DISTRIBUCIONES DE MUESTREO

Distribuciones muestrales

Una distribución muestral es una distribución de posibles valores que una estadística puede tomar cuando se extraen muestras repetidamente de una población.

La idea principal detrás de las distribuciones muestrales es que nos permiten hacer inferencias sobre la población utilizando información de las muestras. Por ejemplo, si queremos estimar la media de una población, podemos tomar varias muestras de la población y calcular la media muestral para cada una de ellas. Las distribuciones muestrales nos proporcionan una idea de cómo varían estas medias muestrales al repetir este proceso de muestreo.

Distribución muestral de la media

Una de las distribuciones muestrales más conocidas es la distribución muestral de la media. Si tomamos muestras de tamaño n de una población con media μ y desviación estándar σ, y calculamos la media muestral para cada muestra, estas medias muestrales seguirán una distribución normal cuando el tamaño de la muestra sea lo suficientemente grande (gracias al Teorema del Límite Central). La media de la distribución muestral de la media será igual a la media de la población (μ), y la desviación estándar de la distribución muestral de la media, conocida como error estándar, será igual a $\sigma/\sqrt{n}$.

Estimadores Puntuales

- Un estimador puntual de un parámetro poblacional θ es una cantidad que se puede calcular a partir de una muestra y se puede usar como una aproximación al valor de θ.
- Este estimador puntual es una variable aleatoria dado que su valor cambia al cambiar la muestra.
- Un estimador puntual se espera que sea insesgado, es decir, $\mathrm{E}(\hat{\theta}) = \theta$

Estimadores más comunes
x es un estimador puntual insesgado de μ
s^2 es un estimador puntual de insesgado σ^2
$p = {}^{x}/_{n}$ es un estimador puntual insesgado de p en la binomial

Todos estos estimadores cumplen también con la propiedad de mínima varianza, es decir, son los estimadores puntuales que como variable aleatoria tienen varianza mínima entre todos los estimadores puntuales del parámetro.

Error estándar
El error estándar de un estimados $\hat{\theta}$ es su desviación estándar $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. Si por si mismo el error tiene que ver con parámetros desconocidos cuyos valores se pueden estimar, la substitución de estas estimaciones en $\sigma_{\hat{\theta}}$, produce el error estándar estimado. El error estándar estimado se puede denotar por $\widehat{\sigma_{\hat{\theta}}}$ o bien por $s_{\hat{\theta}}$.

Distribuciones Muestrales
Llamamos distribución muestral a la distribución de probabilidad de un estimador puntual. Las distribuciones muestrales que vamos a ver son las distribuciones muestrales de los estimadores puntuales $\bar{x}, s^2$ y $\hat{p}$

Teorema del Límite Central
Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño $n \geq 30$, extraída de una población con media μ y varianza σ^2, entonces x tiene distribución normal con media μ y varianza σ^2/n, es decir $X \sim N(\mu, \sigma^2/n)$.

Distribución muestral de proporción
Sea X el número de elementos de una muestra de tamaño n que poseen una caracteristica de interés, $\hat{p} = x/n$ sigue una distribución normal con medida p, la proporción de elementos de la población que poseen la característica de interés y varianza $p(1-p)/n$, es decir, $\hat{p} \sim N(p, p(1-p)/n$.

Distribución Ji(Chi) cuadrada
Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria extraída de una población normal con media μ y varianza σ^2, entonces la variable aleatoria $(n-1)s^2/\sigma^2$ tiene una distribución x^2 con $n-1$ grados de libertad.

Distribución F de Fisher
Si ${s_1}^2$ es la varianza de una muestra aleatoria de tamaño n_1 de una población normal con varianza ${\sigma_1}^2$ y ${s_2}^2$ es la varianza de una muestra alearoria de tamaño n_2 de otra población normal con varianza ${\sigma_2}^2$, entonces $F = \frac{{s_1}^2/{\sigma_1}^2}{{s_2}^2/{\sigma_2}^2}$ tiene una distribución de F de Fisher con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

Distribuciones muestrales basadas en la media
Las distribuciones muestrales basadas en la media son particularmente importantes en estadística inferencial. Estas distribuciones se centran en la estadística de la media muestral, que es una estimación de la media de la población.
Cuando tomamos muestras repetidas de una población y calculamos la media para cada muestra, estas medias muestrales seguirán una distribución específica. En general, si la población de interés tiene una distribución normal o si el tamaño de la muestra es lo suficientemente grande (según el Teorema del Límite Central), la distribución muestral de la media será aproximadamente una distribución normal.

La distribución muestral de la media tiene varias propiedades importantes:

- Media: La media de la distribución muestral de la media será igual a la media de la población. En otras palabras, si la media de la población es μ, entonces la media de la distribución muestral de la media será μ también.
- Variabilidad: La variabilidad de la distribución muestral de la media está determinada por el tamaño de la muestra (n) y la variabilidad de la población (representada por la desviación estándar σ). La desviación estándar de la distribución muestral de la media, conocida como error estándar, se calcula como

$\sigma/\sqrt{n}$. A medida que el tamaño de la muestra aumenta, la variabilidad de la distribución muestral disminuye.

- Forma de la distribución: Cuando la población es normal o el tamaño de la muestra es grande, la distribución muestral de la media será una distribución normal. Si la población no es normal y el tamaño de la muestra es pequeño, la distribución muestral de la media puede aproximarse a una distribución normal mediante el uso del teorema del límite central.

Las distribuciones muestrales basadas en la media son ampliamente utilizadas para realizar inferencias sobre la población.

Por ejemplo, se pueden calcular intervalos de confianza para estimar la media poblacional o realizar pruebas de hipótesis para tomar decisiones sobre la media poblacional. Estas técnicas se basan en el conocimiento de las propiedades de la distribución muestral de la media.

Distribuciones muestrales basadas en la varianza

Las distribuciones muestrales basadas en la varianza son utilizadas en estadística inferencial cuando queremos hacer inferencias sobre la varianza de una población.

Cuando tomamos muestras repetidas de una población y calculamos la varianza muestral para cada muestra, estas varianzas muestrales seguirán una distribución específica. En particular, si la población sigue una distribución normal, la distribución muestral de la varianza se basa en la distribución chi-cuadrado (χ^2).

La distribución muestral de la varianza tiene las siguientes características clave:

- Grados de libertad: La forma de la distribución chi-cuadrado en la distribución muestral de la varianza está determinada por los grados de libertad. Los grados de libertad se calculan como el tamaño de la muestra menos 1 (n-1). Cuanto mayor sea el número de grados de libertad, más se asemejará la distribución muestral de la varianza a una distribución normal.

- Media: La media de la distribución muestral de la varianza es igual a la varianza poblacional. Si la varianza de la población es σ^2, entonces la media de la distribución muestral de la varianza será σ^2.
- Variabilidad: La variabilidad de la distribución muestral de la varianza está determinada por la varianza poblacional y el tamaño de la muestra. En general, a medida que el tamaño de la muestra aumenta, la variabilidad de la distribución muestral de la varianza disminuye.

Las distribuciones muestrales basadas en la varianza son útiles en diversas aplicaciones estadísticas. Por ejemplo, se utilizan en la prueba de hipótesis para comparar las varianzas de dos poblaciones, en la construcción de intervalos de confianza para estimar la varianza poblacional, y en el análisis de la varianza (ANOVA) para comparar las medias de varias poblaciones.

Es importante destacar que las distribuciones muestrales basadas en la varianza asumen que la población de interés sigue una distribución normal. En casos en los que la distribución de la población no es normal o el tamaño de la muestra es pequeño, se deben considerar otros métodos o técnicas estadísticas.

En resumen, las distribuciones muestrales basadas en la varianza nos permiten hacer inferencias sobre la varianza de una población utilizando información de las muestras. La distribución chi-cuadrado es clave en este contexto, y los grados de libertad y la variabilidad son factores importantes a considerar al trabajar con estas distribuciones muestrales.

Las distribuciones muestrales también se utilizan para calcular intervalos de confianza y realizar pruebas de hipótesis. Estos procedimientos nos permiten tomar decisiones sobre la población en función de la información contenida en las muestras.

En resumen, las distribuciones muestrales son una herramienta fundamental en estadística inferencial, ya que nos permiten hacer inferencias sobre la población utilizando información de las muestras.

Referencias:

Agresti, A., & Franklin, C. (2018). Estadística inferencial (2a ed.). Pearson Educación.
Hogg, R. V., McKean, J. W., & Craig, A. T. (2018). Introducción a la estadística matemática (8a ed.). Pearson Educación.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (5th ed.). Wiley.

Casella, G., & Berger, R. L. (2002). Statistical inference (2nd ed.). Duxbury Press.

Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric statistical inference (5th ed.). CRC Press.

Lehmann, E. L., & Romano, J. P. (2005). Testing statistical hypotheses (3rd ed.). Springer.

Wasserman, L. (2004). All of statistics: A concise course in statistical inference. Springer.