Muestreo a partir de una distribución normal

Estadística

Santiago Alférez

Febrero de 2020

Contenidos

Propiedades de la media y varianza de una muestra

Teorema

Sea $X_1, \cdots, X_n$ una muestra aleatoria de una distribución $n(\mu, \sigma^2)$, y sea $\bar{X} = (1/n) \sum_{i=1}^{n} X_i$ y $S^2 = [1/(n-1)] \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$. Entonces

1. $\bar{X}$ y S^2 son independientes.
2. $\bar{X}$ tiene una distribución $n(\mu, \sigma^2/n)$.
3. $(n-1)S^2/\sigma^2$ tiene una distribución cuadrada chi con $n-1$ grados de libertad.

Propiedades de la media y varianza de una muestra

Lema: hechos acerca de la distribución chi cuadrada

Se usa χ_p^2 para denotar una variable aleatoria que sigue una distribución chi-cuadrada con p grados de libertad.

- Si Z es una variable aleatoria que sigue una distribución $n(0,1)$, entonces $Z^2 \sim \chi_1^2$.
- Si $X_1, \cdots, X_n$ son independientes y $X_i \sim \chi_{p_i}^2$, entonces $X_1 + \cdots + X_n$ tiene una distribución chi-cuadrada con $p_1 + \cdots + p_n$ grados de libertad.

Notas

1. Si $X_1, \cdots, X_n$ es una muestra aleatoria de $n(0,1)$, entonces $X_1^2 + \cdots + X_n^2$ tiene una distribución chi-cuadrada con n grados de libertad.
2. Si $X_1, \cdots, X_n$ es una muestra aleatoria de $n(\mu, \sigma^2)$, entonces $\frac{(X_1-\mu)^2}{\sigma^2} + \cdots + \frac{(X_n-\mu)^2}{\sigma^2}$ tiene distribución chi-cuadrada con n grados de libertad.

Propiedades de la media y varianza de una muestra

Ejemplo

Una máquina embotelladora se puede regular para que descargue un promedio de μ onzas por botella. Se ha observado que la cantidad de llenado dispensado por la máquina tiene una variación de $\sigma^2 = 2$. Una muestra de $n = 40$ botellas llenas se seleccionan aleatoriamente de la salida de la máquina en un día determinado y se mide la cantidad de onzas (llenadas) para cada una. Encuentre la probabilidad de que la media de la muestra esté dentro de 0.5 de la media verdadera μ para esta configuración en particular.

Propiedades de la media y varianza de una muestra

Lema

Sea $X_j \sim n\left(\mu_j, \sigma_j^2\right), j = 1, \cdots, n$, independientes. Para unas constantes a_{ij} y b_{rj} $(j = 1, \cdots, n;\ i = 1, \cdots, k;\ r = 1, \cdots, m)$, donde $k + m \leq n$, se define

$$U_i = \sum_{j=1}^{n} a_{ij} X_j;\ i = 1, \cdots, k, \text{ y } V_r = \sum_{j=1}^{n} b_{rj} X_j;\ r = 1, \cdots, m$$

1. U_i, y V_r son independientes si y sólo si $\mathrm{Cov}\left(U_i, V_r\right) = 0$. Además, $\mathrm{Cov}\left(U_i, V_r\right) = \sum_{j=1}^{n} a_{ij} b_{rj} \sigma_j^2$
2. Los vectores $(U_1, \cdots, U_k)$ y $(V_1, \cdots, V_m)$ son independientes si y sólo si U_i es independiente de V_r para todas las parejas $(i, r)(i = 1, \cdots, k;\ r = 1, \cdots, m)$.

Propiedades de la media y varianza de una muestra

Notas sobre el lema anterior

El resultado del Lema (1) implica que para que dos variables aleatorias normales sean independientes, solo necesitamos mostrar que su covarianza es 0. Supongamos que $X_1, \cdots, X_n$ sea una variable aleatoria de una población $n\left(\mu, \sigma^2\right)$. Deseamos saber la distribución de

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{S^2/\sigma^2}} = \frac{U}{\sqrt{V/(n-1)}}$$

donde $U \sim n(0,1)$ y $V \sim \chi^2_{n-1}$ y, U y V son independientes.

Distribución t de student

Definición

Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $n(\mu, \sigma^2)$. La cantidad $\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ tiene una distribución t de Student con $n-1$ grados de libertad. De manera equivalente, una variable aleatoria $T \sim t_p$ si tiene un pdf dado por

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty$$

Distribución t de student

Notas sobre la definición anterior

1. Si $p = 1$, obtenemos una distribución Cauchy(0,1). En la configuración de muestra aleatoria, esto sucede cuando $n = 2$.
2. Si $T_p \sim t_p$, entonces solo existen momentos $p - 1$. En particular,

$$E(T_p) = 0, \;\; p > 1 \text{ y } \operatorname{Var}(T_p) = \frac{p}{p-2} \;\; \text{si } p > 2$$

3. No existen función generadora de momento para la distribución de Student t.
4. En general, si $U \sim n(0,1), V \sim \chi^2_p$ y U y V son independientes, entonces $T_p = U/\sqrt{V/p} \sim t_p$.

Propiedades de la media y varianza de una muestra

Ejemplo (continuando con el anterior

Si se desea encontrar el tamaño de la muestra n de modo que la probabilidad de que la media de la muestra sea menor que 0.5 de la media real μ, sea de al menos 0.95.

Distribución F de Snedecor

Definición

- Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $n(\mu_x, \sigma_x^2)$ y $Y_1, \cdots, Y_m$ sea una muestra aleatoria de una población independiente $n(\mu_y, \sigma_y^2)$.
- La variable aleatoria $F = \frac{S_X^2 / S_\gamma^2}{\sigma_X^2 / \sigma_Y^2}$ tiene una distribución F con $n-1$(grados de libertad del numerador) y $m-1$ (grados de libertad del denominador.)
- De manera equivalente, la variable aleatoria F tiene una distribución F con p y q grados de libertad si tiene una pdf

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1+(p/q)x]^{(p+q)/2}}, 0 < x < \infty$$

Distribución F de Snedecor

Notas sobre la definición anterior

1. En general, si $U \sim \chi_p^2$ y $V \sim \chi_q^2$ y U son independientes, entonces $F_{p,q} = \frac{U/p}{V/q}$ tiene una distribución F con p y q grados de libertad.
2. La distribución F se usa comúnmente en los métodos de análisis de varianza.
3. Si $X \sim F_{p,q}$ entonces $1/X \sim F_{q,p}$.
4. Si $X \sim t_q$, entonces $X^2 \sim F_{1,q}$.