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?

$\overline{CS} = 21.4$

(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) $CS_1 = 19 \quad CS_2\ 23$

$\Delta\hat{TS} = -5.82 \times (23-19)$

$= -23.28$

c) $\overline{CS} = 21.4$

We're interested in $\overline{TS}$

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) $SER = \sqrt{\frac{1}{n-2} \sum_i (\hat{u}_i - \bar{\hat{u}})^2}$

$= \sqrt{\frac{1}{n-2} \sum_i \hat{u}_i^2}$

$= \sqrt{\frac{1}{n-2} SSR}$

$R^2 = \frac{SSR}{TSS}$

$= \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$

$s^2 = \frac{TSS}{n-1}$

$n = 100$

$\frac{1}{n-2} SSR = SER^2$

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

$= (100-2) \cdot 11.5^2$

$TSS = \frac{R^2}{SSR} = \frac{0.08}{(100-2) \cdot 11.5^2}$

2. (SW Exercise 4.3) A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25-65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \ R^2 = 0.023, \ SER = 624.1.$$

(a) Explain what the coefficient values 696.7 and 9.6 mean.

(b) The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)

(c) The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)

(d) What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

(e) Will the regression give reliable predictions for a 99-year-old worker? Why or why not?

(f) Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (*Hint*: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)

(g) The average age in this sample is 41.6 years. What is the average value of AWE in the sample? (*Hint*: Review Key Concept 4.2)

a) 696.7 the intercept

9.6 the marginal effect of age

b) Depends on the dependent variable, here \$

c) Unit free

d) i. $696.7 + 9.6 \cdot 25 = 936.7$

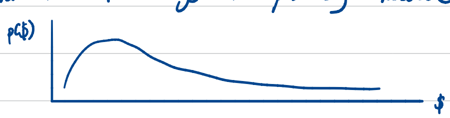
ii. $45 = 1{,}128.7$

e) Out of sample, though it will depend on your theory

But for this example, we expect 95 year olds to behave differently

so no

f) Distribution of earnings are positively skewed, kurtosis > normal



g) $\overline{Age} = 41.6$

Plug into $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\Leftrightarrow \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$

$= 696.7 + 9.6 \cdot 41.6$

$= 1096.06$

3. (SW Exercise 4.5) A professor decides to run an experiment to measure the effect of time pressure on final exam scores. He gives each of the 400 students in his course the same final exam, but some students have 90 minutes to complete the exam, while others have 120 minutes. Each student is randomly assigned on of the examination times, based on the flip of a coin. Let Y_i denote the number of points scored on the exam by the ith student ($0 \leq Y_i \leq 100$), let X_i denote the amount of time that the student has to complete the exam ($X_i = 90$ or 120), and consider the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$.

(a) Explain what the term u_i represents. Why will different students have different values of u_i?

(b) Explain why $E(u_i|X_i) = 0$ for this regression model.

(c) Are the other assumptions in Key Concept 4.3 satisfied? Explain.

(d) The estimated regression is $\hat{Y}_i = 49 + 0.24X_i$.

i. Compute the estimated regression's prediction for the average score of students given 90 minutes to complete the exam. Repeat for 120 minutes and 150 minutes.

ii. Compute the estimated gain in score for a student who is given an additional 10 minutes on the exam.

Coin flip $\xrightarrow{H}$ 90-min test $(X_i = 90)$

$\xrightarrow{T}$ 120-min test $(X_i = 120)$

a) u_i represents unincluded variables that affect/influence Y_i

b) Random assignment: u_i and X_i are thus independent

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

c) Key Concept 4.3: the OLS Assumptions

✓1. $\mathbb{E}[u_i | X_i] = 0$

✓2. (X_i, Y_i) are iid draws from a joint distribution

✓3. Large outliers unlikely

d) $\hat{Y}_i = 49 + 0.24X_i$

i) Plug in $X_i = 90, 120, 150$

$\Rightarrow \hat{Y}_i = 70.6, 77.8, 85.0$

$$\boxed{\text{ii) } \Delta Y_i = 0.24 \cdot 10 = 2.4}$$

4. (SW Exercise 4.9)

(a) A linear regression yields $\hat{\beta}_1 = 0$. Show that $R^2 = 0$.

(b) A linear regression yields $R^2 = 0$. Does this imply that $\hat{\beta}_1 = 0$?

a) $R^2 = \frac{SSR}{TSS}$ $\frac{\text{sum of sq. resids}}{\text{total sum of squares}}$

$= \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y}_i)^2}$

If $\hat{\beta}_1 = 0$, then $\hat{\beta}_0 = \bar{Y}$ (expand on why)

Then $\hat{Y}_i = \bar{\hat{Y}}_i \Rightarrow \hat{Y}_i - \bar{\hat{Y}}_i = 0$

$\Rightarrow R^2 = 0$

b) $R^2 = 0 \Rightarrow SSR = 0$

$\Rightarrow \hat{Y}_i - \bar{\hat{Y}}_i = 0 \quad \forall i$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

$\Rightarrow \bar{Y} = \hat{\beta}_0 + \beta_1 X_i \quad \forall i$ (i.e. RHS is constant)

$\Rightarrow \beta_1 X_i$ is constant $\Rightarrow \beta_1 = 0$ OR $X_i = X \quad \forall i$

OR both

Empirical Exercise 4.2 : Earnings + Height

e) 1 cm $\approx$ 0.394 inches

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$, X is height in inches

$\Rightarrow \hat{\beta}_1 X = \hat{\beta}_1 \times \frac{0.394}{1} \times \tilde{X}$ where $\tilde{X}$ is height in CM

$= \tilde{\beta}_1 \tilde{X}$

$\Rightarrow \tilde{\beta}_1 = \hat{\beta}_1 \frac{X}{\tilde{X}}$

$= \tilde{\beta}_1 \times 0.394$

β_0 will not change since represents earnings at height 0 which is same for both cm and inches (0cm = 0 inches)

R^2 is unit free

SER