



ANÁLISIS CUANTITATIVO DE DATOS

AUTORES

Jyotikumarie Juggernath: African Vision Research Institute (AVRI), Durban, South Africa

Kovin S Naidoo: Brien Holden Vision Institute (BHVI), African Vision Research Institute (AVRI), Durban, South Africa

Urmilla Bob: University of KwaZulu Natal (UKZN), Durban, South Africa

Vadivelu Moodley: African Vision Research Institute (AVRI), Durban, South Africa

PAR REVISOR

James Loughman: Dublin Institute of Technology (DIT), Dublin, Ireland

ANÁLISIS CUANTITATIVO DE DATOS/ANÁLISIS ESTADÍSTICO

Los datos cuantitativos se refieren a los datos que incluyen valores numéricos, que van desde la frecuencia numérica de eventos, a la presentación compleja de datos.

En términos de gráficas y tablas, el análisis estadístico, se lleva a cabo interpretando, información numérica de una forma científica. Los datos encontrados son cuantificados y reducidos a valores numéricos. El análisis estadístico proporciona procedimientos de recolección, resumen, análisis, presentación e interpretación empírica de los datos.

PRINCIPALES TÉCNICAS ESTADÍSTICAS

- **Descriptiva:** Resume, muestra e interpreta los datos
- **Inferencia estadística:** Hace generalizaciones sobre una población basándose en los datos
- **Significancia:** observa si una tendencia o relación es significativa, o resultado del azar
- **Predicción:** los resultados de ciertos procesos pueden ser predichos hasta dentro de ciertos límites

LIMITACIONES DEL ANÁLISIS ESTADÍSTICO

El análisis estadístico no lanza juicios o da conclusiones concretas. Simplemente proporciona probabilidades matemáticas. Los investigadores lanzan juicios sobre los datos, y su interpretación es generalmente influenciada por algunos supuestos, las técnicas de muestreo (incluyendo el tamaño de la muestra) y las diferentes pruebas usadas para analizar los datos. El tipo de técnica estadística usada dependerá de los datos recolectados.

La mayoría de los programas de análisis estadístico son únicamente herramientas para ser usadas en un marco teórico particular y deben ser interpretadas críticamente. Un diseño estadístico bien formulado, incluye datos exactos como prerrequisito para un buen análisis estadístico. La estadística y los métodos cuantitativos en general son herramientas complementarias entre todas las posibilidades disponibles. Es importante que el investigador sea consciente de *la falacia ecológica*, los resultados de un análisis no deben ser extrapolados más allá de la escala y del ambiente que representan los datos originales.

- **Ejemplo:** Los hallazgos de un estudio en estudiantes de Optometría en Singapur indican una prevalencia de miopía del 10%. Este resultado no puede extrapolarse a otras poblaciones, en comunidades rurales con menor

grado de escolaridad. El hallazgo no es representativo de una población como un todo, pero puede relacionarse con estudiantes de medicina de la misma universidad

DATOS

Los datos son la forma más básica de información, en su forma más cruda y sin procesar):

Medidas/Observaciones/Casos individuales (fenómeno numérico o verbal que se está estudiando).

- Lo que se ha medido recibe el nombre de variable (ej: edad, tamaño de la casa, etc) es la característica de cada observación que se ha medido, clasificado o contado. Un grupo de datos consiste en varias variables por cada variable medida. Las variables se marcan como: x_1, x_2, x_3, \dots . Donde x_1 es la primera variable y x_n es la última variable registrada, el número total de variables observadas es 'n'.

Si el número total de observaciones es de $n = 50$, el rango de observaciones será entre 1 y 50.

Ejemplo: Se tomaron las medidas de la agudeza visual en los pacientes de una clínica, observándose:

- 20/20, 20/25, 20/40, 20/120, 20/200. n50

MEDIDAS DE LOS DATOS

- **MEDIDAS DISCRETAS**

- Unidades o enteros (Tamaño de la vivienda, género, número de árboles en un terreno)

- **MEDIDAS CONTINUAS**

- Puede ser un valor incluyendo fracciones (edad, medida de la agudeza visual o error refractivo con equivalente esférico). Nivel de detalle (cuantos decimales) depende de la precisión requerida y de que tan precisos son los instrumentos.

ESCALAS DE MEDIDA DE LOS DATOS

ESCALA NOMINAL	<ul style="list-style-type: none"> • Xategorías que no son clasificables • Las observaciones son clasificadas en categorías mutuamente excluyentes que no tienen relación inherente • E.j.: raza, grupos idiomáticos, género, áreas residenciales, preguntas de si o no
ESCALA ORDINAL	<ul style="list-style-type: none"> • Categorías que se pueden clasificar • Las observaciones se hacen en un orden específico y se clasifican según cierto criterio • Las categorías son medidas que tienen un orden jerárquico • Análisis estadístico no- paramétrico <p>E.j.: ingreso de los grupos (alto, medio, bajo), orden de cables de circuitos, ciudades clasificadas según habitabilidad</p>
ESCALA DE INTERVALOS	<ul style="list-style-type: none"> • Categorías que no son clasificables • Datos continuos, no cero fijo • A cada observación se le da un valor numérico exacto • Se reconoce la diferencia entre una y otra observación • Puede ser subjetiva a pruebas paramétricas <ul style="list-style-type: none"> ◦ E.j.: Temperatura (no existe una entidad física sin temperatura, 0°C Representa una posición relativa en la escala de Celsius)
ESCALA DE PROPORCIONES	<ul style="list-style-type: none"> • Datos continuos con un cero fijo • Cada observación tiene un valor numérico exacto y ya hay un cero verdadero en la escala • la proporción entre cualquiera de los dos valores se mantiene igual, independientemente de la escala que se use • puede ser subjetiva a pruebas paramétrica <p>E.j.: medida de la lluvia o ingresos (0 significa no lluvia o ingreso y es un cero verdadero, no una posición relativa)</p>

MEDIDAS DE TENDENCIA CENTRAL

PROMEDIO ARITMÉTICO

Las medidas de tendencia central dan una idea del promedio del grupo de observaciones , mientras que las medidas de dispersion, indican el grado de variabilidad en un grupo de medidas.

Se calcula adicionando todas las variables y dividiendo por el número total de variables.

- **Ejemplo: $93 + 97 + 65 + 66 + 41 + 56$ dividido por $6 = 418/6 = 69.7$**

La ventaja de la medida , es que es matemática como base del cálculo. La desventaja es que en diario valor muy alto en los datos, el cual lleva a un cambio brusco, o sesgado, la proporción no será un buen indicador del promedio o valor central.

Observando el ejemplo anterior, suponga que la observación 56 fue cambiada por 256, el resultado del cálculo no será. $618/6 = 103$. En estos casos se pueden usar, otras medidas de tendencia central como la moda o la mediana.

MEDIANA

- Punto Central de un grupo de datos clasificados
- Los Valores se ordenan de menor a mayor
 - **Ejemplo:** 10, 70, 20, 50, 20, 30, 40, 40, 10, 60, 70, 80, 90, 90, 90
 - Se organiza así: 10, 10, 20, 20, 30, 40, 40, 50, 60, 70, 70, 80, 90, 90, 90
 - La mediana es 50, que está en el punto central, cuyo valor es 50
- **Si el número de valores es par, la mediana será el promedio de los dos valores centrales**
 - La desventaja de la mediana es que es muy sensible a puntajes extremos. Adicionalmente una condición necesaria para usar la mediana es que el valor debe ser ordinal y la observación tiene que ser clasificada primero de forma significativa.

MODO

- Es la forma menos complicada de describir el centro de una distribución
 - Es el valor que se presenta más veces en un grupo de datos
 - Si se presentan dos medidas de modo, el grupo de datos es bimodal, si hay tres o más modos, este será multimodal
 - El modo no requiere ningún orden de categorías ni de una distancia igual entre categorías
 - Esto aplica para cualquier variable pero se usa típicamente para variables nominales
 - La desventaja del modo es que su valor en ocasiones es difícil de determinar y puede ser en ocasiones ambiguo

Ejemplo: En una distribución de categorías por edades, dos grupos de edades 20 y 24 se presentaron 50 veces, pero otras categorías de edades se pueden presentar casi tan frecuentemente, ej: 28 y 30 años (se presento 51 veces) pero no representaran el modo de ninguna forma.

MEDIDAS DE LA VARIABLE DE DISPERSION

La desviación estándar es la medida de variabilidad más usada, y se relaciona con la distancia entre las observaciones y la media. La desviación estándar es la distancia promedio de la media, y se refiere a la distribución usada frecuentemente, ej: la distribución normal, la cual es importante en la mayoría de las pruebas estadísticas.

La desviación estándar se expresa en las mismas unidades de medida que la misma variable. Por ejemplo, si la variable – peso corporal- se mide en kilogramos, la desviación estándar indica la distancia promedio en kilogramos.

Se considera adecuada solo para intervalos y variables de proporción, y es sensible a los valores típicos (distribuciones de sesgo derecha e izquierda). Es importante darse cuenta de que la desviación estándar relacionada con la distribución de la muestra, no es conocida como la desviación estándar, sino como el error estándar.

ORGANIZACION DE LOS DATOS

VARIABLES CLASIFICABLES

- Los datos se clasifican en orden ascendente o descendente
- La variable más pequeña o más grande se denomina x_1 , la segunda x_2 , la última variable es x_n .

TABLAS DE FRECUENCIA

- Se cuenta cada valor o clase que se presente

DIAGRAMAS DE DISPERSION

- Útil para observar una variable en comparación con otras e identificar tendencias entre los datos
- Se puede identificar el tipo de correlación
- Cada punto tiene dos variables conectadas
- Es importante decidir el tipo de análisis que se realizara

MEDIDAS Y MUESTREO

Requeridas en investigación cuantitativa, es el proceso de asignación de palabras, símbolos o números a personas, objetos o lugares. Involucra observar el error de la medida, por ejemplo la diferencia entre el valor real y el valor de la medición. Consideraciones importantes de la medida son:

- **Confiabilidad:** la introducción de sesgo puede emanar de diferentes fuentes y debe ser minimizada.
- **Validez:** se refiere a los asuntos relacionados con la naturaleza, el significado, o definición del concepto de variable (necesita definiciones operacionales).
- **Precisión:** se refiere al nivel de exactitud asociado a la medida (con frecuencia es determinado por calibración de la medida del instrumento, ej: el tonómetro)
- **Exactitud:** Se refiere a todo el sistema de sesgo, en las medidas del proceso, de modo que no se obtengan medidas que no sean reales.

FUENTES DE ERROR EN LA MEDIDA

Errores de cálculo: errores que se cometen en la simple ejecución de operaciones matemáticas.

Errores de medida: errores sin corregir en el registro de datos (ej: codificación incorrecta, error en el conteo,etc).

Errores de especificidad: uso de supuestos incorrectos, y/o de aplicaciones matemáticas, formulas, o ecuaciones incorrectas.

Errores de muestreo: si la población no se muestrea correctamente, el error es inducido por generalización de la población.

Otras variables no identificadas: se refiere al “ruido” en los datos, o la falta de entendimiento.