**CAPITULO 10. ANÁLISIS DE LOS DATOS CUANTITATIVOS.**

El análisis cuantitativo de los datos de lleva a cabo por computadora. Y nadie lo hace de forma manual, en especial si es un volumen considerable de datos. Se centran ahora en la interpretación de los resultados de los métodos de análisis y no el procedimiento de cálculo.

El proceso de análisis se esquematiza en los siguientes pasos:

Fase 1. Seleccionar un programa estadístico en computadora para analizar los datos. Existen diversos programas. En esencia es similar su funcionamiento. Lo que se hace una vez recolectados los datos, es definir los parámetro de la matriz de datos en el programa e introducir los datos. Cada celda contiene un dato.

Fase 2. Ejecutar el programa: SPSS, Stas, SAS o su equivalente. Statistical Package for the Social Sciences o Paquete Estadistico para las Ciencias Sociales fue desarrollado en la Universidad de Chicago, y es uno de los más difundidos, pues contiene todos los análisis estadísticos.

Fase 3. Explorar los datos: analizar descriptivamente los datos por variable y visualizar los datos por variable. Se inicia el análisis. Si llevábamos a cabo toda la investigación paso a paso, tenemos claridad.

Fase 4. Evaluar la confiabilidad y validez logradas por el instrumento de medición.

Fase 5. Analizar mediante pruebas estadísticas las hipótesis planteadas.

Fase 6. Realizar análisis adicionales.

Fase 7. Preparar los resultados para presentarlos.

Estadística descriptiva para cada variable. La primer tarea es describir los datos, los valores o las puntuaciones obtenidas en cada variable.

Una distribución de frecuencia es un conjunto de puntuaciones ordenadas en sus respectivas categorías. Se pueden presentar utilizando unos histogramas o gráficas de otro tipo como polígonos de frecuencia, los cuales representan curvas útiles para describir datos. Nos indican hacia dónde se concentran los casos en la escala variable.

El porcentaje acumulado es lo que aumenta en cada categoría de manera porcal y progresiva.

Las medidas de tendencia central son puntos en una distribución, los valores medios o centrales de ésta, y nos ayudan a ubicarlas dentro de la escala de medición. Las principales medidas de tendencia central son tres: moda, mediana y media. La moda es la categoría o puntuación que ocurre con mayor frecuencia. La mediana es el valor que divide la distribución por la mitad.

Las medidas de la variabilidad indican la dispersión de los datos en la escala de medición.

El rango, también llamado recorrido, es la diferencia entre la puntuación mayor y la menor, indica el número de unidades en la escala de medición que necesitan incluir para los valores máximo y mínimo.

La desviación estándar o típica es el promedio de desviación de las puntuaciones con respecto a la media. Esta medida se expresa en las unidades originales de medición de la distribución.

La varianza, es la desviación estándar elevada al cuadrado. Es un concepto estadístico muy importante, ya que muchas de las pruebas cuantitativas se fundamentan en él.

Las puntuaciones z, son transformaciones que se pueden hacer a los valores o puntuaciones obtenidas, con el propósito de analizar su distancia respecto a la media, en unidades de desviación estándar.

La razón es la relación entre dos categorías: por ejemplo masculino y femenino.

Fase 4. Evaluar la confiabilidad y validez logradas por el instrumento de medición. La confiabilidad. Se calcula con diversos métodos como son la medida de estabilidad, calculada aplicando a los participantes la misma prueba dos veces y obtiene un coeficiente de correlación entre ambas puntuaciones. El método de formas alternativas o paralelas. A través de un coeficiente de correlación de resultados de dos pruebas supuestamente equivalentes. El método de mitades partidas, a través de un coeficiente de correlación entre las puntuaciones de las mitades del instrumento.

Fase 5. Analizar mediante pruebas estadísticas las hipótesis planteadas. Estadística inferencial; de la muestra de población. Con el propósito de ir más allá de describir las distribuciones de las variables, se generalizan los resultados obtenidos de la muestra.

Nivel de significancia. Es un nivel de la probabilidad de equivocase y que fija de mena a priori el investigador.

Fase 6. Realizar análisis adicionales. Pruebas de hipótesis, análisis paramétricos. Coeficiente de correlación de Pearson: es una prueba estadística para analizar la relación entre dos variables medidas en un nivel por intervalo o de razón. Se simboliza con r.

Regresión Lineal. Es un modelo matemático para estimar el efecto de una variable sobre otra. Sirve para predecir las puntuaciones de una variable tomando las puntuaciones de la otra variable. Entre mayor sea la correlación entre las variables, mayor capacidad de predicción.

La prueba t, es una prueba estadística para evaluar si dos grupos diferentes entre sí de manera significativa respecto a su medias. Simboliza con la letra t.

La varianza unidireccional es una prueba estadística para analizar si más de dos grupos difieren significativamente entre sí en cuanto a sus medias y varianzas.

Las tablas de contingencia además de servir para el cálculo de chií cuadrada y otros coeficientes, son útiles para describir conjuntamente dos o más variables.

Fase 7. Preparar los resultados para presentarlos.

Se recomienda una vez que se han obtenido los resultados de los análisis estadísticos las siguientes actividades:

* Revisar cada resultado.
* Organizar los resultados por variable, luego por confiabilidad y validez.
* Cotejar diferentes resultados.
* Priorizar la información más valiosa.
* Copiar y/o formatear las tablas en el programa con el cual se elaborará el reporte de la investigación.
* Comentar o describir brevemente la esencia de los análisis, valores, tablas, diagramas, gráficas.
* Volver a revisar los resultados.
* Y finalmente, elaborar el reporte de investigación.

**EJERICIOS: ESTADISTICA DESCRITIVA.**

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

n= (Z2α \* p \* q)/i2 = (2.70536704) (.5) (.5) / .04= .67834176/.04 = **16.90.**

1. 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

PREVALENCIA .5

n= (Z2α \* N \* p \* q ) / i2 (N-1) + ( Z2α  \* p \* q) =

n= (2.705367049(350,000)(.5)(.5)/ (0.0001) (349,999) + (2.70536704)(.5)(.5) =

n= 236,719.616 / (34.999) + .67634176 = 236719.616 / 35.67624176

n= **6,635.217285**

PREVALENCIA .7

n= (Z2α \* N \* p \* q ) / i2 (N-1) + ( Z2α  \* p \* q) =

n= (2.705367049)(350,000)(.7)(.3)/ (0.0001) (349,999) + (2.70536704)(.7)(.3) =

n= 198,844.477744 / (34.999) + .5681270784 = 197844.47744 / 35.5680270784

n= **5,590.5399**

1. 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 de 1.5% con un nivel de confiabilidad del 90%

N= 1,176

Se= 1.5% = 0.015

P= 90% = 0.90

S2= p (1-p) = 0.90 (1-0.90) = 0.09

α2= (se)2 = (0.015) 2 = 0.000225

n’= S2/ α2 = 0,09 / 0,000225 = 400

n= n’ / (1 + n’/N) = 400/ (1+ 400/1176) = 400 / 1+ .340136054421769 =

n = **298.47298.**

1. 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



La desviación estándar nos permite ver que elementos están dentro y fuera de los márgenes respecto de la media, es decir los elementos que están dentro de un rango, con el fin de escogerlos para determinada muestra.

1. Considera las estaturas de un padre y su hijo

* Obtén el promedio de estaturas. En ambos casos
* Elabora la gráfica 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*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | PADRES | Hijos |  |  |  |
|  |  | x | y | x \*y | x2 | y2 |
|  | 1 | 1.7 | 1.74 | 2.958 | 2.89 | 3.0276 |
|  | 2 | 1.77 | 1.78 | 3.1506 | 3.1329 | 3.1684 |
|  | 3 | 1.68 | 1.72 | 2.8896 | 2.8224 | 2.9584 |
|  | 4 | 1.75 | 1.77 | 3.0975 | 3.0625 | 3.1329 |
|  | 5 | 1.8 | 1.78 | 3.204 | 3.24 | 3.1684 |
|  | 6 | 1.75 | 1.77 | 3.0975 | 3.0625 | 3.1329 |
|  | 7 | 1.69 | 1.71 | 2.8899 | 2.8561 | 2.9241 |
|  | 8 | 1.72 | 1.76 | 3.0272 | 2.9584 | 3.0976 |
|  | 9 | 1.71 | 1.73 | 2.9583 | 2.9241 | 2.9929 |
|  | 10 | 1.73 | 1.74 | 3.0102 | 2.9929 | 3.0276 |
|  |  | 17.3 | 17.5 | 30.2828 | 29.9418 | 30.6308 |
| Promedio |  | 1.73 | 1.75 |  |  |  |

r = αxy / αx αy

= 17.3 / 10 = 1.73

17.5/ 10 = 1.75

α = = **0.0357777**

α = = **0..0240831**

**covarianza:**

= (30.2828/10) – (1,73)(1,75) = 3.02828 – 3.0275 =

**= .00078**

αxy = (.00078)/ (.035777)(.0240831) = 0.000078 / 0.00086162

**r = .9052. EXISTE UNA CORRELACIÓN POSITIVA.**