Problem 1. Suppose Y_t is the return on an equity portfolio at month t, and X_t is the market return. Their sample means are, respectively, 0.003 and 0.005. Suppose we run an OLS regression

$$Y_t = a + b\,X_t + e_t.$$

(a) Find the estimates for a and b given that

$$\sum_{t=1}^{60} X_t Y_t = 0.005; \qquad \sum_{t=1}^{60} X_t^2 = 0.004.$$

Answer. We find the value of b estimate as follows:

$$\widehat{b} = \frac{\sum_{t=1}^{60} X_t Y_t - 60\overline{X}\,\overline{Y}}{\sum_{i=1}^{60} X_t^2 - 60\overline{X}^2} = \mathbf{1.64}.$$

Since $\widehat{a} = \overline{Y} - \widehat{b}\overline{X}$, we compute to find $\widehat{a} = \mathbf{-0.0052}$. □

(b) Given that the residual sum of squares (RSS) is 5.8×10^{-5}, compute the t statistic of the a estimate under the hypothesis that $H_0 : a = 0$. What inference can be drawn?

Answer. We need to find $\widehat{\sigma}_e^2$ by

$$\widehat{\sigma}_e^2 = \frac{\text{RSS}}{60-2} = 1.00 \times 10^{-6}$$

Now,

$$\mathbb{V}(\widehat{a}) = \widehat{\sigma}_e^2 \left(\frac{1}{60} + \frac{\overline{X}^2}{\sum_{i=1}^{60} X_t^2 - 60\overline{X}^2} \right) = \mathbf{2.67 \times 10^{-8}}$$

Under the null hypothesis,

$$t_{58} = \frac{-0.0052}{\sqrt{2.67 \times 10^{-8}}} = \mathbf{-31.84}.$$

Hence, the hypothesis must be rejected, as the p value is almost zero. □

(c) Do likewise under the null hypothesis that $H_0 : b = 1$. What inference can be drawn?

Answer.

$$\mathbb{V}(\widehat{b}) = \widehat{\sigma}_e^2 \left(\frac{1}{\sum_{i=1}^{60} X_t^2 - 60\overline{X}^2} \right) = 0.0004.$$

□

The hypothesis is $b = 1$. Hence,

$$t_{58} = \frac{1.64 - 1}{\sqrt{0.0004}} = \mathbf{32.00}.$$

The inference is that the hypothesis must be rejected, as the p value is almost zero.

Problem 2. A student runs the following regression of stock i's return r_{it} on market portfolio return r_{mt} based on the market model:

$$r_{it} = a + b r_{mt} + e_{it},$$

where e_{it} is a residual noise that is i.i.d. and independent of r_{mt}.

(a) Is e_{it} independent of r_{it}?

Answer. It is sufficient to examine the covariance:

$$\mathbb{C}(e_{it}, r_{it}) = \mathbb{C}\left(r_{it} - a - br_{mt}, r_{it}\right) = \mathbb{V}(r_{it}) - b\,\mathbb{C}\left(r_{mt}, r_{it}\right)$$

In general, $\mathbb{C}(e_{it}, r_{it})$, the covariance is non-zero. Thus, e_{it} is generally correlated with the dependent variable r_{it}. It follows that **e_{it} is not independent of r_{it}**. □

(b) He performs OLS regression and obtains OLS estimates $\widehat{a}$ and $\widehat{b}$. He interprets $\widehat{b}$ as a parameter estimate that is proportionate to CAPM's notion of systematic risk of stock i, and determines $\widehat{a}$ as Jensen's alpha. Comment if his interpretation is sound.

Answer. $\widehat{b}$ is an estimate of stock i's beta since it is an estimate of $\dfrac{\mathbb{C}\left(r_{it}, r_{mt}\right)}{\mathbb{V}\, r_{mt}}$. But $\widehat{a} \approx r_f\big(1 - \widehat{b}\big)$ according to CAPM, where r_f is the risk-free rate. This is NOT the Jensen's alpha. Hence, **his interpretation about the $\widehat{b}$ is correct but incorrect for $\widehat{a}$**. □

(c) The student selects all stocks with positive $\widehat{a}$ and forms a portfolio. Is this portfolio likely to outperform the market index on average? Provide an explanation for your answer.

Answer. Since $\widehat{a} \approx r_f\big(1 - \widehat{b}\big)$ and the risk-less rate is positive, $\widehat{a} > 0$ implies that $\widehat{b} < 1$. Thus the portfolio beta b_p must also be less than 1. Moreover, according to CAPM,

$$\mathbb{E}(r_p) = r_f + b_p\big(\mathbb{E}(r_m) - r_f\big) < r_f + \big(\mathbb{E}(r_m) - r_f\big) = \mathbb{E}(r_{mt}).$$

Therefore, **the portfolio return is likely to underperform the market on average**. □

Problem 3. Let $\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 6 & 45 \\ 45 & 355 \end{pmatrix}$ and $\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. The covariance between the intercept estimate and the slope estimate is $-\dfrac{3}{28}$.

(a) What is the dimension of matrix $\boldsymbol{X}$?

Answer. The column 1 row 1 element of $\boldsymbol{X}'\boldsymbol{X}$ is the number of observations. Therefore, the dimension of $\boldsymbol{X}$ is **6 by 2** (6 rows and 2 columns). □

(b) What is the slope of the simple linear regression (as an irreducible fraction, e.g., 11/21)?

Answer. The inverse of $\boldsymbol{X}'\boldsymbol{X}$ is computed as

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \begin{pmatrix} \frac{71}{21} & -\frac{3}{7} \\ -\frac{3}{7} & \frac{2}{35} \end{pmatrix}.$$

Thus,

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} \frac{80}{21} \\ -\frac{17}{35} \end{pmatrix}.$$

Therefore, **the slope estimate is $-\dfrac{17}{35}$**. □

(c) What is the average of the explanatory variable (as an irreducible fraction, e.g., 11/21)?

Answer. The row 1 column 2 element of the $\boldsymbol{X}'\boldsymbol{X}$ is the sum of X_i, i.e., $\sum_{i=1}^{6} X_i = 45$. Hence, the average of the explanatory variable is

$$\frac{45}{6} = \boldsymbol{\frac{15}{2}}.$$

□

(d) What is the unbiased variance of the explanatory variable (as an irreducible fraction, e.g., 11/21)?

Answer. The row 2 column 2 element of the $\boldsymbol{X}'\boldsymbol{X}$ is the sum of squared X_i, i.e., $\sum_{i=1}^{6} X_i^2 = 355$. Thus,

$$\widehat{\sigma}_X^2 = \frac{\sum_{i=1}^{6}\left(X_i - \overline{X}\right)^2}{6-1} = \frac{\sum_{i=1}^{6} X_i^2 - 6\overline{X}^2}{5} = \frac{355 - 6 \times \left(\frac{15}{2}\right)^2}{5} = \frac{\frac{35}{2}}{5} = \boldsymbol{\frac{7}{2}}.$$

□

(e) What is the unbiased variance of the residuals (as an irreducible fraction, e.g., 11/21)?

Answer. The covariance between the intercept and the slope estimate is $-\frac{3}{28}$. From the off-diagonal element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, which is $-\frac{3}{7}$, we have

$$\widehat{\sigma}_e^2\left(-\frac{3}{7}\right) = -\frac{3}{28}.$$

Consequently, the unbiased variance of the residuals is

$$\widehat{\sigma}_e^2 = \frac{1}{4}.$$

□

(f) What is the t statistic of the y-intercept estimate (rounded to 2 decimal places)?

Answer. The estimate of the y-intercept is $\frac{80}{21}$. From the first diagonal element of $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, which is $\frac{71}{21}$. It follows that the standard error of the y-intercept estimate is

$$\sqrt{\frac{1}{4} \times \frac{71}{21}} = \sqrt{\frac{71}{84}},$$

and the t statistic of the y-intercept is

$$\frac{\frac{80}{21} - 0}{\sqrt{\frac{71}{84}}} = \mathbf{4.14}.$$

□

(g) Suppose a new observation of the explanatory variable is obtained and its value is 1.5.

(i) What is the point forecast for y (rounded to 2 decimal places)?

Answer. The point forecast is

$$\widehat{y}_7 = \frac{80}{21} - \frac{17}{35} \times 1.5 = \mathbf{3.08}.$$

□

(ii) What is the upper bound of the prediction interval at the 5% level of significance (rounded to 2 decimal places)? (Hint: You need to set $\boldsymbol{x} = \begin{pmatrix} 1 & 1.5 \end{pmatrix}'$ and apply your understanding about Slide 31 of `S4_2_MLR.pdf`)

Answer. First we compute

$$\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{x} = \begin{pmatrix} 1 & 1.5 \end{pmatrix} \begin{pmatrix} \frac{71}{21} & -\frac{3}{7} \\ -\frac{3}{7} & \frac{2}{35} \end{pmatrix} \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} = 2.2238$$

Now we need $t_{4,97.5\%} = 2.776$. Consequently, the upper bound of the prediction interval is

$$3.08 + 2.776 \times \sqrt{\frac{1}{4}} \sqrt{1 + 2.2238} = \mathbf{5.57}.$$

□