2. Guión

2.1. Partes del trabajo práctico de Probabilidades y Estadística II

1. Parte 1: Identificación de Modelo y Muestreo

   a) Ajuste de Modelo

      1) Breve descripción de las variables `sleeptime` y `steps`
      2) Ajustar todos los datos de `sleeptime` a una d. Normal, una d. Gamma y una d. Exponencial. Mostrar los 3 resultados: estimadores de los parámetros respectivos, histograma de los datos con la curva de densidad del modelo correspondiente. Usar el *Test de Kolmogorov-Smirnov* para analizar el ajuste realizado (un p-value menor de 0.1 indica que el ajuste no es bueno).

   b) Muestreo

      1) Se toman muestras del conjunto de datos con tamaño 20 para la variable `Age`
      2) Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Medias muestrales y representar `hist` y `boxplot`. Ajustar a la distribución normal cada uno de los vectores de medias (variable aleatoria muestral)
      3) Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 Varianzas muestrales y representar `hist` y `boxplot`. Ajustar a la distribución normal cada uno de los vectores de varianzas (variable aleatoria muestral)
      4) Con 30, 50 y 100 muestras, es decir, calcular las 30, 50 y 100 proporcioń muestral de Mujeres/Varones y representar `hist` y `boxplot`. Ajustar a la distribución normal cada uno de los vectores de proporciones (variable aleatoria muestral)

2. Parte 2: Estimación Clásica (puntual, intervalos)

   a) Puntual

      1) Estimar media y varianza de las variables `sleeptime` y `steps`. Primero con todos los datos y segundo con una muestra de tamaño 20.
      2) Estimar media y varianza de las variables `sleeptime` y `steps` entre las Mujeres. Primero con todos los datos y segundo con una muestra de tamaño 20.
      3) Estimar media y varianza de las variables `sleeptime` y `steps` entre los Varones. Primero con todos los datos y segundo con una muestra de tamaño 20.

   b) Intervalo

      1) Estimación del intervalo de confianza para la media, varianza, proporción, al nivel de confianza 90 %, 95 % y 99 %, para las variables `sleeptime` y `step` entre según niveles {"M", "V" del factor `Sex`, con una muestra de tamaño 20. Primero suponer normalidad y segundo usar *Bootstrap* para el caso de poblaciones de distribución general o arbitraria. Para la media suponer primero varianza conocida y segundo desconocida.
      2) Estimación del intervalo de confianza para la diferencia de medias, razón de varianzas, proporción, al nivel de confianza 90 %, 95 % y 99 %, para las variables `sleeptime` y `steps` según niveles {"M", "V" del factor `Sex`, con una muestra de tamaño 20. Primero suponer normalidad y segundo usar *Bootstrap* para el caso de poblaciones de distribución general o arbitraria. Para la diferencia de medias suponer primero varianzas conocidas y segundo desconocidas e iguales.

3. Parte 3: Estimación Bayesiana (puntual, intervalos)

   a) La proporción p_e de individuos de nacionalidad española en la población está entre 25 % y el 35 %. En una muestra de 20 personas hay nE españoles. Suponer que la $p_e \sim \beta(\alpha = 5, \beta = 10)$, con función de densidad $f(x; \alpha; \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ y moda $= \frac{\alpha-1}{\alpha+\beta-1}$.

      - Obtener la p_e tras la información aportada por la muestra (la distribución a posteriori)
      - Obtener el IC con 95 % de confianza para la p_e, (los cuantiles que dejan a derecha e izquierda el 0.025 de probabilidad según la distribución a posteriori)
      - La estatura (variable `height`) del grupo de españoles, franceses e italianos sigue una $N(170, 7)$. Estimar con la muestra anterior, la estatura media con varianza conocida.

En esta parte, se trata de implementar en un script de R el procedimiento de inferencia manual, con las formulas de clase y según los ejemplos de clase, no se usarán paquetes R (*https://cran.r-project.org/web/views/Bayesian.html*)

4. # Parte 4: Contrastes (paramétricos y no paramétricos)

   a) Tomar dos muestras de tamaño 20 de la variable `IMC`: $Sample_1$ y $Sample_2$

      - Contrastar si la media μ_1 de $Sample_1$ es $Q_1 \leq \mu_1$, con varianza desconocida (Q_1 : cuartil 1 de la muestra)
      - Contrastar si la media μ_1 de $Sample_1$ es $\mu_1 \leq Q_3$, con varianza desconocida
      - Contrastar si la varianza σ^2 de $Sample_1$ es mayor que 1.0, con media desconocida
      - Contrastar si $\mu_1 - \mu_2 = 0$, con $Sample_1$ y $Sample_2$ respectivamente, con varianzas desconocidas
      - Contrastar si $\sigma_1^2/\sigma_2^2 = 1$, con $Sample_1$ y $Sample_2$ respectivamente

      (Q_1 y Q_3 : cuartiles 1 y 3 de la muestra)

   b) Tomar una muestra de tamaño 20 de la variable `IMC`, $Sample$, con nivel de significación $\alpha = 0{,}05$

      - Contrastar la normalidad de $Sample$, mediante el test de Pearson y el test de Kolmogorov-Smirnov
      - Contrastar la independencia de $Sample$, mediante el test de Durbin-Watson. Se trata de ver si hay dependencia de `IMC` respecto a algunas variables de conjunto de datos. Sugerencia: paquete `lmtest` y función `dwtest()`, es decir, se toma una muestra de tamaño 20 del conjunto de datos y se proponen algunas variables independientes de las que puedea depender `IMC`, y tras hacer el test se saca una conclusión.
      - Contrastar la homogeneidad de $Sample$, mediante el test de Wilcoxon. Se trata de ver si varias muestras provienen de la misma población, es decir, tomamos dos muestras de tamaño 20 de la variable `IMC` (de la misma población) y tras hacer el test se saca una conclusión.
        Sugerencia: paquete `stats` y función `wilcox.test()`

      Nota: en R un modelo de dependencia se define: $W \sim X + Y$, es decir, $IMC \sim height + weight$

5. # Parte 5, Regresión lineal simple (estimación y contraste)

   a) Estimación del modelo de regresión simple
      Con las variables Data$height y Data$weight tomar una muestra de tamaño 20 y estimar un modelo de regesión simple.

      - Hipótesis (linealidad, residuos $\sim N$, homocedasticidad residuos, independencia residuos)
      - Metodología, trasformaciones. Estimación y propiedades
      - Predicciones e Intervalo de Confianza para predicciones

   b) Contraste de regresión
      Para el modelo de regresión Data$weight ∼ Data$height

      - Linealidad, $\beta_1 \neq 0$
      - Hipótesis (linealidad, residuos $\sim N$, homocedastidad residuos, independencia residuos) contrastadas con los residuos
      - Interpretación

2.2. REFERENCIAS:

`https://fhernanb.github.io/Manual-de-R/`
`http://rcompanion.org/documents/RHandbookProgramEvaluation.pdf`
`http://rcompanion.org/handbook/`