

Problem Set 4

Yajie Dong

30th November 2023

1 Question 1

- 1.1 (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`).

Step 1: Installing and loading the 'car' package: This installs and then loads the 'car' package, which provides functions and datasets for Companion to Applied Regression, a book on regression analysis.

Do it in R:

```
1 install.packages("car")
2 library(car)
```

Step 2: Loading the Prestige dataset: the Prestige dataset from the 'car' package is loaded. The `help(Prestige)` line provides documentation on this dataset, which includes data on the prestige of Canadian occupations.

Do it in R:

```
1 data(Prestige)
2 help(Prestige)
```

Step 3: Creating a new variable 'professional': It adds a new column to the Prestige dataset. The new column is named `professional`. It assigns a value of 1 if the type of occupation is "prof" (professional), and 0 otherwise.

Do it in R:

```
1 Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
```

1.2 (b) Run a linear model with prestige as an outcome and income, professional, and the interaction of the two as predictors.

Let's do it in R: A linear regression model is built with prestige as the outcome variable, and income, professional, and the interaction between income and professional as predictor variables.

```
1 model <- lm(prestige ~ income + professional + income:professional, data =
  Prestige)
2 summary(model)
```

Table 1: Linear Model Results

Coefficient	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	21.14226	2.80443	7.539	2.93e-11 ***
Income	0.00317	0.00050	6.351	7.55e-09 ***
Professional	37.78128	4.24827	8.893	4.14e-14 ***
Income:Professional	-0.00233	0.00057	-4.098	8.83e-05 ***

Residual standard error: 8.012 on 94 degrees of freedom

(4 observations deleted due to missingness)

Multiple R-squared: 0.7872, Adjusted R-squared: 0.7804

F-statistic: 115.9 on 3 and 94 DF, p-value: < 2.2e-16

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

The prediction equation from a linear regression model can be formulated using the estimated coefficients from the model's output. Based on your model's summary, the prediction equation for the prestige (Y) in the Prestige dataset is as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \quad (1)$$

Where:

- β_0 (Intercept) = 21.1422589
- β_1 (Coefficient for x_1 , Income) = 0.0031709
- β_2 (Coefficient for x_2 , Professional) = 37.7812800
- β_3 (Coefficient for the interaction between x_1 and x_2) = -0.0023257

In this representation:

- Y is the **prestige** score.
- x_1 represents the **Income**.
- x_2 is the binary variable indicating whether the occupation is professional (1) or not (0).

Putting these values into the equation, we get:

$$Y = 21.1422589 + 0.0031709 \times x_1 + 37.7812800 \times x_2 - 0.0023257 \times x_1 \times x_2 \quad (2)$$

The explanation of the prediction equation:

This equation can be used to predict the **prestige** score (Y) based on the values of **Income** (x_1) and whether the occupation is professional (x_2) or not, along with the interaction effect between **Income** (x_1) and **Professional** (x_2).

1.4 (d) Interpret the coefficient for income.

Coefficient: 0.003170909

Interpretation: This coefficient suggests that, for non-professional occupations (since the effect is modified for professionals by the interaction term), each additional unit of income (presumably \$1) is expected to increase the prestige score by approximately 0.00317 points, holding other factors constant.

1.5 (e) Interpret the coefficient for professional.

Coefficient: 37.78128

Interpretation: This suggests that, all else being equal, professional occupations are associated with a 37.78128 point increase in the prestige score compared to non-professional occupations.

1.6 (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 associated with a \$1,000 increase in income based on your answer for (c).

This question asks about the effect of a \$1,000 increase in income on the prestige score for professional occupations, I'll use the prediction equation and focus on how the increase in income (when $x_2 = 1$, i.e., for professional occupations) affects the prestige score Y .

Prediction Equation:

$$Y = 21.1422589 + 0.0031709 \times x_1 + 37.7812800 \times x_2 - 0.0023257 \times x_1 \times x_2 \quad (3)$$

Calculation for Question(f): We are interested in the marginal effect of income (x_1) when x_2 (professional) is 1.

- The coefficient for income (β_1) is 0.0031709.
- The coefficient for the interaction term (β_3) is -0.0023257.

The marginal effect of income for professional occupations is the income coefficient plus the interaction term coefficient:

$$\text{Marginal Effect for Professionals} = \beta_1 + \beta_3 = 0.0031709 - 0.0023257 = 0.0008452 \quad (4)$$

To find the effect of a \$1,000 increase in income:

$$\text{Effect of \$1,000 increase} = 0.0008452 \times 1000 = 0.8452 \quad (5)$$

Double check it in R

```
1 # Calculating the effect for (f)
2 income_effect_professional <- (coefficients["income"] + coefficients["income:
  professional"]) * 1000
3 cat("Effect of $1,000 increase in income for professionals: ", income_effect_
  professional, "\n")
```

Thus, for professional occupations, a \$1,000 increase in income is associated with an increase of approximately 0.8452 in the prestige score, considering the interaction between income and professional status.

1.7 (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 based on your answer for (c).

This question asks about the effect of changing one's occupation from non-professional to professional when the income is \$6,000, we'll again use the prediction equation. This time, we focus on how the change in professional status (x_2) at a specific income level ($x_1 = 6000$) affects the prestige score (Y).

Prediction Equation:

$$Y = 21.1422589 + 0.0031709 \times x_1 + 37.7812800 \times x_2 - 0.0023257 \times x_1 \times x_2 \quad (6)$$

Calculation for question (g): We are interested in the marginal effect of changing from a non-professional occupation to a professional one when the income (x_1) is \$6,000.

- The coefficient for professional (β_2) is 37.7812800.
- The coefficient for the interaction term (β_3) is -0.0023257.
- The specified income level is \$6,000.

The effect of changing occupation to professional at an income of \$6,000 includes both the professional coefficient and the interaction term adjusted by the income level:

$$\text{Change Effect} = \beta_2 + (\beta_3 \times 6000) \quad (7)$$

$$\text{Change Effect} = 37.7812800 + (-0.0023257 \times 6000) \quad (8)$$

Calculating this:

$$\text{Change Effect} = 37.7812800 - 13.9542 \quad (9)$$

$$\text{Change Effect} \approx 23.82708 \quad (10)$$

Double check it in R:

```

1 # Calculating the effect for (g)
2 change_effect <- coefficients["professional"] + coefficients["income:professional"]
  * 6000
3 cat("Effect of changing to professional with $6,000 income: ", change_effect, "\n")

```

Therefore, changing one's occupation from non-professional to professional with an income of \$6,000 is associated with an increase of approximately 23.82708 in the prestige score. This calculation considers both the inherent prestige of being in a professional occupation and the interaction effect of income and professional status.

2 Question 2

a) Hypothesis Test for Precincts Assigned Lawn Signs

Initial Explanation: We test the null hypothesis that lawn signs have no effect on the vote share against the alternative hypothesis that they do have an effect.

Hypothesis Test

Null Hypothesis H_0 : The coefficient for the effect of lawn signs $\beta_{\text{lawn signs}}$ is zero, implying that lawn signs have no effect on the vote share for Ken Cuccinelli's opponent.

$$H_0 : \beta_{\text{lawn signs}} = 0$$

Alternative Hypothesis H_a : The coefficient is not zero, suggesting that lawn signs do affect the vote share.

$$H_a : \beta_{\text{lawn signs}} \neq 0$$

T-Statistic Calculation The t-statistic for the lawn signs coefficient is calculated as:

$$t_{\text{lawn signs}} = \frac{\hat{\beta}_{\text{lawn signs}}}{SE_{\hat{\beta}_{\text{lawn signs}}}} = \frac{0.042}{0.016}$$

Let calculate t-values in R:

```

1 # Coefficients and Standard Errors
2 coeff_lawn_signs <- 0.042
3 se_lawn_signs <- 0.016
4 coeff_adjacent <- 0.042
5 se_adjacent <- 0.013
6 # Calculating t-values for (a)

```

```

7 t_lawn_signs <- coeff_lawn_signs / se_lawn_signs
8 cat("t-value for precincts with lawn signs: ", t_lawn_signs, "\n")

```

Thus, the calculated t-statistic is:

$$t_{\text{lawn signs}} = 2.625$$

Interpretation of T-Statistic:

The t-statistic of 2.625 indicates the number of standard errors that the estimated coefficient (0.042) is away from zero. Since this t-value is greater than the critical t-value for our degrees of freedom at a 95% confidence level (usually around 2.04 for a large sample), we have sufficient evidence to reject H_0 . This implies that the presence of lawn signs in a precinct significantly increases the vote share for Ken Cuccinelli's opponent, contrary to the null hypothesis of no effect. The specific t-value also suggests that the result is statistically significant at the 5% level, meaning there is less than a 5% probability that such an extreme result is due to chance if the true coefficient were actually zero.

b) Hypothesis Test for Precincts Adjacent to Lawn Signs

Initial Explanation: Similar to part (a), we test the null hypothesis that being adjacent to precincts with lawn signs has no effect on the vote share.

Null Hypothesis H_0 : The coefficient for the effect of being adjacent to precincts with lawn signs β_{adjacent} is zero, implying that adjacency has no effect on the vote share.

$$H_0 : \beta_{\text{adjacent}} = 0$$

Alternative Hypothesis H_a : The coefficient is not zero, suggesting that adjacency to lawn signs affects the vote share.

$$H_a : \beta_{\text{adjacent}} \neq 0$$

T-Statistic Calculation:

$$t_{\text{adjacent}} = \frac{\hat{\beta}_{\text{adjacent}}}{SE_{\hat{\beta}_{\text{adjacent}}}} = \frac{0.042}{0.013}$$

Let calculate t-values in R:

```

1 # Calculating t-values for (b)
2 t_adjacent <- coeff_adjacent / se_adjacent
3 cat("t-value for precincts adjacent to lawn signs: ", t_adjacent, "\n")

```

Thus,

$$t_{\text{adjacent}} = 3.230769$$

Interpretation: The t-statistic of 3.230769 suggests that the coefficient estimate (0.042) is more than 3 standard errors away from zero. This t-value exceeds the critical value for a 95% confidence level and therefore strongly indicates rejection of H_0 . We conclude that being adjacent to precincts with lawn signs also significantly increases the vote share for Ken Cuccinelli's opponent. The strength of this result, indicated by the higher t-value compared to part (a), reinforces the conclusion that adjacency to these signs has a notable impact on voting preferences. Again, the probability that this result is due to random chance is very low (well below 5%), lending confidence to our conclusion.

In both cases, the specific t-statistics are high enough to indicate that the effects are statistically significant, meaning that lawn signs and their visibility in adjacent precincts do have a measurable impact on vote share in this election context.

c) Interpretation of the Constant Term

The coefficient for the constant term in a regression model represents the expected value of the dependent variable when all the independent variables are held at zero. In the context of this study:

Constant: 0.302

Interpretation: The constant term of 0.302 implies that if none of the precincts were assigned lawn signs and none were adjacent to precincts with lawn signs (i.e., the independent variables are zero), the expected proportion of the vote share for Ken Cuccinelli's opponent would be 30.2%. This can be considered the baseline level of support for Cuccinelli's opponent in the absence of the lawn sign campaign.

d) Model Fit Evaluation

R^2 : 0.094

Interpretation: The R^2 value of 0.094 indicates that the model explains only 9.4% of the variance in the vote share for Cuccinelli's opponent. This suggests that while the model captures a statistically significant portion of the vote share variation due to lawn signs and adjacency, a

majority of the variation (90.6%) is due to other factors not included in the model.

What This Tells Us:

The relatively low R^2 value indicates that there are likely many other factors influencing the vote share that are not captured by the presence or adjacency of lawn signs alone. These could include, but are not limited to, factors such as the overall political climate, the effectiveness of other campaign strategies (like advertisements, debates, and canvassing), voter demographics, historical voting patterns, or even broader issues that were important during the election cycle.

While lawn signs have a noticeable impact, they are far from being the only or even the most influential factor in determining voter preferences. Campaigns should consider a holistic approach that integrates various strategies to address the wide range of influences on voter behavior.

The analysis underlines the complexity of voter decision-making processes and the challenges in isolating the effects of individual campaign tactics. It is a reminder for political scientists and campaign managers alike that electoral outcomes are multi-faceted and can seldom be attributed to single causes.