

## Practical 1

### Jumping Rivers

#### Simple Hypothesis Testing

Method A					
78.64	79.01	79.57	79.52	80.71	79.95
78.50	79.10	81.98	80.09	80.29	80.22

Method B					
81.92	81.12	82.47	82.86	82.89	82.45
82.51	81.11	83.07	82.77	82.38	83.14

1. We conducted an experiment and collected the data in the tables above. This data set isn't paired.<sup>1</sup>
  - a) Input the data into <sup>2</sup>. Combine the two data sets into a single data frame.
  - b) Exploratory data analysis. Construct boxplots, histograms and q-q plots for both data sets. Work out the means and standard deviations. Before carrying out any statistical test, what do you think your conclusions will be? Do you think the variances are roughly equal? Do you think the data conforms to a normal distribution.
  - c) Carry out a two sample  $t$ -test. Assume that the variances are unequal.

How does this answer compare with your intuition?

- d) Carry out a two sample  $t$ -test, assuming equal variances.
2. Suppose we are interested whether successful business executives are affected by their zodiac sign. We have collected 4265 samples and obtained the following data

```
data(zodiac, package = "jrAnalytics")
head(zodiac)
```

```
##      sign count
## 1  Aries   348
## 2  Taurus  353
## 3  Gemini  359
```

<sup>1</sup> I intentionally didn't make the data available for download so you would have to think about how to enter the data. You could enter it either Excel and import or directly into R.

<sup>2</sup> Here I would suggest input the data into Excel and using `read_csv()`

```
## 4 Cancer    357
## 5    Leo     350
## 6   Virgo    355
```

- Carry out a  $\chi^2$  goodness of fit test on the zodiac data. Are business executives distributed uniformly across zodiac signs?
- What are the expected values for each zodiac sign?
- The formula for calculating the residuals <sup>3</sup> is given by

$$\frac{\text{observed} - \text{expected}}{\sqrt{\text{expected}}}$$

<sup>3</sup> These residuals are called Pearson residuals. Hint: use `str(m)` to extract the residuals.

Which residuals are large?

### *One way ANOVA tables*

- A pilot study was developed to investigate whether music influenced exam scores. Three groups of students listened to 10 minutes of Mozart, silence or heavy metal before an IQ test. The results of the IQ test are as follows

Mozart	109	114	108	123	115	108	114
Silence	113	114	113	108	119	112	110
Heavy Metal	103	94	114	107	107	113	107

- Construct a one-way ANOVA table. Are there differences between treatment groups?
- Check the standardised residuals of your model.
- Perform a multiple comparison test to determine where the difference lies.