06-Classical Tests

2024-11-26

Preface

The following exercises demonstrate some of the most common classical tests by means of simple examples.

1 A sleep duration study: statistical tests of location

The example is inspired by a classical test data set (Student, 1908) about a study with two groups of persons treated with two different pharmaceutical drugs.

Drug 1: 8.7, 6.4, 7.8, 6.8, 7.9, 11.4, 11.7, 8.8, 8, 10

Drug 2: 9.9, 8.8, 9.1, 8.1, 7.9, 12.4, 13.5, 9.6, 12.6, 11.4

The data are the duration of sleeping time in hours. It is assumed that the normal sleeping time would be **8 hours**.

1.1 One sample t-Test

Let's test whether the drugs increased or decreased sleeping time, compared to 8 hours:

```
x <- c(8.7, 6.4, 7.8, 6.8, 7.9, 11.4, 11.7, 8.8, 8, 10)
t.test(x, mu = 8)
```

Exercise 1: Test the effect of the second drug. Does it change sleeping duration?

1.2 Two sample t-Test

The two sample t-Test is used to compare two groups of data: Related to our example, we test the following hypotheses:

H_0: Both drugs have the same effect.

H_A: The drugs have a different effect, i.e. one of the drugs is stronger.

```
x1 <- c(8.7, 6.4, 7.8, 6.8, 7.9, 11.4, 11.7, 8.8, 8, 10)
x2 <- c(9.9, 8.8, 9.1, 8.1, 7.9, 12.4, 13.5, 9.6, 12.6, 11.4)
t.test(x1, x2)  # Welch-t-test
```

Here, R performs the Welch test by default, that is also valid for samples with different variances.

The classical approach suggested to check homogeneity of variances with the F-Test (`var.test`) first and if the assumption holds, to apply the "ordinary" two sample t-test (`t.test(....., var.equal = TRUE)`). This method is not anymore recommended (Delacre et al., 2017).

```
var.test(x1, x2)                    # F-test as pre-test
t.test(x1, x2, var.equal = TRUE)    # classical t-test
```

Exercise 2: Create a boxplot, and perform the tests. What is the effect size, i.e. by how many hours differs sleep duration?

1.3 Paired t-test

Given is a number of students that passed an examination in statistics. The examination was written two times, one time before and one time after an additional series of lectures. The values represent the numbers of points approached during the examination. Check whether the additional lectures had any positive effect:

```
x1 <- c(69,  77, 35, 34, 87, 45, 95, 83)
x2 <- c(100, 97, 67, 42, 75, 73, 92, 97)
```

Exercise 3: The test was conducted by the same individuals before and after the course, so one can use a paired t-test:

```
t.test(x1, x2, paired = TRUE)
```

Then compare the results with the ordinary two-sample t-test.

1.4 Wilcoxon test (optional)

The Mann-Whitney and Wilxon tests are nonparametric tests of location. "Nonparametric" means, that the general location of the distributions is compared and not a parameter like the mean. This makes the test independent of distributional assumptions, but can sometimes lead to a vague interpretations.

Exercise 4: Now repeat the comparison for the sleep study using the Wilcoxon test for unpaired and paired samples. Note that the unpaired test is often also called "Mann-Whitney U test".

In R both tests can be found as `wilcox.test`. Use the help system of R (`?wilcox.test`) and read the help page about the usage of these tests.

2 Own project: Weight of Clementine fruits

Import the Clementines data set (fruits-2023-hse.csv)[1]. Think about an appropriate data structure and use a suitable statistical test to compare the weights. Check variance homogeneity and normal distribution graphically. Can the weights from each brand be considered as independent samples?

[1] `fruits.csv` available from: https://tpetzoldt.github.io/datasets/data/fruits-2023-hse.csv

2.1 Perform a statistical test for the following hypotheses

H_0: The weight of the fruits bought on Friday (box2) and on Monday (box1) are the same.

H_A: The weight of the fruits is different.

Select a proper statistical test and interpret its results.

2.2 Calculate absolute and relative effect sizes

1. Calculate the mean values of both samples $\bar{x}_1, \bar{x}_2$ and the **absolute effect size**:

$$\Delta = \bar{x}_1 - \bar{x}_2$$

.

2. Calculate the pooled standard deviation (N_1, N_2 = sample size, s_1, s_2= standard deviation):

$$s_{1,2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2)}{N_1 + N_2 - 2}}$$

4. Calculate the relative effect size as

$$\delta = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{1,2}}$$

.

5. Read in Wikipedia about Cohen's d and other measures of effect size.

3 Chi-squared test and Fisher's exact test (optional)

3.1 Introduction

Taken from Agresti (2002), Fisher's Tea Drinker:

"A British woman claimed to be able to distinguish whether milk or tea was added to the cup first. To test, she was given 8 cups of tea, in four of which milk was added first. The null hypothesis is that there is no association between the true order of pouring and the woman's guess, the alternative that there is a positive association (that the odds ratio is greater than 1)."

The experiment revealed the following outcome: With tea first, the tea taster identified three times the correct answer and was one time wrong and the same occurred with milk first (3 true, 1 wrong).

For a TV show this would be sufficient, but how big was the probability to get such a result just by chance?

3.2 Methods and Results

We put the data into a matrix:

```
x <- matrix(c(3, 1, 1, 3), nrow = 2)
x
fisher.test(x)
```

This tests for an association between truth and guess, but if we want only positive associations, we should perform a one-sided test:

```
fisher.test(x, alternative = "greater")
```

A similar test can be performed with the chi-squared test, but this is not precise for small data sets, so the Fisher test should be preferred.

3.3 Discussion

Exercise 5: Compare the results of Fisher's exact test with `alternative="two.sided"` (the default) with `alternative = "greater"` and `less` and discuss the differences and their meaning. Which option is the best in this case?

Exercise 6: How many trials would be necessary to get a significant statistical result with $p < 0.05$ for the tea taster experiment, given that we allow one wrong decision for tea first and milk first?

3.4 Background

Read the Wikipedia articles "Lady tasting tea" about Fisher's experiment and Fisher's exact test and "The Lady Tasting Tea" about a popular science book on the "statistical revolution" in the 20th century.

The odds ratio describes the strength of association in a two-by-two table, see explanation of "Odds ratio" in Wikipedia or a statistics text book.

References

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology*, *30*(1), 92–101. https://doi.org/10.5334/irsp.82

Student. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1–25. https://doi.org/10.2307/2331554