

# Mètodes Bayesians: Sessions Pràctiques

---

## Sessió 5: Regressió lineal

---

Objectius:

- Implementar un model de regressió lineal amb el Win BUGS

**Exercici 5.1. Regressió lineal simple. Pes del cervell.** Tenim dades de diferents animals, on tenim el pes del cervell i el pes del cos, i volem veure si a partir del pes del cos podem conèixer el pes del cervell.

- a) Formula des de la perspectiva freqüentista un model de regressió, i dóna un interval de predicció per a un animal amb un pes del cos de 100kg.
- b) Repeteix el mateix des de la perspectiva bayesiana
  1. Formula el model
  2. Defineix distribucions a priori per als paràmetres sota el supòsit de no informació
  3. Mostra les distribucions a posteriori per als paràmetres
  4. Dona un interval de predicció per a una animal de 100 kg.
  5. T'atreviries a posar distribucions dels paràmetres més informatives ?

Les dades estan als fitxers *Cervell.xls* (el pes del cos està expressat en kg., i el pes del cervell en grams).

**Exercise 5.2 Weight and height.** In the file *WeightHeight.txt* there are data about the weight, height and sex of several students. Implement the next models:

- a) A model where the weight is explained as a function of the height.
- b) A model where the weight is explained as a function of the height and the sex.
- c) A model where the weight is explained as a function of the height, the sex and their interaction.

**Exercici 5.3 Notes finals d'Inferència Bayesiana.** Els professors de l'assignatura d'Inferència Bayesiana que s'imparteix al Màster d'estadística volen saber si hi ha diferències significatives en el rendiment dels estudiants a l'assignatura en funció del sexe de l'estudiant. Les notes finals a l'assignatura del curs 2007-2008, excloent els no presentats són:

| Nois | Noies |
|------|-------|
| 9.6  | 6.1   |
| 7.0  | 9.1   |
| 5.0  | 8.8   |
| 8.0  | 5.7   |
| 8.4  | 8.9   |
| 6.4  | 6.1   |
|      | 6.5   |

Per això formulen el model estadístic:

$$y | \beta_0, \beta_1, \sigma \sim \text{Normal}(\beta_0 + \beta_1 \text{sexe}, \sigma),$$

on la variable *sexe* val 1 per als nois i 0 per les noies.

Respon a les següents preguntes:

- Escriu el model bayesià que utilitzaries, raonant el perquè de les distribucions a priori escollides.
- Actualitza el model i dibuixa les distribucions a posteriori marginals de cada paràmetre i les distribucions bivariants de totes les parelles de paràmetres.
- Calcula per a cada paràmetre la mitjana, la desviació i els percentils 5%, 25%, 50%, 75% i 95%.
- A partir dels resultats del model creus que hi ha diferències significatives entre els nois i les noies?.
- Hi ha una alumna que no va poder fer l'examen per que es va quedar atrapada a un ascensor. Com calcularies un interval de credibilitat del 90% per la nota final d'aquesta alumna?
- Sabent que la nota final de l'assignatura es calcula com:

$$\text{Nota Final} = 0.1 * \text{ExamenParcial} + 0.2 * \text{Pràctiques} + 0.2 * \text{Projecte} + 0.5 * \text{ExamenFinal}$$

i que l'alumna que no ha fet l'examen té les següents notes:

|                 |     |
|-----------------|-----|
| Examen Parcial: | 7.0 |
| Pràctiques:     | 8.5 |
| Projecte:       | 7.5 |

donar una predicció puntual per la nota de l'examen final i calcular un interval de credibilitat del 90% per aquesta nota.

**Exercici 5.4. Yield of Potatoes.** A researcher is investigating the relationship between yield of potatoes ( $y$ ) and level of fertilizer ( $x$ ). She divides a field into eight plots of equal size and applied fertilizer at a different level to each plot. The level of fertilizer and yield for each plot is recorded below:

| Fertilizer level | Yield         |
|------------------|---------------|
| $x$              | $y$           |
| 1                | 25            |
| 1.5              | 31            |
| 2                | 27            |
| 2.5              | 28            |
| 3                | 36            |
| 3.5              | 35            |
| 4                | Not available |
| 4.5              | 34            |

Suppose that we know that yield given the fertilizer level is  $Normal(\beta_0 + \beta_1 x, \sigma)$ .

- Using non informative priors for the parameters find the posterior distribution of  $\beta_1$ .
- Find a 95% credible interval for  $y$  given  $x=4$ .

**Exercici 5.5. Regressió lineal amb variable categòrica.** Using the data in BirthWeight.txt from Hosmer and Lemeshow (1989), build a model for the weight of a newborn baby, birhtweight, using as explanatory variables the age of the mother, age, and a variable indicating her smoking status during pregnancy, smoke, which is 1 if the mother smokes and 0 if the mother doesn't smoke. What does the model tell you about the birth weight of babies?

**Exercici 5.6. Regressió lineal múltiple.** Al fitxer de dades *graduates.txt* hi trobem dades d'una mostra de 22 universitats, en que per a cada una tenim el percentatge promig de graduats en 6 anys, el percentatge promig de SAT (similar a la nota de selectivitat, SAT pot prendre valors entre 600 i 2400), el cost promig per alumne i el tipus d'escola (1 de nois o noies, 0 mixta).

- Formula des de la perspectiva bayesiana un model de regressió lineal múltiple utilitzant les tres variables explicatives, escriu el model en codi WinBugs,
- Selecciona un model utilitzant els següents criteris:
  - les a posteriors dels paràmetres,
  - el criteri d'informació de la deviança, DIC.
- Interpreta el model.