Problem Set 1.

Question 1.

The correlation coefficient between X and Y (r_{XY}) is given by

$$r_{XY} = \frac{n\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sqrt{\sum X^2 - (\sum X)^2/n}\sqrt{\sum Y^2 - (\sum Y)^2/n}},$$

where we used the fact that

$$\sum xy = \sum XY - (\sum X)(\sum Y)/n.$$

Multiplying both the numerator and the denominator by n gives

$$r_{XY} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}.$$

Problem Set 2.

Question 1.

Solve exercise 1.12 in Johnston and Dinardo.

In what follows we will denote $X' = aX + b$, and $Y' = cY + d$.

The correlation coefficient between X' and Y' (r') is given by

$$r' = \frac{\sum x'y'}{\sqrt{\sum (x')^2}\sqrt{\sum y'^2}}.$$

It can be seen that

$$x' = X' - \overline{X'} = ax, \text{ and } y' = cy.$$

Thus, we have

$$r' = \frac{\sum axcy}{\sqrt{\sum (ax)^2}\sqrt{\sum (cy)^2}} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = r.$$

Problem Set 3.

Question 1.

Note that

$$d[\ln(y)] = \frac{\partial[\ln(y)]}{\partial y} dy = \frac{1}{y} dy, \text{ and } d[\ln(x)] = \frac{1}{x} dx.$$

Thus, we have

$$\frac{d[\ln(y)]}{d[\ln(x)]} = \frac{x}{y}\frac{dy}{dx} = \frac{dy/y}{dx/x} = e_x^y.$$

where e_x^y denotes the elasticity of y with respect to x.

Question 2.

The main equation is:

$$\frac{100}{100 - y} = \alpha + \frac{\beta}{x}.$$

The two asymptotes are:

$$\begin{aligned} y &\to \infty \Rightarrow 0 = \alpha + \frac{\beta}{x} \Rightarrow x = -\frac{\beta}{\alpha}, \\ x &\to \infty \Rightarrow \frac{100}{100 - y} = \alpha \Rightarrow y = 100(1 - \frac{1}{\alpha}). \end{aligned}$$

Next, we denote $y' = \frac{100}{100-y}$ and $x' = \frac{1}{x}$ and we estimate the following regression:

$$y' = \alpha + \beta x'.$$

The estimates of α and β are $a = 2.068(.159)$ and $b = 16.266(1.323)$ with standard errors inside the parentheses.

Problem Set 4.

Question 1a.

The likelihood function is given by

$$\begin{aligned} L(\theta; x) &= \theta^{x_1}(1-\theta)^{1-x_1} \times \cdots \times \theta^{x_n}(1-\theta)^{1-x_n} = \\ &\theta^{\sum x}(1-\theta)^{n-\sum x}. \end{aligned}$$

The log-likelihood function is given by

$$\ln[L(\theta; x)] = \sum x \ln(\theta) + (n - \sum x) \ln(1 - \theta).$$

Thus, the first derivative with respect to θ is given by

$$\frac{\partial\{\ln[L(\theta;x)]\}}{\partial\theta} = \frac{\sum x}{\theta} - \frac{(n-\sum x)}{1-\theta}.$$

We set the first derivative equal to 0 and we solve with respect to θ:

$$\widehat{\theta} = \frac{\sum x}{n} = \overline{x}.$$

Question 4.

We express the regression in deviations from the means:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

The vector b with the estimated parameters is given by

$$b = (x'x)^{-1}x'y,$$

where

$$x'x = \begin{bmatrix} \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_1x_2 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 \end{bmatrix} = \begin{bmatrix} 10 & 10 & 5 \\ 10 & 30 & 15 \\ 5 & 15 & 20 \end{bmatrix},$$

and

$$x'y = \begin{bmatrix} \sum x_1y \\ \sum x_2y \\ \sum x_3y \end{bmatrix} = \begin{bmatrix} 7 \\ -7 \\ -26 \end{bmatrix}.$$

Thus

$$b = \begin{bmatrix} 10 & 10 & 5 \\ 10 & 30 & 15 \\ 5 & 15 & 20 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ -7 \\ -26 \end{bmatrix} = \begin{bmatrix} 1.4 \\ 0.2 \\ -1.8 \end{bmatrix}.$$

Next, the RSS are given by

$$\text{RSS}=e'e = (y - xb)'e = y'e - b'x'e = y'e,$$

and

$$y'e = y'(y - xb) = y'y - y'xb = y'y - b'x'y = 60 - 55.2,$$

since

$$b'x'y = \begin{bmatrix} 1.4 & 0.2 & -1.8 \end{bmatrix} \begin{bmatrix} 7 \\ -7 \\ -26 \end{bmatrix} = 55.2,$$

and $y'y = \sum y^2 = 60$.

Thus, it follows that

$$s^2 = \frac{\text{RSS}}{n-k} = \frac{4.8}{20} = 0.24.$$

Finally, the variance-covariance matrix of the three estimates is

$$s^2(x'x)^{-1} = 0.24 \begin{bmatrix} 0.15 & -0.05 & 0 \\ -0.05 & 0.07 & -0.04 \\ 0 & -0.04 & 0.08 \end{bmatrix}.$$

Testing the hypothesis $\beta_1 = 1$ gives:

$$t = \frac{b_1 - 1}{\text{SE}(b_1)} = \frac{0.4}{\sqrt{0.24 \times 0.15}} = 2.108.$$

Problem Set 5.

Question 4.

The relationship between the β's and the α's is:

$$\beta_2 = \alpha_2 + \beta_3,\ \beta_3 = \alpha_3 \Rightarrow \beta_2 = \alpha_2 + \alpha_3.$$

Testing the hypothesis that $\beta_2 = 0$ (average speed has no effect on fatalities) gives:

$$t = \frac{(\alpha_2 + \alpha_3)}{\sqrt{\text{var}(\alpha_2 + \alpha_3)}} = \ldots = -0.226.$$

It is insignificant so we accept the null.