# Ejercicios estadística

1. Considerar el siguiente conjunto de datos $S = \{2, 5, 6, 2, 8, -1\}$. Calcular la media, mediana, moda, rango, varianza y rango intercuartil.

2. Considerar el siguiente conjunto de datos obtenido de entrevistar a 11 personas

| Nombre | Edad | Sexo | Estatura | Nivel educativo |
|---|---|---|---|---|
| Pedro | 20 | M | 1.73 | Secundaria |
| Arturo | 23 | M | 1.75 | Primaria |
| Juan | 24 | M | 1.63 | Primaria |
| Iván | 24 | M | 1.95 | Secundaria |
| Roberto | 32 | M | 1.52 | Preparatoria |
| Carmen | 32 | F | 1.60 | Preparatoria |
| Alejandra | 32 | F | 1.65 | Maestría |
| Gabriela | 45 | F | 1.59 | Doctorado |
| Carlos | 45 | M | 1.82 | Licenciatura |
| María | 51 | F | 1.54 | Preparatoria |
| Rocío | 57 | F | 1.77 | Preparatoria |

   *a*) Para cada una de las variables, identifique si es categórica o numérica. Para las variables categóricas identifique si es nominal u ordinal.

   *b*) Para la variable Edad calcular: media, mediana, moda, $p_{25}$, $p_{75}$, varianza, rango y rango intercuartil.

3. Sea $S = \{X_1, X_2, X_3, X_4, X_5\}$ una muestra aleatoria de una población $Poi(\lambda)$, con $\lambda$ -desconocida. Recordar que si $X \sim Poi(\lambda)$ entonces:

$$Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots, \qquad E(X) = \lambda \qquad \text{y} \qquad V(X) = \lambda.$$

   Considerar las siguientes funciones de la muestra:

$$\hat{\lambda}_1(S) = X_1 + X_2 + X_3 + X_4 + X_5, \quad \hat{\lambda}_2(S) = X_1 + \lambda X_2, \quad \hat{\lambda}_3(S) = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} \quad \text{y} \quad \hat{\lambda}_4(S) = \frac{X_1 + 4X_2}{5}.$$

   *a*) Indicar cuáles pueden ser considerados estimadores de $\lambda$ y cuáles no.

   *b*) Indicar cuáles estimadores son insesgados y cuáles sesgados.

   *c*) De los estimadores sesgados, calcular el sesgo.

   *d*) De los estimadores sesgados, indicar cuál tiene el menor ECM y concluya cuál es el mejor.

4. Sea $S = \{X_1, X_2, X_3, X_4\}$ una muestra aleatoria de una población $Bin(10, p)$, con $p$ desconocido. Recordar que si $X \sim Bin(n, p)$, entonces

$$Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n, \qquad E(X) = np \qquad \text{y} \qquad V(X) = np(1-p).$$

   Considerar las siguientes funciones de la muestra:

$$\hat{p}_1(S) = \frac{X_1 + X_2 + X_3 + X_4}{4}, \quad \hat{p}_2(S) = p + X_1 + X_4, \quad \hat{p}_3(S) = \frac{X_1 + X_2 + X_3 + X_4}{40} \quad \text{y} \quad \hat{p}_4(S) = \frac{3X_1 + X_4}{40}.$$

   *a*) ¿Cuál es el espacio parametral de este modelo?

   *b*) Indicar cuáles pueden ser considerados estimadores de $p$ y cuáles no.

*c*) Indicar cuáles estimadores son insesgados y cuáles sesgados.

*d*) De los estimadores sesgados, calcular el sesgo.

*e*) De los estimadores sesgados, indicar cuál tiene el menor ECM y concluya cuál es el mejor.

5. Sea $S = \{X_1, X_2, X_3\}$ una muestra aleatoria de una población $F(x)$ tal que $E(X_i) = \mu$ y $V(X_i) = 1$. Calcular el ECM de los siguientes estimadores de $\mu$.

$$\hat{\mu}_1(S) = X_1 \qquad \text{y} \qquad \hat{\mu}_2(S) = \frac{3X_1 - 2X_2 + X_3}{6}$$

¿Para qué valores de $\mu$ se tiene que $\hat{\mu}_2$ es mejor estimador que $\hat{\mu}_1$?

6. Se quiere verificar que la estatura promedio de las mujeres es menor que la estatura promedio de los hombres. Suponga que la distribución de la estatura de mujeres y hombres puede ser modelada con una distribución normal con la misma varianza conocida para ambas poblaciones. Es decir, $X \sim N(\mu_x, \sigma^2)$ y $Y \sim N(\mu_y, \sigma^2)$, donde $X$ modela la estatura de las mujeres y $Y$ la estatura de los hombres.

Se observó una muestra aleatoria de tamaño 100 en cada una de las poblaciones y se obtivieron los siguientes resultados

$$\bar{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = 169 \text{ cm} \qquad \text{y} \qquad \bar{y} = \frac{1}{100}\sum_{i=1}^{100} y_i = 175 \text{ cm.}$$

Suponer que la varianza en ambas poblaciones es $\sigma^2 = 25$. Construir un intervalo de confianza 95 % para la diferencia de medias $\mu_x - \mu_y$. Concluya si en efecto tenemos evidencia para decir que la estatura promedio de las mujeres es menor a la estatura promedio de los hombres.

7. Se tiene interés en realizar un muestreo para estimar el peso promedio (en kg) de los hombres mayores de 18 años en México. Si se asume que el peso de esta población se puede modelar con una distribución $N(\mu, \sigma^2)$ con $\sigma^2$ conocida igual a 36. Calcular el tamaño de muestra necesario para estimar la media $\mu$ con un error $d = 0.5$ kg y una confianza de 95 %.

8. La UNAM actualmente utiliza bombillas de la Compañía A para iluminar los distintos salones de clases, sin embargo hace un par de meses la Compañía B afirmó que ellos fabrican bombillas con mayor durabilidad. La UNAM desea tomar una decisión para saber si cambia de proveedor o no, para ello se realizó un experimento y se tomó una muestra de tamaño 10 bombillas de ambas compañías y se midió el tiempo que permanecieron encendidas hasta que fallaron. Los datos de este experimento se muestran en la siguiente tabla:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Compañía A | 1035 | 906 | 1393 | 1026 | 931 | 991 | 1425 | 1005 | 1146 | 1291 |
| Compañía B | 1199 | 1283 | 1199 | 1058 | 1295 | 1187 | 1397 | 1381 | 1359 | 1051 |

Se puede asumir que $X \sim N(\mu_x, \sigma^2)$ y $Y \sim N(\mu_y, \sigma^2)$, donde $X$ modela el tiempo de vida en horas de las bombillas de la Compañía A y $Y$ el tiempo de vida en horas de las bombillas de la compañía B. Se debe notar que ambas poblaciones tienen la misma varianza pero ésta es desconocida.

Construir un intervalo de confianza 95 % para la diferencia de medias $\mu_x - \mu_y$ y con dicho intervalo concluya si recomendaría a la UNAM cambiar de proveedor.