

Problem Set 4

Applied Stats/Quant Methods 1

Name: Darragh McGee (18319331)

Question 1: Economics

In this question, use the `prestige` dataset in the `car` library. First, run the following commands:

```
install.packages(car)
library(car)
data(Prestige)
help(Prestige)
```

We would like to study whether individuals with higher levels of income have more prestigious jobs. Moreover, we would like to study whether professionals have more prestigious jobs than blue and white collar workers.

- (a) Create a new variable `professional` by recoding the variable `type` so that professionals are coded as 1, and blue and white collar workers are coded as 0 (Hint: `ifelse`).

Create new variable `professional` using `ifelse()`:

```
1 Prestige$professional <- as.factor(ifelse(Prestige$type == "prof", 1, 0))
2 print(Prestige$professional)
```

- (b) Run a linear model with `prestige` as an outcome and `income`, `professional`, and the interaction of the two as predictors (Note: this is a continuous \times dummy interaction.)

Linear Regression Model with `prestige` as outcome variable:

```
1 model_1 <- lm(prestige ~ income*professional, data = Prestige)
2 summary(model_1)
```

Table 1: Summary of Regression Results for Prestige Model

Variable	Estimate	Std. Error	t value	Significance
Intercept	21.1423	2.8044	7.539	***
income	0.0032	0.0005	6.351	***
professional1	37.7813	4.2483	8.893	***
income:professional1	-0.0023	0.0006	-4.098	***
Model Summary				
Observations	98			
Residual Std. Error	8.012 (df = 94)			
Multiple R-squared	0.7872			
Adjusted R-squared	0.7804			
F-statistic	115.9*** (df = 3; 94)			

Significance Codes:

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

(c) Write the prediction equation based on the result.

Prediction Equation for **prestige** model:

$$\text{prestige} = 21.142 + 0.003 \cdot \text{income} + 37.781 \cdot \text{professional} - 0.002 \cdot (\text{income} \times \text{professional})$$

(d) Interpret the coefficient for **income**.

- An interaction term moderates the individual effect of each explanatory variable by capturing how the relationship between the explanatory variable and outcome depends on another explanatory variable. Therefore, the coefficients of individual explanatory variables represent the isolated effects when the interacting variable is 0.
- When the individual does not work as a professional (professional = 0) a one unit (i.e. one dollar) increase in income is associated with a 0.003 increase in the individuals prestige score.

(e) Interpret the coefficient for **professional**.

- When income is equal to zero, having a professional job (professional = 1) is associated with a 37.781 increase in prestige score. However, in reality, it is highly unlikely for an individual to hold a professional job with no income.
- Nonetheless, it represents the baseline difference between professionals and blue collar and white collar workers controlling for income.

- (f) What is the effect of a \$1,000 increase in income on prestige score for professional occupations? In other words, we are interested in the marginal effect of income when the variable **professional** takes the value of 1. Calculate the change in \hat{y} associated with a \$1,000 increase in income based on your answer for (c).

Predicted Prestige Score for Professional Occupations with 1000 dollar Income

```
1 answer_f <- 21.142 + 0.003*(1000) + 37.781*(1) - 0.002*(1000)*(1)
2 print(answer_f)
```

59.923

Baseline Prestige Score for Professional Occupations with 0 dollar Income

```
1 baseline_prestige <- 21.142 + 37.781
2 print(baseline_prestige)
```

58.923

Difference in Prestige Score from a 1,000 dollar increase in Income for professionals

```
1 prestige_increase <- answer_f - baseline_prestige
2 print(prestige_increase)
```

1

- A professional's prestige score is equal to 58.923 ($21.142 + 37.781$) when their income is zero. Therefore, the increase in prestige score associated with an increase in income of 1000 dollars for a professional can be estimated as 1 ($59.923 - 58.923$).
- The marginal effect for each dollar increase in income for professionals can be calculated by dividing 1 by 1000, which gives a value of 0.001.
- The marginal increase in prestige for each dollar increase in income for professionals can be directly verified from the prediction equation itself: $(0.003 - 0.002) = 0.001$.

- (g) What is the effect of changing one's occupations from non-professional to professional when her income is \$6,000? We are interested in the marginal effect of professional jobs when the variable `income` takes the value of 6,000. Calculate the change in \hat{y} based on your answer for (c).

Predicted Prestige Score for Non-Professional Occupations with 6000 dollar Income

```
1 answer_g_non_professional <- 21.142 + 0.003*(6000) + 37.781*(0) - 0.002*(6000)*  
  (0)  
2 print(answer_g_non_professional)
```

39.142

Predicted Prestige Score for Professional Occupations with 6000 dollar Income

```
1 answer_g_professional <- 21.142 + 0.003*(6000) + 37.781*(1) - 0.002*(6000)*(1)  
2 print(answer_g_professional)
```

64.923

Difference in Prestige Score between Professionals and Non-Professionals earning 6000 dollars

```
1 prestige_increase_g <- answer_g_professional - answer_g_non_professional  
2 print(prestige_increase_g)
```

25.781

- A non-professional's prestige score is 39.142 when their income is 6,000 dollars. A professional's prestige score is 64.923 at the same income level. Thus, the difference in prestige between a professional and a non-professional with a 6,000 dollar income is 25.781 (64.923 - 39.142), representing the effect of professional occupation on prestige at this income level.

Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting preferences.¹ Working with a campaign in Fairfax County, Virginia, 131 precincts were randomly divided into a treatment and control group. In 30 precincts, signs were posted around the precinct that read, “For Sale: Terry McAuliffe. Don’t Sellout Virginia on November 5.”

Below is the result of a regression with two variables and a constant. The dependent variable is the proportion of the vote that went to McAuliff’s opponent Ken Cuccinelli. The first variable indicates whether a precinct was randomly assigned to have the sign against McAuliffe posted. The second variable indicates a precinct that was adjacent to a precinct in the treatment group (since people in those precincts might be exposed to the signs).

Impact of lawn signs on vote share	
Precinct assigned lawn signs (n=30)	0.042 (0.016)
Precinct adjacent to lawn signs (n=76)	0.042 (0.013)
Constant	0.302 (0.011)

Notes: $R^2=0.094$, $N=131$

¹Donald P. Green, Jonathan S. Krasno, Alexander Coppock, Benjamin D. Farrer, Brandon Lenoir, Joshua N. Zingher. 2016. “The effects of lawn signs on vote outcomes: Results from four randomized field experiments.” *Electoral Studies* 41: 143-150.

- (a) Use the results from a linear regression to determine whether having these yard signs in a precinct affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Linear Regression Assumptions

- **Linear Relationship:** There is a linear relationship between the outcome and explanatory variables.
- **Independence of Errors:** The errors (residuals) are independent of each other.
- **Normality of errors:** For any given value of the explanatory variable, the errors (residuals) are assumed to follow a normal distribution.
- **Constant variance (Homoscedasticity):** The variance of the errors is constant across all values of the explanatory variable.
- **No Perfect Multi-Collinearity:** The explanatory variables should not be perfectly correlated with each other.

Setting Up Hypothesis

- **Null Hypothesis:** Having yard signs in a precinct DOES NOT affect vote share ($\beta = 0$)
- **Alternative Hypothesis:** Having yard signs in a precinct DOES affect vote share ($\beta \neq 0$)

Calculate the Test Statistic

```
1 TS_1 <- (0.042 - 0) / 0.016
2 print(TS_1)
```

2.625

Calculate the p-value

```
1 p_value_1 <- 2*pt(abs(TS_1), 131-2, lower.tail = F)
2 print(p_value_1)
```

0.009711646

Conclusion

- Reject the Null Hypothesis as the p-value is less than 0.05 (below the significance level).
- There is sufficient evidence to conclude that having yard signs in a precinct does effect vote share.

- (b) Use the results to determine whether being next to precincts with these yard signs affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Setting Up Hypothesis

- **Null Hypothesis:** Being next to precincts with yard signs DOES NOT affect vote share ($\beta = 0$).
- **Alternative Hypothesis:** Being adjacent to precincts with yard signs DOES affect vote share ($\beta \neq 0$).

Calculate the Test Statistic

```
1 TS_2 <- (0.042 - 0) / 0.013
2 print(TS_2)
```

3.230769

Calculate the p-value

```
1 p_value_2 <- 2*pt(abs(TS_2), 131-2, lower.tail = F)
2 print(p_value_2)
```

0.001566685

Conclusion

- Reject the Null Hypothesis as the p-value is less than 0.05 (below the significance level).
- There is sufficient evidence to conclude that being adjacent to precincts with yard signs DOES affect vote share.

- (c) Interpret the coefficient for the constant term substantively.

Prediction Equation for Proportion of Vote Share

$$\begin{aligned} \text{Proportion of Vote Share} = & 0.302 + 0.042 \cdot \text{Precincts Assigned Lawn Signs} \\ & + 0.042 \cdot \text{Precincts Adjacent to Lawn Signs} \end{aligned} \quad (1)$$

- The constant (0.302) represents the expected proportion of the vote share that went to McAuliff's opponent, Ken Cuccinelli, in precincts that neither had yard signs nor were adjacent to precincts with yard signs.

- (d) Evaluate the model fit for this regression. What does this tell us about the importance of yard signs versus other factors that are not modeled?
- The R-squared value (i.e., the proportion of variance in the outcome variable explained by the explanatory variables in the model) is 0.094. This means that less than 10 percent of the variance in vote share is explained by the presence of yard signs in the precinct or being adjacent to a precinct with yard signs.
 - The relatively low R-squared value suggests that other factors likely have a more substantial impact on vote share.