STAT 1221/2220
Homework 6

Name: Stephen Skaslen

Problem 1e requires you to print the model summary and 1f requires you to use R to obtain summary statistics. Several other subparts throughout the assignment require R as well to obtain critical values and p-values, but these do not need to be included in the submitted output. You know the drill by now - download, fill in the appropriate code, and upload with the rest of the assignment.

1. Recall the dataset from Homeworks 3, 4, and 5 that compared the relationship between GDP growth in a presidential election year and the percentage of the two-party vote that the incumbent party received between 1928 and 2020. Read in the dataset **elections.csv**, which contains the percentage of growth in GDP during the presidential election year (X) and the percentage of the two-party vote that the incumbent party received (Y). Recall that the sample correlation between these variables is $r = 0.65$.

   (a) Test if there is a significant correlation between GDP growth and two-party vote percentage for the incumbent party using a 5% level of significance.

      i. State the hypotheses.

      H_0: $\rho = 0$ vs. H_A: $\rho \neq 0$

      ii. Calculate the test statistic by hand and show the calculation.

      $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, $r = 0.65$

      $t = \frac{0.65\sqrt{24-2}}{\sqrt{1-(0.65)^2}} = 4.012$

      iii. Use R to obtain and report the critical value and p-value. Report the degrees of freedom.

      Degrees of Freedom: 22 - Critical Value: ± 2.074 - P-value: 0.000586

      iv. Write a conclusion in the context of the problem.

      Reject H_0 and conclude that the correlation between GDP Growth and the two-party vote percentage is Significantly Correlated.

   (b) Test if the correlation between GDP growth to predict two-party vote percentage is significantly greater than 0.25 using a 5% level of significance.

      i. State the hypotheses.

      H_0: $\rho = 0.25$ vs. H_A: $\rho > 0.25$

      ii. Calculate the test statistic.

      $z = \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{1/\sqrt{n-3}} = \frac{\frac{1}{2}\ln\left(\frac{1+0.65}{1-0.65}\right) - \frac{1}{2}\ln\left(\frac{1+0.25}{1-0.25}\right)}{1/\sqrt{24-3}}$ $\rho = 0.25$, $r = 0.65$

      $= 2.38$

iii. Use R to obtain and report the critical value and p-value. Report the degrees of freedom.

Degrees of Freedom: 22 - Critical Value: −1.645 - P-value: 0.9913

iv. Write a conclusion in the context of the problem.

Fail to Reject H_0 and conclude that 0.25 is a plausible correlation for GDP Growth to predict the two-party vote percentage.

(c) Calculate a 95% confidence interval for the correlation between GDP growth and two-party vote percentage for the incumbent party. Show the calculations.

$$\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \pm Z_{1-\alpha/2}\left(\frac{1}{\sqrt{n-3}}\right) = (L_Z, U_Z)$$

$$\frac{1}{2}\ln\left(\frac{1+0.65}{1-0.65}\right) \pm 1.96\left(\frac{1}{\sqrt{24-3}}\right) = (0.35, 1.2)$$

$$\frac{e^{2(0.35)}-1}{e^{2(0.35)}+1} = 0.34 \Rightarrow (0.34, 0.83)$$

$$\frac{e^{2(1.2)}-1}{e^{2(1.2)}+1} = 0.83$$

(d) Are the results of the tests in parts (a) and (b) consistent with the confidence interval in part (c)? Briefly justify your answer.

Yes, since we fail to reject and conclude that there is a 25% plausibility correlation between GDP Growth and the two-party vote percentage, meaning its contained.

(e) Use R to fit the simple linear regression model that uses GDP growth to predict the two-party vote percentage. Print the model summary.

(f) Use R to obtain and print the sample mean and sample standard deviation for GDP growth percentage.

(g) Set up but do not evaluate a 95% confidence interval for the two-party vote percentage for the incumbent party when the percentage of GDP growth is 3%.

$\hat{y} = 49.6505 + 0.7969(3) = 52.04$

$$\hat{Y} \pm t_{n-2} \times S_{Y|X} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)S_X^2}}$$

$$52.04 \pm 2.07(4.535) \cdot \sqrt{\frac{1}{24} + \frac{(3-3.30)^2}{(24-1)(4.76)^2}} = (50.12, 53.96)$$

(h) Set up but do not evaluate a 95% prediction interval for the two-party vote percentage for the incumbent party when the percentage of GDP growth is 3%.

$$\hat{Y} \pm t_{n-2} \times S_{Y|X} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)S_X^2}}$$

$$52.04 \pm 2.07(4.535) \cdot \sqrt{1 + \frac{1}{24} + \frac{(3-3.30)^2}{(24-1)(4.76)^2}} = (42.44, 61.64)$$

(i) The confidence interval from part (g) is $(50.12, 53.96)$. Interpret this interval in the context of the problem.

We are 95% confident that the true mean two-party vote percentage of GDP Growth at 3% is between 50.12% and 53.96%

(j) The prediction interval from part (h) is $(42.44, 61.64)$. Interpret this interval in the context of the problem.

We are 95% confident that the true two-party vote percentage GDP Growth at 3% is between 42.44% and 61.64%.