# Problem Set 4

## Applied Stats/Quant Methods 1

### Due: November 18, 2024

## Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on GitHub.

- This problem set is due before 23:59 on Monday November 18, 2024. No late assignments will be accepted.

## 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`).

```
1  install.packages("car")
2  library(car)
3  data(Prestige)
4  # Question 1
5  Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
```

(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 × dummy interaction.)

Table 1: Linear Model Output

|  | *Dependent variable:* |
| --- | --- |
|  | Prestige |
| Income | 0.003*** |
|  | (0.0005) |
|  |  |
| Professional | 37.781*** |
|  | (4.248) |
|  |  |
| Income:Professional | −0.002*** |
|  | (0.001) |
|  |  |
| Constant | 21.142*** |
|  | (2.804) |
|  |  |
| Observations | 98 |
| $R^2$ | 0.787 |
| Adjusted $R^2$ | 0.780 |
| Residual Std. Error | 8.012 (df = 94) |
| F Statistic | 115.878*** (df = 3; 94) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

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

$$\hat{\text{prestige}} = 21.1423 + 0.0032 \cdot \text{income} + 37.7813 \cdot \text{professional} - 0.0023 \cdot (\text{income} \cdot \text{professional}) + \epsilon$$

(d) Interpret the coefficient for `income`.

According to Table 1, for each one unit increase in income, the prestige level increases by 0.0032 percentage points. This is when the 'professional' level is zero.

(e) Interpret the coefficient for `professional`.

According to Table 1, there is a strong positive coefficient for professional occuations. This means that professional occupations (those equal to one) are associated with a significantly higher prestige, 37.781***, or 38 percentage points, than non-professional ones.

(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).

To attain the marginal effect of income we need to calculate the partial derivative of the prediction equation which is as follows:

$$\frac{\partial \hat{y}}{\partial \text{income}} = \beta_{\text{income}} + \beta_{\text{income:professional}} \cdot \text{professional}$$

This becomes:
$$\Delta \hat{y} = (0.00317 + (-0.00232)) \cdot 1000 = 0.8452$$

This result indicates that, for professionals, a 1,000 increase in income is associated with an increase of about 0.85 percentage points in the prestige score.

```
1  #Extract coefficients from the model
2  coef_income <- coef(model)["income"]
3  coef_interaction <- coef(model)["income:professional"]
4  #Calculate the marginal effect of income when professional = 1
5  marginal_effect_professional <- coef_income + coef_interaction
6  View(marginal_effect_professional)
7  # Calculate the change in prestige for a $1,000 increase in income
8  change_in_prestige <- marginal_effect_professional * 1000
9  View(change_in_prestige)
```

(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).

Here we need to know the difference when professional changes from the value 0 to the value 1 when income has the value $6,000$ The effect of switching to professional requires the difference between the predicted values of professional to non-professional. We can use the following equation to determine this:

$$\Delta \hat{y} = \hat{y}_{\text{professional}} - \hat{y}_{\text{non-professional}}$$

We can enter the coefficients from the table and it becomes:

$$\Delta \hat{y} = \hat{y}_{\text{professional}} - \hat{y}_{\text{non-professional}} = 37.781 + (-0.00232 \cdot 6000) = 23.861$$

We can therefore conclude that when income is 6,000 dollars per annum, the prestige score increases by 23.98 percentage points if the person switches to a professional occupation.

```r
# Given coefficients based on the answer for (c)
coef_professional <- 37.7813
coef_interaction <- -0.0023
# Calculate the marginal effect of changing from non-professional to
    professional with an income of = 6000
change_in_prestige_professional <- coef_professional + (coef_interaction
    * 6000)
# Display the result
change_in_prestige_professional
```

# Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting prefer-ences.[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 Virgina 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

(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 $\alpha = .05$).

I will use a Student's t-test and I am interested in the partial effect of one individual regression coefficient. I want to know if having yard signs in a precinct effects the vote share.

**Null Hypothesis** ($H_0$): The coefficient is zero, meaning that yard signs have no effect on vote share (for Cuccinelli) in the precincts where they were posted.

$$H_0 : \beta = 0$$

**Alternative Hypothesis** ($H_1$): The coefficient is not zero, meaning that yard signs do affect the vote share in those precincts.

---

[1] Donald 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.

$$H_1 : \beta \neq 0$$

The $t$-statistic for testing whether a coefficient is significantly different from zero is calculated as:

$$t = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}$$

In the case of this regression, we can calculate the t-statistic using the above formula:

$$t = \frac{0.042}{0.016} = 2.625$$

Since our calculated $t$-statistic is 2.625, which exceeds the critical value of 1.96, we can reject the null hypothesis. This implies that the presence of yard signs in a precinct has a statistically significant effect on the vote share for Cuccinelli at the 0.05 significance level, suggesting that the signs likely influenced voting behavior.

(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 $\alpha = .05$).

Similarly to the previous question, I will use a Student's t-test as I am interested in the partial effect of one individual regression coefficient. I want to know whether being adjacent to precincts with yard signs affects vote share.

Step 1: Set up the Null

**Null Hypothesis** ($H_0$): The coefficient is zero, meaning that being adjacent to precincts with yard signs has no effect on Cuccinelli's vote share.

$$H_0 : \beta = 0$$

**Alternative Hypothesis** ($H_1$): The coefficient is not zero, meaning that being adjacent to precincts with yard signs does affect Cuccinelli's vote share.

$$H_1 : \beta \neq 0$$

As above, the $t$-statistic for testing whether a coefficient is significantly different from zero is calculated as:

$$t = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}$$

We can calculate the t-statistic using the above formula:

$$t = \frac{0.042}{0.013} = 3.231$$

7

Once again, we observe a t-statistic of 3.231, which is greater than the critical value of 1.96, so we reject the null hypothesis. This suggests that being adjacent to precincts with yard signs has a statistically significant effect on Cuccinelli's vote share.

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

The constant is 0.302 with a standard error of 0.011. This tells us the baseline vote share for Ken Cuccinelli without there being any effect from the signs placed against his opponent, McAuliff. The baseline vote share tells us that within presincts that do not have any signs against McAuliff in place, Cucinelli can expect to recieve 30.2 percent of the vote share. This is further supported by the low standard error term beneath it (0.011)

(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? To evaluate the goodness of

fit for this model, I'll first examine the $R^2$, which in this case is **0.094**. This indicates that the model explains approximately **9.4% of the variation in vote share**. This relatively low $R^2$ suggests that a large portion of the variation in vote share remains unexplained by the model. This means that there are probably a large number of other factors that explain varience in Cucinelli's vote share perhaps with greater precision than the anti-McAuliff signs. For example, a strong canvassing campaign, hisorical family association, etc.