BS2280 - Econometrics 1

Lecture 12 - How To Write An Econometrics Report

Dr. Yichen Zhu

R Workshop Feedback Survey

Use QR code above or click the following link:
`https://forms.office.com/e/RGtS610xV0`

Structure of today's lecture

R Programming

- All R workshop worksheets (BB Week 2, 4, 6, 9 and 11)
- All R workshop Walk-through videos (BB Week 2, 4, 6, 9 and 11)
- Stargazer (BB Week 11 and 12)
- Weekly slide codes
- Online resources

Interpretations

Simple linear model:

$$Life\_Expectancy_i = \beta_1 + \beta_2 Alcohol_i + \varepsilon_i \quad (1)$$

Carefully interpret the regression coefficients and their statistical significance. Besides, discuss the Goodness-of-Fit of the model. Explain why the Alcohol coefficient of this simple regression model may be biased.

- Week 2 Part 2
- Week 3 Part 2
- Week 4 Part 2
- Homework 2, 3, 4 and 5

Interpretations

Multiple linear regression model I:

$$Life\_Expectancy_i = \beta_1 + \beta_2 Alcohol_i + \beta_3 Schooling_i + \beta_4 BMI_i + \varepsilon_i \quad (2)$$

Carefully interpret the regression coefficients and their statistical significance. Discuss the Goodness-of-Fit. Undertake an F-test to find out if adding a new set of variables to the model improves the explanatory power.

- Week 5 Part 1 and Part 2
- Week 6 Part 1 and Part 2
- Homework 5 and 6

Interpretations

Multiple linear regression model II:

$$Life\_Expectancy_i = \beta_1 + \beta_2 Alcohol_i + \beta_3 Schooling_i + \beta_4 Schooling_i^2 + \beta_5 BMI_i + \varepsilon_i \quad (3)$$

Carefully interpret the regression coefficients (especially the coefficients of the Schooling variables), and discuss the Goodness-of-Fit. Using adj. R squared, compare the fit of models (2) and (3). Does the additional squared term improve the fit of the model?

- Week 5 Part 1 and Part 2
- Week 6 Part 1 and Part 2
- Week 10 Part 2
- Week 11 Part 2
- Homework 10

Interpretations

Logarithmic model:

$$log(Life\_Expectancy_i) = \beta_1 + \beta_2 Alcohol_i + \beta_3 log(Schooling_i) + \beta_4 BMI_i + \varepsilon_i \quad (4)$$

Carefully interpret the regression coefficients and discuss the Goodness-of-Fit of the model.

- Week 5 Part 1 and Part 2
- Week 6 Part 1 and Part 2
- Week 9 Part 2
- Week 10 Part 1 and Part 2
- Homework 9

Title page

- The title page should include the report title, the module code / name and your candidate number.
- Do not give your name

Introduction

- Provide an outline of the topic of your report and highlight its relevance. Top tip: You can make references here to newspaper articles or journal papers. (**approx. 100 words, 5 marks**).

- Quite likely requires re-write at the very end!
- You can write this once you've completed other sections

Methodology

- Provide a brief description of the data sources and methods employed in your report (**approx. 100 words, 5 marks**).

- **Data sources**
- When introducing data sources, provide a brief overview mentioning the types of data used, their origin, and their relevance to the topic or study. This sets the context for the information you'll be referencing or analysing.
- Give enough details such that someone else could get the same data as you
- If your data have shortcomings this is a good place to say so

Methodology

- Provide a brief description of the data sources and methods employed in your report (**approx. 100 words, 5 marks**).

- **Methods**
- Think of your audience: module leader, moderator and this BS2280 class
- Do not merely copy stuff from the lecture notes; There's no necessity to reintroduce simple/multiple regression models. We all know it!
- Simply focus on detailing the type of your model and clearly present each model using equations.
- It will be better to discuss any shortcomings of the methodology within your specific setting.

Regression Analysis

- This will be the main part of your report. Do not use bullet points - this is a report not a power point presentation.
- This section has to include the regression analysis. Add here the well-formatted results table that you have created in section 3.2. Interpret the coefficients and their statistical significance of **each regression model** and comment on the goodness of fit of each model. State here all your responses to the tasks given in section 3.2 (**approx. 1,100 words, 80 marks**).

- Offer interpretations of the statistical outcomes, considering aspects like **t-values, p-values, statistical significance, goodness-of-fit R^2, adj.R^2, F-tests, etc**
- Provide economic interpretations of the results, considering **coefficients and their implications**
- Excellent report reveals some economic implications and includes critical discussions with references for support

Conclusions and limitations

- Conclude your report by summarising your findings, provide an outlook for further investigations and / or policy recommendations. Furthermore, good report will also highlight any limitations of the analysis. (**approx. 200 words, 10 marks**).

- Clearly summarise your main findings, think of someone who reads nothing but the introduction and the conclusion
- Summarise shortcomings of your work

References

- You have to provide a reference section at the end of the report conforming to the (British) Harvard-style referencing. Present here all the source you have used throughout this report.

- You have to use consistent referencing and add a reference section at the end (after the conclusion but before the appendix).
- A commonly used reference style is British Harvard Style. Aston University has an excellent guide on referencing: `https://libguides.aston.ac.uk/referencing`

Appendix

- Include here all the R commands that you have used and all R-output that you have used in your report. Furthermore, you should add here any further statistics and graphs that could provide further information for your study but did not fit into the main section of your report.

- Both R codes and outputs
- Good example:

Appendix

```
# Preparing the workspace
setwd("/Users/Desktop/Aston/Econometrics/Rworkshop5")


# Import data
load("~/Desktop/Aston/Econometrics/Rworkshop5/wages.Rdata")


# Task 1
wages$lnexper <- log(wages$exper)     # log transformation of
exper to get lnexper

sum(is.infinite(wages$lnexper))
[1] 0
```

Equation

- Use separate lines for equations rather than embedding equations within the text.
- Write the equations correctly
- Number your equations
- Good examples:

$$LGEARN_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + \beta_4 SEXP_i + u_i \tag{5}$$

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i \tag{6}$$

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i \tag{7}$$

$$\frac{\partial \widehat{LGEARN}_i}{\partial EXP_i} \text{ or } \frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i \tag{8}$$

Table

- Important elements
  1. Clear (high resolution)
  2. Number/Title
  3. Good row and column labelling (Variable name)
  4. Put all regression model outputs into a single table
  5. Not all R outputs are reported, only report those we are interested in
  6. Just copy estimation output from R is not appropriate
  7. Stargazer output can be a good starting point (R Workshop 4 and 5)
  8. Do not include R codes into your report write-up

Table

- Good example

Table 3: Results

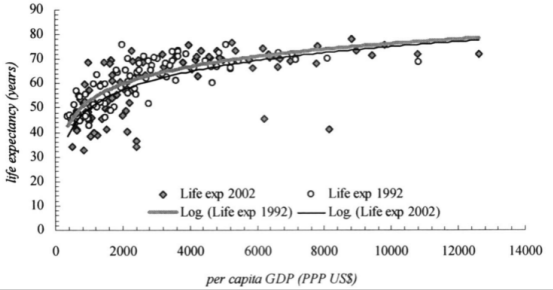
	Dependent variable:		
	rating		high.rating
	OLS		*probit*
	(1)	(2)	(3)
complaints	0.692***	0.682***	
	(0.149)	(0.129)	
privileges	−0.104	−0.103	
	(0.135)	(0.129)	
learning	0.249	0.238*	0.164***
	(0.160)	(0.139)	(0.053)
raises	−0.033		
	(0.202)		
critical	0.015		−0.001
	(0.147)		(0.044)
advance			−0.062
			(0.042)
Constant	11.011	11.258	−7.476**
	(11.704)	(7.318)	(3.570)
Observations	30	30	30
R^2	0.715	0.715	
Adjusted R^2	0.656	0.682	
Log likelihood			−9.087
Akaike Inf. Crit.	210.375	206.412	26.175
Bayesian Inf. Crit.	220.183	213.418	31.780
Residual Std. Error	7.139($df = 24$)	6.863($df = 26$)	
F statistic	12.063***($df = 5; 24$)	21.743***($df = 3; 26$)	
Note:			*p<0.1; **p<0.05; ***p<0.01

- Note: No need to report Log likelihood, Akaike Inf. Crit. and Bayesian Inf. Crit. for this report

Figure

- Important elements
  1. Clear (high resolution)
  2. Number/Title
  3. Good axis labelling (Variable name and description)
  4. Notes with enough details to make the figure almost self-explanatory
- Good example

FIGURE 2 CHANGE IN LIFE EXPECTANCY AND INCREASING INCOME



Bad example

- Just copy estimation output from R
- Equation is embedded in the text

Table 3

Call:
lm(formula = FDI2122_data_set_1$YR2016FDI ~ FDI2122_data_set_1$YR2016GDP, data = FDI2122_data_set_1)

	min	1Q	median	3q	Max
Residuals:	-89632	-9175	-8261	-5717	255027

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8417.254	4231.864	1.989	0.0493 *
FDI2122_data_set_1$YR2016GDP	22.786	1.874	12.162	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 42200 on 107 degrees of freedom
Multiple R-squared: 0.5802, Adjusted R-squared: 0.5763
F-statistic: 147.9 on 1 and 107 DF, p-value: < 2.2e-16

From the out above the estimated regression line equation can be written as FDI= 8417.2554 + 22.786*GDP . The intercept (b0) is 8417.2554. It can be interpreted as the predicted FDI unit for a zero gross domestic product as a unit. Let say GDP is 1000 units using the equation FDI will equal 8417.2554 + 22.786*1000= $31,203.25 millions of dollars.

Bad example

- Table is not clear, poor resolution
- No title

	Dependent variable:
	YR2016FDI
YR2016GDP	22.786***
	(1.874)
Constant	8,417.254**
	(4,231.864)
Observations	109
R2	0.580
Adjusted R2	0.576
Residual Std. Error	42,195.430 (df = 107)
F Statistic	147.907*** (df = 1; 107)
Note:	*p<0.1; **p<0.05; ***p<0.01

In this model, the GDP coefficient is 22.786, which is a positive coefficient, meaning that every 1 percent rise in GDP, the GDP co-efficient increases by 22.786.