

# MATH 523 - EXERCISES 1

*These exercises are not for assessment.*

1 In a linear model with a single continuous predictor,  $X$ , the mean model

$$\mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x \quad (1)$$

is considered. Estimation of the parameters  $(\beta_0, \beta_1)^\top$  via least-squares is carried out under an assumption of homoscedastic errors (ie a constant error variance  $\sigma^2$ ), from  $n$  independent pairs  $(x_i, y_i), i = 1, \dots, n$ .

(a) Compare the forms of the estimators, and their statistical properties, derived from the model in (1) and from the model using instead the centered covariate  $z = x - \bar{x}$ .

(b) Suppose now the model

$$\mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (2)$$

is considered. Does centering the covariate have a similar effect as in (a) ? Justify your answer.

(c) Finally, consider the model

$$\mathbb{E}[Y|X = x] = \beta_0 + \beta_2 x^2. \quad (3)$$

Comment on the difficulty on interpreting the results from this model, in light of the results from (a) and (b).

2 Using the `glm` function in R (or another suitable computing package), fit a Poisson GLM (regression) model to the data in the `Accidents.csv` and interpret the results. The response data  $y$  record the number of reports of fatal traffic accidents in each of 48 districts of a city recorded in a five year period, and the two predictors are

- $x_1$ : a continuous measure of deprivation (increasing  $x_1$  corresponding to more deprivation).
- $x_2$ : a factor predictor taking two levels
  - 0: stricter than normal speed restrictions (that is, maximum allowable speed is lower than normal).
  - 1: normal speed restrictions

Use different combinations of the predictors, and base your conclusions on the parameter estimates and standard errors.

*Example code to fit the Poisson regression model can be found on page 2. The `glm` function inherits many of the properties from the linear model function `lm`. To read the data from the file into a data frame `data.set`, use the command*

```
data.set<-read.csv("Accidents.csv",header=T)
```

## R COMMANDS FOR FITTING GENERALIZED LINEAR MODELS.

The function `glm` is the basic fitting model for generalized linear models in R. It performs **maximum likelihood estimation** of the coefficient parameters  $\beta$ . Specifically, the default Poisson model assumes that for data point  $i$

$$\mathbb{E}[Y_i|X_i = x_i] = \exp\{x_i\beta\} = \exp\left\{\sum_{j=1}^p x_{ij}\beta_j\right\} = \mu_i$$

say, and a consequence of this is that  $\text{Var}[Y_i|X_i = x_i] = \mu_i$  also.

If the data are contained in vectors `Y`, `X1` and `X2`, then the command

```
glm(Y~X1+X2,family=poisson)
```

fits the model

$$1 + X_1 + X_2$$

and specifies that  $Y_i \sim \text{Poisson}(\mu_i)$ , with

$$\log \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

- To remove the intercept:

```
summary(glm(Y~-1+X1+X2,family=poisson))
```

- To fit an interaction:

```
summary(glm(Y~X1+X2+X1:X2,family=poisson))
```

or

```
summary(glm(Y~X1*X2,family=poisson))
```

- The parameter estimates and standard errors etc. can be obtained using the `summary` function

```
summary(glm(Y~X1+X2,family=poisson))
```

- To obtain the parameter estimates alone, use `coef`:

```
coef(glm(Y~X1+X2,family=poisson))
```

- To obtain the fitted values, use `fitted`:

```
fitted(glm(Y~X1+X2,family=poisson))
```

- The `glm` function can be assigned to an object in R:

```
model1<-glm(Y~X1+X2,family=poisson))
summary(model1)
coef(model1)
fitted(model1)
```

- If the data are in an R data frame, called `data.set`, then `glm` can be used on the data frame

```
glm(Y~X1+X2,family=poisson,data=data.set)
```