IE 3301-004   Fall 2017   PROJECT

Part I: Due Friday, November 17 by 11:59pm via Blackboard SafeAssignment

Part II: Due Wednesday, December 6 by 11:59pm via Blackboard SafeAssignment

Students will complete 2 parts to a project for the semester. Each submission is individual and should not be shared with other classmates. **Any form of copying and pasting from other sources and projects will be reported to the UT Arlington Office of Student Conduct**.

Aim: The overall aim of these projects is to analyze real-world data. The specific objectives are:

1. To sample two sets of data from the real-world.
2. To summarize each set of data statistically.
3. To perform statistical chi-square tests on each set of data.
4. To describe the above steps, data, and results in a report.

On the cover of <u>each</u> Project Part report, please transcribe the following statement:

> "I _____ did not give or receive any assistance on this project, and the report submitted is wholly my own."

Write your name in the blank and sign below it. You may use an electronic signature, such as Adobe Sign**.**

<u>Tasks for Part I</u>

Data Collection: Students will collect two sets of data from the real world. Set 1 will be collected from a large number of observations (at least 100) for a <u>continuous</u> random variable from a population that is <u>suspected to be</u> <u>normally</u> distributed. Examples of such data include the body weight of people, the circumferences of oranges, the extension length of rubber bands at the point at which they burst, etc. Set 2 will be the inter-arrival time of a sequence of 100 or more events. First, record the time of each (to the nearest second) of at least 100 consecutive events, such as the checkout time at a cashier in a supermarket. Then, determine the <u>interval between occurrences</u> by taking the difference between successive event times. Consequently, Set 2 will comprise of at least 99 inter-arrival times. You may use 'second' as a unit of time. You must get instructor permission before starting to collect data for each project to ensure the relevance, quality and accuracy of your data.

Descriptive Statistics: For both Sets 1 and 2, use software to do the following:
- Calculate the sample mean and sample standard deviation.
- Calculate the quartiles Q1, Q2, and Q3.
- Construct a box-and-whisker plot.
- Construct a frequency table.
- Construct a frequency histogram.

Report: The project report is to be written in clear English with complete sentences. Be sure to define all notations and include descriptions of all tables and figures in the text. To improve your writing, you should consider taking your report to the UTA Writing Center. Your report should include a cover page, the following sections, and two appendices:

I.   Data. Describe the data collection process for Sets 1 and 2 with enough detail that the reader could replicate the process. Appendices I and II should include tables of your raw data for Sets 1 and 2, respectively.

II. Descriptive Statistics: Include your descriptive statistics analysis and interpret the results of the box-and-whisker plot and the histogram. Does Set 1 appear to follow a Normal Distribution? Does Set 2 appear to follow an Exponential Distribution?

Tasks for Part II

Chi-Square Goodness-of-Fit Test: Using a Chi-Square Goodness of Fit Test with a significance level of 0.05, test the hypothesis that Set 1 is sampled from a Normal Distribution with a population mean equal to the sample mean and a population standard deviation equal to the sample standard deviation. Similarly, test the hypothesis with a significance level of 0.05 that Set 2 is sampled from an Exponential Distribution with a population mean equal to the sample mean. For each test, start with the data classes from your histogram and merge them to ensure each class has a sufficient number of observations. Then, for each data class, calculate the following:

- Numbers of observations in the data.
- Class probability.
- Class expected value.
- Chi-square component values.

Finally, for each test, calculate the chi-square value, describe the degrees of freedom, and explain your conclusion. Note that since you are using the sample means and standard deviations, you will need to adjust your degrees of freedom accordingly.

EXAMPLE SETUP

Class	Observed Frequency (o_i)	Class Probability	Expected Frequency (e_i)	χ^2 Class Component
$X \leq 2$ $2 < X \leq 7$ $7 < X \leq 12$ $X > 12$	Count observations based on your collected data.	Calculate using the assumed probability distribution.	For each class, take its probability and multiply by n.	$\dfrac{(o_i - e_i)^2}{e_i}$
Total	n	1.0	n	χ^2 statistic

Report: The project report is to be written in clear English with complete sentences. Be sure to define all notations and include descriptions of all tables and figures in the text. To improve your writing, you should consider taking your report to the UTA Writing Center. Your report should include a cover page and the following additional section:

III. Goodness-of-Fit Tests: Describe the chi-square tests with tables for the calculated values and clearly stated conclusions.