

Entregable 5: Model Lineal General

El fitxer de dades conté estimacions del percentatge de greix corporal determinat pesant sota l'aigua, a més de diferents mesures de circumferències corporals per 200 individus. Una mesura acurada del greix corporal és costosa i és desitjable disposar de mètodes d'estimació amb menys cost, basats en mesures indirectes.

D'altra banda, aquests individus, que 3 mesos abans tenien un percentatge de greix similar, havien estat assignats a quatre grups: un de control i tres tipus de tractaments que combinen dieta amb pautes d'entrenament físic de diferents nivell. Al cap de 3 mesos, es recull el seu estat actual.

Els camps són:

- **Density:** Densitat determinada a partir del pesatge submarí (g/cm^3)
- **bodyfat:** Percentatge de greix corporal de l'equació de Siri (1956)
- **Age:** Edat (anys)
- **Weight:** Pes (lliures)
- **Height:** Alçada (polzades)
- **Neck:** Circumferència del coll (cm)
- **Chest:** Circumferència toràcica (cm)
- **Abdomen:** Circumferència abdominal (cm)
- **Hip:** Circumferència de maluc (cm)
- **Thigh:** Circumferència de la cuixa (cm)
- **Knee:** Circumferència de genoll (cm)
- **Ankle:** Circumferència del turmell (cm)
- **Biceps:** Circumferència de bíceps (estesa) (cm)
- **Forearm:** Circumferència de l'avantbraç (cm)
- **Wrist:** Circumferència de canell (cm)
- **Diet:** Grup de tractament (A-(No Diet),B,C,D)

Hi ha dos objectius:

1. Trobar una model estadístic que permeti predir la mesura del greix corporal (**Density** o **bodyfat**) en base a algunes de les mesures indirectes recollides.
2. Comparar els grups de tractament en quant a la mesura de greix corporal.

1 Anàlisis a realitzar

1. Per a la variable resposta proposada en l'apartat 1) de l'enunciat individual i amb el predictors indicats:
 - (a) Anomenem *M1* el model complet. Pren nota del valor de la *F* i del p-valor del test Omnibus d'aquest model, així com dels coeficients estimats i de la R^2 del model.
 - (b) Analitza la possible existència de multicolinealitat a partir dels VIFs del model
 - (c) Simplifica el model eliminant les variables predictores no significatives una a una ($\alpha = 0.05$). Anomenem aquest model *M2*.
 - (d) Valida les premises del model *M2*, analitzant els residus estudentitzats per veure si hi ha tendència i/o valors estranys. Analitzeu també els valors influents.
 - (e) Fent servir el model *M2*, calcula la predicció puntual i els intervals de predicció i confiança pels valors de les covariables indicats en l'enunciat individual.
2. Considerem ara el model per a la variable indicada en l'apartat 2) de l'enunciat individual respecte només al factor tractament **Diet**
 - (f) Fent servir el contrast de tipus baseline **contr.treatment**, calcula el model associat. Anomenem aquest model *M3*.
3. Considerem ara el model en el que s'inclou només la covariable **Weight** i el factor **Diet** amb la seva interacció
 - (g) Calcula el model associat. L'anomenem *M4*
4. Simplifica el model *M4* traient els termes no significatius. Anomenem aquest model *M5*

Batlle Casellas, àlex

- e) Per a la regressió Lineal Múltiple, considera la variable resposta **bodyfat** fent servir com a variables predictores:
Weight, Neck, Chest, Abdomen, Knee

El cas proposat per predir és:

(Weight=175,3, Neck=38, Chest=100,7, Abdomen=91,4, Knee=38,5)

Per a la segona part, considera la variable resposta bodyfat respecte el factor tractament Diet i la covariable Weight

Les preguntes que haureu de contestar al qüestionari d'Atenea, les teniu a continuació agrupades pels apartats de l'exercici.

1.

1) Coeficient de determinació R^2 del model total (M1):
2) Variància residual del model M1:
3) Valor de l'ordenada a l'origen (intercept) del model M1:
4) Estadístic F del test òmnibus del model M1:
5) P-valor de la taula de coeficients (summary) per a la primera variable explicativa (Weight) en el model M1:
6) Màxim dels VIFs en el model M1:
7) Número de variables predictores en el model M2:
8) P-valor del test que compara els models M1 i M2:
9) Percentatge de residus estudentitzats amb valor absolut superior a 2 en el model M2:
10) Valor màxim de les distàncies de Cook del model M2:
11) Límit inferior de la regió de predicció al 95% del cas proposat (model M2):
12) Límit superior de la regió de predicció al 95% del cas proposat (model M2):
13) Límit inferior de l'interval de confiança al 95% del cas proposat (model M2):
14) Límit superior de l'interval de confiança al 95% del cas proposat (model M2):

2.

15) Valor de l'intercept del model només amb la variable Diet amb contrast tipus treatment (model M3):
16) Coeficient de determinació R^2 del model M3:
17) Error Estàndard dels coeficients de les variables indicadores (dummy) per a les dietes en el model M3
18) Límit inferior de l'interval de confiança al 95% per la mitjana del grup Diet="A" (emmeans en el model M3):
19) Límit superior de l'interval de confiança al 95% per la mitjana del grup Diet="A" (emmeans en el model M3):
20) Desviació estàndard residual del model M4:
21) R^2 Ajustada pel model M4:
22) Estadístic F per decidir la significació de la interacció en el model M5: