Estadística Inferencial

Jackson M'coy Romero Plasencia

Universidad Nacional de San Cristóbal de Huamanga

Departamento Académico de Matemática y Física

Ayacucho 2020

Muestreo en poblaciones normales

Teorema: Sea $X \sim N(\mu, \sigma^2)$ y sea $X_1, X_2, ..., X_n$ una muestra aleatoria de tamaño n, entonces: $\overline{X} \sim N(\mu, \sigma^2/n)$

Sea $\overline{X} = f(X_1, X_2, ..., X_n) = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ como X_i son v.a iid $\forall_i$

Como $\overline{X}$ es combinación lineal de v.a. normales $\longrightarrow$ es normal

$\overline{X} \sim N(\mu, \sigma^2/n)$

Distribuciones derivadas de la distribución normal

Teorema: La suma de los cuadradros de v.a. $N(0,1)$ se distribuye como una $\chi^2_{(k)}$

$$Z_i \longrightarrow N(0,1) \quad i = \overline{1,k}$$

$$\sum_{i=1}^{k} Z_i^2 \longrightarrow \chi^2_{(k)}$$

$$X_i \longrightarrow N(\mu, \sigma^2) \quad i = \overline{1,k}$$

$$\sum_{i=1}^{k} \left(\frac{X_i - \mu}{\sigma} \right)^2 \longrightarrow \chi^2_{(k)}$$

Distribución de la Varianza Muestral

Sea una muestra aleatoria $X_1, X_2, ..., X_n$ con $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y

$$s^2 = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\begin{aligned}
(n-1)s^2 &= \sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= \sum_{i=1}^{n}[X_i + \mu - \mu - \overline{X}]^2 \\
&= \sum_{i=1}^{n}[(X_i - \mu) - (\overline{X} - \mu)]^2 \\
&= \sum_{i=1}^{n}(X_i - u)^2 - 2\sum_{i=1}^{n}(X_i - \mu)(\overline{X} - \mu) + \sum_{i=1}^{n}(\overline{X} - \mu)^2
\end{aligned}$$

$$(n-1)s^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - 2\underbrace{\sum_{i=1}^{n}(X_i-\mu)(\overline{X}-\mu)}_{n(\overline{X}-\mu)^2} + \sum_{i=1}^{n}(\overline{X}-\mu)^2$$

$$(n-1)s^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2$$

$$\left(\frac{n-1}{\sigma^2}\right)s^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\chi^2_{(n-1)} \qquad\qquad \chi^2_{(n)} \qquad\qquad \chi^2_{(1)}$$

$$\frac{(n-1)}{\sigma^2}s^2 \longrightarrow \chi^2_{(n-1)}$$

Media y Varianza de la Varianza muestral

Cuasivarianza $n \geq 30$

$$\mathbb{E}(s^2) = \mathbb{E}\left(\frac{\sigma^2\,\chi^2_{(n-1)}}{n-1}\right) = \sigma^2$$

$$\mathbb{V}(s^2) = \mathbb{V}\left(\frac{\sigma^2\,\chi^2_{(n-1)}}{n-1}\right) = 2\sigma^4/(n-1)$$

Varianza $n < 30$

$$\mathbb{E}(s^2) = \mathbb{E}\left(\frac{\sigma^2\,\chi^2_{(n-1)}}{n}\right) = \left(\frac{n-1}{n}\right)\sigma^2$$

$$\mathbb{V}(s^2) = \mathbb{V}\left(\frac{\sigma^2\,\chi^2_{(n-1)}}{n}\right) = \left(\frac{2(n-1)}{n^2}\right)\sigma^4$$

Distribución t- student

Teorema: Si Y y Z son variables aleatorias, tal que: $Y \longrightarrow \chi^2_{(k)}$ y $Z \longrightarrow N(0,1)$, entonces la distribución de:

$$t = \frac{Z}{\sqrt{Y/K}} \quad \text{está dada}$$

$$f_T(t) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\ \Gamma(k/2)} \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}$$

se llama distribución t-student con k grados de libertad.

$$\mathbb{E}(t) = 0 \qquad \mathbb{V}(t) = \frac{k}{k-2}$$

Teorema: Si $\overline{X}$ y s^2 son la media y la varianza de una muestra aleatoria de tamaño n de una población normal con media μ y varianza σ^2, entonces:

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \longrightarrow t_{(n-1)}$$

$$t = \frac{Z}{\sqrt{Y/K}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\frac{\overline{X}-\mu}{\cancel{\sigma}/\sqrt{n}}}{\sqrt{\frac{\cancel{(n-1)}s^2}{\cancel{\sigma^2}}/\cancel{(n-1)}}}$$

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \longrightarrow t_{(n-1)}$$

La población debe ser normal, la varianza desconocida y $n < 30$

Distribuciones Muestrales en Poblaciones no Normales

Definición: Definamos una variable aleatoria X con 2 posibles valores identificados con 0 y 1, con probabilidades asociadas $1-p$ y p respectivamente.

$$f_X(x,p) = f_X(x) = p^x\,(1-p)^{1-x} \qquad x = 0,1$$

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria tomada de esta población

$$\begin{aligned} f_{X_1,...,X_n}(x_1,...,x_n) &= \prod_{i=1}^{n} f_{X_i}(x_i) \\ &= p^{\sum X_i}\,(1-p)^{n-\sum X_i} \end{aligned}$$

Los X_i son 1 o 0, entonces:

$$\sum X_i = \text{número de éxitos en n ensayos}$$

Denominemos: $S = \sum_{i=1}^{n} X_i$, entonces la distribución de S depende de n y p

$$S \longrightarrow Bin(n, p)$$

Para n suficentemente grande:

$$\hat{p} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$\hat{p} \longrightarrow N\left(p, \frac{p(1-p)}{n}\right)$$