Chapter 9: Hypothesis Testing

Answers to Exercises

Brian Fogarty

Contents

Exercise 1

a. There is no relationship between individuals' residential location and belief in the cause of climate change.

b. There is no relationship between companies' workforce diversity and profits.

c. There is no relationship between cities' pollution and violent crime.

d. There is no relationship between countries' health care cost and economic growth.

e. There is no relationship between rat bounty programs and the number of rats in a city.

Exercise 2

a. There is a statistically significant relationship between individuals' residential location and belief in the cause of climate change. Or, individuals' residential location has a statistically significant effect on beliefs in the cause of climate change.

b. There is a statistically significant relationship between companies' workforce diversity and profits. Or, companies' workforce diversity has a statistically significant effect on profits.

c. There is a statistically significant relationship between cities' pollution and violent crime. Or, cities' pollution has a statistically significant effect on violent crime.

d. There is a statistically significant relationship between countries' health care cost and economic growth. Or, countries' health care cost has a statistically significant effect on economic growth. Or, whether a country has free health care or not has a statistically significant effect on economic growth.

e. There is a statistically significant relationship between rat bounty programs and the number of rats in a city. Or, whether a city has a rat bounty program or not has a statistically significant effect on the number of rats.

Exercise 3

Exercise 3.a

Before running the t-test, we'll label the categories of `urban`.

```
library(tidyverse)
simd <- read_csv("simd2020.csv", na = "*")

simd <- simd %>%
          mutate(urban_fct = recode(urban, `1` = "Urban", `0` = "Rural"))

t.test(crime_rate ~ urban_fct, data = simd)
```

```
    Welch Two Sample t-test

data:  crime_rate by urban_fct
t = -19.321, df = 6430.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Rural and group Urban is not equal to 0
95 percent confidence interval:
 -171.3931 -139.8180
sample estimates:
mean in group Rural mean in group Urban
           200.0550            355.6605
```

Since $p \leq 0.05$, we conclude that there is a statistically significant difference in the mean crime rate between urban and rural Scottish datazones.

Exercise 3.b

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

```
t.test(crime_rate ~ urban_fct, alternative = "less", data = simd)
```

```
    Welch Two Sample t-test

data:  crime_rate by urban_fct
t = -19.321, df = 6430.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Rural and group Urban is less than 0
95 percent confidence interval:
      -Inf -142.3568
sample estimates:
mean in group Rural mean in group Urban
           200.0550            355.6605
```

Since $p \leq 0.05$, we conclude that rural datazones' mean crime rate is statistically significantly smaller than urban datazones' mean crime rate.

Exercise 3.c

```
wilcox.test(crime_rate ~ urban_fct, data = simd)
```

```
    Wilcoxon rank sum test with continuity correction

data:  crime_rate by urban_fct
W = 2904970, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Since $p \leq 0.05$, we conclude that there is a statistically significant difference between crime rates between urban and rural Scottish datazones. Therefore, our significance results are the same as they were with the t-test.

Exercise 3.d

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

```
wilcox.test(crime_rate ~ urban_fct, alternative = "less", data = simd)
```

```
    Wilcoxon rank sum test with continuity correction

data:  crime_rate by urban_fct
W = 2904970, p-value < 2.2e-16
alternative hypothesis: true location shift is less than 0
```

We see that $p \leq 0.05$ and thus the distribution for rural datazones' crime rate is statistically significantly smaller than for urban datazones. Therefore, our significance results are the same as they were with the t-test.

Exercise 4

Exercise 4.a

```
t.test(simd$SMR, simd$DRUG, paired = TRUE)
```

```
    Paired t-test

data:  simd$SMR and simd$DRUG
t = 1.0418, df = 6973, p-value = 0.2975
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.359277  4.442840
sample estimates:
mean of the differences
               1.541782
```

Since $p > 0.05$, we can conclude that there is not a statistically significant difference between the means of the standardised mortality and standardised hospital stays from drug use.

Exercise 4.b

We see that the mean of standardised mortality is greater than the mean of standardised hospital stays from drug use. Thus, we need to use the option `alternative = "greater"`.

```
t.test(simd$SMR, simd$DRUG, alternative = "greater", paired = TRUE)
```

```
    Paired t-test

data:  simd$SMR and simd$DRUG
t = 1.0418, df = 6973, p-value = 0.1488
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.8927644        Inf
sample estimates:
mean of the differences
               1.541782
```

Since $p > 0.05$, the mean of standardised mortality is not statistically significantly greater than the mean of standardised hospital stays from drug use.

Exercise 4.c

```
wilcox.test(simd$SMR, simd$DRUG, paired = TRUE)
```

```
    Wilcoxon signed rank test with continuity correction

data:  simd$SMR and simd$DRUG
V = 15477177, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Since $p \leq 0.05$, we conclude that the distributions for standardised mortality and standardised hospital stays from drug use across all Scottish datazones are statistically significantly different. Therefore, our significance result is different from the t-test.

Exercise 4.d

```
wilcox.test(simd$SMR, simd$DRUG, alternative = "greater", paired = TRUE)
```

```
    Wilcoxon signed rank test with continuity correction

data:  simd$SMR and simd$DRUG
V = 15477177, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
```

Since $p \leq 0.05$, the distribution for standardised mortality is statistically significantly greater than the distribution for standardised hospital stays from drug use. Therefore, our significance result is different from the t-test.