

IE 3301 - 004  
ENGINEERING PROBABILITY

---

## **REAL-WORLD DATA: ANALYSIS**

*Part I*

---

November 19, 2017

Joe Cloud  
Student ID: 1000921236  
University of Texas at Arlington

*I, Joe Cloud, did not give or receive any assistance on  
this project, and the report submitted is wholly my own.*

# Contents

Introduction . . . . .	2
Motivation . . . . .	2
Data . . . . .	2
Set One . . . . .	2
Set Two . . . . .	2
Descriptive Analysis . . . . .	3
Set One . . . . .	3
Set Two . . . . .	4
References . . . . .	5

## INTRODUCTION

The aim of this project, as stated in the project brief is to analyze real-world data using techniques covered in the course. For this portion of the project, this includes data collection of two independent data sets: one is a sample of a random variable that is suspected to be normally distributed, while the second is a measure of the time interval between events; then summarizing each dataset statistically.

For my project, I chose to collect a set of resistor values from a batch of  $100\ 1\text{k}\Omega \pm 5\%$  resistors, for my first set. For the second set, I compiled login logs for users from a compute cluster and ran preprocessing scripts. The result was a dataset of the login time (in seconds) interval between user logins.

## MOTIVATION

### DATA

#### Set One

Data collection began with the process of measuring the resistance of each resistor and recording the value. The measurements were conducted with a calibrated 4.5 digit digital multimeter (see appendix 4 for equipment information), each measurement was entered into a spreadsheet and later outputted as CSV file for further processing by data analysis scripts.

Before data analysis was performed,  $0.33\Omega$  was subtracted from each value, this done to account for the resistance added by the multimeter probes. Although this minor offset does not impact the distribution of measurements for the purposes of the project, it does provide us with truer results. The values are offset by approximately 0.03%.

In an ideal world, the resistors would measure identically to the  $1\text{k}\Omega$  label. In reality, the cost of highly accurate manufacturing processes is expensive, tolerances guarantee a range of possible [random] values. This dataset was created to explore the distribution of these values, within the specified, yet continuous range.

#### Set Two

The data collection for Set Two revolved around login access to a compute cluster and may be difficult to produce outside the context I outlined above. Though, it is possible for the reader to perform similar analysis with data from their own machine using their own login log. On a Unix-like operating system this can be performed with the use of the `last` command. One of the benefits of computer-triggered collection is that it is much less susceptible to timing errors as opposed to that of a 'human polling' based approach.

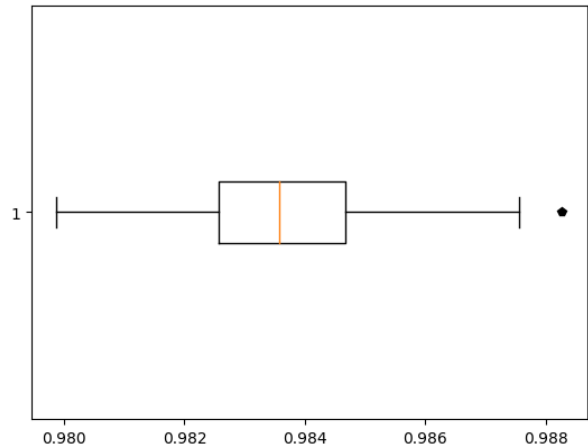
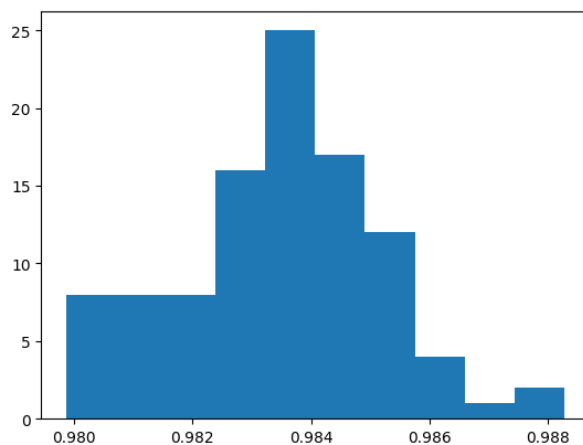
I included the source code necessary to collect the data in appendix REPLACEME. All source code written for this part of the project is included.

## DESCRIPTIVE ANALYSIS

This section details the results of analysis performed on both datasets one and two.

### Set One

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.



Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy

pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

**Set Two**

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

## REFERENCES

Weiss, S., 1989. Tissue Destruction by Neutrophils, *New England Journal of Medicine*, 320, pp. 365-376.