

Una guía rápida para jueces y fiscales.

Basado en la publicación especial del NIST (NIST SP) - 1500-28, adaptación técnica al español por AICEF. Traducido con permiso, cortesía del Instituto Nacional de Estándares y Tecnología (NIST); no es una traducción oficial del Gobierno de los Estados Unidos. Todos los derechos reservados, Secretaría de Comercio de los Estados Unidos. Cualquier desviación del contenido original ha sido realizada con fines pedagógicos y está claramente identificada en el texto por el carácter “**”.

DEFINICIONES

La estadística es la ciencia que se ocupa de desarrollar y estudiar métodos para recopilar, analizar, interpretar y presentar datos empíricos. Los datos se describen o se utilizan para realizar una inferencia sobre una población.

La ciencia de la estadística puede centrarse en datos provenientes de una población o de una muestra. Una población es un grupo completo sobre el cual se obtendrá información específica.



GENERALIDADES

Una población puede ser finita (p. ej., el peso de todos los estudiantes de segundo grado de la Escuela Primaria Hipotética el 1/3/2022), o puede ser muy grande, infinita o estar en constante crecimiento (p. ej., el peso de todos los estudiantes de segundo grado del mundo).

POBLACIONES VERSUS MUESTRAS

Las características del tamaño que puede tener una población, hacen poco práctico o imposible contar con valores para todos los ítems de la población. Si el conjunto de datos incluye menos que la población completa, entonces el conjunto de datos es un subconjunto de la población y se denomina muestra.

ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

La estadística descriptiva reduce un gran número de valores a unos pocos manejables o presenta los valores de una manera que facilite su visualización. Esta se calcula utilizando datos (de una muestra o de una población) y puede presentarse de forma gráfica, tabular o numérica. Ejemplos de resúmenes gráficos y tabulares incluyen histogramas, gráficos de probabilidad y distribuciones de frecuencia. Ejemplos de estadísticas descriptivas numéricas incluyen la media, la varianza y la desviación estándar. La estadística inferencial utiliza una muestra para generalizar sobre una población. La muestra debe ser representativa de la población y el proceso para seleccionar la muestra debe ser repetible.



MEDIDA DE TENDENCIA CENTRAL

Una medida de tendencia central es un método para representar una población o un conjunto de datos mediante un único punto en el centro de sus valores de datos. La media (valor promedio) es la estadística más ampliamente utilizada para describir la tendencia central de un conjunto de datos.

OTROS TÉRMINOS

- La mediana (valor central) es la siguiente estadística más utilizada para expresar la tendencia central de un conjunto de datos. Para conjuntos de datos simétricos, la media y la mediana serán iguales.
- Otras estadísticas que se utilizan para expresar la tendencia central de un conjunto de datos son el rango medio (el punto medio entre los valores mínimo y máximo) y la moda (el valor que ocurre con mayor frecuencia). Un conjunto de datos no tendrá moda si cada valor ocurre solo una vez en el conjunto de datos, o el conjunto de datos puede ser bimodal o multimodal si los valores ocurren con la misma frecuencia.

MEDIDA DE DISPERSIÓN

- A menudo existe el deseo de representar una población o una muestra mediante un único número que exprese la dispersión de los datos (variabilidad, dispersión o amplitud) alrededor del centro de la distribución (véase la Figura 1).
- El rango (la diferencia entre los valores máximo y mínimo) es la estadística más simple de calcular. Debido a la sensibilidad de la estadística de rango a los valores extremos, puede ser preferible proporcionar los valores máximo y mínimo o el rango intercuartílico, que se enfoca en el 50 % central de los valores del conjunto de datos. Un rango intercuartílico grande significa que existe una gran dispersión, o mayor variación, entre esos valores.
- La desviación estándar es la medida de dispersión más utilizada. Esta compara cada punto de datos con la media de todos los puntos de datos.
- La varianza también mide la desviación de los datos respecto a la media y es el cuadrado de la desviación estándar.

Una guía rápida para jueces y fiscales.

Basado en la publicación especial del NIST (NIST SP) - 1500-28, adaptación técnica al español por AICEF. Traducido con permiso, cortesía del Instituto Nacional de Estándares y Tecnología (NIST); no es una traducción oficial del Gobierno de los Estados Unidos. Todos los derechos reservados, Secretaría de Comercio de los Estados Unidos. Cualquier desviación del contenido original ha sido realizada con fines pedagógicos y está claramente identificada en el texto por el carácter “**”.

DISTRIBUCIONES

La frecuencia de cada valor puede mostrarse gráficamente con los valores en el eje x y la frecuencia de ocurrencia de cada valor en el eje y. Si los datos se limitan a valores distintos y separados, conocidos como una variable aleatoria discreta, los datos se mostrarán como los diferentes valores posibles dentro del intervalo (p. ej., los datos de control de calidad de una balanza con una resolución de 0,01 gramo se mostrarían en intervalos de 0,01 g cada uno). Si las entradas del conjunto de datos dentro del intervalo establecido pueden tomar cualquier valor con un número infinito de decimales posibles, es decir, una variable aleatoria continua, entonces la distribución, o curva de densidad, se vuelve continua (p. ej., datos de control de calidad para un método cuantitativo).



LA DISTRIBUCIÓN NORMAL

Muchas personas están familiarizadas con la distribución normal, también denominada distribución gaussiana o curva de campana. Los datos provenientes de muchos fenómenos naturales, así como de procesos de medición, pueden aproximarse mediante una distribución normal. Esta es una distribución simétrica en la que la mayoría de los valores de los datos se concentran alrededor de la media, y es menos probable que se observen valores de datos en los márgenes externos de la distribución. En una distribución normal, la media, la mediana y la moda tienen el mismo valor. Si los datos de un proceso se aproximan a una distribución normal con una media poblacional μ y una desviación estándar poblacional σ , entonces la probabilidad de que el siguiente valor de ese mismo proceso se encuentre dentro de una desviación estándar de la media es aproximadamente del 68 %. La probabilidad de que el siguiente valor se encuentre dentro de dos desviaciones estándar de la media es aproximadamente del 95 %, y la probabilidad de que el siguiente valor se encuentre dentro de tres desviaciones estándar de la media es aproximadamente del 99 % (véase la Figura 1). Si los datos no están distribuidos normalmente, pueden hacerse afirmaciones respecto al “siguiente valor”, pero los porcentajes pueden diferir.

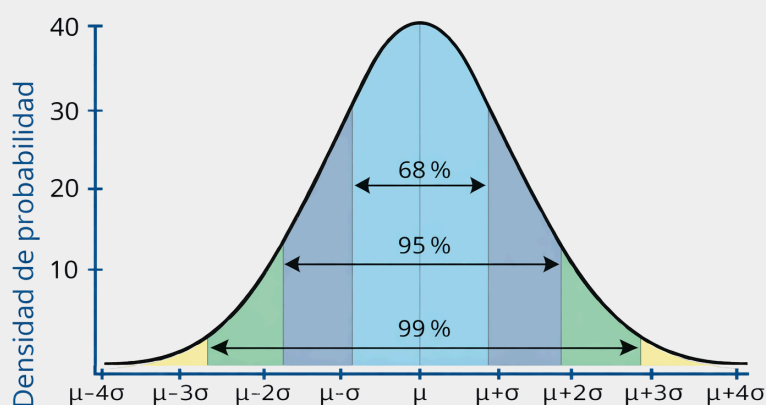
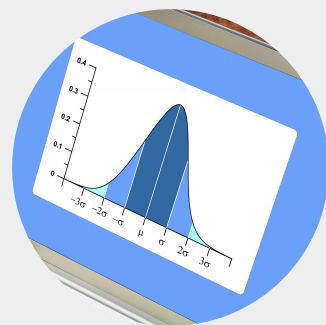


Figura 1. Distribución normal (μ = media, σ = desviación estándar)

PUNTOS CLAVE

- Una muestra de uno o más ítems es un subconjunto de una población que contiene todos los ítems bajo consideración.
- La media, la mediana, la moda y el rango medio son medidas de tendencia central.
- La desviación estándar, la varianza, el rango y el rango intercuartílico son medidas de dispersión.
- Se requiere tanto un descriptor de tendencia central como uno de dispersión para describir adecuadamente un conjunto de datos.