STAT 3202: Practice 07

Spring 2019, OSU

Exercise 1

Suppose that a researcher is interested in the effect of caffeine on typing speed. A group of nine individuals are administered a typing test. The following day, they repeat the typing test, this time after taking 400 mg of caffeine. (Note: This is not recommended.) The data gathered, measured in words per minute, is

```
decaf = c(98,  124, 107, 105, 80, 43, 73, 68, 69)
caff  = c(104, 128, 110, 108, 86, 53, 72, 73, 72)
```

```
##   decaf caff
## 1    98  104
## 2   124  128
## 3   107  110
## 4   105  108
## 5    80   86
## 6    43   53
## 7    73   72
## 8    68   73
## 9    69   72
```

Note that these are paired observations.

Use the **sign test** with a significance level of 0.05 to assess whether or not caffeine has an effect on typing speed. That is, test

$$H_0\colon\ m_D = m_C - m_N = 0 \quad \text{vs} \quad H_A\colon\ m_D = m_C - m_N \neq 0$$

where

- m_C is the median typing speed in words per minute of individuals using caffeine
- m_N is the median typing speed in words per minute of individuals not using caffeine

Since it is possible that the caffeine makes typing speed worse, use a two-sided test. (Also note that this is a silly experience, we aren't considering typing accuracy!)

Report:

- The value of the **test statistic** for the observed data.
- The **distribution** of the test statistic under the null hypothesis.
- The **p-value** of the test.
- A **decision** when $\alpha = 0.05$.
- A **conclusion** in words.

Exercise 2

Does meditation have an effect on blood pressure. A group of six college aged individuals were given a routine physical examination including a measurement of their systolic blood pressure. (Measured in millimeters of

mercury.) A week after their physicals, the same six individuals returned for a guided meditation session. Immediately afterwords there (systolic) blood pressure was measured. The data gathered is

```
physical    = c(125, 108, 185, 135, 112, 133)
meditation  = c(120, 114, 160, 131, 124, 125)
```

```
##   physical meditation
## 1      125        120
## 2      108        114
## 3      185        160
## 4      135        131
## 5      112        124
## 6      133        125
```

Note that these are paired observations.

Use the **sign test** with a significance level of 0.10 to assess whether or not meditation has an effect on blood pressure. That is, test

$$H_0\colon\ m_D = m_M - m_P = 0 \quad \text{vs} \quad H_A\colon\ m_D = m_M - m_P \neq 0$$

where

- m_P is the median systolic blood pressure in millimeters of mercury measured without meditation
- m_M is the median systolic blood pressure in millimeters of mercury measured with meditation

Since it is possible that the meditation makes blood pressure worse, use a two-sided test.

Report:

- The value of the **test statistic** for the observed data.
- The **distribution** of the test statistic under the null hypothesis.
- The **p-value** of the test.
- A **decision** when $\alpha = 0.10$.
- A **conclusion** in words.

Exercise 3

Return to the sleep data in Exercise 2. This time test

- H_0: The distribution of systolic blood pressure is **the same** with and without meditation
- H_A: The distribution of systolic blood pressure is **different** with and without meditation

To do so, use a **permutation test** that permutes the *statistic*

$$\bar{x}_D$$

where $\bar{x}_D$ is the sample mean difference. Assume that the distribution of blood pressure with and without meditation has the same shape, but may have different locations. Use at least 10000 permutations.

```
physical    = c(125, 108, 185, 135, 112, 133)
meditation  = c(120, 114, 160, 131, 124, 125)
```

- Create a histogram that illustrates the distribution of the statistic used.
- Report the p-value of the test.

Example 4

Which profession pays more? Data Scientist or Actuary? A (far too small) survey of junior (less than three years experience) data scientists and actuaries resulted in the following data:

```
data_sci = c(88000, 121000, 91000, 50000, 78000, 95000)
actuary = c(63000, 75000, 81000, 75000, 85000)
```

Use a **permutation test** that permutes the *statistic*

$$t = \frac{(\bar{x} - \bar{y}) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

to test

- H_0: The distribution of salaries is **the same** for junior data scientists and actuaries
- H_A: The distribution of salaries is **different** for junior data scientists and actuaries

Assume that the distribution of salaries for both has the same shape, but may have different locations. Use at least 10000 permutations.

- Create a histogram that illustrates the distribution of the statistic used.
- Report the p-value of the test.

Exercise 5

Repeat exercise 3, but use an appropriate test available in the `R` function `wilcox.test()`.

Report:

- The **p-value** of the test
- A **decision** when $\alpha = 0.05$.
