Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística
EYP1016-1 – Introducción a la Estadística

# Guía 2

**Introducción a la Estadística - Laboratorio (01-2024)**

Profesor: Cristian Capetillo Constela

Ayudante: Andrés Díaz

1. La figura 1 corresponde al histograma de una simulación de 100 variables aleatorias $Beta(\alpha, \beta)$.

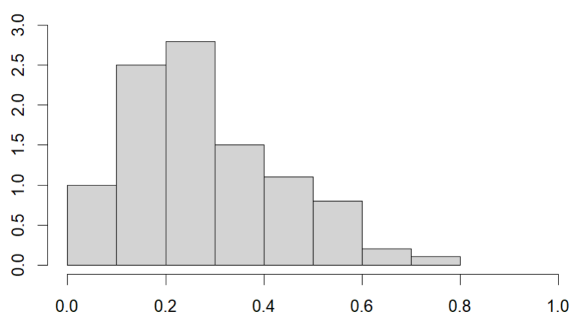


Figura 1: Histograma de simulaciones $Beta(\alpha, \beta)$.

El objetivo es encontrar los valores $\alpha$ y $\beta$ que fueron utilizados para su simulación. Para llevarlo a cabo, responda.

a) Replique el histograma en R utilizando el siguiente código:

Código 1: Figura 1

```r
frecuencias <- c(1, 2.5, 2.8, 1.5, 1.1, 0.8, 0.2, 0.1)*100
casillas = c(rep(0.05, frecuencias[1]), rep(0.15, frecuencias[2]),
  rep(0.25, frecuencias[3]), rep(0.35, frecuencias[4]),
  rep(0.45, frecuencias[5]), rep(0.55, frecuencias[6]),
  rep(0.65, frecuencias[7]), rep(0.75, frecuencias[8]))
hist(casillas, breaks = 5, freq = F, xlim = c(0,1), ylim = c(0,3))
```

b) Graficando curvas sobre tales histogramas, proponga valores para $\alpha$ y $\beta$. Deje un comentario en el código con su propuesta.

c) Según los parámetros que propuso en (b), ¿cuál es la probabilidad de que la variable tome un valor mayor que 0.3?

2. Supongamos que la altura de estudiantes de una universidad se distribuye normalmente con una media ($\mu$) de 170 cm y una desviación estándar ($\sigma$) de 10 cm.

a) Calcular la probabilidad de que un estudiante seleccionado al azar tenga una altura menor de 160 cm.

b) Calcular la probabilidad de que un estudiante tenga una altura entre 165 cm y 175 cm.

c) Determinar la altura que deja por debajo al 90 % de los estudiantes (utilizar función cuantil *qnorm*).

3. El tiempo (en horas) que tarda en fallar un componente electrónico sigue una distribución exponencial con una tasa de fallo ($\lambda$) de 0.5 fallos por hora.

   a) Calcular la probabilidad de que un componente tarde más de 3 horas en fallar.

   b) Calcular la probabilidad de que un componente falle entre 2 y 5 horas.

   c) Determinar el tiempo que deja por debajo al 95 % de los componentes en cuanto a tiempo hasta el fallo (utilizar función cuantil *qexp*).

4. Supongamos que la proporción de defectos en la producción de una fábrica sigue una distribución Beta con parámetros $\alpha = 2$ y $\beta = 5$.

   a) Calcular la probabilidad de que la proporción de defectos sea menor que 0.3.

   b) Calcular la probabilidad de que la proporción de defectos esté entre 0.2 y 0.6.

   c) Determinar el valor que deja por debajo al 75 % de las proporciones de defectos (utilizar función cuantil *qbeta*).

5. Un resultado de probabilidad dice que si $X \sim Gamma(\alpha, \lambda)$ e $Y \sim Gamma(\beta, \lambda)$, entonces

$$\frac{X}{X+Y} \sim Beta(\alpha, \beta).$$

   Utilizando esta relación, simule 300 variables aleatorias provenientes de una distribución $Beta(6, 2)$. Cree un histograma normalizado ($freq = F$) de sus resultados y la curva correspondiente a tal densidad.

6. En estadística, es muy común aproximar probabilidades mediante la simulación de datos. En particular, si $X_1, ..., X_n$ es una muestra aleatoria de una distribución $F$ (por ejemplo $\mathcal{N}(0, 1)$), e interesa la distribución de una transformación de $X_1, ..., X_n$ (como lo sería el promedio) entonces simular $k$ veces dicha muestra con $k$ lo suficientemente grande y aplicarle a cada una de las $k$ simulaciones la transformación respectiva podría ayudarnos a encontrar su distribución.

   Entendamos lo anterior con un ejemplo. Suponga que se tiene $X_1, ..., X_{10} \sim Gamma(3, 3)$, e interesa la distribución de $Y = \frac{1}{10} \sum \log(X_i)$. Responda,

   a) Simule 200 veces una muestra de tamaño 10 de la distribución $Gamma(3, 3)$, y a cada una de estas simulaciones aplíquele la transformación correspondiente. Usted tendrá así 200 simulaciones de $Y$

   b) Cree un histograma normalizado ($freq = F$) para las simulaciones de $Y$.

   c) En R, para calcular probabilidades de una muestra se utiliza *mean*(condición). Por ejemplo $mean(Y < 3)$ corresponde a la probabilidad de que $Y$ sea menor a 3. Con esto, ¿cuál es la probabilidad de que $Y$ esté entre 2 y 4?