Classwork 8: Midterm Review

Consider the new linear model $y_i = \beta_1 x_i + u_i$. In this model, the intercept is omitted (thus forced to be 0). In this classwork, we'll build the econometrics of this model from scratch.

1. Recall that OLS minimizes the sum of squared residuals. Write down what that means in this context and take first order conditions. Show that $\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$. (2 points)

2. Next, show that as long as the model is correctly specified (assume it reflects the true data generating process), we'll have that $\hat{\beta}_1 = \beta_1 + \frac{\sum_i x_i u_i}{\sum_i x_i^2}$. (2 points)

3. Let $w_i = \frac{x_i}{\sum_i x_i^2}$ and show that $\hat{\beta}_1$ is an unbiased estimator of β_1 under the key assumption of exogeneity. (1 point)

4. Consider this dataset. Sketch the data points and a line of best fit (with the intercept forced to be 0). Fill out the table for the model with the omitted intercept $y_i = \beta_1 x_i + u_i$. Note that when an intercept β_0 is not included in the model, the parameters which minimize the sum of the squared residuals may not generate residuals that sum to zero. (2 points)

x	y	$\hat{y}$	e
1	3	__	__
2	3	__	__
3	3	__	__
4	4	__	__
5	2	__	__

5. For this model, the standard error for $\hat{\beta}_1$ can be shown to be $\sqrt{\frac{\sum_i e_i^2}{(n-1)\sum_i x_i^2}}$. Calculate this standard error using the data from the previous question. (1 point)

6. Continuing from the previous question, conduct a formal hypothesis test at the .05 significance level for $\hat{\beta}_1$. Hint: the degrees of freedom in the general case for a linear regression model are $n - k$ where k is the number of model parameters, which is 1 here because there's only one β to estimate. You should show your work, but you can also check it using functions in R. (2 points)

   - Null hypothesis: ____________________________
   - Alternative hypothesis: ____________________________
   - Confidence interval: ____________________________
   - T-statistic and critical value: ____________________________
   - P-value: ____________________________
   - Hypothesis test conclusion: ____
   - Interpretation: do X and Y really seem to be related, or is it just as likely that the correlation between the two comes from sampling error? ____

   Here is some potentially useful information:

   `pt(3.45, df = 4)` is 0.987

   `qt(.025, df = 4)` is -2.78