Homework Week 4 (Multiple regression)

For this assignment, select two explanatory variables (X_1, X_2) and an outcome variable (Y) from a dataset of your choosing (see: datasets.zip). Be sure that the variables you choose seem to have a meaningful relationship – one you might be actually interested in for a research question. Also make sure that all of the variables are quantitative/ continuous.

1) In **Homework Week 2**, you calculated separate least squares regressions for the regression of Y on X_1 (Problem 2e) and Y on X_2 (Problem 3). In this problem, *using the same variables*, you will **use SPSS** and **ALL OF THE DATA** to calculate the multiple regression of Y on X_1 and X_2.
   (a) Write out the regression relationship. Explain what each of the coefficients mean.
   (b) Find the least squares estimates a, b_1, and b_2 for the regression of Y on X_1 and X_2. Also find the standard error of the regression, s_e, and R^2.
   (c) Interpret each of the coefficients (a, b_1, b_2, s_e, R^2).
   (d) Compare the estimates for b_1 and b_2 of the multiple regression with the slope estimates of the simple regressions you performed in problems 2(e) and 3 of homework 2.1. If the coefficients differ (or if they do not differ at all), explain why.
   (e) Calculate the standardized regression coefficients. Explain how each of these coefficients is interpreted. Is the effect of X_1 or X_2 larger? Explain.

2) Using **SPSS** and **ALL OF THE DATA**, calculate the multiple regression of Y on X_1 and X_2. *Note that this homework builds on the previous homework; use the same Y, X_1, and X_2 variables from last week.*
   (a) Assuming independently distributed errors, find the standard error of the slope coefficients for X_1 and X_2 (i.e. SE(b_1) and SE(b_2)).
   (b) Construct an analysis-of-variance table for the regression, testing the omnibus null hypothesis H_0: $\beta_1 = \beta_2 = 0$. Interpret the results.
   (c) Separately test the null hypothesis H_0: $\beta_1 = 0$ and H_0: $\beta_2 = 0$. Interpret the results. (Note: this is asking for two individual t-tests from your multiple regression model)
   (d) Are the hypotheses tested in (b) the same as that tested in (c)? Explain.
   (e) Compute an incremental F-test of the hypothesis H_0: $\beta_2 = 0$ (be sure to state both H_0 and H_a). Verify that the F-statistic for this test is equal to the square of the t-value obtained. Why does this matter?

1. a) $SATVerbal_i = \alpha + \beta_1 TeacherPay_i + \beta_2 PercentTaking_i + \varepsilon_i$

α = intercept. = average value of SAT verbal when TeacherPay = PT = 0.

β_1 : slope. = effect of a 1-unit increase in teacher pay on sat verbal, holding PT.

β_2 : effect of a 1-unit increase in PT on Satverbal. holding constant teacher pay.

ε_i = $Y_i - E(Y_i)$ is the residual for i.

b) $\alpha = 555.7$ $\beta_1 = 0.54$ $\beta_2 = -1.21$ $Se = 15.9$

$R^2 = 0.778$ (555.7)

c) α = average value for sat verbal when teacher pay and percent taking = 0

β_1 : teacher pay is not significant.
for every 1 unit increase in teacher pay, sat verbal increase 0.54, holding constant these other variables.

β_2 : teacher pay is significant
for every 1 unit increase in percent taking, sat verbal decrease 1.21, holding constant these other variables.

Se : on average, the model is incorrect by 15.9 when predicting sat verbal

R^2 : the model explained 77.8% of variation in sat verbal!

d) b in (2e) and b in (3) are different than b_1 and b_2.

Because the coefficient estimated is a partial coefficient. Holding constant. But in simple regression models, they are not.

e) $b_1 = 0.1$

$b_2 = -0.944$

X_2 has longer effect on Y.

Because for 1 sd increase in teacher pay, Sat Verbal increase 0.1 sd unit, controlling for PT.

for 1 sd increase in PT, Sat Verbal decrease 0.944 sd unit, controlling for teacher pay.

2. a) $SE(b_1) = 0.45$

$SE(b_2) = 0.11$

b) $p = 0.01 < 0.05$. Ho was rejected.

So at least one of the covariates in the model have effect on Y.

c) 1. $t = 1.18$ $p = 0.24 < 0.05$ so not sufficient evidence to reject Ho. So, there is no evidence that shows X_1 and Y has relationship

2. $t = -11.3$ $p = 0.01 < 0.05$ Ho was rejected.

So X_2 and Y has relationship

d) They are different. Omnibus test could be rejected, but the individual tests can not.

They are using different tests.

e) $H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Teacher pay $t = -3.755$ $F = 14.1 = t^2$

PT $t = -11.31$ $F = 127.9 = t^2$

$P = 0.01 < 0.05$ H_0 was rejected

so the full model is significantly better.