

SPSS: Estadística descriptiva

Ciclo de Investigación Cuantitativa Centro para la Excelencia Académica

Víctor E. Bonilla Rodríguez, Ph.D.
Programa de Maestría en Investigación y Evaluación Educativa (INEVA)
Departamento de Estudios Graduados
Facultad de Educación

Centro de Cómputos, Facultad de Ciencias Sociales
15 de abril de 2011, 8:30am – 12:00pm

Definición

- La estadística descriptiva incluye aquellos análisis o técnicas utilizadas para organizar, resumir, presentar y describir los datos recopilados en una investigación (King & Minium, 2008).

Tipos de análisis

Tendencia central

- Índice que mejor resume/describe el centro de un grupo de puntuaciones.

Variabilidad o dispersión

- Indica cuán dispersas se encuentran las puntuaciones en una distribución.

Gráficas

- Permiten describir la forma de la distribución.

Nota. King y Minium (2008) indican que para describir una distribución de puntuaciones debe proveer al menos un tipo de análisis de cada renglón. Otras medidas importantes al describir una distribución son el sesgo (asimetría) y la curtosis (altura).

Tipos de análisis (continuación)

Tablas de frecuencia

- Presenta la cantidad de observaciones en cada categoría o valor en el conjunto de datos.

Posición o ubicación

- Permite describir dónde se encuentra una puntuación respecto a las demás.

Análisis cubiertos en esta presentación

Tendencia central	Variabilidad o dispersión	Gráficas	Tablas de frecuencia	Posición o ubicación
Moda Mediana Promedio	Rango Rango semi-intercuartil Varianza Desviación estándar	Diagrama de sectores ("Pie") Barra Histograma Polígono de frecuencia Diagrama de dispersión	Frecuencia Proporción Porcentaje	Percentil Puntuación z

Tendencia central

Moda (M_o)

- La más simple de las medidas de tendencia central.
- Es la puntuación que ocurre con mayor frecuencia.
- Útil cuando se tienen variables cualitativas (e.g., género).
- No es estable de muestra en muestra.
- Puede existir más de una moda.

Tendencia central (continuación)

Mediana (Mdn o P₅₀)

- Puntuación bajo la cual se encuentra el 50% de las puntuaciones.
- Para calcular la mediana se deben organizar las puntuaciones.
- Es más resistente ante la presencia de algunos casos extremos por lo cual puede ser el mejor índice de tendencia central en distribuciones con un sesgo marcado.
- Es el mejor índice de tendencia central para distribuciones en las cuales no se pueden calcular los límites con exactitud.

Tendencia central (continuación)

Promedio

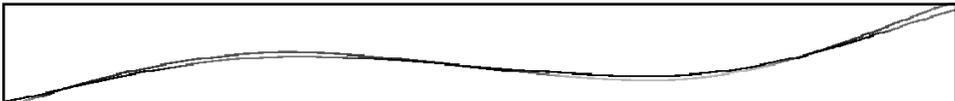
- Es la suma de las puntuaciones dividida entre el total de observaciones.
- Considera la posición exacta de cada puntuación en la distribución.
- Es más sensible a la presencia (o ausencia) de puntuaciones en los extremos de la distribución que la moda y la mediana.
- Se presta para manipulación algebraica en formas que no se pueden utilizar otros índices.
- Más estable de muestra en muestra lo cual la convierte en un indicador valioso en relación a la tendencia central de la población.



Tendencia central (continuación)

Promedio

- Si se conoce la posición relativa de la mediana y el promedio se puede determinar la dirección del sesgo sin necesitar observar la distribución completa. El promedio se mueve en la dirección en que se encuentran las puntuaciones extremas.



Variabilidad o dispersión

- Rango
 - Es el más simple de los índices de variabilidad.
 - Es la diferencia entre la puntuación mayor y la puntuación menor. Es severamente afectado por valores extremos porque sólo toma en consideración el valor más alto y el más bajo de la distribución.
 - Tiende a variar de muestra en muestra más que otros índices de variabilidad.

Variabilidad o dispersión(continuación)

Rango semi-intercuartil

- Índice más sofisticado que el rango, pues depende de puntuaciones más estables que se encuentran en medio de la distribución.
- Sólo utiliza el 50% de las puntuaciones en medio de la distribución, o sea, de P25 hasta P75. Por tal razón, es menos afectado por la presencia de valores extremos en la distribución y es útil en el caso en que se tienen distribuciones abiertas.
- Es la diferencia entre el cuartil 3 y el cuartil 1 dividido entre 2. Los cuartiles son los tres puntos que dividen la distribución en cuatro partes iguales (i.e., P25, P50, P75).

Variabilidad o dispersión(continuación)

Varianza

- Promedio de dispersión de las puntuaciones alrededor de la media aritmética.
- Utiliza todos los valores de la distribución. Por tanto es más estable de muestra en muestra.
- No se encuentra en la misma métrica de las puntuaciones originales.

Desviación estándar

- Raíz cuadrada de la varianza.
- Se encuentra en la misma métrica de las puntuaciones originales.

Gráficas

- Diagrama de sectores (“Pie chart”)
 - Útil para variables discretas.
- Barra
 - Las barras se encuentran separadas indicando que la variable que presenta es discreta.
- Histograma
 - Las barras se encuentran unidas indicando que la variable que presenta es continua.

Gráficas

Polígono de frecuencia

- Puntos unidos por líneas.
- Representa variables continuas.
- Útil para comparar distribuciones.

Diagrama de dispersión

- Cada punto representa la intersección entre dos puntuaciones de un mismo elemento (sujeto u objeto).
- Útil para examinar tendencias.

Sesgo y curtosis

Sesgo

- A la derecha (positivo) o a la izquierda (negativo) de una distribución.
- Se debe a la existencia de casos extremos a los costados de la distribución.
- El sesgo en una distribución normal es cero.

Curtosis

- Describe la altura de la distribución.
- Puede ser puntiaguda, moderada o plana.
- Depende de la concentración de las puntuaciones alrededor de la zona central de la distribución.
- El coeficiente se compara contra el valor de cero, que se asume como la curtosis de una distribución normal.

Sesgo y curtosis

Hinkle, Wiersma y Jurs (2002)

- Una distribución se puede considerar aproximadamente normal si ambos valores (sesgo y curtosis) se encuentran entre -1 y $+1$.

Tabachnick y Fidell (2007)

- Presentan en su libro una prueba de z para determinar si el sesgo y la curtosis son significativamente diferentes de cero. Puede realizar este análisis en caso de que alguno de los valores de sesgo o curtosis no se encuentre entre -1 y $+1$.

Tablas de frecuencia

- Resume y presenta la cantidad de observaciones (frecuencia), la proporción ($f \div n$) y el porcentaje ($f \div n \times 100$) que esa cantidad representa en cada categoría o valor.
- Puede presentar también la frecuencia acumulada, la proporción acumulada y el porcentaje acumulado.
- La proporción y el porcentaje se conocen como frecuencias relativas ya que para calcularlos se toma en consideración el total de casos en la distribución. Son útiles si se desea comparar distribuciones donde el número total de casos es diferente (King & Minium, 2008).

Posición o ubicación

- Percentil
 - Punto bajo el cual se encuentra un porcentaje específico de casos en una distribución.
 - El sistema de percentil es comúnmente utilizado en el ambiente educativo para informar dónde se encuentra un individuo con relación a las destrezas de un grupo. Está fundamentado en la distribución de porcentaje acumulado.

Posición o ubicación (continuación)

- Percentil
 - Se puede determinar el sesgo de la distribución al examinar los percentiles P25, P50 y P75. Estos tres percentiles dividen la distribución en 4 partes iguales. En una distribución con sesgo negativo, la distancia entre P25 y P50 es mayor que la distancia entre P50 y P75. En una distribución con sesgo positivo se observa lo contrario.

Posición o ubicación (continuación)

- Puntuación z
 - Una puntuación no se puede interpretar de forma aislada. Necesitamos un marco de referencia para determinar si la puntuación constituye un buen o un mal valor.
 - Los percentiles proveen un buen marco de referencia pues nos permiten conocer cuántas puntuaciones se encuentran por debajo y por encima de nuestra puntuación. Sin embargo, no nos permiten conocer cuán lejos se está del centro de la distribución.
 - La puntuación z es una puntuación estándar que establece cuán lejos se está del promedio en desviaciones estándar.

Posición o ubicación (continuación)

- Puntuación z
 - La puntuación z nos puede ayudar a identificar valores extremos en una distribución.
 - Según Hinkle, Wiersma y Jurs (2002), aquellos valores con una puntuación z mayor de 3.0 pueden considerarse extremos.
 - Por otro lado, Tabachnick y Fidell (2007) consideran extremos aquellos valores con puntuaciones z mayores de 3.29.

Posición o ubicación (continuación)

- Puntuación z
 - La puntuación z nos permite también comparar puntuaciones que se encuentran en dos distribuciones diferentes. Esto se debe a que al calcular la puntuación de z, los dos valores se llevan a una misma métrica.
 - Sin embargo, se pueden comparar puntuaciones de distintas distribuciones siempre y cuando las distribuciones posean aproximadamente la misma forma (King & Minium, 2008).

Referencias

Hinkle, D. E., Wiersma, W. & Jurs, S. G. (2002). *Applied statistics for the behavioral sciences (5a. ed.)*. Boston, MA: Houghton Mifflin Company.

King, B. M. & Minium, E. W. (2008). *Statistical reasoning in the behavioral sciences (5a. ed.)*. Hoboken, NJ: John Wiley & Sons, Inc.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics (5a. ed.)*. Boston, MA: Allyn and Bacon.