

**The TEFL University PHD Course**

**Module 2**

**Module Notes:**

Make sure that you save your course-work as you go along!

You must also add your full name to the end of the file name of this document before you start. For example; with this module which is currently saved as:

phd-module2-TEFL-711-yourname

To change the file name, simply go to the top left of your screen, and click on **File**, **Save As**, then change the name of the file by simply adding your name to the end of the existing file name and save the module to a folder on your computer (‘My Documents’ for example). As an example; if your name were **Jane Mortimer**, you would save the file as:

phd-module2-TEFL-711-Jane-Mortimer

**Submitting your completed modules**

Once you have completed a module, you may use the link below to re-visit the course download/upload area, where you will find the **drop-box** to send your completed work back to us. Please note that grading will not commence until all modules are fully completed:

<http://www.thetefluniversity.com/phd-course-download.html>

**TEFL 711—Quantitative Research**

****

**Course Description**

Quantitative Research Methods is designed to provide Ph. D. students with tools and skills necessary in academic research. This course focuses on linear regression techniques. A good understanding will enable students to apply these techniques, as well as acquire on their own additional multivariate statistical techniques rooted in linear methodology, such as discriminant analysis, factor analysis, or structural equations.

This course presents single-equation regression models with two and three variables, including estimation and inference. It also examines how regression is used and interpreted when data do not conform to some of its basic assumptions, such as normality or homoscedasticity of errors. The course identifies: the nature of the deviations from assumptions; the resulting estimation and interpretation problems; practical consequences; detection; and, some remedies.

**Course Goals**

By the end of this course, students will be able to:

1. understand the nature of the basic linear regression model and its statistical underpinnings
2. and basic assumptions;
3. formulate research questions that require the use of linear regression;
4. set up, solve and evaluate a regression model;
5. interpret results obtained using a statistical package, as well as critique results reported in
6. professional journals by other researchers.

 **Required reading:**

**Textbook (required):**

Greene, William H. *Econometric Analysis*. New York, NY: Prentice Hall, 1999. ISBN: 9780130132970.

Available for download at:

<http://stat.smmu.edu.cn/DOWNLOAD/ebook/econometric.pdf>

Alternative download link:

<https://drive.google.com/open?id=1WgajaX408szbNpjsVP4cKKYCs2C7d8uf>

Gujarati, D. *Essentials of Econometrics*, New York: McGraw Hill Irwin, 2006 (3d edition).

Available for download at:

<http://homepages.wmich.edu/~corder/Table%20of%20Contents.pdf>

Alternative download link:

<https://drive.google.com/open?id=1T9kQIwC1SkFICbT-U0JdiWwNC3hAQ7lu>

Balnaves, M. & Caputi, P. *Introduction to Quantitative Research Methods: An Investigative Approach*. Sage Publications, 2001. ISBN: 0761968040.

Available for download at:

<http://www.sagepub.com/upm-data/36869_muijs.pdf>

Alternative download link:

<https://drive.google.com/open?id=1scV8j8Re3-wdyyHJQeASt9N4EIv1qnXy>

**Software and Tutorials**

You can use any statistical software program to complete the coursework. **R** is an open-source computing language that is very widely used in statistics. You can download it for free from:

<http://www.r-project.org> The web provides many great tutorials and resources to learn.

A list of these is provided at:

<http://wiki.math.yorku.ca/index.php/>

**Getting started**

A nice way to start you off are the two video tutorials provided by Dan Goldstein available at:

<http://www.dangoldstein.com/flash/Rtutorial1/Rtutorial1.html>

and

<http://www.dangoldstein.com/flash/Rtutorial2/Rtutorial2.html>.

Another good resource is the set of tutorials provided by DataCamp available at:

 <https://www.datacamp.com>

This runs on a wide variety of UNIX platforms, Windows and MacOS. **R** makes programming very easy, has strong graphical capabilities, and also contains canned functions for most commonly used estimators.

For those using **R** for the first time, you are expected to work through one of the following free tutorials. All three tutorials cover similar material, just pick the one you like best: Owen. The **R** Guide. At:

<http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>

Alternative download link:

<https://drive.google.com/open?id=1HQ0sLmlnAH4prdX-Q5OqFOnZD0D7YXUT>

Venables and Smith. An Introduction to R. At:

<http://cran.r-project.org/doc/manuals/R-intro.pdf>

Alternative download link:

<https://drive.google.com/open?id=11xE_KmUAdTjNxdjtSHm1U3eIGVGnYCe->

Verzani. Simple R. At:

<http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

Alternative download link:

<https://drive.google.com/open?id=1GivFV0n-D9mE3ooFB_chSTP_82LZ5rcO>

**Optional Reading:**

Kennedy, J. *A Guide to Econometrics*. Malden, MA: Blackwell, 1998. ISBN: 0631200886.

Available for download at:

<http://facweb.knowlton.ohio-state.edu/pviton/courses/crp8703/kennedy-1-3.pdf>

Alternative download link:

<https://drive.google.com/open?id=10uI0btbtJf9c69TDuYH2kx1hpehPLxqe>

**Useful Summary Articles**

The following papers summarize the main methods learned in this course. They are dense and detailed and you might not understand all of the details the first time you read through them. However, if you plan to conduct applied empirical work that involves causal inference, you should revisit these again and again as reference.

* Guido W. Imbens and Jeffrey Wooldridge. 2008. Recent Developments in the Econometrics of Program Evaluation. NBER Working Paper No. 14251.
* Joshua D. Angrist and Alan B. Krueger. 1999. Empirical Strategies in Labor Economics. In Handbook of Labor Economics, ed. O. Ashenfelter and D. Card: Elsevier Science.

**Useful Websites**

<http://www.cas.lancs.ac.uk/glossary_v1.1/basicdef.html> Statistics Glossary for basic definitions

<http://www.statgraphics.com/Statlets> and <http://www.statcrunch.com/Statcrunch>

<http://www.cne.gmu.edu/modules/dau/math/dau1_frm.html> for DAU Math refresher

**Academic Integrity**

All work submitted must be your own or must be properly documented. Collaborative work and work previously submitted in another class must be approved in advance.

**Assignment Descriptions**

The course is practice intensive and operates under the philosophy ‘learning by doing’. With this in mind, students will be required to work through the following questions.

**Please complete the following exam. Write your answers below.**

**Question 1 (3 marks)**

At a hospital nursing station the following information is available about a patient.

* Temperature: 30.2
* Blood Type: A
* Response to treatment: Excellent

Indicate the level of measurement (nominal, ordinal or interval) of each variable.

**Question 2 (2 marks)**

At a used car lot the following information is obtained about one of the cars on the lot.

* Model year: 2003
* Gas consumption (per 100 Km): 9.8 liters

Indicate if the data is discrete or continuous for each variable.

**Question 3 (4 marks)**

Indicate what research strategy is used in each of the following situations.

1. To determine the effect of a new fertilizer on productivity of tomato plants one group of plants is treated with the new fertilizer while a second group is grown without such treatment. The number of ripe tomatoes produced by each group is counted.
2. A sociologist joins a group of homeless people to study their way of life.
3. A linguist studies the content of different Web sites and rates their level of English.
4. A company sends a satisfaction questionnaire to its current customers at the end of the year.

**Question 4 (3 marks)**

Consider the following cross-tabulation:

|  |  |
| --- | --- |
| **Residence**  | **Type of Health Worker**  |
| Doctors  | Nurses  |
| Rural  | 22  | 69  |
| Urban  | 51  | 113  |

**Table 4.1: Number of health workers by type of residence**

1. a)  According to these data, can we say that Nurses have a stronger tendency than Doctors to establish themselves in urban area?
2. b)  In the context of the previous question, which is the independent variable and which is the dependent variable?
3. c)  What percent of the sample live in urban area?

**Question 5 (2 marks)**

Consider the following data: 17, 22, 41, 12, 60, 19, 26, 74, 22, 30, 51
Find (a) the mean, (b) the sample standard deviation, (c) the population standard deviation, and (d) the percentile rank of 45.

**Question 6 (2 marks)**

|  |  |
| --- | --- |
| **Level of Schooling** | **Total**  |
| Total population aged 20-34 | 3140 |
| % of the population aged 20-34 with a BA/BS | 6.1 |
| Total population aged 35-44 | 2865 |
| % of the population aged 35-44 with a BA/BS | 9.4 |
| Total population aged 45-64 | 4730 |
| % of the population aged 45-64 with a BA/BS | 8.4 |

Using the above data, determine what percentage of the population aged **20-64** has a trade certificate. Show all your calculations.

**Question 7 (6 marks)**

Consider the following data:

***Class Interval*** f

40-46 4

33-39 8

26-32 12

19-25 9

12-18 7

**Total:** 40

1. a)  Determine the **cumulative percentage** of the class interval 26-32.
2. b)  Determine the **midpoint** of the class interval 26-32.
3. c)  Determine the **proportion** of the class interval 26-32.
4. d)  Determine the **percentile rank** of the value 39.

**Question 8 (4 marks)**

Consider the following data:

***Class Interval*** f

49-56 15

41-48 9

33-40 12

25-32 4

**Total:** 40

1. a)  Draw the corresponding **frequency polygon**.
2. b)  Draw the corresponding **histogram**.

**Question 9 (4 marks)** Consider the following data:

40, 52, 21, 48, 21, 14, 31, 48, 34, 16, 28, 51, 41, 30, 18, 22, 43, 23, 19, 21

Construct a frequency table for this data by using 4 class intervals.

* Each class interval must have the same size.
* The first and the last class interval must contain at least one data.