

Práctica 2 de Estadística para Data Science

Problema 1 *Considera un examen de 10 preguntas con 5 posibles respuestas para cada pregunta en el que se responde al azar. Considera la variable aleatoria X = número de preguntas acertadas en el examen.*

- i) ¿Cuánto valen $\mathbb{E}(X)$ y $\text{SD}(X)$?*
- ii) Calcula la probabilidad de acertar 0, 1, 2, ..., 10 preguntas. ¿Cuánto vale la suma de las once probabilidades que acabas de calcular?*
- iii) Calcula las probabilidades acumuladas, es decir la probabilidad de obtener 0 aciertos o menos, 1 acierto o menos, 2 aciertos o menos ... 10 aciertos o menos.*
- iv) Representa las probabilidades de los apartados anteriores en una gráfica con dos paneles. Guarda la figura en un archivo. ¿Cuál es el número de aciertos más probable?*

Problema 2 *En el archivo GOT.csv figura el número de muertes por cada uno de los 67 primeros capítulos de la serie Juego de Tronos.*

- i) Razona qué distribución usarías para calcular probabilidades asociadas a esta variable. Representa gráficamente los datos correspondientes al número de muertes de forma que te ayude a decidir qué distribución usar.*
- ii) Estima la probabilidad de que en el siguiente capítulo haya más de 4 muertes. Explica qué suposiciones has hecho. ¿Podría fallar alguna de las suposiciones?*
- iii) Calcula la probabilidad de que en el siguiente capítulo haya exactamente 0, 1, 2, ..., 10 muertes. Representa las probabilidades en una gráfica. ¿Cuál es el número de muertes más probable? ¿Cuánto vale la suma de las once probabilidades? ¿Por qué?*
- iv) ¿Cuánto valen $\mathbb{E}(X)$ y $\text{SD}(X)$? Explica su significado en este contexto.*

Problema 3 *El colesterol se mide en una población con media 180 mg/dL y desviación típica 25 mg/dL.*

- i) Calcula la probabilidad de que una persona tenga más de 225 mg/dL de colesterol en sangre.*
- ii) Calcula el percentil del 95% de colesterol en sangre en dicha población.*
- iii) Representa la función de densidad y la función de distribución en una gráfica con dos paneles. Guarda la figura en un archivo.*
- iv) ¿Cuánto valen $\mathbb{E}(X)$ y $\text{SD}(X)$?*

Problema 4 En el archivo `results.csv` figuran los resultados de 40838 partidos de fútbol internacionales. En la columna `home_score` figuran los goles marcados por los equipos locales en cada partido. En la columna `home_team` figura el nombre del equipo local en cada partido.

- i) Considera los datos de goles de España como equipo local. Razona qué distribución usarías para calcular probabilidades asociadas a esta variable. Representa gráficamente los datos de forma que te ayude a decidir qué distribución usar.
- ii) Estima la probabilidad de que en el siguiente partido como local España marque más de 3 goles. Explica qué suposiciones has hecho. ¿Podría fallar alguna de las suposiciones?
- iii) Calcula la probabilidad de que en el siguiente partido como local España marque exactamente 0, 1, 2, ..., 10 goles. Representa las probabilidades en una gráfica. ¿Cuál es el número de goles más probable? ¿Cuánto vale la suma de las once probabilidades? ¿Por qué?
- iv) ¿Cuánto valen $\mathbb{E}(X)$ y $\text{SD}(X)$? Explica su significado en este contexto.
- v) Una casa de apuestas paga 4 Euros por cada Euro apostado (ganancia neta = 3 Euros) si España marca más de 3 goles en el próximo partido como local. En caso de que marcase 3 goles o menos perderíamos el Euro apostado. Razona matemáticamente si es rentable apostar.