7. It is claimed in the text that in the case of simple linear regression of Y onto X, the R^2 statistic (3.17) is equal to the square of the correlation between X and Y (3.18). Prove that this is the case. For simplicity, you may assume that $\bar{x} = \bar{y} = 0$.

7. For a simple linear regression with $\bar{x}=\bar{y}=0$, $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=0$, $\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$ ← Official Ans

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{\beta}_1x_i)^2}{\sum_{i=1}^{n}(y_i)^2}$$

$$\operatorname{cov}(x,y)=\frac{1}{n-1}\sum_{i=1}^{n}(x_iy_i)$$

$$\operatorname{var}(x)=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2;\quad \operatorname{var}(y)=\frac{1}{n-1}\sum_{i=1}^{n}y_i^2$$

$$=1-\frac{\sum_{i=1}^{n}y_i^2-2\hat{\beta}_1\sum_{i=1}^{n}x_iy_i+\hat{\beta}_1^2\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(y_i)^2}$$

$$=1-1+\frac{2\hat{\beta}_1\sum_{i=1}^{n}x_iy_i-\hat{\beta}_1^2\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(y_i)^2}$$ (substitute $\hat{\beta}_1$ into it)

$$=\frac{2(\sum_{i=1}^{n}x_iy_i)^2-(\sum_{i=1}^{n}x_iy_i)^2}{\sum_{i=1}^{n}(x_i)^2\sum_{i=1}^{n}(y_i)^2}=\frac{\sum_{i=1}^{n}(x_iy_i)^2}{\sum_{i=1}^{n}(x_i^2)\sum_{i=1}^{n}(y_i)^2} \quad ①$$

We have $r_{xy}=\frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(y)}}\Rightarrow r_{xy}^2=\frac{[\operatorname{cov}(x,y)]^2}{\operatorname{var}(x)\operatorname{var}(y)}=\frac{(n-1)^2(\frac{1}{n-1})^2[\sum_{i=1}^{n}x_iy_i]^2}{\frac{n-1}{n-1}\sum_{i=1}^{n}x_i^2\frac{n-1}{n-1}\sum_{i=1}^{n}y_i^2}=\frac{\sum_{i=1}^{n}(x_iy_i)^2}{\sum_{i=1}^{n}(x_i^2)\sum_{i=1}^{n}(y_i^2)}$ ①

Then, in this case, $R^2=r_{xy}^2$ □

Q7. $$\operatorname{cor}(X,Y)^2=\frac{\left(\sum_{i=1}^{n}x_iy_i\right)^2}{\left(\sum_{i=1}^{n}x_i^2\right)\cdot\left(\sum_{i=1}^{n}y_i^2\right)}$$ ← My Sol

$$R^2=\frac{SSreg}{TSS}=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{\sum_{i=1}^{n}\hat{y}_i^2}{\sum_{i=1}^{n}y_i^2}$$

$$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1\cdot x_i \qquad \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=0 \qquad \hat{\beta}_1=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$$

$$\therefore\ \hat{y}_i=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}\cdot x_i$$

$$\therefore\ \sum_{i=1}^{n}\hat{y}_i^2=\sum_{i=1}^{n}\frac{(\sum_{i=1}^{n}x_iy_i)^2}{(\sum_{i=1}^{n}x_i^2)^2}\cdot x_i^2=\frac{(\sum_{i=1}^{n}x_iy_i)^2}{(\sum_{i=1}^{n}x_i^2)^2}\left(\sum_{i=1}^{n}x_i^2\right)=\frac{(\sum_{i=1}^{n}x_iy_i)^2}{\sum_{i=1}^{n}x_i^2}$$

$$\therefore\ R^2=\frac{(\sum_{i=1}^{n}x_iy_i)^2}{\sum_{i=1}^{n}x_i^2}\cdot\frac{1}{\sum_{i=1}^{n}y_i^2}=\frac{\left(\sum_{i=1}^{n}x_iy_i\right)^2}{\left(\sum_{i=1}^{n}x_i^2\right)\cdot\left(\sum_{i=1}^{n}y_i^2\right)}=\operatorname{cor}(X,Y)^2$$

(a)

In the simple regression model[redacted], suppose the value of the predictor X is replaced by $Z = aX + b$, where $a \neq 0$ and b are constants. Thus, we are considering 2 simple regression models,

← See the "Q4_sol.pdf" for sol.

$$\text{I: } \mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$$

$$\text{II: } \mathrm{E}(Y|Z = z) = \gamma_0 + \gamma_1 z = \gamma_0 + \gamma_1(ax + b)$$

Find the relationships between β_0 and γ_0; between β_1 and γ_1; between the estimates of variance in the 2 regressions, and between the t-tests of $\beta_1 = 0$ and of $\gamma_1 = 0$.

(b) Suppose each value of the response Y is replaced by $V = dY$, for some $d \neq 0$, so we consider the two regression models

$$\text{I: } \mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x$$

$$\text{III: } \mathrm{E}(V|X = x) = \delta_0 + \delta_1 x$$

Find the relationships between β_0 and δ_0; between β_1 and δ_1; between the estimates of variance in the 2 regressions, and between the t-tests of $\beta_1 = 0$ and of $\delta_1 = 0$.

Consider the simple linear regression.

The t-test for slope as a function of the correlation Show that the t-statistic for testing the slope $\beta_1 = 0$ can be written as a function of sample size and the sample correlation r_{xy},

where $SXX = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{\mathrm{SXX}}} = \sqrt{n-2}\,\frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \qquad (2.26)$$