Resumen distribuciones muestrales e intervalos de confianza

Sergio Nava

2023-04-18

Resumen de distribuciones muestrales

Distribución muestral para la media:
Cuando la desviación estándar poblacional es conocida, la distribución muestral para la media se puede aproximar a una distribución normal con media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$, donde μ es la media poblacional, σ es la desviación estándar poblacional y n es el tamaño muestral. La fórmula para la distribución muestral de la media es:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

o

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Cuando la desviación estándar poblacional es desconocida, se utiliza la distribución t de Student. La distribución muestral para la media se puede aproximar a una distribución t con $n - 1$ grados de libertad, donde n es el tamaño muestral. La fórmula para la distribución muestral de la media es:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Donde $\bar{x}$ es la media muestral, μ es la media poblacional, s es la desviación estándar muestral y n es el tamaño muestral.

Distribución muestral para la proporción:
La distribución muestral para la proporción también se puede aproximar a una distribución normal con media p y desviación estándar $\sqrt{\frac{p(1-p)}{n}}$, donde p es la proporción poblacional y n es el tamaño muestral. La fórmula para la distribución muestral de la proporción es:

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

Distribución muestral para la varianza:
La distribución muestral para la varianza se puede aproximar a una distribución chi-cuadrado con $n - 1$ grados de libertad, donde n es el tamaño muestral. La fórmula para la distribución muestral de la varianza es:

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

Donde s^2 es la varianza muestral y σ^2 es la varianza poblacional.

Resumen del cálculo de intervalos de confianza de una muestra

Cálculo de intervalos de confianza para la media:
Los intervalos de confianza son un rango de valores que, con cierto nivel de confianza, se espera que contengan el parámetro de interés en una población.
Para el caso de la distribución muestral para la media cuando se desconoce la desviación estándar poblacional, el intervalo de confianza para la media poblacional μ se puede calcular como:

$$\bar{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Donde $\bar{x}$ es la media muestral, s es la desviación estándar muestral, n es el tamaño muestral y $t_{n-1,\frac{\alpha}{2}}$ es el valor crítico de la distribución t de Student con $n-1$ grados de libertad y un nivel de significancia α.
Para el caso de la distribución muestral para la media cuando se conoce la desviación estándar poblacional, el intervalo de confianza para la media poblacional μ se puede calcular como:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Donde σ es la desviación estándar poblacional y $z_{\frac{\alpha}{2}}$ es el valor crítico de la distribución normal estándar con un nivel de significancia α.
Cálculo de intervalos de confianza para la proporción:
Para el caso de la distribución muestral para la proporción, el intervalo de confianza para la proporción poblacional p se puede calcular como:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Donde $\hat{p}$ es la proporción muestral y $z_{\frac{\alpha}{2}}$ es el valor crítico de la distribución normal estándar con un nivel de significancia α.
Cálculo de intervalos de confianza para la varianza:
Para el caso de la distribución muestral para la varianza, el intervalo de confianza para la varianza poblacional σ^2 se puede calcular como:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \right)$$

Donde s^2 es la varianza muestral y $\chi^2_{n-1,\frac{\alpha}{2}}$ y $\chi^2_{n-1,1-\frac{\alpha}{2}}$ son los valores críticos de la distribución chi-cuadrado con $n-1$ grados de libertad y niveles de significancia $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$, respectivamente.

Resumen del cálculo de intervalos de confianza de una muestra

Intervalos de confianza para la diferencia de medias con desviación estándar conocida:
El intervalo de confianza para la diferencia de medias con desviación estándar conocida se calcula como:

$$\bar{x_1} - \bar{x_2} \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

donde $\bar{x_1}$ y $\bar{x_2}$ son las medias muestrales de las dos muestras, σ_1 y σ_2 son las desviaciones estándar poblacionales de las dos poblaciones, n_1 y n_2 son los tamaños muestrales de las dos muestras, y $z_{\alpha/2}$ es el valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.

Intervalos de confianza para la diferencia de medias con desviación estándar desconocida:
El intervalo de confianza para la diferencia de medias con desviación estándar desconocida se calcula como:

$$\bar{x_1} - \bar{x_2} \pm t_{\alpha/2, n_1+n_2-2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

donde $\bar{x_1}$ y $\bar{x_2}$ son las medias muestrales de las dos muestras, s_1 y s_2 son las desviaciones estándar muestrales de las dos muestras, n_1 y n_2 son los tamaños muestrales de las dos muestras, y $t_{\alpha/2, n_1+n_2-2}$ es el valor crítico de la distribución t de Student con $n_1 + n_2 - 2$ grados de libertad correspondiente al nivel de confianza deseado.

Intervalos de confianza para la diferencia de proporciones:
El intervalo de confianza para la diferencia de proporciones se calcula como:

$$\hat{p_1} - \hat{p_2} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

donde $\hat{p_1}$ y $\hat{p_2}$ son las proporciones muestrales de las dos muestras, n_1 y n_2 son los tamaños muestrales de las dos muestras, y $z_{\alpha/2}$ es el valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.

Intervalos de confianza para el cociente de varianzas:
El intervalo de confianza para el cociente de varianzas se calcula como:

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2, n_2-1, n_1-1}$$

donde s_1^2 y s_2^2 son las varianzas muestrales de las dos muestras, σ_1^2 y σ_2^2 son las varianzas poblacionales de las dos poblaciones, n_1 y n_2 son los tamaños muestrales de las dos muestras, y $F_{\alpha/2, n_1-1, n_2-1}$ es el valor crítico de la distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad en los percentiles $\alpha/2$ y $1 - \alpha/2$ correspondientes al nivel de confianza deseado.

Valores críticos para diferentes niveles de confianza

Valores críticos $z_{\alpha/2}$ para niveles de confianza $(1-\alpha)\%$:

Nivel de confianza	0.90	0.95	0.99
Valor crítico	1.645	1.960	2.576

Table 1: Valores críticos de la distribución normal estándar para intervalos de confianza con diferentes niveles de confianza.

En R para calcular los valores críticos de la distribución z para niveles de confianza $1-\alpha$ se utiliza la función `qnorm`. Para el siguiente ejemplo se tiene un nivel de confianza de 90% se usa $\alpha = 0.10$ para que $1-\alpha = 0.9$.

```
alfa=0.10 # el nivel de confianza es (1-alfa)
qnorm(alfa/2,lower.tail = FALSE)
```

```
## [1] 1.644854
```

Valores críticos $t_{\alpha/2,\nu}$ para niveles de confianza $(1-\alpha)\%$:

En R para calcular los valores críticos de la distribución `t` para niveles de confianza $1-\alpha$ y distintos grados de libertad $n-1$ se utiliza la función `qt`. Para el siguiente ejemplo se tienen $n = 16$ datos, por lo tanto son $n-1 = 15$ grados de libertad, para un nivel de confianza de 90% se usa $\alpha = 0.10$ para que $1-\alpha = 0.9$.

```
n <- 16
alfa=0.10 # el nivel de confianza es (1-alfa)
qt(alfa/2,df=(n-1),lower.tail = FALSE)
```

```
## [1] 1.75305
```

Valores críticos $\chi^2_{\alpha/2,\nu}$ para niveles de confianza $(1-\alpha)\%$:

En R para calcular los valores críticos de la distribución χ^2 para niveles de confianza $1-\alpha$ y distintos grados de libertad se utiliza la función `qchisq`. Para el siguiente ejemplo se tienen $n = 16$ datos, por lo tanto son $n-1 = 15$ grados de libertad, para un nivel de confianza de 90% se usa $\alpha = 0.10$ para que $1-\alpha = 0.9$.

```
n <- 16
alfa=0.10 # el nivel de confianza es (1-alfa)
qchisq(alfa/2,df=(n-1),lower.tail = FALSE)
```

```
## [1] 24.99579
```

```
qchisq(alfa/2,df=(n-1),lower.tail = TRUE)
```

```
## [1] 7.260944
```

Valores críticos $F_{\alpha/2,n_1,n_2}$ para niveles de confianza $(1-\alpha)\%$:

En R para calcular los valores críticos de la distribución F para niveles de confianza $1-\alpha$ y distintos grados de libertad (n_1, n_2) se utiliza la función `qchisq`. Para el siguiente ejemplo se tienen $n_1 = 16$ datos en la primer muestra y $n_2 = 20$ en la segunda muestra, por lo tanto son $n_1 - 1 = 15$ y $n_2 - 1 = 19$ grados de libertad, para un nivel de confianza de 90% se usa $\alpha = 0.10$ para que $1-\alpha = 0.9$.

```
n1 <- 16
n2 <- 20
alfa=0.10 # el nivel de confianza es (1-alfa)
qf(alfa/2,df1=(n1-1),df2=(n2-1),lower.tail = FALSE)
```

```
## [1] 2.234063
```

```
qf(alfa/2,df1=(n1-1),df2=(n2-1),lower.tail = TRUE)
```

```
## [1] 0.4273834
```