FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

EYP1113 - Probabilidad y Estadística

Taller R - 05: Distribuciones Muestrales

Ricardo Aravena C. - Cristian Capetillo C. - Ingrid Guevara R.
Ricardo Olea O. - Bladimir Morales T., Daniel Saavedra M.

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2024

Contenido I

Motivación

Si X es una variable aleatoria, entonces $h(X)$ también lo es. Con esta lógica, el promedio muestral, varianza muestral, cuantiles muestrales, etc. son variables aleatorias. **Distribución muestral** se refiere a la distribución de estadisticos calculados sobre una muestra.

Dependiendo de la distribución de la muestra y el estadístico calculado, la distribución muestral puede ser desconocida o **difícil de obtener** analíticamente. En otras ocasiones, se pueden obtener de forma explícita.

En este taller, estudiaremos las distribuciones que subyacen de una muestra aleatoria $\mathcal{N}(\mu, \sigma^2)$, muy útiles al momento de una **inferencia estadística**.

De utilidad: ciclo for

En un punto de esta clase tendremos que utilizar lo que se denomina un ciclo `for`.

Un ciclo `for` permite repetir una acción una **cantidad determinada de veces**. Tambien nos permite recorrer vectores, matrices u objetos con distintas dimensiones.

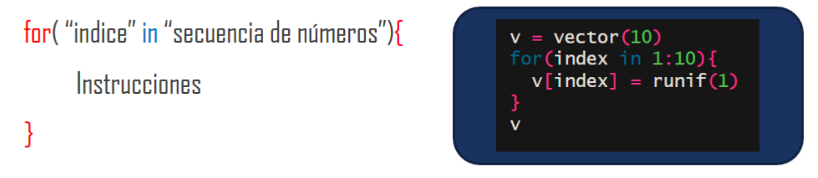

Figura 1: Ciclo for y ejemplo de uso.

Distribuciones Muestrales

Si $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ y $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}(\theta, \tau^2)$, entonces

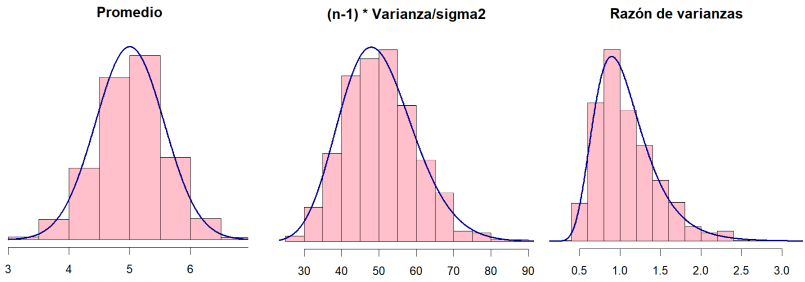

Figura 2: Distribución de estadísticos calculados sobre una muestra Normal.

Distribuciones Muestrales

Distribución Normal

Si $X \sim \text{Normal}(\mu, \sigma^2)$, $x \in \mathbb{R}$, entonces:

- Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dnorm(x,mean=`μ`,sd=`σ`)`.
- Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pnorm(q,mean=`μ`,sd=`σ`)`.
- Para calcular cuantiles de la distribución se usa el comando `qnorm(p,mean=`μ`,sd=`σ`)`.
- Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rnorm(n,mean=`μ`,sd=`σ`)`.

Distribuciones Muestrales

Distribución χ^2

Si $X \sim \chi^2_n(n)$, $x > 0$, entonces:

- Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dchisq(x, df = n)`.
- Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pchisq(q, df = n)`.
- Para calcular cuantiles de la distribución se usa el comando `qchisq(p, df = n)`.
- Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rchisq(n, df = n)`.

Distribuciones Muestrales

Distribución t-Student

Si t-Student(n), ,$x \in \mathbb{R}$, entonces:

- Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dt(x, df = n)`.
- Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pt(q, df = n)`.
- Para calcular cuantiles de la distribución se usa el comando `qt(p, df=n)`.
- Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rt(n, df = n)`.

Si $X \sim F(n, m), \; x > 0$, entonces:

- Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `df(x, df1 = n, df2 = m)`.
- Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pf(q, df1 = n, df2 = m)`.
- Para calcular cuantiles de la distribución se usa el comando `qf(p, df1 = n, df2 = m)`.
- Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rf(n, df1 = n, df2 = m)`.

Distribuciones Muestrales

Resultados

Si $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces:

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

Distribuciones Muestrales

Resultados

Si $X_1, ..., X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces:

1. $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
2. $\dfrac{X_i - \mu}{\sigma} \overset{iid}{\sim} \mathcal{N}(0, 1)$ y $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.

Distribuciones Muestrales

Resultados

Si $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces:

1. $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
2. $\dfrac{X_i - \mu}{\sigma} \overset{iid}{\sim} \mathcal{N}(0,1)$ y $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
3. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ y $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

Distribuciones Muestrales

Resultados

Si $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces:

1. $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
2. $\dfrac{X_i - \mu}{\sigma} \overset{iid}{\sim} \mathcal{N}(0,1)$ y $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
3. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ y $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.
4. $\dfrac{\overline{X} - \mu}{S_x/\sqrt{n}} \sim t_{n-1}$, donde $S_x^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$.

Distribuciones Muestrales

Resultados

Si $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces:

1. $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
2. $\dfrac{X_i - \mu}{\sigma} \stackrel{iid}{\sim} \mathcal{N}(0,1)$ y $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
3. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ y $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.
4. $\dfrac{\overline{X} - \mu}{S_x/\sqrt{n}} \sim t_{n-1}$, donde $S_x^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$.
5. Si tenemos otra muestra $Y_1, \ldots, Y_m \stackrel{iid}{\sim} \mathcal{N}(\theta, \tau^2)$, entonces se sostiene que $\dfrac{S_x^2/\sigma^2}{S_y^2/\tau^2} \sim F_{n-1,m-1}$

Distribuciones Muestrales

Ejemplo

- Simule una muestra $\mathcal{N}(\mu, \sigma^2)$ de tamaño $n = 50$ con $\mu = 5$ y $\sigma^2 = 16$ y calcule el promedio $\overline{X}$ y la suma de cuadrados $\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$.
- Repita el proceso anterior 1000 veces (mediante ciclo `for`) guardando cada uno de los promedios y suma de cuadrados.
- Cree un histograma para los promedios y agregue una curva $\mathcal{N}(5, 16/100)$. ¿Se respeta la curva?
- Cree un histograma para las sumas de cuadrados y agregue una curva χ^2_n. ¿Se respeta la curva?

Distribuciones Muestrales

Ejemplo

- ¿Cómo corroboraría que $\dfrac{\overline{X}-\mu}{S_x/\sqrt{n}} \sim t_{n-1}$?
- Mediante el uso de otra muestra de tamaño $m = 30$ de la distribución $\mathcal{N}(\theta, \tau^2)$, con $\theta = 10$ y $\tau^2 = 25$, corrobore el hecho que $\dfrac{S_x^2/\sigma^2}{S_y^2/\tau^2} \sim F_{n-1,m-1}$.

Aplicación

Utilizaremos lo aprendido sobre la base de datos `Abalon.xlsx`. Para esto, responda

a) Será de interés el alto de los abalones, tanto del centro Caldera como Puerto Montt. Cree un histograma de esta variable en ambos centros. ¿Parecen Normales?

b) Calcule $\frac{\overline{X}-\mu}{S_x/\sqrt{n}} \sim t_{n-1}$ para ambos centros con valores de $\mu = 3.2, 3.4, 3.6$ y 3.8. Compare estos con la distribución t_{n-1}.

c) ¿Parece posible que $\mu = 3.2$ en Puerto Montt? ¿Y en Caldera? ¿Que hay de un valor $\mu = 3.6$?

Aplicación

d) Ambos histogramas de (a) júntelos en uno solo. ¿Cuál es la diferencia entre las distribuciones?

e) Calcule $\frac{S_x^2/\sigma^2}{S_y^2/\tau^2}$ para $\sigma^2 = \tau^2$ y compárelo con la distribución $F_{n-1,n-1}$. ¿Cree que podría haber diferencias entre las varianzas de los grupos?