Chapter 9: Hypothesis Testing

Exercises

Brian Fogarty

15 May 2018

Contents

EXERCISE I **1**

ANSWERS FOR EXERCISE I **2**

EXERCISE II **2**

ANSWERS FOR EXERCISE II **2**

EXERCISE III **2**

ANSWERS FOR EXERCISE III **3**
Question 3.1 3
Question 3.2 3
Question 3.3 3
Question 3.4 4

EXERCISE IV **4**

ANSWERS FOR EXERCISE IV **4**
Question 4.1 4
Question 4.2 5
Question 4.3 5
Question 4.4 5

EXERCISE I

Write the null hypotheses for the following alternative hypotheses:

1. Individuals who are liberal are more likely to vote for the liberal party/candidate than individuals who are conservative.
2. Cities with high poverty rates are expected to have high murder rates compared to cities with low poverty rates.
3. Countries with free university are expected to have more first time university students than countries without free university.
4. US states with high sales tax are expected to have lower economic growth than US states with low sales tax.
5. Individuals with a high education level are expected to have a good health status compared to individuals with a low education level.

ANSWERS FOR EXERCISE I

1. There is no relationship between individuals' political ideology (or liberal/conservative) and vote choice.
2. There is no relationship between cities' poverty rates and murder rates.
3. There is no relationship between countries' tuition fees and the number of first time university students.
4. There is no relationship between US states' sales tax rate and economic growth.
5. There is no relationship between individuals' education level and health status.

EXERCISE II

Write a statement signifying a statistically significant relationship for each of the previous alternative hypotheses.

ANSWERS FOR EXERCISE II

1. There is a statistically significant relationship between individuals' political ideology (or liberal/conservative) and vote choice. Or, political ideology has a statistically significant effect on individuals' vote choice.
2. There is a statistically significant relationship between cities' poverty rates and murder rates. Or, poverty rates have a statistically significant effect on cities' murder rates.
3. There is a statistically significant relationship between countries' tuition fees and the number of first time university students. Or, whether a country has tuition fees or not has a statistically significant effect on the number of first time university students.
4. There is a statistically significant relationship between US states' sales tax rate and economic growth. Or, sales tax rates have a statistically significant effect on states' economic growth.
5. There is a statistically significant relationship between individuals' education level and health status. Or, education level has a statistically significant effect on individuals' health status.

EXERCISE III

Using `pcnt_unemployed` at the outcome variable and `urban` as the grouping variable from the Scottish postcode data (`depdata.csv`), carry out the following **independent samples** tests.

1. Perform a non-directional t-test. Are the groups significantly different?
2. Based on your findings from the previous question, perform an appropriate directional t-test. Is there a significant difference?
3. Perform a non-directional Wilcoxon Rank-Sum Test. How are the significance results different from the t-test using this non-parametric test?
4. Based on your findings from the previous question, perform an appropriate directional Wilcoxon Rank-Sum Test. How are the significance results different from the directional t-test using this non-parametric test?

ANSWERS FOR EXERCISE III

Question 3.1

```
setwd("C:/QSSD/Exercises/Chapter 9 - Exercises/")
getwd()
```

```
[1] "C:/QSSD/Exercises/Chapter 9 - Exercises"
```

```
depdata <- read.csv("depdata.csv")

t.test(pcnt_unemployed~urban, data=depdata)
```

```
    Welch Two Sample t-test

data:  pcnt_unemployed by urban
t = -12.317, df = 929.41, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.340914 -1.697466
sample estimates:
mean in group Rural mean in group Urban
           5.198791            7.217981
```

Since $p \leq .05$, we conclude that there is a statistically significant difference in the mean unemployment percentage between urban and rural Scottish postcodes.

Question 3.2

Since rural postcodes have a smaller mean than urban postcodes, we need to use the option `alternative="less"` in the t-test.

```
t.test(pcnt_unemployed~urban, alternative="less", data=depdata)
```

```
    Welch Two Sample t-test

data:  pcnt_unemployed by urban
t = -12.317, df = 929.41, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -1.749274
sample estimates:
mean in group Rural mean in group Urban
           5.198791            7.217981
```

Since $p \leq .05$, we conclude that rural postcodes' mean unemployment percentage is statistically significantly smaller than for urban postcodes.

Question 3.3

```
wilcox.test(pcnt_unemployed~urban, data=depdata)
```

```
    Wilcoxon rank sum test with continuity correction

data:  pcnt_unemployed by urban
W = 76650, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

We see that `p < 2.2e-16`, which tells us that there is a statistically significant difference between unemployment percentages of rural and urban postcodes. Therefore, our significance results are the same as they were with the t-test.

Question 3.4

Since we know that rural postcodes have a smaller mean than urban postcodes, we need to use the option `alternative="less"` in the Wilcoxon Rank-Sum Test.

```
wilcox.test(pcnt_unemployed~urban, alternative="less", data=depdata)
```

```
    Wilcoxon rank sum test with continuity correction

data:  pcnt_unemployed by urban
W = 76650, p-value < 2.2e-16
alternative hypothesis: true location shift is less than 0
```

We see that `p < 2.2e-16` and thus the rural postcode distribution is significantly less than the urban postcodes. Therefore, our significance results are the same as they were with the t-test.

EXERCISE IV

Using `owned` and `rent` from the Scottish postcode data (`depdata.csv`), carry out the following **dependent samples** tests.

1. Perform a non-directional t-test. Are the groups significantly different?
2. Based on your findings from the previous question, perform an appropriate directional t-test. Is there a significant difference?
3. Perform a non-directional Wilcoxon Signed Test. How are the significance results different from the t-test using this non-parametric test?
4. Based on your findings from the previous question, perform an appropriate directional Wilcoxon Signed Test. How are the significance results different from the directional t-test using this non-parametric test?

ANSWERS FOR EXERCISE IV

Question 4.1

```
t.test(depdata$owned,depdata$rent, paired=TRUE)
```

```
    Paired t-test
```

```
data:  depdata$owned and depdata$rent
t = 19.56, df = 1011, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 506.0198 618.8715
sample estimates:
mean of the differences
               562.4457
```

Since $p \leq .05$, we conclude that there is a statistically significant difference between the means of the number of people who own and who rent.

Question 4.2

Since the differences of means value is positive, it implies that the mean of people who own is greater than the mean of people who rent.

```
t.test(depdata$owned,depdata$rent, alternative="greater", paired=TRUE)
```

```
    Paired t-test

data:  depdata$owned and depdata$rent
t = 19.56, df = 1011, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 515.1049      Inf
sample estimates:
mean of the differences
               562.4457
```

Since $p \leq .05$, we conclude that the mean of people who own is significantly greater than the mean of people who rent.

Question 4.3

```
wilcox.test(depdata$owned,depdata$rent, paired=TRUE)
```

```
    Wilcoxon signed rank test with continuity correction

data:  depdata$owned and depdata$rent
V = 431910, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Based on the p-value, we conclude that the distributions for people who own and people who rent across all Scottish postcodes are statistically significantly different. Therefore, our significance results are the same as they were with the t-test.

Question 4.4

We need to use the option `alternative="greater"`.

```r
wilcox.test(depdata$owned,depdata$rent, alternative="greater", paired=TRUE)
```

```
    Wilcoxon signed rank test with continuity correction

data:  depdata$owned and depdata$rent
V = 431910, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
```

We see that `p < 2.2e-16` and thus the owned distribution is significantly greater than the rent distribution. Therefore, our significance results are the same as they were with the t-test.