MATH 423/533 – BIVARIATE NORMAL MODELLING

We may decide to treat the predictor x as a random variable, and make a bivariate Normal distribution assumption for the data pairs $(x_i, y_i), i = 1, \ldots, n$. We specify that

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \right) \equiv \text{Normal}\left(\mu, \Sigma\right).$$

Writing

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

the joint density is

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

- σ_{XY} is the covariance parameter
- ρ is the correlation parameter.

We may factorize the joint density

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$$

where $X \sim \text{Normal}(\mu_X, \sigma_X^2)$ and

$$Y|X = x \sim \text{Normal}\left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_X), \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \right) \equiv \text{Normal}\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2) \right)$$

Equating the conditional expectation of Y given $X = x$

$$\mathbb{E}_{Y|X}[Y|x] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

with a simple linear regression

$$\mathbb{E}_{Y|X}[Y|x] = \beta_0 + \beta_1 x$$

we identify

$$\beta_0 = \mu_Y - \mu_X \rho \frac{\sigma_Y}{\sigma_X} \qquad\qquad \beta_1 = \rho \frac{\sigma_Y}{\sigma_X}.$$

The sample correlation is

$$r = \frac{\sum\limits_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^n (x_i - \overline{x})^2 \sum\limits_{i=1}^n (y_i - \overline{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}\text{SS}_\text{T}}}$$

so that

$$\widehat{\beta}_1 = \left(\frac{\text{SS}_\text{T}}{S_{xx}} \right)^{1/2} r \qquad \Longrightarrow \qquad r^2 = \frac{\text{SS}_\text{R}}{\text{SS}_\text{T}} = R^2.$$

We may test the hypothesis $\text{H}_0 : \rho = 0$ using the statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

which is compared against the relevant quantile of the Student$(n-2)$ distribution.

```
set.seed(234)
library(MASS)
n<-10000
muX<-50;sigmaX<-20;muY<-100;sigmaY<-50;rho<-0.5
Sig<-matrix(c(sigmaX^2,rho*sigmaX*sigmaY,rho*sigmaX*sigmaY,sigmaY^2),2,2)
XY<-mvrnorm(n,mu=c(muX,muY),Sigma=Sig)
be0<-muY-rho*(sigmaY/sigmaX)*muX;be1<-rho*(sigmaY/sigmaX)
```

Fitting a linear model to the simulated bivariate (x_i, y_i) pairs yields estimates that are close to the true values.

```
> c(be0,be1)
[1] 37.50  1.25
> coef(lm(XY[,2] ~ XY[,1]))
(Intercept)     XY[, 1]
  37.269659    1.253241
```

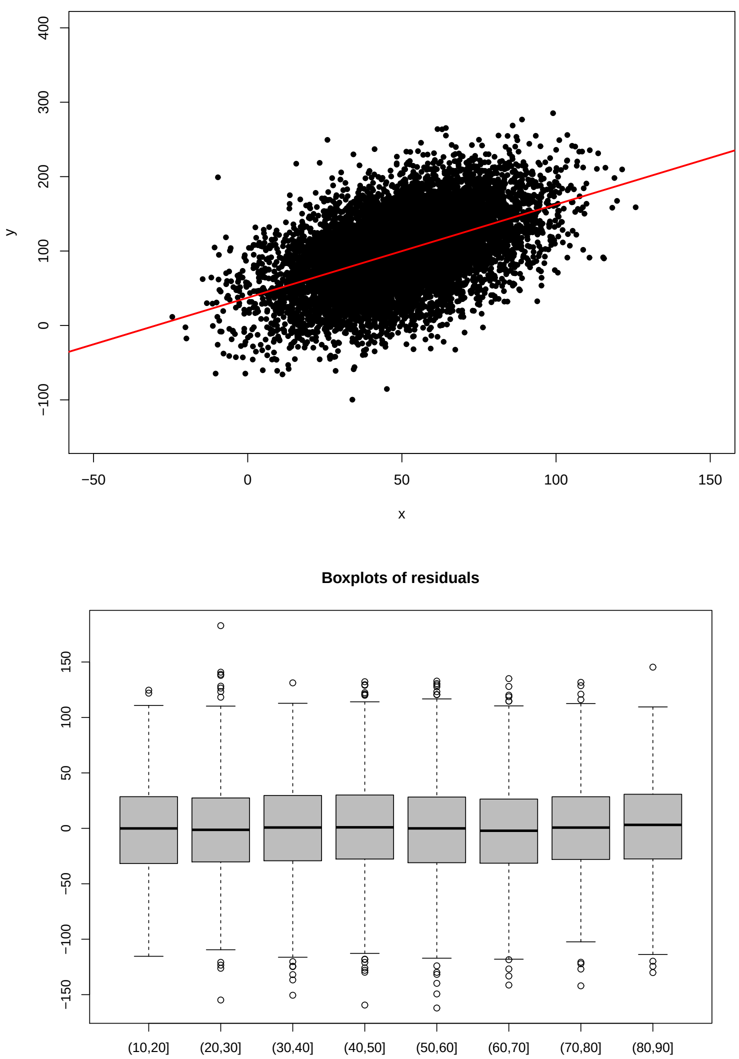

Figure 1: The residual plots indicate that the fit of the model is adequate.