



Estadística: Introducción y Normalidad

Dr. Joel De La Cruz Oller
2-2025-2026

Importante



Al final de esta presentación hay un enlace a una prueba corta de *Microsoft Forms*. Debe obtener al menos un 80% en la prueba para aprobarla.

Le exhortamos a que estudie el módulo con detenimiento antes de tomar la prueba.

Objetivos



Después de estudiar este módulo, los estudiantes podrán:

- Comprender el propósito y alcance de la estadística inferencial.
- Conocer el concepto de poder estadístico y su interpretación.
- Definir y caracterizar la distribución normal

Objetivos



Después de estudiar este módulo, los estudiantes podrán:

- Entender y medir la asimetría (skewness) y la curtosis para evaluar la normalidad.
- Conocer las limitaciones de medidas y pruebas según tamaño muestral.
- Aprender las pruebas de normalidad Shapiro–Wilk y Kolmogorov–Smirnov (KS): propósito, estadísticos e interpretación.

Introducción

- La estadística inferencial es un método que contribuye a apreciar y evaluar parámetros de la población usando muestras representativas de población.
- Es con esto que podemos enunciar hipótesis y decir cuál se acepta y cuál se rechaza basada en los datos cuantitativos.
- Los análisis estadísticos utilizados acatarán los tipos de variables en los datos que componen la muestra en general.

Introducción

- La estadística inferencial la vamos a utilizar cuando hacer un cálculo o análisis de la población total sería casi imposible o demasiado costo efectivo.
- Pero su propósito principal es ayudar a contribuir a poder tomar decisiones basadas en resultados cuantitativos que permiten evaluar una hipótesis.

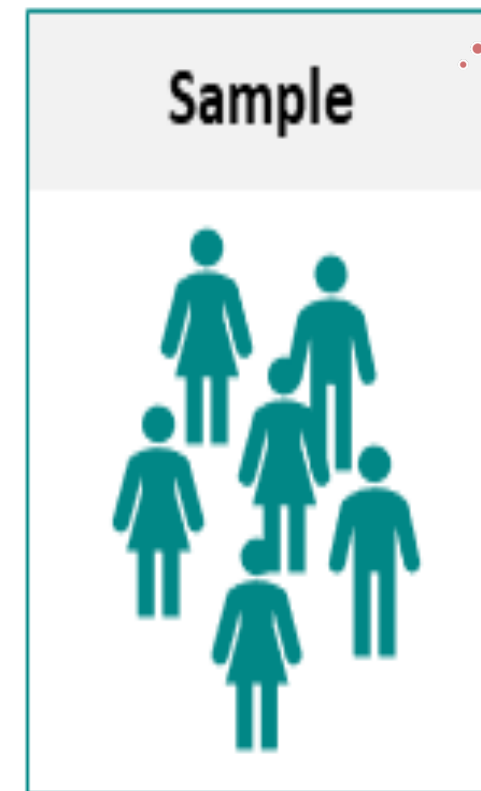
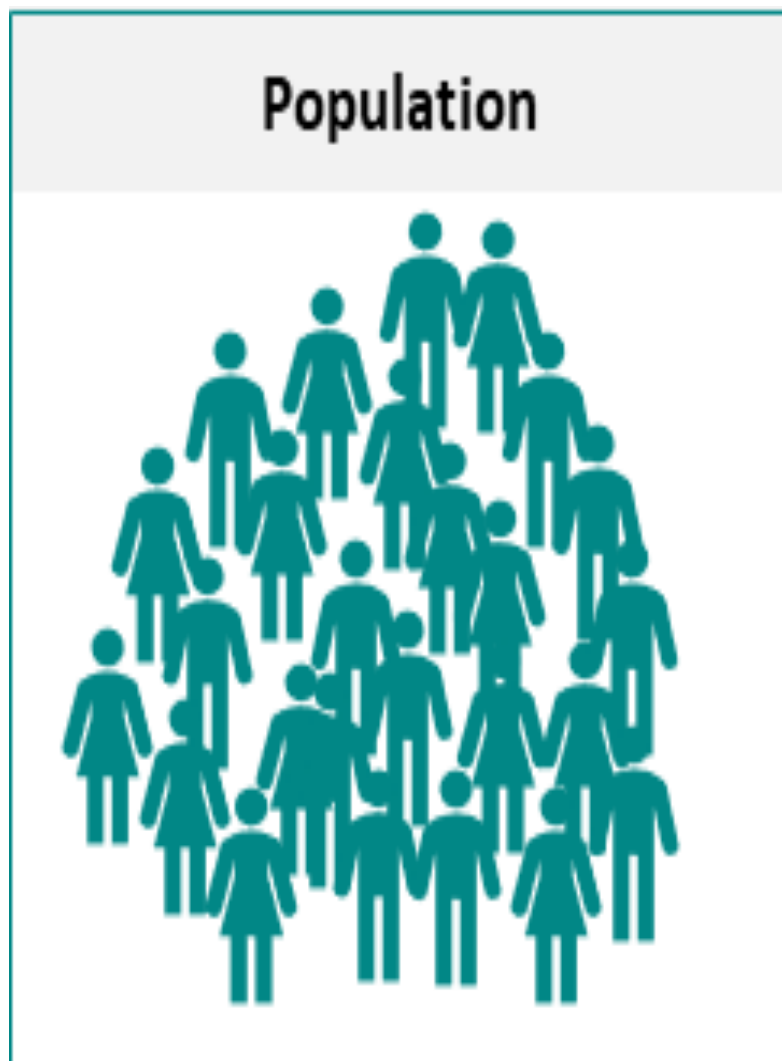
Introducción

- Las estadísticas inferenciales aplican análisis de la muestra de participantes o variables de la investigación para poder medir los grupos que se están comparando.
- Es de esta manera que se puede presentar resultados que generalicen sobre la población total o objetivos.
- Tener presente que hay varios tipos de estadísticas inferenciales y cada uno es adecuado para un esquema de examen determinado y rasgos de la muestra.

Introducción



- En resumen, podemos concluir que tomando una muestra de una forma correcta las conclusiones que lleguemos sobre los resultados se llamarán estadísticas inferenciales.



Descriptive statistics

Poder estadístico

- El poder estadístico es la probabilidad de que la prueba rechace la hipótesis nula cuando la nula es falsa.
- En otras palabras, es la probabilidad de no cometer una falsa decisión negativa.

LA DISTRIBUCIÓN NORMAL



- La distribución normal es una distribución unimodal y simétrica especial.
- Se caracteriza por su media, μ y desviación estándar.
- Casi todos los datos se encuentran dentro de tres desviaciones estándar de la media; aproximadamente el 67% de las observaciones se encuentran dentro de una desviación estándar alrededor de la media, el 95% se encuentran dentro de dos desviaciones estándar y el 99,7% se encuentran dentro de tres desviaciones estándar de la media.

LA DISTRIBUCIÓN NORMAL

- Una medida, llamada asimetría, refleja una desviación asimétrica de una distribución normal como se ve en una distribución sesgada.
- Un valor cercano a cero indica una distribución normal.
- En una distribución unimodal, un valor de asimetría positiva indica una distribución sesgada a la derecha; de manera similar, la asimetría negativa indica una distribución sesgada a la izquierda.

LA DISTRIBUCIÓN NORMAL



- Otra medida, la curtosis, también es útil para evaluar la normalidad de una distribución.
- La curtosis cuantifica el pico (ancho del pico) y el peso de las colas en relación con una distribución normal.
- Los valores cercanos a tres son indicativos de normalidad.

LA DISTRIBUCIÓN NORMAL



- En una distribución unimodal y simétrica, la curtosis positiva refleja un aumento del pico y colas pesadas, mientras que la curtosis negativa sugiere planitud y colas más claras.
- Tanto la asimetría como la curtosis son sensibles al tamaño de la muestra.
- En muestras pequeñas, los valores de curtosis y / o asimetría pueden sugerir que la distribución no es normal, aunque la distribución en la población de la que se extrajo la muestra puede ser bastante normal.

LA DISTRIBUCIÓN NORMAL

- También se pueden utilizar pruebas estadísticas específicas, como las de Shapiro-Wilk y Kolmogorov-Smirnov, para evaluar la normalidad de una distribución;
- sin embargo, al igual que con cualquier prueba estadística, dependen del tamaño de la muestra y pueden tener poca potencia para muestras pequeñas.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)



- ¿Qué son y para qué sirven?
- Ambos son pruebas de bondad de ajuste usadas para evaluar si una muestra de datos sigue una distribución específica (casi siempre se usa para ver si los datos son aproximadamente normales).
- Resultado típico: una estadística de prueba y un valor p . Si p es pequeño ($p < \alpha$, por ejemplo 0.05) rechazamos la hipótesis de que los datos provienen de la distribución propuesta (p. ej. normal).

Shapiro–Wilk y Kolmogorov–Smirnov (KS)



- Shapiro–Wilk
- Propósito: probar normalidad en una muestra.
- Ventajas: buena potencia (es decir, detecta con más facilidad desviaciones de la normalidad), especialmente con muestras pequeñas a moderadas.
- Supuestos: datos independientes; ideal para n entre ~ 3 y 2000 (muy usado cuando $n \leq 50\text{--}200$).

Shapiro–Wilk y Kolmogorov–Smirnov (KS)

- Shapiro–Wilk
- Estadístico: W (va de 0 a 1). Valores cercanos a 1 indican más concordancia con la normalidad.
- Interpretación: $p > 0.05 \rightarrow$ no hay evidencia suficiente para decir que los datos no son normales; $p \leq 0.05 \rightarrow$ los datos difieren significativamente de la normal.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)



- Shapiro–Wilk
- Ejemplo:
- Muestra: [12, 11, 13, 12, 14, 13, 12] ($n = 7$).
- Aplicando Shapiro–Wilk obtenemos $W \approx 0.98$, $p \approx 0.95 \rightarrow$ no rechazamos normalidad (los datos son compatibles con una distribución normal).
- Si la muestra fuera [2, 50, 3, 100, 4, 60, 3], W sería pequeño y p muy bajo $\rightarrow$ rechazamos la normalidad.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)

- Kolmogorov–Smirnov (KS)
- Propósito: comparar la distribución empírica de la muestra con una distribución teórica (por ejemplo, normal, exponencial) o comparar dos muestras (versión de dos muestras).
- Versión más usada para normalidad: Lilliefors (corrección del KS cuando parámetros de la normalidad se estiman de la muestra). El KS “clásico” asume que la distribución teórica es completamente especificada (media y desviación conocidas).

Shapiro–Wilk y Kolmogorov–Smirnov (KS)

- Kolmogorov–Smirnov (KS)
- Estadístico: D — la máxima diferencia absoluta entre la función de distribución empírica (EDF) y la función de distribución teórica (CDF).
- Interpretación: p pequeña ($p \leq 0.05$) $\rightarrow$ rechazamos que la muestra provenga de la distribución teórica.
- Ventajas/desventajas:
- KS es general (puede usarse para muchas distribuciones) y visualmente intuitivo.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)

- Kolmogorov–Smirnov (KS)
- Menos sensible que Shapiro–Wilk para detectar ciertas desviaciones de la normalidad, especialmente en las colas.
- Requiere ajuste (Lilliefors) si los parámetros teóricos no son conocidos.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)



- Kolmogorov–Smirnov (KS)
- Ejemplo simple:
- Muestra: [1.2, 0.5, 0.8, 1.1, 0.9] y queremos saber si es $N(\mu=1, \sigma=0.2)$ (completamente especificada).
- Calculamos EDF y CDF normal con esos parámetros; D = máximo $|EDF - CDF|$. Si D es pequeño y $p > 0.05$ no rechazamos.
- Si en cambio comparamos la muestra con una normal usando media y desviación estimadas, debemos usar Lilliefors; KS clásico podría dar p engañosamente alto.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)

- ¿Cuándo usar cada uno?
- Para probar normalidad en muestras pequeñas o moderadas: preferir Shapiro–Wilk.
- Para comparar con otra distribución o para comparar dos muestras: KS (o su versión de dos muestras).
- Si estima la media/desviación desde la muestra, use pruebas con corrección (Lilliefors) o Shapiro–Wilk para normalidad.

Shapiro–Wilk y Kolmogorov–Smirnov (KS)



Recomendaciones prácticas:

- Siempre mirar primero un gráfico: histograma y Q-Q plot (gráfico cuantiles vs cuantiles). Las pruebas dan p, pero el gráfico ayuda a entender el tipo de desviación (asimetría, colas pesadas, outliers).
- No depender solo del p: con muestras muy grandes, pequeñas desviaciones producen p pequeñas; con muestras pequeñas, la prueba puede no detectar desviaciones importantes. Combine prueba, gráfico y conocimiento del contexto.
- Si los datos no son normales y va a usar métodos que requieren normalidad, considere transformaciones (log, sqrt) o métodos no paramétricos (por ejemplo, pruebas de Mann–Whitney en lugar de t-test).

LA DISTRIBUCIÓN NORMAL



- Diferencia entre sesgo (Bias) y factor de confusión (Cofounder)
- El sesgo (bias) es un error sistemático que conduce a una medición falsa de la variable dependiente y por lo tanto a una conclusión errónea.
- Un factor de confusión es una variable que se relaciona con la variable dependiente e independiente, no se encuentra en la vía causal (que se llamaría una variable intermedia) y conduce al supuesto de una relación causal entre la variable dependiente e independiente que, sin embargo, se sobreestima o induce a error.

REFERENCIAS



- Emmer, M. (2006). Mathematics and Cinema. En. M Emmer (Ed.), *The Visual Mind II* (pp. 569-601). MIT Press.
- Grima, C., & Fernández Borja, E. (2017). *Las matemáticas vigilan tu salud: Modelos sobre epidemias y vacunas*. eISBN 978-84-946669-9-5 (ebook) . Next Door Publishers.
- Mackey, M. C., & Santillán, M. (2005). Mathematics, Biology and Physics, Interaction and Interdependence. *Notices of the American Mathematical Society*, 52(8), 834-840.
- Martin, E. (2015). *Basic College Mathematics*. Pearson Education.

REFERENCIAS



Paulos, J. A. (2001). *Innumeracy: Mathematical Illiteracy and Its Consequences*. McMillan.

Paulos, J. A. (1999). *Once Upon a Number: The Hidden Mathematical Logic of Stories*. Basic Books.

Paulos, J. A. (2013). *A Mathematician Reads the Newspaper*. Basic Books.

Para finalizar, haga clic sobre el enlace para tomar la prueba y enviar el informe a su profesor(a):

[Estadística Introducción y Normalidad – Fill out form](#)

¡Gracias por utilizar los recursos del  !

Conozca más sobre nuestros servicios:

<http://generales.uprrp.edu/competencias-linguisticas/>

Para revisiones: <https://forms.office.com/r/f2pYYSQnHp>