Fundamentals of Econometrics

Homework Solution Sample Template

This template provides a guide of how I anticipate that your homework solutions will look. Where possible, please place your answers directly below the questions. This will make it easier for us to grade your work. Otherwise, please feel free to exercise your creativity when producing your solutions.

1. **Unless stated, you are required to display the R chunk that produced your results. Hence the `echo = TRUE` argument in the `setup` chunk.**
2. Please ensure you do a quick spell check of your document. Press **F7** on your keyboard.
3. An appropriate title must accompany all tables, graphs, and figures. Graph axes must be labeled where appropriate.

- **Whenever appropriate, please try to refer to the output.** Please take a look at the `.Rmd` file in this template to see how I am able to:

i. Add captions to my plots,
ii. hyperlink and reference the plots automatically, and
iii. embed the results of variables computed and stored in the R chunks into my text. Gone are the days when you had to memorize the result and then type it over in your Word document. *Once you have it stored as a variable, you can directly extract it in the document text.*

Question 1: Generating random variables

You are given the following model:

$$y_i = \beta_0 + \varepsilon_i$$

Obtain the OLS estimator of β_0 by minimizing the sum of squared residuals. Show that the OLS estimator of β_0 is $\hat{\beta}_0 = \bar{y}$.

$$\min \hat{\varepsilon}^2 = \min \sum_{i=1}^{n} (y_i - \beta_0)^2 \tag{1}$$

$$\frac{\partial \hat{\varepsilon}^2}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0) = 0 \tag{2}$$

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 = 0$$

$$\beta_0 = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y} \tag{3}$$

Answer: As seen in Equation (3), the OLS estimator of β_0 is $\hat{\beta}_0 = \bar{y}$. This means that if we do not have additional regressors, the OLS estimator of intercept, β_0, is the sample mean of the dependent variable.

Question 2: Stargazer in action

The `stargazer` package is a powerful tool for presenting regression results in a professional manner. For this example, I will use an example from that package's vignette.

Estimate the following models using the `attitude` dataset:

$$\begin{aligned}
\text{Model 1:} \quad \text{rating} &= \beta_0 + \beta_1 \cdot \text{complaints} + \beta_2 \cdot \text{privileges} + \beta_3 \cdot \text{learning} + \beta_4 \cdot \text{raises} + \beta_5 \cdot \text{critical} + u \\
\text{Model 2:} \quad \text{rating} &= \beta_0 + \beta_1 \cdot \text{complaints} + \beta_2 \cdot \text{privileges} + \beta_3 \cdot \text{learning} + u \\
\text{Model 3:} \quad \text{rating} &= \beta_0 + \beta_1 \cdot \text{complaints} + u
\end{aligned}$$

a. Store the results in `linear.1`, `linear.2`, and `linear.3`, respectively.

```
linear.1 <- lm(rating ~ complaints + privileges + learning + raises +
                 critical,
               data = attitude)
linear.2 <- lm(rating ~ complaints + privileges + learning,
               data=attitude)
linear.3 <- lm(rating ~ complaints, data = attitude)
```

b. Use the `stargazer` package to present the results in a single table.

```
stargazer(linear.1, linear.2, linear.3,
          title = "OLS Regression Results",
          type = "latex", header = FALSE, digits = 3,
          font.size = "small", notes.align = "l")
```

Table 1: OLS Regression Results

	Dependent variable:		
	rating		
	(1)	(2)	(3)
complaints	0.692***	0.682***	0.755***
	(0.149)	(0.129)	(0.098)
privileges	−0.104	−0.103	
	(0.135)	(0.129)	
learning	0.249	0.238*	
	(0.160)	(0.139)	
raises	−0.033		
	(0.202)		
critical	0.015		
	(0.147)		
Constant	11.000	11.300	14.400**
	(11.700)	(7.320)	(6.620)
Observations	30	30	30
R^2	0.715	0.715	0.681
Adjusted R^2	0.656	0.682	0.670
Residual Std. Error	7.140 (df = 24)	6.860 (df = 26)	6.990 (df = 28)
F Statistic	12.100*** (df = 5; 24)	21.700*** (df = 3; 26)	59.900*** (df = 1; 28)
Note:	*p<0.1; **p<0.05; ***p<0.01		

c. Interpret the coefficient on `complaints` in the first two models.

Answer: The coefficient on `complaints` in Model 1 is 0.692 while in Model 2, it is 0.682. This implies that for a 1% increase in the favorable responses to handling of employee complaints, the rating increases by 0.692 in Model 1 and 0.682 in Model 2.

Notice: Here, decided not to copy and paste but instead, used the \`r\` command to extract the coefficients from the models and display them in the text. This is a powerful tool that you can use to refer to the results directly in your text. Think of all the time I saved (and errors I avoided) if the data were to be updated later on.

---

d. Visualize the model fit for Model 3.

```
attitude |> mutate(fitted3 = fitted(linear.3)) |>
  ggplot(aes(x = complaints, y = rating)) +
  geom_point(color = "goldenrod") +
  geom_line(aes(y = fitted3), color = "red",lwd = 1.1) +
  labs(title = "Model 3 Fit",
       x = "Complaints",
       y = "Rating")
```