WEEK 11 REPORT – LINEAR REGRESSION

Group Number: 25
Date and Time of Tutorial Session: Wednesday 4:00 – 5:30 pm
Group Members (First and Last name – no student numbers):

1. Katarzyna Fraser
2. Annika Tran
3. Ella Pustil
4.
5.

Graded out of 10 marks.

1. What type of relationship, if any, is apparent between expert portrayals in the media and number of seats held by women?
(1 mark)

There is a weak positive relationship between the number of expert portrayals in the media and the number of seats won by women. The points are roughly scattered, with no clear upward or downward trend.

Additionally, the slope of the regression line (0.163), which is positive but small. Furthermore, the relationship shown is not statistically significant, as the p-value is 0.237, which is less than 5%.

2. What are the null and alternative hypotheses for your linear regression?
(1 mark)

Null Hypothesis ($H0$) - There is no relationship between expert portrayals in the media and the number of seats won by women

Alternative Hypothesis (Ha) - There is a relationship between expert portrayals in the media and the number of seats won by women

3. What is the equation of your linear regression?
(1 mark)

Seats=17.67+0.163×Expert

17.67 is the intercept which represents the expected number of seats when there are no expert portrayals.

0.163 is the slope which indicates the predicted increase in the number of seats for every additional expert portrayal

4. Include your scatterplot with a regression line. Make sure the predictor variable is on the x-axis and that the response variable is on the y-axis. Provide a descriptive figure caption.
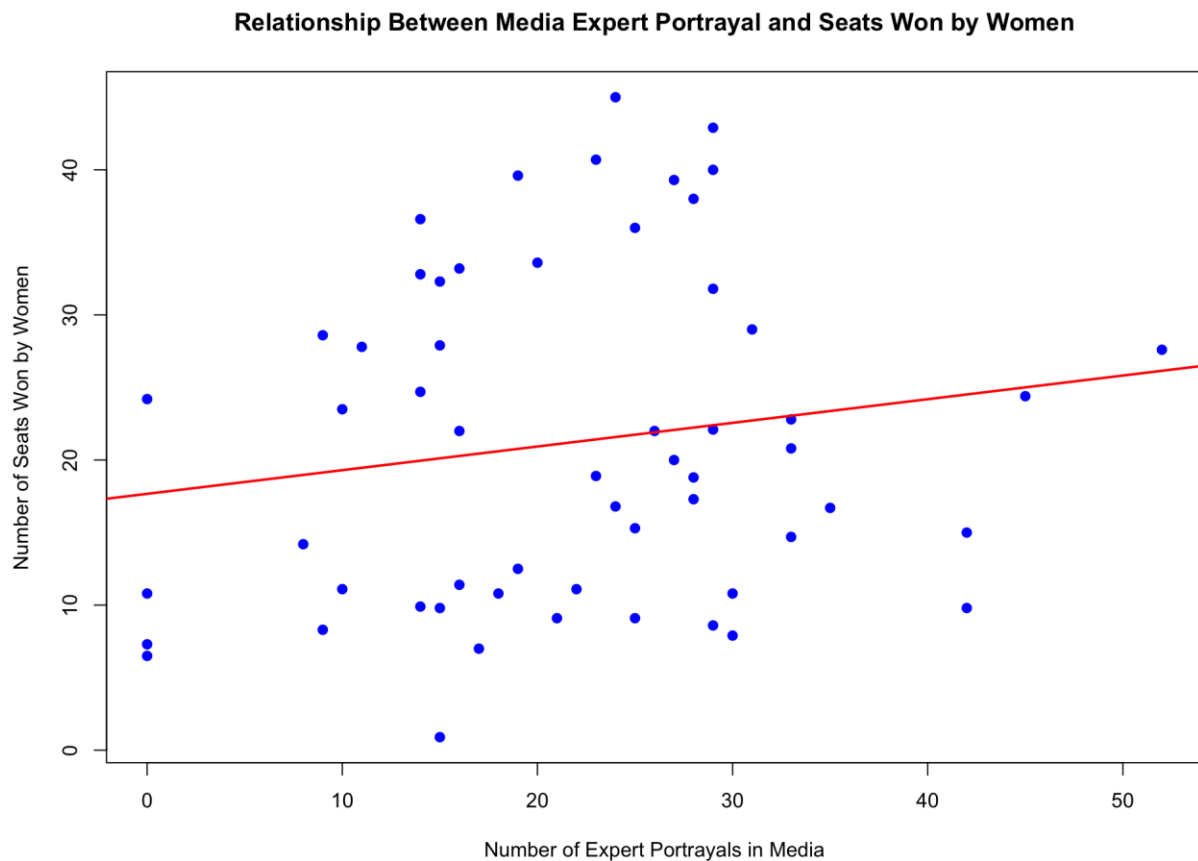(1.5 marks)

Figure 1. A scatter plot that represents the relationship between the number of news stories portraying women as political experts (predictor variable) and the number of seats won by women (response variable) with a red linear regression line displayed.

5. For your regression analysis, why is there a t-score and p-value beside each of the statistical parameters?
(1 mark)

The t-score measures how many standard deviations the regression coefficient ($\beta 1$) is away from zero. The p-value indicates the probability of observing such a t-score (or more extreme one) under the null hypothesis ($\beta 1 = 0$). These are used to see whether the explanatory variable significantly influences the response variable.

6. What is the *statistical* conclusion of your test (i.e. do you reject or fail to reject the null hypothesis for your linear regression)? Report your adjusted R^2 value, df, alpha value, and p-value in brackets.

(1 mark)

We will fail to reject the null hypothesis, as the p-value for the slope is 0.237, which is greater than the conventional α=0.05. This means there is poor statistical evidence to conclude that expert portrayals in the media significantly predict the number of seats won by women. Overall, since the p-value (0.237) is greater than the significance level of (α=0.05), we fail to reject the null hypothesis.

(Adjusted $R2$= 0.008 Degrees of Freedom= 55 Alpha= 0.05 p-value= 0.237)

7. Examine a plot of the residuals vs. predicted/fitted values and conduct a Shapiro-Wilk test. Do the model assumptions appear to be met? Does this examination of assumptions suggest your model is appropriate for the given data? Explain why or why not.
(1.5 mark)
The model assumptions do not appear to be met as the red line in the graph, the moving average of the residuals is not a linear line. The residuals do not appear to follow a clear random pattern around the horizontal red line at zero, which suggests violations of homoscedasticity. This examination of assumptions suggest that are model is not appropriate for the given data because the moving average was not linear. The Shapiro- Wilk test statistic is 0.938, and the p-value is 0.006. Since the p-value is less than 0.05, we reject the null hypothesis that the residuals are normally distributed. The violations of normality and homoscedasticity in this suggest that the assumptions of linear regression are not fully met. Alternative approaches such as using non-linear models might give better results.

8. What is the *scientific* conclusion of your test? What do your results suggest about the media and women's political success?
(2 marks)

The results suggest that media portrayals of women candidates as experts have a weak and statistically insignificant relationship with the number of seats won by women. While a positive relationship is present, it is not significant enough to support a predictive relationship. The statistical evidence is weak, with a non-significant p-value (0.237), and a low adjusted $R2$ value (0.008), and violations of key linear regression assumptions such as normality of residuals. These may imply that other factors which are beyond media portrayals likely play a more significant role in women's political success.

Notes:

- **Only one group member submits the report**
- **The report must be a Word .DOC, .DOCX or .PDF file**
- **Make sure everyone in the group has a copy of the report**
- **Double check what you have submitted!!**
  - **view it on OnQ to make sure everything is there and visible**
- **Lastly, everyone in the group needs to submit their own version of the Rscript file.**