

Parcial I

Curso IMSER 2012

Instrucciones:

La carpeta comprimida “parcial-I.zip” tiene un script de R (“parcial.R”) en el cual usted deberá guardar todos los comandos del ejercicio, siguiendo la demarcación que se muestra en el archivo mismo.

La hoja de cálculo llamada ‘datos.xls’ contiene tres variables muestreadas. La primera variable son los tamaños corporales (en Kg) de 40 competidores del último Mundial de Magic, 20 hombres y 20 mujeres. La segunda variable es el sexo de los mismos, codificados como 1 para mujeres y 2 para varones. La tercera variable es la altura de los participantes, en metros.

Nota: los ejercicios del parcial son dependientes de los anteriores en el sentido de que utilizan objetos creados, pero no implica que no se puedan tratar de resolver independientemente.

Una vez terminado el parcial usted deberá comprimir la carpeta y subirla al EVA en la [página correspondiente](#). Nótese que debe contener al menos dos archivos:

- “parcial.R”
- Un archivo de texto plano (.txt o .csv) con los datos de “datos.xls”, según lo indicado en la parte **a**.

a. Importar la tabla de la hoja de cálculo a R; el objeto resultante debe ser una data.frame llamada **magic** y sus columnas deben llamarse **body.size**, **sex** y **height** (i.e.: los valores que R asigna por defecto).

Debe guardarse en la carpeta del parcial el archivo de texto (.txt o .csv) utilizado para la importación.

b. Generar un vector aleatorio llamado **age**, compuesto por números enteros positivos entre 18 y 35 aprox. (este rango puede tener cierta flexibilidad). Nota: debe ser generado con una función creadora de valores aleatorios, y se deben permitir ocurrencias repetidas del mismo valor.

c. Agregar la variable **age** al data.frame **magic**.

- d. La variable `sex` del `data.frame` presenta los valores 1 y 2. Transformar esta variable en factor. Luego modificar los nombres de los niveles del mismo a "mujer" y "hombre" (correspondientes a los valores originales 1 y 2 respectivamente).
- e. Crear la variable factorial `agef` basada en `age` de forma tal que cuente con 3 niveles (dividiendo el rango en 3 franjas etarias de igual amplitud). Agregar dicha variable a la `data.frame` `magic`.
- f. Nombrar a cada uno de los niveles de la variable `agef` (dentro de `magic`) como "novatos", "intermedios", "expertos", en ese orden.
- g. Cambiar los nombres de las columnas del `data.frame` a '`peso`', '`sexo`', '`altura`', '`edad`' y '`edadf`'.
- h. Graficar: peso en función de sexo. El gráfico debe ser un diagrama de cajas y tener las etiquetas "Sexo", "Peso (Kg)" en los ejes correspondientes, así como el título: "Peso en función del sexo".
- i. Realizar un anova con la variable de respuesta `peso` y la variable explicativa `sexo`; guardar el resultado en el objeto `peso.sexo`.
- j. Crear los objetos `peso.hombre` y `peso.mujer` con los valores esperados de peso para los sexos respectivos, según los resultados del modelo `peso.sexo` creado en el punto anterior (considere la interpretación de los coeficientes del anova dada en la lección correspondiente).
- k. Graficar: $altura^2$ (altura al cuadrado) en función del peso.
- l. Realizar una regresión lineal entre estas dos variables ($altura^2 \sim \text{peso}$) sin intercepto. El modelo obtenido debe guardarse en el objeto `altura.peso`.
- m. Realizar una regresión lineal entre estas dos variables ($altura^2 \sim \text{peso}$) sin intercepto, pero esta vez excluyendo a los outliers de peso (es decir, aquellos tales que $\text{peso} > 120$ Kg). Guardar el modelo en el objeto `altura.peso2`.

n. Utilizando los coeficientes obtenidos en este segundo modelo, determine la altura esperada para la secuencia de pesos:

```
p <- seq(40, 120, by = 0.5)
```

Los valores de altura esperados se deben guardar en un objeto llamado **ae**.

(Nótese que el modelo establece la relación del peso con la *altura*², pero no con la altura per se).

o. Guardar en el objeto **r2** el valor del R cuadrado (no ajustado) del modelo (es decir, el coeficiente de determinación o “proporción de varianza explicada”).

p. Agregar al último gráfico creado anteriormente las líneas correspondientes a ambas regresiones lineales, utilizando diferentes trazos y/o colores para diferenciarlas.

q. Agregar una línea vertical indicando el valor 120 en el eje del peso; utilice un estilo de línea diferente a los anteriores (con color y/o trazo diferente).

r. Superponer al mismo gráfico los puntos de los outliers en peso, utilizando un símbolo diferente (y opcionalmente, un color distinto), de forma tal que se puedan diferenciar a simple vista.

s. Agregar una última variable, llamada **IMC**, a la data.frame **magic**: el **índice IMC** correspondiente a cada participante, calculado como:

$$IMC = \frac{Peso(Kg)}{Altura^2(m)}$$

t. Exportar la data.frame **magic** a un archivo de texto plano (.txt o .csv), conteniendo todas las columnas agregadas y los encabezados, pero excluyendo los nombres de fila.