USP 634 Spring 2018
Assignment 4: Association and Linear Regression
Due **Sunday, June 10th, 2018**

The objectives of the assignment are to

- Calculate and interpret measures of association and
- Interpret linear regression model

1. In question 1.3 from Assignment 1 you examined the relationship between two continous variables in your dataset. Calculate two measures of association for your two variables, interpret their meaning, and show how the values are derived by presenting the correct computational formula – although you don't have to compute the measures manually. (20 points).

2. In question 3 from Assignment 3 you tested a hypothesis involving a bivariate relationship between two categorical variables. Calculate two measures of association for your variables, interpret their meaning, and show how the values are derived by presenting the correct computational formula – although you don't have to compute the measures manually (20 points).

3. Consider the following regression output from R. The data consists of individuals drawn from the Panel Study of Income Dynamics (PSID).

```
> summary(lm(income92~age+educ, data=psid92, subset=subset))

Residuals:
   Min     1Q Median     3Q    Max
-24744  -9826   1324   1971 476389

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1162.21     601.42  -1.932  0.05340 .
age           -53.93      19.52  -2.763  0.00576 **
educ         1590.49      66.53  23.905  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17550 on 2858 degrees of freedom
  (66 observations deleted due to missingness)
Multiple R-squared: 0.2118,     Adjusted R-squared: 0.2113
F-statistic: 384.1 on 2 and 2858 DF,  p-value: < 2.2e-16
```

The dependent variable is total labor income ($) in 1992. The independent variables are age and years of schooling (educ).

- In these data, what is the predicted increase in earnings from a one-year increase in years of schooling (holding age constant)? (5 points)

- What is the estimated standard error of the coefficient on years of schooling? What is the likelihood of observing an estimated effect of education of this magnitude if education is not truly associated with earnings? (5 points)
- What proportion of the variation in earnings can be explained by variation in education and age? (5 points)
- Your neighbor is 20 years old and has 10 years of formal schooling. Suppose you wished to use the regression results above to predict how much someone like your neighbor would have earned in 1992. How would you do it? You do not need to perform the calculation, just make it clear which numbers would get added to, subtracted from, multiplied by or divided by which other numbers (5 points).

4. In question 5 from Assignment 3, you conducted an ANOVA of a continous variable by a categorical variable. Present a univariate regression analysis with the continous variable as your dependent (response) variable and the categorical variable as your independent (explanatory) variable. Make sure your categorical is of a character or factor class before running regression.

1) Postulate and present a research question and corresponding hypothesis for your regression;
2) Present and interpret your regression results;
3) Explain how your regression results relate to the results of ANOVA;
4) In a brief paragraph, what do the results tell you about your research question?

In responding to these questions, be concise! Also, note that your grade on this question will depend not on the size of your R^2, but on the quality and conciseness of your presentation and interpretations. (40 points)