# TRABAJO R

## **1.Carga en memoria el fichero CSV como tibble, asegurándote de que las variables cualitativas sean leídas como factores.**

Para cargar el archivo lo primero que haré será cargar la librería readr para cargar el archivo como tibble, en esta importación además asignare a cada variable su tipo de variable.

**Imagen que contiene Diagrama

Descripción generada automáticamente**

**Resultado:**

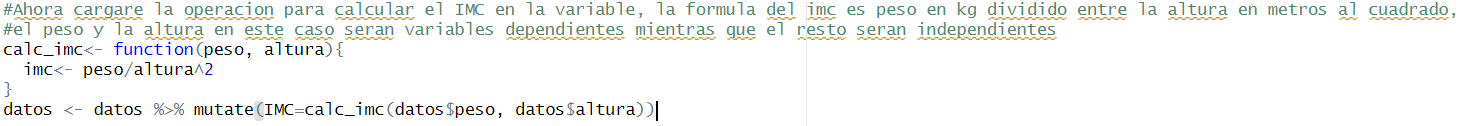
**Imagen que contiene Texto

Descripción generada automáticamente**

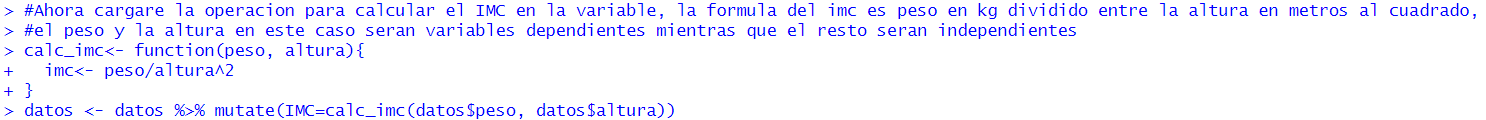
**Se han añadido las variables correctamente**

**2. Construye una nueva columna llamada IMC que sea igual al peso dividido por la altura al cuadrado. La variable explicada será IMC, las variables explicatorias serán el resto de 12 variables exceptuando peso y altura.**

**Para crear la columna de IMC creo una función con la formula del IMC que es peso en kg dividido entre altura al cuadrado y realizo un mutate para añadir la columna a la tabla.**

****

**Resultado:**

****

**Como no salen errores, puedo saber que la columna se ha añadido correctamente**

## **3. Elimina completamente las filas que tengan algún valor NA en una de sus columnas.**

Para eliminar las filas con datos no disponibles uso na.omit

****

****

**La salida no ha producido ningun**

## **4. Calcula las medias y desviaciones típicas (no cuasidesviación) de todas las variables numéricas.**

Primero para quedarme con las variables numéricas aprovechare la función is.numeric con la función keep, luego aplico map con la función mean para la media, en el caso de la desviación típica aplico con el map una función sin nombre que calcula la desviación típica directamente con la corrección de Bessel aplicada

Texto

Descripción generada automáticamente con confianza baja

**Resultado:**

**Texto

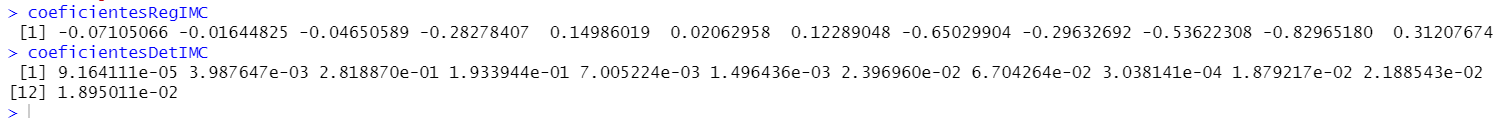
Descripción generada automáticamente con confianza baja**

**Los resultados obtenidos me parecen bastante acordes por lo que puedo deducir que mi calculo esta bien, en la media podemos apreciar valores que cuadran perfectamente y en la desviación apreciamos una dispersión coherente respecto a la media para cada variable.**

## **5. Calcula los coeficientes de regresión y el coeficiente de determinación para las 12 regresiones lineales unidimensionales.**

Para hallar los coeficientes he tenido que primero aislar las variables independientes, luego crear una función que me calculase cada coeficiente con la entrada del dataset, la variable dependiente y la variable independiente, y finalmente con un map he llamado a estas funciones creadas por mi .

## **Texto, Aplicación Descripción generada automáticamente**



La salida aparenta ser correcta , hay valores negativos que indican un decrecimiento de la regresión y viceversa.

## **6. Representa los gráficos de dispersión en el caso de variables numéricas y los boxplots en el caso de variables cualitativas. En el caso de las variables numéricas (y sólo en ese caso) el gráfico debe tener sobreimpresa la recta de regresión simple correspondiente.**

## **7. Separa el conjunto original de datos en tres conjuntos de entrenamiento, test y validación en las proporciones 60%, 20% y 20%.**

## **8. Selecciona cuál de las 12 variables sería la que mejor explica la variable IMC de manera individual, entrenando con el conjunto de entrenamiento y testeando con el conjunto de test.**

## **9. Selecciona un modelo óptimo lineal de regresión, entrenando en el conjunto de entrenamiento, testeando en el conjunto de test el coeficiente de determinación ajustado y utilizando una técnica progresiva de ir añadiendo la mejor variable.**

## **10.Evalúa el resultado en el conjunto de validación.**

## **11. Lee el dataframe de evaluación que te habrá llegado (eval.csv) y utiliza el modelo creado para añadirle una nueva columna con el valor de la variable IMC y, a continuación, otra columna con el valor de la variable Peso. Salva el resultado como evalX.csv para enviarlo como parte de la solución al trabajo.**

## **12.Expresa tus conclusiones sobre el modelo creado. Incluyendo, al menos, respuestas a las siguientes cuestiones:**

## **• Que utilidad podría tener el modelo matemático que has obtenido.**

## **• Que se puede deducir a partir del modelo sobre la relación entre las variables. • Problemas que has encontrado en el desarrollo.**

## **• Qué te ha llamado la atención en el proceso.**

## **• Qué más podría hacerse y cómo plantearlo.**