

ACTL 2131/5101 Assignment, T2 2020

Submission deadline:

Monday, 10 August, 11 am sharp

via Turnitin (Link will be available on Moodle)

Background

You are an actuary working for a general insurance company. You were given a task to investigate unemployment insurance (UI)¹ data for a particular state in the U.S. The data can be downloaded from the United States Department of Labour available on

<https://oui.doleta.gov/unemploy/claimssum.asp>

You should download the data available on a monthly basis for a particular state UI starting from the earliest available date until end of June 2020. You have been assigned one of the 53 states in the document StateStudent.pdf uploaded to Moodle. The data you have downloaded consists of 7 variables for the assigned state. These variables are:

- Initial claims
- First payments
- Weeks claimed
- Weeks compensated
- Average weekly benefit
- Benefits paid
- Final payments

Detailed description of the variables is available on the website of the Department of Labour.

You are asked to complete the following tasks:

¹For detailed information about UI refer to <https://oui.doleta.gov/unemploy/aboutui.asp>

For each question below, you are required to describe the methodology (including commentaries on the R code where appropriate), present results from your analysis and provide a discussion on your findings. Unless specified, use the entire sample to answer the questions.

1. (a) Present summary statistics for all 7 variables, such as mean, variance, skewness, kurtosis, and other descriptive statistics you may find interesting (e.g. correlation between variables). Comment on your findings.
(b) Perform graphical analysis on the 7 variables. You may use time series graphs (with x-axis representing calendar month and y-axis representing the variable of interest) to identify any time trends, or scatter plots (with one variable shown in the x-axis and another variable in the y-axis) to help you identify relationship between two variables, or any other graphs you might find worth including in the data analysis. Comment on your findings.
2. Perform normality test for the following transformed variables $\log(\text{Weeks compensated})$ and $\log(\text{Benefits paid})$. Is log-normal distribution appropriate to describe the distribution of "Weeks compensated" and "Benefits paid"? Use an appropriate test, and support your evidence using appropriate graphs.
3. Graph histogram and empirical cumulative distribution function for $\log(\text{Weeks compensated})$ and $\log(\text{Benefits paid})$.
4. Test if $\log(\text{Weeks claimed})$ and $\log(\text{Weeks compensated})$ have equal mean, at 5% significance level. Comment on the test and the results.
5. Test the hypothesis at 5% significance level that the average value of $\log(\text{Benefits paid})$ is greater than the value corresponding to the 55% empirical quantile of this variable.
6. (a) You are interested in knowing whether there is a linear relationship between the number of Weeks compensated and the number of Weeks claimed in the time frame from January 1971 (or earliest date for which your data is available) to December 2010. Set up an appropriate univariate regression model, obtain parameter estimates and comment on your findings. Your discussion should include, but not limited to, the significance of the estimated coefficients, model fit, analysis on the residuals and other findings you might find interesting.
(b) Use the model you estimated in (a) to predict the values of the dependent variable using the independent variable for the time frame from January 2011 to June 2020. Compare your prediction with actual values. Comment on the quality of your model.
7. Assume that now you are interested in explaining Benefits paid using other variables available in your data set. Set up an appropriate multivariate regression model, obtain parameter estimates for the model and comment on your findings. Your discussion should include, but not limited to, the significance of variables, model fit, residual statistics and other findings you might find interesting.

Learning outcomes

The assignment aims at assessing the program goals “Knowledge”, “Problem solving and critical thinking”, as well as “Communication”. It is based on the application of the technical concepts introduced in the course. You are expected to demonstrate your ability to analyse a real problem, apply appropriate theories and logic to interpret the problem, and develop solutions and conclusions. The communication of those will also be assessed.

Additional instructions

You must submit the following two items:

- Your submission, which should address each of the questions above.

The maximum number of pages is five (excluding the title page and references). The first four pages must be a self-contained report and the fourth page can be used as a technical appendix. If the length exceeds 5 pages, the pages beyond page 5 will not be marked.

- Your R code (in a separate document, as one file)

Note that we must be able to assess your work without running the R code. Your R code will be run in a random number of cases to check that you have done the work, and in suspected cases of plagiarism (if any). **Students will risk failing the assignment if the code cannot be run or the output provided in the report is inconsistent with the output generated by the code.**

You should **not**

- Include programming codes in the main body of your report
- Have figures or tables that are not referred to or analysed in the main body of your report
- Include material that are not highly relevant in the main body of your report

Communication skills

Your report must be written in form of a report. To seek further help about writing skills, please consult Learning Support provided by UNSW Business School ([click here](#)) .

Assignment submission

Assignment must be submitted via the Turnitin submission box that is available on the course Moodle website. Turnitin reports on any similarities between their own cohort's assignments, and also with regard to other sources (such as all assignments submitted all around the world

via Turnitin). More information is available [here](#). Please read this page, as we will assume that you are familiar with its content.

Please note that the School of Risk and Actuarial Studies will apply the following policy on **late assignments**. A penalty of 25% of the mark the student would otherwise have obtained, for each full (or part) day of lateness. (e.g., 0 day 1 min = 25% penalty, 2 days 21 hours = 75 = % penalty).

As long as the due date is in future, you can resubmit your work and the previous version of your assignment will be replaced by the new version. You need to check your document once it is submitted. **We will not mark assignments that cannot be read on screen.**

Students are reminded of the risk that technical issues may delay or even prevent their submission (such as internet connection and/or computer breakdowns). Students should **allow enough time (at least 24 hours is recommended) between their submission and the due time**. Please note, Turnitin will still allow you to upload a late submission, but penalty will be applied.

Plagiarism awareness

Students are reminded that the work they submit must be their own. While we have no problem with students discussing assignment problems if they wish, the material students submit for assessment must be their own. In particular, this means that any code you present are from your own computer, which you yourself developed, without any reference to any other student's work.

While some small elements of code are likely to be similar, big patches of identical code (even with different variable names, layout, or comments, Turnitin picks this up) will be considered as plagiarism. The best strategy to avoid any problem is not to share bits and pieces of code with other student outside your group.

Students should make sure they understand what plagiarism is as cases of plagiarism have a very high probability of being discovered. For issues of collective work, having different persons marking the assignment does not decrease this probability. For more information on plagiarism, see [here](#).

Assessment criteria

Please see the file **Rubic**.