

Entrega de evaluación  
Inferencia  
Grado en Ingeniería Biomédica. Tercer Curso

Jorge Calero Sanz

Fecha final de entrega: 10 de mayo de 2019

## Problema 1

En el Excel 'VEF.xls' contiene mediciones sobre 654 niños de edades entre 3 y 19 sobre el VEF (Volumen Espiratoria Forzado) que mide el volumen (en litros) del aire expulsado durante el primer segundo de la espiración máxima. La tabla contiene los siguientes datos:

- ID del sujeto observado
- Edad (años)
- VEF (en litros)
- Altura (en pulgadas)
- Sexo (0 = femenino / 1 = masculino)
- Fumador (0 = no fumador / 1 = fumador)

Se pide:

- a. ) Cargar la tabla en Python usando Pandas, y convertir la altura de pulgadas a centímetros.
- b. ) Utilizar el método de estadística descriptiva que se crea más conveniente para representar las columnas 'Edad', 'FEV' y 'Altura'.
- c. ) Utilizar el método de estadística descriptiva que se crea más conveniente para comparar los niveles de FEV por género y por estatus de fumador. ¿Se puede obtener alguna conclusión?

## Problema 2

Un programa de detección de neuroblastoma fue llevado a cabo en Alemania para niños nacidos entre el 1 de noviembre de 1993 y el 30 de junio del año 2000 mientras tenían entre 9 y 18 meses. Participaron un total de 1.475.773 niños en el programa, de los cuales 204 fueron diagnosticados con dicha enfermedad. Los investigadores esperaban que el ratio de incidencia fuera de 7.3 por 100.000 niños, que es el ratio conocido en ausencia del programa de detección.

- a). Realizar un contraste para comprobar si el número de casos detectados fue mayor que el esperado.
  - a.1) ¿Cuáles son las hipótesis nula  $H_0$  y alternativa  $H_1$ ?
  - a.2) ¿Es un test de una alternativa, o de doble alternativa?
  - a.3) ¿Qué test estadísticos podemos usar? ¿Cuál es más conveniente?
  - a.4) Aplicar un contraste para  $\alpha = 0.05$ . ¿Qué podemos decir con el resultado del test?
  - a.1) ¿Cuál es el  $p$ -valor?
- b). Obtener un intervalo de confianza del 95 % para el ratio de incidencia. ¿Está el ratio de incidencia esperado en ese intervalo? (Expresar el intervalo de confianza de la forma  $(p_1, p_2)$  donde  $p_1$  y  $p_2$  estén expresados en casos por 100.000 observaciones). Comentar el resultado

### Problema 3

Se plantea un estudio para analizar la efectividad de la aspirina reduciendo la temperatura corporal. Para ello, se observa a 12 pacientes con gripe. Se toman dos observaciones: antes y después de una hora de haberse administrado un aspirina. Los resultados se pueden observar en la siguiente tabla (donde la temperatura está en  $^{\circ}F$ ).

Paciente	Antes	Después
1	102.4	99.6
2	103.2	100.1
3	101.9	100.2
4	103	101.1
5	101.2	99.8
6	100.7	100.2
7	102.5	101
8	103.1	100.1
9	102.8	100.7
10	102.3	101.1
11	101.9	101.1
12	101.4	100.2

Suponiendo que los datos se distribuyen como una normal, y que queremos comprobar si la aspirina reduce la temperatura:

- Definir la hipótesis nula y la alternativa
- Las dos muestras: ¿Son independientes o no? ¿Qué test debemos utilizar?
- Realizar dicho test estadístico para  $\alpha = 0.05$  y comentar el resultado.

### Problema 4

Para los datos del problema 1, comparar los niveles medios de VEF entre fumadores y no fumadores de chicos entre 10 y 14 años y entre chicas de 10 y 14 años.

### Problema 5

En unos test de aptitud se conoce que las notas suelen seguir una distribución normal de media 500 y desviación típica 100. En la tabla IQTest.xls aparecen los datos de este año.

- Realizar el test de Kolmogorov-Smirnov con el supuesto de que los datos siguen la normal  $N(500, 100)$ . Realizar el test de Lillefors. Comentar ambos resultados.
- Realizar un contraste de hipótesis para verificar que la media de este año ha sido más baja con una significación del  $\alpha = 0.05$ .
- Si fijamos como hipótesis alternativa  $\mu_1 = 450$  ¿Cuál es la potencia del test?
- Con  $H_1$  como  $\mu_1 = 450$  y queremos un test de potencia  $1 - \beta = 0.8$  ¿Cuál debe ser el tamaño de la muestra?