POLS 6481, Spring 2021
Professor Scott Basinger

Homework Assignment 1
Distributed Wednesday, January 27
Due Tuesday, February 2 at 10 AM *if possible*

In 2020, some political leaders urged re-opening the economy in spite of the Coronavirus. For example, on March 26, President Trump stated the following, "You have suicides over things like this when you have terrible economies. You have death. Probably — and I mean definitely — would be in far greater numbers than the numbers that we're talking about with regard to the virus."

To support this claim, one might cite the article "An Economic Theory of Suicide," published in the *Journal of Political Economy* in 1974. Hamermesh and Soss write: "When unemployment rises, individuals' expectations of future incomes are revised downward. … People will believe future prospects to have diminished and will commit suicide." To test this claim, examine the following table:

Year	Unemployment rate	Suicide rate (per million population)
1968	3.6	107
1969	3.5	110
1970	4.9	115
1971	5.9	117
1972	5.6	120
1973	4.9	120
1974	5.6	121
1975	8.5	127
1976	7.7	125
1977	7.0	133
1978	6.0	125

a. Load the data (*suicide.csv*), and use R to find the sample means, sample variances, and sample standard deviations of suicide rate (y), and find the same quantities for unemployment rate (x).

$\bar{y}$ = _____ $\bar{x}$ = _____

$var(y) = s_y^2$ = _____ $var(x) = s_x^2$ = _____

s_y = _____ s_x = _____

If you wish, you can download the optional file "homework 1 problem 1.pdf" and fill it out.

b. Use R to find the covariance between the unemployment and suicide rates, and then calculate the correlation (Pearson's r) between the unemployment and suicide rates.

$cov(x,y)$ = _____

r_{xy} = _____

c. Using the quantities you computed for a. and b., obtain the intercept and the slope estimates in for a linear model in which the suicide rate is the dependent variable and the unemployment rate is the independent variable. That is, obtain the intercept and slope estimates in the sample regression function: $\widehat{suicide\ rate} = \widehat{\boldsymbol{\beta}}_{\mathbf{0}} + \widehat{\boldsymbol{\beta}}_{\mathbf{1}} unemployment\ rate$

$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ = *cov*(*x*,*y*) ÷ *var*(*x*) = _____

$\widehat{\boldsymbol{\beta}}_{\mathbf{0}} = \bar{y} - \widehat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x}$ = _____

d. Use R to verify your calculations for the slope and intercept: run the linear regression model, and display the results.

e. Verify the property that the regression line passes through the point whose Cartesian coordinates are the average unemployment rate and the average suicide rate. You can do this by computing the predicted suicide rate when the unemployment rate is set at its mean.

$\widehat{\boldsymbol{\beta}}_{\mathbf{0}} + \widehat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x}$ = _____

f. Use R to compute the fitted values ($\hat{y}_i$) and residual ($\hat{u}_i$) for each observation, filling in the table below – especially the bottom two rows (sums and averages). Round to one decimal point. Verify two important properties of the residuals: E($\hat{u}$) = 0, and E($\hat{u}$·x) = 0.

y_i	x_i	$\hat{y}_i$	$\hat{u}_i$	$x_i \cdot \hat{u}_i$
107	3.6			
110	3.5			
115	4.9			
117	5.9			
120	5.6			
120	4.9			
121	5.6			
127	8.5			
125	7.7			
133	7.0			
125	6.0			
Sum (Σ·) =				
Average ($\frac{1}{n}\Sigma\cdot$) =				

Are countries that are more inclusive in terms of political representation also more equitable in terms of the income distribution? Download "*inequality and representation.csv*" which contains data on 20 OECD countries – advanced, industrialized countries that accept the principles of a free-market economy and representative democracy. ***Inequality*** refers to the ratio of the richest 10%'s wealth to the poorest 10%'s wealth, and ***Representation*** refers to the percent of seats in parliament occupied by women.

Country	Inequality	Representation
Australia	12.2	28
Austria	6.7	32
Belgium	8.0	36
Canada	9.2	25
Denmark	7.9	37
Finland	5.4	38
France	8.9	14
Germany	6.7	31
Greece	10.0	13
Ireland	9.2	14
Italy	11.4	16
Japan	4.2	11
Netherlands	9.0	34
NewZealand	12.2	32
Norway	6.0	38
Portugal	14.6	21
Spain	10.0	31
Sweden	6.0	45
Switzerland	8.8	25
UnitedKingdom	13.6	19

a. Using R, estimate the OLS regression in which ***Representation*** is the dependent variable, and ***Inequality*** is the independent variable. Is the slope statistically significant? The slope coefficient, _____, tells us that for every one-unit increase in the inequality ratio (e.g., from 5 to 6) the percent of seats occupied by women decreases by roughly _____.

b. To continue interpreting the results, the intercept, _____, tells us that _____ percent of seats would be occupied by women if the richest 10% held none of the wealth, which is obviously illogical. Compute a more logical (if implausible) prediction: in a perfectly equitable country, in which the top 10% and bottom 10% hold equal wealth (i.e., the ratio equals 1), roughly _____ percent of seats would be occupied by women.

c. The next page provides a worksheet for calculating the variances of *x* and *y*, and the covariance; it is **not** necessary to fill in the first twenty rows in the first five empty columns, but at least use R like a calculator to find seven entries at the bottoms of the columns:

Means: $\overline{\boldsymbol{Inequality}}$ = _____ $\overline{\boldsymbol{Representation}}$ = _____

Variances: *var*(***Inequality***) = _____ *var*(***Representation***) = _____

Covariance: *cov*(***Inequality***, ***Representation***) = _____

While you are at it, calculate the sums of squared deviations for *x* and *y* too.

d. Using the quantities that you computed in d., obtain the intercept and the slope estimates in for a linear model in which ***Representation*** is the dependent variable and ***Inequality*** is the independent variable. That is, compute $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ = *cov*(***x***,***y***) ÷ *var*(***x***) = _____

$$\widehat{\boldsymbol{\beta}}_{\mathbf{0}} = \bar{y} - \widehat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x} = \_\_\_\_\_$$

e. Look at the regression results provided by R, and find the standard error of the regression, *sigma* = _____ ; and the standard error of the regression slope, s.e.($\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$) = _____ . Fill in the orange circles on the next page to show where these values come from. Fill in the last three columns of the table on page 5, using R as needed to make life easier.

f. Use R to construct a scatterplot with ***Inequality*** on the horizontal axis and ***Representation*** on the vertical axis. Include a fitted line if you can. Next, examine the scatterplot and **identify which case is an outlier**: _____

g. Use R to calculate *hat values*, *studentized residuals*, and *DfFits*; report these values **only for the apparent outlier**.

Hat-value (leverage) = _____ for the outlier

Studentized residual (discrepancy) = _____ for the outlier

DfFits (influence) = _____ for the outlier

Use the Table on page 6 and the blank space below it to calculate these values by hand **only for the apparent outlier**.

h. Use R to calculate the values of DfBeta for the outlier, and calculate what the intercept and slope should be if you had (i) omitted the outlier or (ii) included a dummy variable for the outlier. i.e., without the outlier, the slope would have equaled _____ and the intercept would have equaled _____

i. Adapt the code from the lab for week 2 and/or lecture 4, and either omit the outlier from the analysis or include a dummy variable for the outlier. Show your results! Then use the results to answer these three queries:

What is the new intercept without the outlier? _____

What is the new slope without the outlier? _____

Is the slope statistically significant without the outlier? _____

j. **Extra Credit!** Use the alternative model you estimated for i., and fill in the last three columns on page 5 for all countries. Attempt to calculate the value of Cook's D **for the outlier** based on these columns. (The lecture slides show the equation.) Use R to confirm your calculation.

Country	Inequality	Representation							(Fitted)	(Residual)	(Residual²)
i	x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\bullet(y_i-\bar{y})$	$\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_i$	$\widehat{u_i} = y_i - \widehat{y_i}$	$\widehat{u_i}^2 = (y_i - \widehat{y_i})^2$
Australia	12.2	28								
Austria	6.7	32								
Belgium	8.0	36								
Canada	9.2	25								
Denmark	7.9	37								
Finland	5.4	38								
France	8.9	14								
Germany	6.7	31								
Greece	10.0	13								
Ireland	9.2	14								
Italy	11.4	16								
Japan	4.2	11								
Netherlands	9.0	34								
NewZealand	12.2	32								
Norway	6.0	38								
Portugal	14.6	21								
Spain	10.0	31								
Sweden	6.0	45								
Switzerland	8.8	25								
UnitedKingd	13.6	19								
Σ			**SST_x =**		**SST_y =**				**SSR =**	
Σ/n	**$\bar{x}$ =**	**$\bar{y}$ =**								
Σ/(n-1)			**var(x) =**		**var(y) =**		**= cov(x,y)**			
Σ/(n-2)									**$\widehat{\sigma^2}$ =**	

Intercept	$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$	Std. Error of Intercept	$se(\widehat{\beta_0}) = \frac{\hat{\sigma}}{\sqrt{SST_x}} \cdot \sqrt{\frac{\sum {x_i}^2}{n}}$
Slope	$\widehat{\beta_1} = \frac{cov(x,y)}{var(x)}$	Std. Error of Slope	$se(\widehat{\beta_1}) = \frac{\hat{\sigma}}{\sqrt{SST_x}}$
Regression Line	$\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_i$	Root Mean-Squared Error	$sigma = \sqrt{\widehat{\sigma^2}} = \sqrt{\frac{SSR}{n-2}}$

Country	Inequality	Representation	Leverage	Studentized residuals	DfFits	Alternative fitted values	Differ.	Squared Differ.
i	x_i	y_i	$\frac{1}{n}+\frac{(x_i-\bar{x})^2}{SST_x}$	$t_i=\frac{\hat{u}_i}{\hat{\sigma}\cdot\sqrt{1-h_i}}$	$t_i \times \sqrt{\frac{h_i}{1-h_i}}$	$\widetilde{y_i}=\widetilde{\beta_0}+\widetilde{\beta_1}x_i+\tilde{\delta}$	$\widehat{y_i}-\widetilde{y_i}$	$(\widehat{y_i}-\widetilde{y_i})^2$
Australia	12.2	28						
Austria	6.7	32						
Belgium	8.0	36						
Canada	9.2	25						
Denmark	7.9	37						
Finland	5.4	38						
France	8.9	14						
Germany	6.7	31						
Greece	10.0	13						
Ireland	9.2	14						
Italy	11.4	16						
Japan	4.2	11						
Netherlands	9.0	34						
NewZealand	12.2	32						
Norway	6.0	38						
Portugal	14.6	21						
Spain	10.0	31						
Sweden	6.0	45						
Switzerland	8.8	25						
UnitedKingd	13.6	19						
Σ								
Σ/n								