1. (SW Exercise 4.1) A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS,\ R^2 = 0.08,\ SER = 11.5.$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (*Hint*: Review the formulas for the OLS estimators.)

(d) What is the sample standard deviation of test scores across the 100 classrooms? (*Hint*: Review the formulas for the R^2 and SER.)

(a) $\hat{TS} = 520.4 - 5.82 \times 22$

$= 392.36$

(b) $CS_1 = 19$

$CS_2 = 23$

$\Delta\hat{TS} = (520.4 - 5.82 \times 23)$

$- (520.4 - 5.82 \times 19)$

$= -5.82(23-19)$

$= -23.28$

(c) $n = 100$ classrooms

$\overline{CS} = 21.4$

What is $\overline{TS}$?

Recall the formula for $\hat{\beta}_0$:

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{x}$

$\Rightarrow \overline{TS} = 520.4 - 5.82 \times 21.4$

$= 395.85$

d) $n = 100$, $k = 1$, $R^2 = 0.08$, $SER = 11.5$

We want s_y

Formulas: ① $s_y^2 = \dfrac{\sum (y_i - \bar{y})^2}{n-1} = \dfrac{TSS}{n-1}$

② $R^2 = 1 - \dfrac{SSR}{TSS}$

③ $SER = \sqrt{\dfrac{SSR}{n-k-1}}$

From ③: $SER^2 = \dfrac{SSR}{n-k-1}$

$\Rightarrow SSR = (n-k-1)\, SER^2$

From ②: $\dfrac{SSR}{TSS} = 1 - R^2$

$\Rightarrow TSS = \dfrac{SSR}{1-R^2}$

Plug ③ into ②:

$TSS = \dfrac{(n-k-1)\, SER^2}{1-R^2}$

Plug into ①:

$s_y^2 = \dfrac{TSS}{n-1}$

$= \dfrac{(n-k-1)\, SER^2}{(1-R^2)(n-1)}$

$= \dfrac{(100-2)(11.5^2)}{(1-0.08)(100-1)}$

$= 142.298 \approx 142.30$

$\Rightarrow s_y = \sqrt{142.298}$

$= 11.92887 \approx 11.93$

2. [Practice question, not graded] Let $KIDS$ denote the number of children born to a woman, and let $EDUC$ denote years of education for the woman. A simple model relating fertility to years of education is

$$KIDS = a + b * EDUC + u,$$

where u is the unobserved residual.

(a) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

(b) Will simple regression of kids on $EDUC$ uncover the ceteris paribus ('all else equal') effect of education on fertility? Explain.

(a) Things that affect # of children that aren't captured by education:

- income
- age
- family background

Note that these can be correlated with education but as long as they're not perfectly correlated, it adds information that education alone does not

(b) No if the other factors are correlated with education. Then someone having higher education is likely to also have higher income. So the regression will be influenced by both but attribute all of it to education. Instead, we'd like to estimate the relationship between them holding, e.g., income fixed so we know we are only capturing the education effect.

Technically:

- u is correlated with X
- Violates OLS Assumption

$$\mathbb{E}[u_i | X_i] = 0$$

5.1 Suppose a researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression

$$\widehat{TestScore} = \underset{(20.4)}{520.4} - \underset{(2.21)}{5.82} \times CS, R^2 = 0.08, SER = 11.5.$$

a. Construct a 95% confidence interval for β_1, the regression slope coefficient.

b. Calculate the p-value for the two-sided test of the null hypothesis $H_0: \beta_1 = 0$. Do you reject the null hypothesis at the 5% level? At the 1% level?

c. Calculate the p-value for the two-sided test of the null hypothesis $H_0: \beta_1 = -5.6$. Without doing any additional calculations, determine whether -5.6 is contained in the 95% confidence interval for β_1.

d. Construct a 99% confidence interval for β_0.

a) 95% CI : $\hat{\beta}_1 \pm 1.96 \times SE_{\hat{\beta}_1}$

$= -5.82 \pm 1.96 \times 2.21$

$\Rightarrow -10.152 \leq \beta_1 \leq -1.4884$

b) T-statistic

$t^{act} = \dfrac{\hat{\beta}_1 - \beta_1^{Null}}{SE(\hat{\beta}_1)}$

$= \dfrac{-5.82 - 0}{2.21}$

$= -2.6335$

p-value for $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

p-value $= 2\Phi(-|t^{act}|)$

$= 2\Phi(-2.6335)$

$= 2 \times 0.0042$

$= 0.0084$

$p < 0.01$ ∴ reject null at both 5% and 1% significance levels

c) $H_0: \beta_1 = -5.6$

$H_1: \beta_1 \neq -5.6$

$t^{act} = \dfrac{\hat{\beta}_1 - (-5.6)}{SE(\hat{\beta}_1)}$

$= \dfrac{-5.82 - (-5.6)}{2.21}$

$= 0.10)$

p-value $= 2\Phi(-|t^{act}|)$

$= 2\Phi(-0.10)$

$= 0.92$

$p > 0.05$ ∴ cannot reject H_0 at 1% or 5% significance level

Since cannot reject at 5% level, -5.6 must be in the 95% CI for β_1

(d) 99% CI for β_0

$520.4 \pm 2.58 \times 20.4$

$\Rightarrow 467.7 \leq \beta_0 \leq 573.0$