

# Capitolo 2

## Ancora sul modello lineare semplice

### 2.1 Analisi dei dati BRAINBOD.DAT

Consideriamo il file `brainbod.dat` che contiene i seguenti dati sul peso del corpo (kg) e peso del cervello (g) di 15 mammiferi terrestri.

species	bodywt	brainwt
afeleph	6654.00	5712.00
cow	465.00	423.00
donkey	187.00	419.00
man	62.00	1320.00
graywolf	36.33	119.50
redfox	4.24	50.40
narmadillo	3.50	10.80
echidna	3.00	25.00
phalanger	1.62	11.40
guineapig	1.04	5.50
eurhedghog	.79	3.50
chinchilla	.43	4.00
ghamster	.12	1.00
snmole	.06	1.00
lbbat	.01	.25

Acquisiamo il file (**NB**: specificare il percorso per arrivare al file!), dopo aver ripulito la memoria degli oggetti

```
> rm(list=ls())
> brainbod <- read.table("I:/modelli/brainbod.dat" , header=T)
```

**NB**: l'opzione `header=T` indica che la prima riga del file deve essere utilizzata per dare il nome alle variabili.

Per usarlo nelle analisi successive, diamo il comando:

```
> attach(brainbod)
```

Facciamo qualche analisi grafica dei dati.

```
> hist(bodywt)
> boxplot(bodywt)
> hist(brainwt)
> boxplot(brainwt)
```

Per creare un grafico dell'andamento del peso del cervello rispetto al peso del corpo, si usa:

```
> plot(bodywt, brainwt)
> identify(bodywt, brainwt, species)
```

Questi grafici non risultano particolarmente leggibili, perché il peso dell'elefante è così elevato che “schiaccia” tutti gli altri punti. Probabilmente è meglio esplorare i dati su scala logaritmica:

```
> hist(log(bodywt))
> boxplot(log(bodywt))
> hist(log(brainwt))
> boxplot(log(brainwt))
> plot(log(bodywt), log(brainwt))
```

Ora i grafici appaiono più interessanti. Si intuisce la presenza di una relazione lineare tra i logaritmi del peso e del peso del cervello, con forse un paio di punti un po' distanti dalla retta interpolante.

**Esercizio:** stimare i coefficienti della retta di regressione  $\log(\text{brainwt}) = \alpha + \beta * \log(\text{bodywt})$  e calcolare gli intervalli di confidenza di livello 0.99 per  $\alpha$  e  $\beta$ .