****

**Actividad 4. Ejercicios de Estadística Descriptiva - Inferencial y control de lectura**

**CAPITULO 10. ANALISIS DE LOS DATOS**.

SINTESIS: EI capítulo presenta Ios procedimientos generales para efectuar análisis estadísticos por computador. Así mismo, se comentan, analizan y ejemplifican Ias pruebas y análisis estadísticos mas utilizados en ciencias sociales; incluyendo estadísticas descriptivas, análisis paramétricos, no paramétricos y muItivariados. En la mayoría de estos análisis el enfoque del capítulo se centra en Ios usos y Ia interpretación de Ia prueba más que en el procedimiento de caIcuIar estadísticas, debido a que actualmente Ios análisis se realizan con ayuda de computador y no manualmente, muy pocas veces es necesario que el investigador haga sus caIcuIos a mano basándose en Ias fórmulas disponibles. En Ia actualidad, Ias fórmulas ayudan a entender Ios conceptos estadísticos, pero no caIcuIar estadísticas. EI capítulo también proporciona una introducción general a Ios análisis muItivariados.

**Noveno paso del proceso de investigación: Analizar Ios Datos. Se compone de:** Decidir que pruebas estadísticas son apropiadas para analizar Ios datos, dependiendo de Ias hipótesis formuladas y Ios niveles de medición de Ias variables.

* Elaborar el programa de computadora para analizar Ios datos: utilizando un paquete estadístico o generando un programa propio.
* Correr el programa.
* Obtener Ios análisis requeridos.
* Interpretar Ios análisis.
* EI anaIisis de Ios datos se efectúa utilizando Ia matriz de datos, Ia cual está guardada en un archivo.
* EI tipo de análisis o pruebas estadísticas a realizar depende del nivel de medición de Ias variables, Ias hipótesis y el interés del investigador.
* Los análisis estadísticos que pueden realizarse son: estadística descriptiva para cada variable (distribución de frecuencias, medidas de tendencia central y medidas de Ia variabilidad), Ia transformación a puntuaciones “Z”, razones y tasas, caIcuIos de estadística inferenciaI, pruebas paramétricas, pruebas no paramétricas y análisis muItivariados.
* Las distribuciones de frecuencias contienen Ias categorías, códigos, frecuencias absolutas (numero de casos), frecuencias relativas (porcentajes) y frecuencias acumuladas (absolutas o relativas).
* Las distribuciones de frecuencias (particularmente hablando de Ias frecuencias relativas) pueden presentarse gráficamente.
* Una distribución de frecuencias puede representarse a través del polígono de frecuencias o curva de frecuencias.
* Las medidas de tendencia central son Ia moda, mediana y media.
* Las medidas de Ia variabilidad son el rango (diferencia entre el máximo y el minimo), Ia desviación estándar y Ia varianza.
* Otras estadísticas descriptivas de utilidad son Ia asimetría a y Ia curtosis.
* Las puntuaciones “Z” son transformaciones de Ios valores obtenidos a unidades de desviación estándar.

Una razón es Ia relación entre dos categorías y una tasa es Ia relación entre eI numero de casos de una categoría a y el numero total de casos, multiplicada por un numero múltiplo de 10.

La estadística inferenciaI es para efectuar generalizaciones de Ia muestra a Ia población. Se utiliza para probar hipótesis y estimar parámetros. Asimismo, se basa en el concepto de distribución muestraI.

La curva o distribución normal es un modelo teórico sumamente útil, su media es cero (0) y su desviación estándar es uno (1).

EI nivel de significancia y el intervalo de confianza son niveles de probabilidad de cometer un error o equivocarse en la prueba de hipótesis o Ia estimación de parámetros. Los niveles mas comunes en ciencias sociales son Ios deI 0.05 y 0.01.

Los análisis o pruebas estad sticas paramétricas más utilizadas son:

|  |  |
| --- | --- |
| **PRUEBA** | **TIPOS DE HIPOTESIS** |
| Coeficiente de correlación de Pearson | CorreIacionaI |
| Regresión LineaI | CorreIacionaI / CausaI |
| Prueba “t” | Diferencia de grupos |
| Contraste de Ia diferencia de proporciones | Diferencia de grupos |
| Análisis de varianza (ANOVA): | Diferencia de grupos / CausaI |
| Análisis de covarianza (ANCOVA) | CorreIacionaI / CausaI |

En todas Ias pruebas estadísticas paramétricas Ias variables están medidas en un nivel por intervalos o razón.

Los análisis o pruebas estad sticas no paramétricas más utilizadas son:

|  |  |
| --- | --- |
| **PRUEBA** | **TIPOS DE HIPOTESIS** |
| Ji cuadrada (z2). | Diferencias de grupos para establecer correlación. |
| Coeficientes de correlación e independencia para tabulaciones cruzadas: Phi, C de Pearson, V de Cramer, Lambda, Gamma, Taub, D de Sommers y Eta. | CorreIacionaI |
| Coeficientes Spearman y KendaII. | CorreIacionaI |

Las pruebas no paramétricas se utilizan con variables nominales u ordinales.

Los análisis muItivariados mas utilizados son:

|  |  |  |
| --- | --- | --- |
| **PRUEBA** | **VARIABLES INVOLUCRADAS Y**  **NIVELES DE MEDICION** | **TIPOS DE HIPOTESIS** |
| Regresión múltiple | Una dependiente (intervalos o razón) y dos o más independientes (cualquier nivel de medición). | CorreIacionaI / Causal X1  X2 Y  XK |
| Análisis Lineal path | Varias, secuencia causaI (cuaIquier niveI de medicion). | CorreIacionaI / Causal  X W L  Y  Z N H |
| Análisis de factores | Varias (intervalos o razón). | CorreIacionaI / Causal  X1 X2 XK |
| Análisis multivariado de varianza (MANOVA) y correlación canoníca. | Varias independientes y varias dependientes (intervalos o razón). | X1 Y1  X2 Y2  XK YK |
| Análisis discriminante | Varias independientes (intervalos o razón) y una dependiente (nominaI u ordinaI). | X1  X2 Y  X3 |

Los análisis estad sticos se IIevan a cabo mediante programas para computador, utilizando paquetes estadísticos.

Los paquetes estadísticos más conocidos son: BMDP, ESP, OSIRIS, SAS y SPSS. Estos paquetes se utilizan consultando el manual respectivo.

**Enfoque Cuantitativo**: Se fundamenta en el método hipotético deductivo. Establece teorías y preguntas iniciales de investigación, de las cuales se derivan hipótesis. Estas se someten a prueba utilizando diseños de investigación apropiados. Mide las variables en un contexto determinado, analiza las mediciones, y establece conclusiones. Si los resultados corroboran las hipótesis, se genera confianza en la teoría, si no es refutada y se descarta para buscar mejores. Reduccionista. Utiliza medición numérica, conteo, y estadística, encuestas, experimentación, patrones, recolección de datos.

**¿QUE PROCEDIMIENTO SE SIGUE PARA ANALIZAR LOS DATOS?**

Una vez que Ios datos se han codificado, transferido a una matriz y guardado en un archivo, el investigador puede proceder a analizarlos. EI énfasis se centra en Ia interpretación de Ios métodos de análisis cuantitativo y no en Ios procedimientos de cálculo. EI análisis de Ios datos se efectúa sobre Ia matriz de datos utilizando un programa de computadora.

¿**QUE ANALISIS PUEDEN EFECTUARSE EN LOS DATOS?**

Los análisis dependen de tres factores:

* EI nivel de medición de Ias variables.
* La manera como se hayan formulado Ias hipótesis.
* EI interés del investigador.

EI investigador busca, en primer término, describir sus datos y posteriormente efectuar análisis estadísticos para relacionar sus variables. Es decir, realizar análisis de estadística descriptiva para cada una de sus variables y Iuego describe Ia relación entre estas. Los principales análisis que pueden efectuarse son:

Estadística descriptiva para las variables, tomadas individualmente.

* Puntuaciones “Z”.
* Razones y Tasas.
* CaIcuIos y Razonamientos de estadística inferenciaI.
* Pruebas paramétricas.
* Pruebas no paramétricas.
* Análisis muItivariados.

ESTADISTICA DESCRIPTIVA PARA CADA VARIABLE.

La primera tarea es describir Ios datos, valores o puntuaciones obtenidas para cada variable.

¿Cómo pueden describirse estos datos?

Describiendo Ia distribución de Ias puntuaciones o frecuencias.

**Ejercicio 1**

* **Calcula la muestra para una población desconocida con un 96% de confianza y 4% error. Para una prevalencia de .5**

|  |  |
| --- | --- |
| Tamaño universo | 100 |
| Heterogeneidad | 96 % |
| Margen de error | 4 %| |
| Nivel de confianza | .5 % |
| Muestra | 1 |
| **Si encuestas a 1 persona, el .50% de las veces el dato real que buscas estará en el intervalo + 4 % respecto al dato que observas en la encuesta.** | |

**Ejercicio 2**

* **Calcula la muestra para una población de 350,000 familias, con un 99% de confianza y 1% error. Para una prevalencia de .5 y .7**

|  |  |
| --- | --- |
| Tamaño universo | 350,000 |
| Heterogeneidad | 99 % |
| Margen de error | 1 %| |
| Nivel de confianza | .5 % |
| Muestra | 1 |
| **Si encuestas a 1 persona, el .50% de las veces el dato real que buscas estará en el intervalo + 1 % respecto al dato que observas en la encuesta.** | |

**Ejercicio 3**

* **De una población de 1,176 padres de familia de la ciudad de Tuxtla Gutiérrez, se pretende conocer la aceptación de los programas educativos mediante caricaturas. Se pretende obtener una muestra para saber el número de entrevistas y con ello obtener información estadísticamente confiable. Se asume un** [**error standard**](Error%20estándar.docx) **de 1.5% con un nivel de confiabilidad del 90%**

|  |  |
| --- | --- |
| Tamaño universo | 1176 |
| Heterogeneidad | 0 |
| Margen de error | 1.5 %| |
| Nivel de confianza | 90 % |
| Muestra | 0 |
| **Si encuestas a 0 persona, el 90 % de las veces el dato real que buscas estará en el intervalo + 1.5 % respecto al dato que observas en la encuesta.** | |

**Ejercicio 4  
Con los siguientes datos**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | * **Son los resultados de preguntarle la estatura a 60 trabajadores del departamento de limpia municipal de SCLC.** * **Obtén la media aritmética (para datos agrupados)** * **Obtén la desviación estándar y la varianza (para datos agrupados)** * **Interpreta los resultados** | | | |
|  | | | |  | | | |
| **Rango** | **ti** | **Xi** | **fi . Xi** | | **2**  **(Xi- X˜)** | **2**  **(Xi- X˜)** | **ti** | |
| **1.52--- 1.58** | **15** | **1.35** | **23.25** | | **0.2459267281** | **3.6889009215** |  | |
| **1.6---1.66** | **16** | **1.63** | **26.08** | | **0.1729811281** | **2.7676980496** |  | |
| **1.7---1.77** | **27** | **2.585** | **69.795** | | **0.2906180281** | **7.8466867587** |  | |
| **1.8---1.83** | **2** | **1.815** | **3.63** | | **0.05331944281** | **0.1066388562** |  | |
| **∑** | **60** | **∑** | **122.755** | |  |  |  | |

**∑= 14.409924586**

**\_ ∑ Xiti = 122.755 = 2.04591 media aritmética**

**X n 60**

**2 \_ 2**

**Ơ ∑ (Xi-X) .f = 14.4099 = 0.2401654098 varianza**

**N 60**

**\_\_\_\_\_\_**

**Ơ = √ =** √ **0.2401654098 =0.4900667401**



**. Desviación estándar**

* **Considera las estaturas de un padre y su hijo.**

|  |  |
| --- | --- |
| **Padre (m)** | **1.70, 1.77, 1.68, 1.75, 1.80, 1.75, 1.69, 1.72, 1.71, 1.73** |
| **Hijo (m)** | **1.74, 1.78, 1.72, 1.77, 1.78, 1.77, 1.71, 1.76, 1.73, 1.74** |

* **Obtén el promedio de estaturas. En ambos casos**
* **Elabora la grafica de dispersión correspondiente**
* **Obtén el coeficiente de correlación de pearson a partir de**
* **Interpretación**
* **Si r = 0 no existe ninguna correlación**
* **Si r = 1 existe una correlación positiva perfecta**
* **Si 0 < r <1, existe una correlación positiva**
* **Si r = -1 existe una correlación negativa perfecta**
* **Si -1 < r < 0 existe una correlación negativa**

**Procedimiento**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X** | **Y** | **X2** | **Y2** | **XY** |
| 1.70 | 1.74 | 2.89 | 3.0276 | 2.958 |
| 1.77 | 1.78 | 3.1329 | 3.1684 | 3.1506 |
| 1.68 | 1.72 | 2.8224 | 2.9584 | 2.8896 |
| 1.75 | 1.77 | 3.0625 | 3.1329 | 3.0975 |
| 1.80 | 1.78 | 3.24 | 3.1684 | 3.204 |
| 1.75 | 1.77 | 3.0625 | 3.1329 | 3.0975 |
| 1.69 | 1.71 | 2.8561 | 2.9241 | 2.8899 |
| 1.72 | 1.76 | 2.9584 | 3.0976 | 3.0272 |
| 1.71 | 1.73 | 2.9241 | 2.9929 | 2.9583 |
| 1.73 | 1.74 | 2.9929 | 3.0276 | 3.0102 |
| ∑ 17.3 | ∑ 17.5 | ∑ 29.9418 | ∑ 30.6308 | ∑ 30.2828 |

**V** = \_\_∑30.2828\_\_\_\_\_\_\_\_ = \_\_\_\_\_30.2828\_\_\_\_\_= \_\_30.2828\_\_ = 0.9999

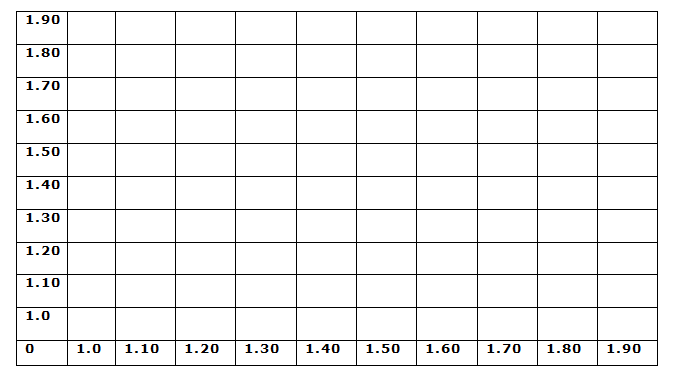
(\sqrt{\ }29.9418) .(\sqrt{\ }30.6308) (5.4719) . (5.5345) 30.2842

**V= 0.9999**

Si 0 < v < 1, existe una correlación positiva, entonces como R= 0.9999, tenemos una correlación positiva, casi apuntando a una perfecta

**Grafica del promedio de estatura**

2

****

**Hijo (m)**

****

* **Considera las calificaciones de 10 alumnos en las asignaturas de matemáticas y física.**

|  |  |
| --- | --- |
| **Matemáticas** | **2, 4, 5, 5, 6, 6, 7, 7, 8, 9** |
| **Física** | **2, 2, 5, 6, 5, 7, 5, 6, 7, 10** |

* **Obtén el promedio de calificaciones**
* **Elabora la grafica de dispersión correspondiente**
* **Obtén el coeficiente de correlación de una persona a partir de**
* **Interpretación**
* **Si r = 0 no existe ninguna correlación**
* **Si r = 1 existe una correlación positiva perfecta**
* **Si 0 < r <1, existe una correlación negativa perfecta**
* **Si -1 < r < 0 existe una correlación negativa**

**Procedimiento**

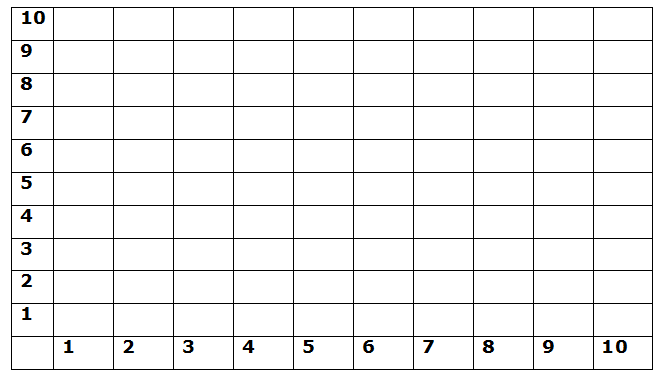
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X** | **Y** | **X2** | **Y2** | **XY** |
| 2 | 2 | 4 | 4 | 4 |
| 4 | 2 | 16 | 4 | 8 |
| 5 | 5 | 25 | 25 | 25 |
| 5 | 6 | 25 | 36 | 30 |
| 6 | 5 | 36 | 25 | 30 |
| 6 | 7 | 36 | 49 | 42 |
| 7 | 5 | 49 | 25 | 35 |
| 7 | 8 | 49 | 64 | 56 |
| 8 | 7 | 64 | 49 | 56 |
| 9 | 10 | 81 | 100 | 93 |
| ∑ 59 | ∑ 57 | ∑ 385 | ∑ 381 | ∑ 376 |

**V=** \_\sqrt{\ }x y\_\_ = \_\_\_\_\_376\_\_\_\_\_\_\_ = \_\_\_\_\_376\_\_\_\_\_\_\_ = \_\_379\_\_ = 0.9817

\sqrt{\ }x \sqrt{\ }y \sqrt{\ }(385) .\sqrt{\ }(381) (19.6214) . (19.5192) 382.994

* **Existe una correlación positiva**

**Grafica del promedio calificaciones**



**Matemáticas**

**Física**