de la referencia. |

|  |
| --- |
| Cuadro de texto |
| En el siguiente video podrá observar la aplicación de la teoría de prueba de hipótesis en un caso práctico real, tomando la información de la base de datos libres de la Alcaldía de Medellín, referente a créditos de consumo otorgados a hombres y mujeres cuyas edades están entre 18 años y 84 años.  Se utilizará Google Colab y Python con sus diferentes librerías: Plotly para gráficos dinámicos, statsmodels. stats librería estadística para Prueba de hipótesis, entre otras.  Link a video: <https://drive.google.com/file/d/1Xp9cluKuhvMIB8rDIfhRGxxeb_oYvf7E/view?usp=sharing>  Video: v228131\_1-CF3 Ver2 |

|  |
| --- |
| Cuadro de texto |
| Para una mejor comprensión del video, el manejo de Google Colab y el documento jupyter Notebook con extensión ipynb, el cual se trabaja en el video, los invitamos a descargar los 3 archivos adjuntos.   |  |  |  |  | | --- | --- | --- | --- | | **Tema** | **Tipo de archivo** | **Nombre de archivo** | **Enlace Web** | | Manejo de Google Colab | .PDF | Manejo de Google Colab.pdf |  | | Créditos de consumo | .CSV | creditos\_de\_consumo\_otorgados.csv |  | | Gráficos ploty | .IPYNB | Prueba\_de\_hipótesis\_diferencias\_de\_medias.ipynb |  |   SE ENCUENTRAN EN LA CARPETA CF03\_228131 - Material descargable |

**SÍNTESIS**

|  |  |
| --- | --- |
| **Tipo de recurso** | Síntesis |
| Gestión de analítica de datos  Síntesis: Probabilidad y prueba de hipótesis | |
| **Introducción** | En el siguiente mapa podrá apreciar todos los conceptos relacionados con la probabilidad, hipótesis y desviación estándar, de modo que se adquiera una visión global de la temática desarrollada en este ítem : |
| Imagen: 228131\_i357  mapa conceptual CF 3: Probabilidad y Prueba de Hipótesis  Fuente online: <https://cmapscloud.ihmc.us:443/rid=1YBSTTZJC-13J9LJP-CVHTTD> | |

**ACTIVIDAD DIDÁCTICA:**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Actividad didáctica. v erdadero y falso | |
| Con esta actividad se pretende repasar los conceptos estadísticos trabajados en el módulo y determinar qué se ha aprendido al respecto.  Muchos éxitos. | | Imagen: 228131\_i358  **Imagen sugerida:**  <https://www.freepik.es/vector-gratis/manos-sosteniendo-carteles-cruz-marca-verificacion-incorrecto-correcto-cancelar-o-aprobar-voto-ilustracion-vectorial-plana-persona-examen-encuesta-concepto-respuesta-banner-diseno-sitio-web-o-pagina-web-inicio_27572576.htm#query=falso%20y%20verdadero&position=3&from_view=search> |
| La probabilidad se define como: | | **Presentar fórmula como imagen para la pregunta** |
| **Verdadero** | | **Falso (correcto)** |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  La afirmación es falsa, dado que la probabilidad se define como | | **Retroalimentación a respuesta errónea:**  INCORRECTO:  La afirmación es falsa, dado que la probabilidad se define como |
| La probabilidad puede ser menor a cero: | | Imagen: 228131\_i359  **Imagen de ejemplo** |
| Verdadero | | **Falso (correcto)** |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  Puede tomar cualquier valor real en el intervalo [0,1] | | **Retroalimentación a respuesta errónea:**  INCORRECTO:  La afirmación es falsa, dado que P puede tomar cualquier valor real en el intervalo [0,1] |
| El espacio muestral se refiere a las veces que se repite un experimento: | | Imagen: 228131\_i360  **Imagen sugerida**  Fuente: <https://www.shutterstock.com/es/image-vector/sample-population-statistics-research-survey-methodology-688050838> |
| Verdadero | | **Falso (correcto)** |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  Espacio muestral sería el conjunto de valores que toma un experimento | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  El espacio muestral sería el conjunto de valores que toma un experimento |
| El promedio se calcula sumando los datos y dividiendo por las veces que se repite el dato: | | Imagen: 228131\_i362  **Imagen sugerida... poner una imagen que indique promedio de datos**  <https://www.shutterstock.com/es/image-vector/colorful-level-satisfaction-icon-emotion-customer-2161599321> |
| Verdadero | | **Falso (correcto)** |
| **Retroalimentación a respuesta acertada:**  CORRECTO:  El promedio se calcula sumando los datos y dividiendo por el número total de los mismos | | **Retroalimentación a respuesta errónea:**  INCORRECTO:  El promedio se calcula sumando los datos y dividiendo por el número total de los mismos |
| La desviación estándar mide qué tanto se alejan los datos de la moda | | Imagen: 228131\_i363  **Imagen sugerida que represente una desviación en Campana de Gauss**  <https://www.shutterstock.com/es/image-photo/businessman-growing-statistic-curve-graph-gaussian-181632143> |
| **Verdadero** | | **Falso (correcto)** |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  La desviación estándar mide que tanto se alejan los datos del promedio. | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  La desviación estándar mide que tanto se alejan los datos del promedio. |
| La distribución normal estandarizada tiene media cero y desviación estándar 1 | | **Ejemplo de imagen sugerida**  Imagen: 228131\_i364 |
| **Verdadero (correcto)** | | Falso |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  **Efectivamente, l**a distribución normal estandarizada tiene ; =1 | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  Recuerde que la distribución normal estandarizada tiene ; =1 |
| La fórmula que se utiliza para obtener la distribución normal estandarizada es: | | **Representar fórmula como imagen** |
| **Verdadero (correcto)** | | Falso |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  La distribución normal estandarizada se obtiene restando a x el valor de la media y dividiéndolo por la desviación estándar. | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  Recuerde que la distribución normal estandarizada se obtiene restando a x el valor de la media y dividiéndolo por la desviación estándar. |
| En la distribución normal estandarizada z en el intervalo [-1,1] tiene el 68.2% de los datos | | Imagen: 228131\_i365  **Imagen que represente valor de Z** |
| **Verdadero (correcto)** | | Falso |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  Efectivamente, la distribución normal estandarizada z en los intervalos (-1, 1), (-2,2), (-3, 3) en el primer intervalo estaría el 68.2 % de los datos, en el segundo intervalo en el 95.4% de los datos y en el último intervalo en el 99.6%. | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  La distribución normal estandarizada z en los intervalos (-1, 1), (-2,2), (-3, 3) en el primer intervalo estaría el 68.2 % de los datos, en el segundo intervalo en el 95.4% de los datos y en el último intervalo en el 99.6%. |
| En la prueba de hipótesis el error tipo 1 (α) ocurre cuando se rechaza la hipótesis nula siendo ésta verdadera. | | Imagen: 228131\_i366  **Imagen sugerida o se puede reemplazar por una persona pensando…**  <https://www.estadistica.net/Aeronautica2016/contrastes.pdf> |
| **Verdadero (correcto)** | | Falso |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  Así es. En un intervalo de confianza del 95%, el resto sería el error, es decir, el error tipo 1 es α = 5%= 0.05, α ocurre cuando rechazo la hipótesis nula siendo esta verdadera. | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  Recuerde que en un intervalo de confianza del 95%, el resto sería el error, es decir, el error tipo 1 es α = 5%= 0.05, α ocurre cuando rechazo la hipótesis nula siendo esta verdadera. |
| En una prueba de hipótesis la hipótesis nula (Ho) y la hipótesis alternativa(H1) son opuestas | | Imagen: 228131\_i367  **Imagen sugerida**  <https://www.iartificial.net/contraste-de-hipotesis-1-como-no-aceptar-lo-falso/> |
| **Verdadero (correcto)** | | Falso |
| **Retroalimentación a respuesta acertada:**  **CORRECTO**  Efectivamente, en una prueba de hipótesis se trabaja con hipótesis nula (Ho) y con hipótesis alternativa (H1) y estas son opuestas. Por ejemplo, si la hipótesis nula es que el promedio de una persona es 80 kg, las alternativas pueden ser promedio mayor a 80 kg., promedio menor a 80 kg. o promedio diferente a 80 kg. | | **Retroalimentación a respuesta errónea:**  **INCORRECTO**  Recuerde que en una prueba de hipótesis se trabaja con hipótesis nula (Ho) y con hipótesis alternativa (H1) y estas son opuestas. |

**MATERIAL COMPLEMENTARIO**

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de recurso | Material complementario | | |
| Tema | Referencia APA del material | tipo | Enlace |
| Estadística descriptiva | Aguilar, J.E. (2021). Estadística *Descriptiva, regresión y probabilidad con aplicaciones* (1ra ed.). Ediciones de la U | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=15156&pg=202> |
| Python práctico | Moreno, A., Córcoles, S. (2020) *Python práctico* (1ra ed.) Ediciones de la U. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=10285&pg=1> |
| Estadística | Triola, M. (2018). *Estadística* (12 ed.) Pearson Educación | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=7386&pg=149> |
| Learning Jupyter 5 | Toomey, D. (2018). *Learning Jupyter 5: Explore interactive computing using Python, Java, JavaScript, R, Julia, and JupyterLab*, (2nd Ed) Packt Publishing Ltd. | Libro digital | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1881491&lang=es&site=ehost-live> |
| Jupyter for Data Science | Toomey, D. (2017) *Jupyter for Data Science. Exploratory analysis, statiscal modeling, machine learning and data visualization with Jupyter*. (1ra. Ed.) Packt Publishing Ltd. | Libro digital | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1637911&lang=es&site=ehost-live> |

**GLOSARIO**

|  |  |
| --- | --- |
| **Tipo de recurso** | Glosario |
| Censo | Es un conjunto de actividades destinadas a recoger, recopilar, el padrón o lista de la población de una nación o pueblo en un momento determinado, la información demográfica social y económica, para luego evaluar, analizar y publicar o divulgar. |
| Continua | Es aquella que puede tomar un número infinito de valores entre dos valores cualesquiera de una característica. La altura de cinco personas: 1.85, 1.89, 1.93, 1.59, 1.65. En la práctica se mide la altura con dos decimales, pero también se podría dar con tres decimales o más. |
| Cualitativa | Son aquellas variables en la que los resultados posibles no son valores numéricos. Por ejemplo: color del pelo, tipo de ropa preferida, lugar de veraneo, etc. |
| Cuantitativa | Se trata de características cuyas variaciones pueden ser expresadas de forma numérica. Por ejemplo: edad, peso, estatura, número de hijos, años de servicio, etc. |
| Desviación estándar | Mide la dispersión de una distribución de datos. Entre más dispersa está, más grande es su desviación estándar. |
| Discreta | Hace referencia cuando no se puede tomar ningún valor entre dos consecutivos, y es continua cuando puede tomar cualquier valor dentro de un intervalo. Ejemplos de variables discretas: número de empleados de una fábrica, número de hijos, número de cuentas en un banco. |
| Distribución Normal | Es un modelo teórico capaz de aproximar satisfactoriamente el valor de una variable aleatoria a una situación ideal. |
| Escala Cronológica | El orden cronológico de los sucesos, es la ubicación de éstos en una línea temporal. |
| Escala n ominal | Es una medición en la cual los números sirven como “etiquetas,” solamente para identificar o clasificar un objeto. Normalmente trata sólo con variables no numéricas (no cuantitativas). |
| Escala o rdinal | Es la que permite evaluar la actitud de un encuestado con respecto a un tema utilizando un conjunto de respuestas ordenadas. Por ejemplo, las respuestas pueden incluir: “Muy satisfecho”, “Satisfecho”, “Insatisfecho” y “Muy insatisfecho”. |
| Estadísticos | Es cualquier función real medible de la muestra de una variable aleatoria. |
| Media | Resultado que se obtiene al dividir la suma de varias cantidades por el número de sumandos. |
| Muestreo | Selección de un conjunto de personas o cosas que se consideran representativos del grupo al que pertenecen, con la finalidad de estudiar o determinar las características del mismo. . |
| Probabilidad | Es el valor numérico que sirve para determinar la ocurrencia o no de una situación dada. Por ejemplo, cuando se lanza una moneda al aire, existe una probabilidad de 0,50 0 50% de obtener cara o sello. |
| Prueba de hipótesis | Es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos. |

**Referencias bibliográficas:**

|  |  |
| --- | --- |
| **Tipo de recurso** | Bibliografía |
| Aguilar, J.E. (2021). Estadística Descriptiva regresión y probabilidad con aplicaciones (1ra ed.). Ediciones de la U. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=15156&pg=202> | |
| Martínez, C. (2019). Estadística y muestreo (14 ed.) ECOE Ediciones. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9278&pg=387> | |
| Martínez, C. (2019). Estadística básica aplicada (5ta ed.) ECOE ediciones. Recuperado de: <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=8975&pg=82> | |
| Lind, D., Marchal, W. and Wathen, S., (2019). Estadística aplicada a los negocios y la economía. 17a ed. México: McGraw-Hill Interamericana. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9246&pg=315> | |
| Gutiérrez, E., Vladimirovna, O. (2016). Estadística inferencial 1 para ingeniería y ciencias. Ed. México: Grupo Editorial Patria. <https://elibro-net.bdigital.sena.edu.co/es/ereader/senavirtual/40474?page=69> | |