

# Inferencia Estadística

## Segundo Parcial

13 de Mayo de 2024

### Instrucciones

- Justifica todas las respuestas.
- Entrega los folios de tus respuestas en tres bloques separados: teoría; problema 1; problema 2.
- Tienes hasta una hora para resolver la teoría y entregar tu respuesta.
- Una vez que entregues la teoría, podrás encender el ordenador exclusivamente para:
  - usar R mientras resuelves los problemas;
  - acceder al campus virtual de la asignatura para consultar documentos.
- No hay una “tarea” en el campus virtual para subir documentos. Has de entregar todas tus respuestas en papel. Si has de mostrar un gráfico, esbózalo a mano alzada dejando claro lo que quieras destacar. Si quieres mostrar código fuente, escríbelo con la mayor precisión posible o, al menos, un pseudocódigo.

### Teoría

1. (1 punto) Enuncia y demuestra el lema de Neyman-Pearson para tests puros.
2. (1,5 puntos) Test de Lilliefors.
  - a) Explica en qué se basa el test de Lilliefors de bondad de ajuste.
  - b) Sea  $X$  una variable aleatoria con distribución exponencial  $\text{Exp}(\lambda)$  con parámetro  $\lambda$  desconocido, de la que se extrae una muestra aleatoria simple  $(X_1, \dots, X_n)$ . Demuestra que la distribución del estadístico  $D_n = \sup_x |F_n(x) - F_{\hat{\lambda}}(x)|$  no depende de  $\lambda$ .
3. (1,5 puntos) Explica el test de Kruskal-Wallis. Demuestra que bajo la hipótesis nula el estadístico es de libre distribución.

## Problema 1

El fichero `eficacia.rda` recoge datos sobre tres variables medidas en varias personas durante un estudio de la eficacia de un tratamiento para cierta enfermedad:

*Grupo*: pertenencia al grupo de control o al de tratamiento.

*Estado*: estado general del paciente (malo, regular, bueno).

*Marcador*: nivel de colesterol en ayunas.

- a) (1 punto) Comprueba, a un nivel de significación  $\alpha = 0.1$ , si la distribución del estado en la población de control es la misma que en la población de personas bajo tratamiento.
- b) (1 punto) Comprueba, a un nivel de significación  $\alpha = 0.05$ , si la distribución del marcador en la población de control es la misma que en la población de personas bajo tratamiento.
- c) (1 punto) Calcula, a un nivel de confianza  $1 - \alpha = 0.9$ , un intervalo de confianza para la mediana del marcador en la población de control.

## Problema 2

El fichero `LasAguas.RData` contiene datos registrados sobre la temperatura del agua en distintas playas. El fichero de datos cuenta con las siguientes variables:

*Temperatura*: Temperatura del agua (en grados centígrados);

*Playa*: Nombre de la playa en la que se midió la temperatura.

Las temperaturas en cada una de las playas fueron registradas en días distintos, asumiendo independencia entre las distintas mediciones. Basándote en los datos recogidos y trabajando a nivel de significación  $\alpha = 0.05$ , responde a las siguientes preguntas:

- a) (1 punto) Justifica mediante un estudio inferencial que la distribución de la temperatura en cada playa es aproximadamente una normal con la misma varianza.
- b) (1 punto) ¿Hay diferencias en la temperatura en las distintas playas? Realiza un estudio descriptivo previo (tanto numérico como gráfico), basa tus conclusiones en los resultados de un contraste de hipótesis y concluye con un estudio a posteriori.
- c) (1 punto) Considerando en ambos casos el tamaño muestral homogéneo e igual a 30 en cada grupo y la varianza común de 9 grados centígrados cuadrados, estima mediante simulación por el método de Monte Carlo la potencia del test ANOVA cuando: (1) las medias en los cinco grupos son iguales a 23 grados centígrados; (2) las medias en cuatro de los cinco grupos son iguales a 23 grados centígrados y en el quinto grupo es igual a 21 grados centígrados. Interpreta los resultados obtenidos.

AYUDA: En R puedes acceder al p-valor de un modelo ANOVA al que se le ha asignado el nombre `modelo` por medio del código `summary(modelo)[[1]][1,5]` .