

## STATISTICAL ESSENTIALS

### Laboratorio de Aplicación N°03:

Se entrega el siguiente set de datos, grasacorporal.csv, el cual es un dataset que contiene las estimaciones del porcentaje de grasa corporal determinado por pesaje de ciertas **características bajo el agua y varias mediciones de circunferencia corporal para 252 hombres**. La medición precisa de la grasa corporal es en algunos casos un poco difícil de cuantificar además de costosa y es deseable tener métodos sencillos para estimar la grasa corporal que sean precisos y sobre todo automáticos. Para mayor detalle ingresar: <http://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Bodyfat.html>.

Las variables que se disponibilidad son:

Variable	Descripción de la Variable
densidad	Densidad determinada por pesaje bajo el agua
edad	Años de edad)
peso	Peso libras)
estatura	Altura (pulgadas)
cuello	Circunferencia del cuello (cm)
pecho	Circunferencia del pecho (cm)
abdomen	Abdomen 2 circunferencia (cm)
cadera	Circunferencia de la cadera (cm)
muslo	Circunferencia del muslo (cm)
rodilla	Circunferencia de la rodilla (cm)
tobillo	Circunferencia del tobillo (cm)
biceps	Bíceps (extendido) circunferencia (cm)
antebrazo	Circunferencia del antebrazo (cm)
muneca	Circunferencia de la muñeca (cm)
grasacorp	Porcentaje de grasa corporal de la ecuación de Siri (1956)

### **Entregables del Laboratorio:**

1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio (Objetivos de ciencias de la salud) bien definidos.



2.- ¿Qué tipo de variables se utilizan en el problema? De un ejemplo de cada uno de los tipos de variables estudiados en clase.

3.- Realizar un análisis completo de gráficos multivariados de acuerdo a las tipologías de variables encontradas. Si corresponde que todas las variables son de la misma naturaleza podría plantearse un gráfico de matriz de asociaciones.

4.- Se puede o podría responder la pregunta de negocio, ¿Cuáles son las características o factores que más influyen en el porcentaje de grasa corporal de los pacientes? Justifique su respuesta.

5.- Realizar un informe sobre las principales medidas de tendencia central y de dispersión. Que conclusión pueden sacar respecto al porcentaje de grasa corporal en los pacientes.

6.- Evaluar los outliers o casos atípicos univariados para las variables cuantitativas. Muestre los principales gráficos para identificarlos.

7.- Evaluar la colinealidad o multicolinealidad en las variables, identificando aquellas en las cuáles su coeficiente de asociación supere el **punto de corte de  $\pm 0.3$ .**

8.- Aplicar un ACP para las variables independientes (Excluir el target: grasacorp) y mencione con cuántos componentes o nuevas variables se quedaría adjuntando la evidencia.

9.- Después de aplicar al ACP, muestre y nombre las nuevas variables dependiendo de las cargas factoriales estudiadas en clases. Muestre cuáles son las principales variables representadas en los componentes principales.

10.- Al set de datos inicial fusiónelo con las puntuaciones factoriales en un nuevo dataset: grasacorporal\_mod.csv y adjúntelo en la presentación de su tarea.

### **Entregables del Laboratorio:**

- **Archivo pdf o word** con el desarrollo de las preguntas.
- **Set de datos** modificado para su corroboración.

