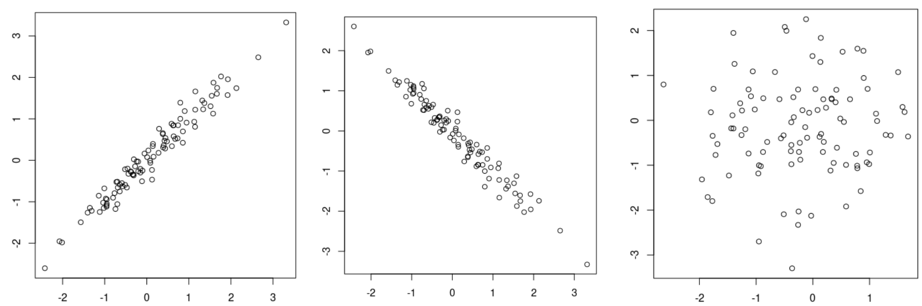
Bioestadística **Universidad de la República**
Curso 2022 **Facultad de Ciencias**

Práctico 6: Coeficiente de Correlación, Ley Fuerte de los Grandes Números, Teorema Central del Límite.

1. Consideremos los siguientes gráficos de parejas de puntos (X_i, Y_i), correspondientes a sorteos de variables aleatorias X e Y, con $i = 1, \dots, 100$. Para cada gráfico, indicar si aparenta ser un escenario



de correlación positiva, correlación negativa, o sin correlación. Justifique su respuesta.

2. Probar que el coeficiente de correlación verifica

- $\rho(aX, bY) = \rho(X, Y)$ para todo $a > 0, b > 0$. Esta propiedad dice que el coeficiente de correlación es invariante por cambios de escala en las variables.
- $\rho(X + a, Y + b) = \rho(X, Y)$.
- $\rho(X, aX + b) = 1$.

3. Sean $x_1, x_2, \dots, x_n$ una muestra aleatoria simple (M.A.S.) provenientes de las variables aleatorias X_1, $X_2, \dots, X_n$ i.i.d. Se define la media empírica de la muestra $\overline{x} = \overline{x_n}$ y la varianza empírica de la muestra s_n^2 como:

$$\overline{x_n} := \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Probar que se cumple:

(a) $\overline{x + y} = \overline{x} + \overline{y}$.

(b) $\overline{ax + b} = a\overline{x} + b \;\forall\, a, b \in \mathbb{R}$.

(c) $s_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(x_i^2 - \overline{x}^2\right) = \frac{1}{n-1}\sum\limits_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2$.

(d) Supongamos que las v.a. X_i son tales que $\mu = \mathbb{E}(X_i)$ y $\sigma^2 = var(X_i)$ son finitas. Deducir a partir de la Ley de los Grandes Números que podemos estimar μ y σ^2 a partir de $\overline{x_n}$ y s_n^2 respectivamente.

4. Asumiendo que la suma de variables aleatorias normales independientes es también una v.a. con distribución normal, deducir quién es la esperanza y la varianza del promedio:

$$\overline{X_n} := \frac{X_1 + X_2 + \dots + X_n}{n}$$

si las v.a. X_1, $X_2, \dots, X_n$ son i.i.d. $N(\mu, \sigma^2)$.

5. Consideremos la muestra de datos: $-1/2, -3, 2, 0, 5, -1$ de una cierta variable aleatoria X.

1 Estimar a partir de la muestra anterior $P(X \leq 0)$.

2 Estimar a partir de la muestra el valor $E(X)$.

3 Graficar la función de distribución empírica para la muestra anterior.

6. Si se tienen $(X_1, Y_1), \ldots, (X_n, Y_n)$ datos independientes e idénticamente distribuidos, con la misma distribución de (X, Y).

   (a) Probar a partir la Ley de los grandes números que r_{xy} es un estimador razonable de $\rho(X, Y)$, siendo r_{xy} [1] :
   $$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2}} \tag{1}$$

   (b) Probar que (1) es equivalente a
   $$r_{xy} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}_n \overline{Y}_n}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}_n^2}}$$

   (c) Probar que si X e Y son v.a. independientes entonces $\operatorname{cov}(X, Y) = \rho(X, Y) = 0$. Mostrar con un ejemplo que el recíproco no es cierto en general. (Sugerencia: Sean U y V dos v.a. independientes pero con la misma distribución. Considerar $X = U + V$ e $Y = U - V$).

7. Suponga que la duración de una canción se distribuye de manera uniforme entre 2 y 3.5 minutos. Armamos una playlist de 43 canciones.

   (a) Estime la duración de la playlist.

   (b) Calcule aproximadamente la probabilidad de que la playlist dure menos de 2 horas.

8. Se sabe que la lamparita de un proyector de un salón de la facultad tiene una vida promedio de $\mu = 100$ horas y una desviación estándar $\sigma = 75$. Durante un período de 5 años se usa ese proyector unas 9000 horas.

   (a) El encargado de compras compró 100 lamparitas para todo el período. Calcule la probabilidad de que no haya que comprar más lamparitas durante esos 5 años.

   (b) Utilice el TCL para estimar el número de lamparitas que debe comprar para que la probabilidad de quedarse corto de lamparitas durante ese período sea del 1%.

[1] En **R** se calcula el estimador r_{xy} para dos muestras del mismo tamaño $X = (x_1, \ldots, x_n)$ e $Y = (y_1, \ldots, y_n)$ con la función $\mathbf{cor}(X, Y)$.