UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Problem Sheet 1 **Semester 1, 2022–2023**

Work on Questions 1, 2 and 6 in the workshop.

1. The observations y_1, y_2 and y_3 were taken on the random variables Y_1, Y_2 and Y_3, where

$$Y_1 = \theta + e_1 \quad Y_2 = 2\theta - \phi + e_2 \quad Y_3 = \theta + 2\phi + e_3$$

$$\mathrm{E}(e_i) = 0, \quad \mathrm{var}(e_i) = \sigma^2 \ (i = 1, 2, 3), \quad \mathrm{cov}(e_i, e_j) = 0 \ (i \neq j).$$

   Minimise the function

$$Q = (y_1 - \theta)^2 + (y_2 - (2\theta - \phi))^2 + (y_3 - (\theta + 2\phi))^2,$$

   with respect to θ and ϕ, and thus find the least squares estimates of θ and ϕ.

2. A set of n independent pairs of regression data $(x_1, y_1), \ldots, (x_n, y_n)$ are observed. Assuming that the data can be adequately modelled by

$$\mathrm{E}(Y_i|x_i) = \gamma + \beta(x_i - \bar{x}) \ \text{ and } \ \mathrm{var}(Y_i|x_i) = \sigma^2 \ (i = 1, \ldots, n)$$

   obtain the least squares estimators of γ and β, and also their variances and covariance.

   In an experiment with $n = 18$ it was found that $\bar{x} = 20$, $\bar{y} = 5$, $\frac{1}{n}\sum_{i=1}^n x^2 = 3456$, $\frac{1}{n}\sum_{i=1}^n xy = 576$ and $\frac{1}{n}\sum_{i=1}^n y^2 = 352$. Assuming Normality

   (a) Show that the hypothesis test of $\beta = 0$ is rejected at the 5% significance level.

   (b) Obtain a two-sided 95% symmetric confidence (prediction) interval for the value of a *future* observation on Y to be taken at the value $x^* = 34$.

3. In a simple linear regression model with

$$\mathrm{E}(Y|x) = \gamma + \beta(x - \bar{x}) \quad \text{and} \quad \mathrm{var}(Y|x) = \sigma^2 \text{ (constant)}$$

   show that the expectation of the *residual sum of squares*

$$RSS = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2, \quad \text{where } \widehat{Y}_i = \widehat{\gamma} + \widehat{\beta}(x_i - \bar{x}) = i\text{th fitted value},$$

   is $(n-2)\sigma^2$. Deduce that $\widehat{\sigma}^2 = \frac{RSS}{n-2}$ is an unbiased estimator of σ^2.

4. Consider a linear regression model in which responses Y_i are uncorrelated and have expectations βx_i and common variance σ^2 $(i = 1, \ldots, n)$.

   (a) Show that the least squares estimator of β is

$$\widehat{\beta} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2}.$$

   (b) Derive the expectation and variance of $\widehat{\beta}$ under the model.

(c) An alternative to the least squares estimator in this case is

$$\widetilde{\beta} = \frac{\sum_i Y_i}{\sum_i x_i} = \frac{\bar{Y}}{\bar{x}}.$$

Show that this estimator is unbiased (if $\bar{x}$ is not zero), and obtain its variance.

(d) Show that the variance of $\widetilde{\beta}$ is at least as large as the variance of the least squares estimator, $\widehat{\beta}$.
[Hint: Use $\sum_{i=1}^{n}(x_i - \bar{x})^2 \geq 0 \Leftrightarrow \sum_{i=1}^{n} x_i^2 \geq n\bar{x}^2$.]

5. For a linear regression model in which responses Y_i are uncorrelated and have expectations βx_i and common variance σ^2 $(i = 1, \ldots, n)$, consider an estimator $\check{\beta}$ of β which is a linear function of the responses, i.e. which has the form

$$\check{\beta} = \sum_i a_i Y_i,$$

with $a_1, \ldots, a_n$ constants.

(a) Show that $\check{\beta}$ is unbiased for β if and only if $\sum_i a_i x_i = 1$, and use a Lagrange multiplier to find the values of $a_1, \ldots, a_n$ which minimize the variance of $\check{\beta}$ among unbiased estimators of β.

(b) Verify that the resulting estimator is the least squares estimator in Question 4.
[Note: This minimum-variance property is a particular case of a more general result about least squares estimators called the *Gauss-Markov Theorem*.]

6. To test a type of plastic conduit pipe, sections 30 cm long were held in a rig, and a striker with a given weight was dropped on each of them from a height of 2 metres. Sections were tested at each of 5 weights: the numbers tested and the numbers which broke were as follows.

Weight of striker (kg)	5.0	5.2	5.5	5.7	6.0
Number of sections tested	12	12	12	12	12
Number which broke	2	6	9	11	12

Plot the data and examine whether or not the two variables appear to be roughly linearly related. If they are not, try to transform one or both so that the relationship between the transformed variables is closer to linearity.

7. Suppose that observations $y_1, \ldots, y_n$ are drawn independently from a distribution with probability density function $f(y;\theta)$.

For (a) the exponential distribution: $f(y;\theta) = \theta e^{-y\theta}$, $y > 0$, $\theta > 0$,
and (b) the binomial distribution: $f(y;\theta) = \binom{m}{y}\theta^y(1-\theta)^{m-y}$, $y = 0, 1, \ldots, m$, $0 \leq \theta \leq 1$,
show from first principles that

$\mathrm{E}(U) = 0$, and

$\mathrm{var}(U) = \mathrm{E}(U^2) = -\mathrm{E}(U')$,

where $l(\theta)$ is the log likelihood function for parameter θ, $U = \frac{dl}{d\theta}$ is the score function for parameter θ, and $U' = \frac{dU}{d\theta}$.