

A tutorial on comparing two dependent samples in R

Version 1.0.0

Konstantinos I. Bougioukas

04/11/2021

Contents

Objectives	3
Packages for these notes:	3
1 Introduction	4
2 Paired t-test	4
2.1 Research question	4
2.2 H0 and H1 Hypotheses	4
2.3 Prerparation of data	5
2.4 Assumptions	6
2.5 Explore the characteristics of distribution of differences	6
2.5.1 Visualization of the distribution of differences	6
2.5.2 Summary statistics	7
2.5.3 Shapiro-Wilk test for normality	10
2.6 Run the paired t-test	11
2.7 Present the results in a summary table	12

3	Wilcoxon Signed-Rank test	13
3.1	Research question	13
3.2	H0 and H1 Hypotheses	13
3.3	Prerparation of data	14
3.4	Explore the characteristics of distribution of differences	15
3.4.1	Visualization of the distribution of differences	15
3.4.2	Summary statistics	16
3.4.3	Shapiro-Wilk test for normality	19
3.5	Run the Wilcoxon Signed-Rank test	20
3.6	Present the results in a summary table	21

Objectives

- Applying hypothesis testing
- Compare two dependent samples
- Interpret the results

Packages for these notes:

We need to download the `rstatix`, `dlookr` and `EnvStats` packages for the notes.

Next, we need to load the following packages:

```
library(rstatix)      # Pipe-Friendly Framework for Basic Statistical Tests
library(here)         # A Simpler Way to Find Your Files
library(tidyverse)    # Easily Load the 'Tidyverse'
```

1 Introduction

Examples of paired study designs are:

1. measurements collected **before and after** an intervention in an experimental study
2. twins, husbands and wives, brothers and sisters, pairs of eyes
3. **matched** cases and controls
4. a **cross-over trial** in which each patient has two measurements on the variable, one while taking active treatment and one while taking placebo

2 Paired t-test

A paired t-test is used to estimate whether the means of two related measurements are significantly different from one another.

2.1 Research question

The dataset `weight` contains the birth and discharge weight of 25 newborns. We might ask if the **mean difference** of the weight in birth and in discharge equals to zero or not. If the **differences** between the pairs of measurements are normally distributed, a paired t-test is the most appropriate statistical test.

2.2 H0 and H1 Hypotheses

- H_0 : the mean difference of weight equals to zero ($\mu_d = 0$)
- H_1 : the mean difference of weight does not equal to zero ($\mu_d \neq 0$)

2.3 Prerparation of data

We import the data:

```
library(readxl)
weight <- read_excel(here("data", "weight.xlsx"), col_names=TRUE)
weight
```

Table 1: Weight Data (first and last 5 rows)

birth_weight	discharge_weight
3250	3220
2680	2640
2960	2940
3420	3350
3210	3140
...	...
3150	3115
2687	2644
2960	2880
3422	3393
3210	3240

We inspect the data:

```
glimpse(weight)
```

```
## Rows: 25
```

```
## Columns: 2
```

```
## $ birth_weight    <dbl> 3250, 2680, 2960, 3420, 3210, 2740, 3250, 3170, ~
```

```
## $ discharge_weight <dbl> 3220, 2640, 2940, 3350, 3140, 2730, 3220, 3150, ~
```

2.4 Assumptions

The differences are **normally** distributed.

We calculate the differences using the function `mutate()` :

```
weight <- weight %>%
  mutate(dif_weight = birth_weight - discharge_weight)
glimpse(weight)

## Rows: 25
## Columns: 3
## $ birth_weight    <dbl> 3250, 2680, 2960, 3420, 3210, 2740, 3250, 3170, ~
## $ discharge_weight <dbl> 3220, 2640, 2940, 3350, 3140, 2730, 3220, 3150, ~
## $ dif_weight      <dbl> 30, 40, 20, 70, 70, 10, 30, 20, 80, 103, 84, 42, ~
```

2.5 Explore the characteristics of distribution of differences

The distributions of the differences can be explored with appropriate plots and summary statistics.

2.5.1 Visualization of the distribution of differences

We can explore the data visually for symmetry with a density plot (a smoothed version of the histogram):

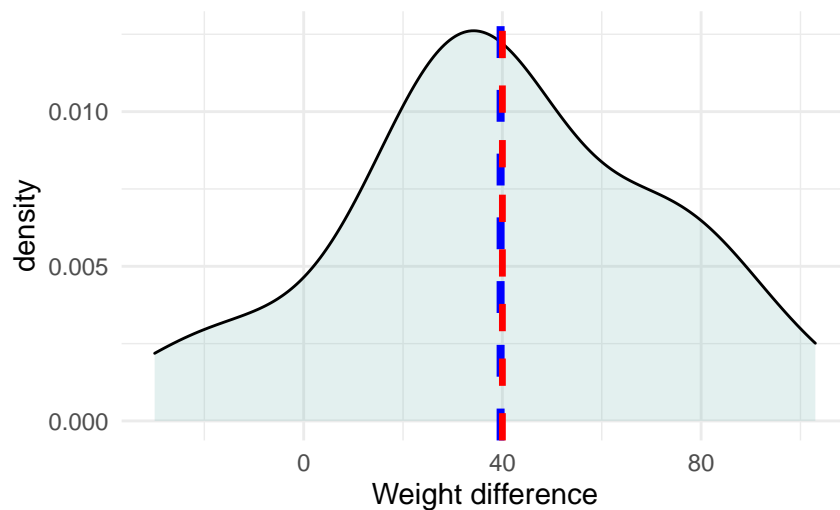
```
weight %>%
  ggplot(aes(x = dif_weight)) +
  geom_density(fill = "#76B7B2", color="black", alpha = 0.2) +
  geom_vline(aes(xintercept=mean(dif_weight))),
```

```

    color="blue", linetype="dashed", size=1.4) +
  geom_vline(aes(xintercept=median(dif_weight)),
    color="red", linetype="dashed", size=1.2) +
  labs(x = "Weight difference",
    title = "Density plot of the weight differences (n = 25)") +
  theme_minimal() +
  theme(plot.title.position = "plot")

```

Density plot of the weight differences (n = 25)



The above figure shows that the data are following an approximately symmetrical distribution. Note that the arithmetic mean (blue vertical dashed line) is very close to the median (red vertical dashed line) of the data.

2.5.2 Summary statistics

Summary statistics can also be inspected for the variables. We can utilize the new `across()` function to obtain the results across the three variables simultaneously:

```

summary_weight <- weight %>%
  dplyr::summarise(across(
    .cols = c(dif_weight, birth_weight, discharge_weight),
    .fns = list(
      N = ~n(),
      Min = min,
      Q1 = ~quantile(., 0.25),
      median = median,
      Q3 = ~quantile(., 0.75),
      Max = max,
      Mean = mean,
      Sd = sd,
      Skewness = EnvStats::skewness,
      Kurtosis= EnvStats::kurtosis
    ), na.rm = TRUE,
    .names = "{col}_{fn}")
  )

# present the results
summary_weight <- summary_weight %>%
  mutate(across(everything(), round, 2)) %>% # round to 3 decimal places
  pivot_longer(1:30, names_to = "Stats", values_to = "Values") # long format

```


Stats	Values
dif_weight_N	25.00
dif_weight_Min	-30.00
dif_weight_Q1	20.00
dif_weight_median	40.00
dif_weight_Q3	70.00
dif_weight_Max	103.00
dif_weight_Mean	39.64
dif_weight_Sd	32.27
dif_weight_Skewness	-0.16
dif_weight_Kurtosis	-0.08
birth_weight_N	25.00
birth_weight_Min	2680.00
birth_weight_Q1	2960.00
birth_weight_median	3150.00
birth_weight_Q3	3210.00
birth_weight_Max	3440.00
birth_weight_Mean	3075.72
birth_weight_Sd	248.00
birth_weight_Skewness	-0.29
birth_weight_Kurtosis	-0.93
discharge_weight_N	25.00
discharge_weight_Min	2640.00
discharge_weight_Q1	2880.00
discharge_weight_median	3100.00
discharge_weight_Q3	3220.00
discharge_weight_Max	3400.00
discharge_weight_Mean	3036.08
discharge_weight_Sd	248.06
discharge_weight_Skewness	-0.22
discharge_weight_Kurtosis	-1.02

As it was previously mentioned, the mean of the differences (39.64) is close to median (40). Moreover, both the skewness and the kurtosis are approximately zero indicating a symmetric and mesokurtic distribution for the weight differences.

Alternatively, we can use the `describe()` from `{dlookr}` package:

```
weight %>%
  dlookr::describe(dif_weight, birth_weight, discharge_weight) %>%
  select(variable, n, mean, sd, p25, p50, p75, skewness, kurtosis)
```

```
## # A tibble: 3 x 9
```

	variable	n	mean	sd	p25	p50	p75	skewness	kurtosis
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	dif_weight	25	39.6	32.3	20	40	70	-0.157	-0.0752
## 2	birth_weight	25	3076.	248.	2960	3150	3210	-0.291	-0.935
## 3	discharge_weight	25	3036.	248.	2880	3100	3220	-0.219	-1.02

2.5.3 Shapiro-Wilk test for normality

Additionally, we can check the statistical test for normality of the differences.

```
weight %>%
  shapiro_test(dif_weight)
```

variable	statistic	p
dif_weight	0.974	0.742

The Shapiro-Wilk test suggests that the weight differences are normally distributed ($p=0.74 > 0.05$).

2.6 Run the paired t-test

We will perform a paired t-test to test the null hypothesis that the mean differences of weight equals to zero.

Our data are in a wide format. However, we are going to use only the differences of the weight, the `dif_weight` variable, inside the `t_test()`:

```
weight %>%  
  t_test(dif_weight ~ 1, detailed = T)
```

estimate	.y.	group1	group2	n	statistic
39.64	dif_weight	1	null model	25	6.142752

p	df	conf.low	conf.high	method	alternative
<0.001	24	26.32139	52.95861	T-test	two.sided

We can also use the `t.test` form base R:

```
t.test(weight$dif_weight)
```

or we can use the initial variables, `birth_weight` and `discharge_weight` :

```
t.test(weight$birth_weight, weight$discharge_weight, paired = TRUE)
```

Caution!

It is possible to perform a paired t-test on long format data with formula notation value group. This requires us to be sure that the data pairs are in order. More often than not, long format data is unpaired and using formula notation to perform a paired test is a mistake.

2.7 Present the results in a summary table

Characteristic	birth_weight, N = 25 ¹	discharge_weight, N = 25 ¹	p-value ²
weights (grams)	3,076 (248)	3,036 (248)	<0.001

¹Mean (SD)

²Paired t-test

There was a significant reduction in weight (39.6 g) after the discharge (p-value <0.001 that is lower than 0.05; reject H_0). Note that the 95% confidence interval (26.3 to 52.9) doesn't include the null hypothesized value of 0. However, is this reduction of clinical importance?

3 Wilcoxon Signed-Rank test

A non-parametric significance test that is carried out to compare paired continuous measurements when the assumptions for the paired t-test can not be met (non-normal distribution of the differences). It is based on the signs of the differences and the magnitude of the rank of the differences between pairs of measurements, rather than the actual values. The null hypothesis is that there is no tendency for values under each paired variable to be higher or lower.

Note Paired observations for which the difference is zero do not contribute to the calculation of the test statistic and therefore the effective sample size for the test may be smaller than the actual total sample size (in other words, the number of pairs contributing to the test may be smaller than the total number of pairs).

Note It is often thought of as a test for small samples. However, if the sample is smaller than 6, then statistical significance is impossible.

3.1 Research question

The dataset `eyes` contains thickness of the cornea (microns) in patients with one eye affected by glaucoma; the other eye is unaffected. We investigate if there is evidence for difference in corneal thickness in affected and unaffected eyes.

3.2 H0 and H1 Hypotheses

- H_0 : The distribution of the differences in thickness of the cornea is symmetrical about zero
- H_1 : The distribution of the differences in thickness of the cornea is not symmetrical about zero

Note If we are testing the null hypothesis that the **median** of the paired rank differences is zero, then the paired rank differences must all come from a symmetrical distribution. Note that we do not have to assume that the distributions of the original populations are symmetrical - two very positively skewed distributions that differ only by location will produce a set of paired rank differences that are symmetrical.

3.3 Prerparation of data

We import the data:

```
library(readxl)
eyes <- read_excel(here("data", "eyes.xlsx"), col_names=TRUE)
eyes
```

Eyes Data		
id	affected_eye	unaffected_eye
1	488	484
2	478	478
3	480	492
4	426	444
5	440	436
6	410	398
7	458	464
8	460	476

We inspect the data:

```
glimpse(eyes)
```

```
## Rows: 8
## Columns: 3
## $ id          <dbl> 1, 2, 3, 4, 5, 6, 7, 8
## $ affected_eye <dbl> 488, 478, 480, 426, 440, 410, 458, 460
## $ unaffected_eye <dbl> 484, 478, 492, 444, 436, 398, 464, 476
```

We calculate the differences using the function `mutate()`:

```
eyes <- eyes %>%
  mutate(dif_thickness = affected_eye - unaffected_eye)
glimpse(eyes)
```

```
## Rows: 8
## Columns: 4
## $ id          <dbl> 1, 2, 3, 4, 5, 6, 7, 8
## $ affected_eye <dbl> 488, 478, 480, 426, 440, 410, 458, 460
## $ unaffected_eye <dbl> 484, 478, 492, 444, 436, 398, 464, 476
## $ dif_thickness <dbl> 4, 0, -12, -18, 4, 12, -6, -16
```

3.4 Explore the characteristics of distribution of differences

The distributions of the differences can be explored with appropriate plots and summary statistics.

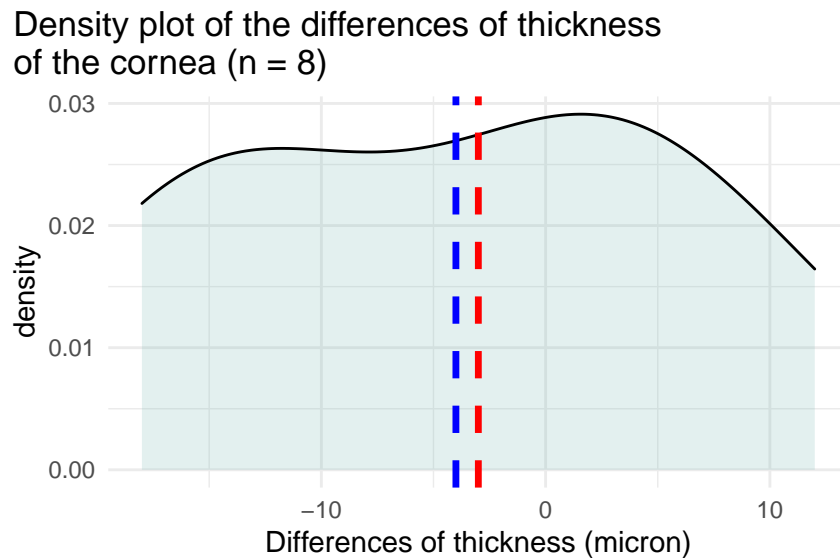
3.4.1 Visualization of the distribution of differences

We can explore the data visually for symmetry with a histogram.

```

eyes %>%
  ggplot(aes(x = dif_thickness)) +
  geom_density(fill = "#76B7B2", color="black", alpha = 0.2) +
  geom_vline(aes(xintercept=mean(dif_thickness)),
             color="blue", linetype="dashed", size=1.2) +
  geom_vline(aes(xintercept=median(dif_thickness)),
             color="red", linetype="dashed", size=1.2) +
  labs(x = "Differences of thickness (micron)",
       title = "Density plot of the differences of thickness \nof the cornea (n = 8)") +
  theme_minimal() +
  theme(plot.title.position = "plot")

```



3.4.2 Summary statistics

Summary statistics can also be inspected for the differences and for the original data using the `across()` function:


```

summary_eyes <- eyes %>%
  dplyr::summarise(across(
    .cols = c(dif_thickness, affected_eye, unaffected_eye),
    .fns = list(
      N = ~n(),
      Min = min,
      Q1 = ~quantile(., 0.25),
      median = median,
      Q3 = ~quantile(., 0.75),
      Max = max,
      Mean = mean,
      Sd = sd,
      Skewness = EnvStats::skewness,
      Kurtosis= EnvStats::kurtosis
    ), na.rm = TRUE,
    .names = "{col}_{fn}")
  )

# present the results
summary_eyes <- summary_eyes %>%
  mutate(across(everything(), round, 2)) %>% # round to 3 decimal places
  pivot_longer(1:30, names_to = "Stats", values_to = "Values") # long format

```

Stats	Values
dif_thickness_N	8.00
dif_thickness_Min	-18.00
dif_thickness_Q1	-13.00
dif_thickness_median	-3.00
dif_thickness_Q3	4.00
dif_thickness_Max	12.00
dif_thickness_Mean	-4.00
dif_thickness_Sd	10.74
dif_thickness_Skewness	0.03
dif_thickness_Kurtosis	-1.37
affected_eye_N	8.00
affected_eye_Min	410.00
affected_eye_Q1	436.50
affected_eye_median	459.00
affected_eye_Q3	478.50
affected_eye_Max	488.00
affected_eye_Mean	455.00
affected_eye_Sd	27.69
affected_eye_Skewness	-0.49
affected_eye_Kurtosis	-0.98
unaffected_eye_N	8.00
unaffected_eye_Min	398.00
unaffected_eye_Q1	442.00
unaffected_eye_median	470.00
unaffected_eye_Q3	479.50
unaffected_eye_Max	492.00
unaffected_eye_Mean	459.00
unaffected_eye_Sd	31.31
unaffected_eye_Skewness	-1.11
unaffected_eye_Kurtosis	0.79

The differences seems to come from a population with a symmetrical distribution and the skewness is close to zero (0.03). However, the (excess) kurtosis equals to -1.37 (platykurtic) and the sample size is small. Therefore, the data may not follow the normal distribution.

Alternatively, we can use the `describe()` from `{dlookr}` package:

```
eyes %>%
  dlookr::describe(dif_thickness, affected_eye, unaffected_eye) %>%
  select(variable, n, mean, sd, p25, p50, p75, skewness, kurtosis)
```

```
## # A tibble: 3 x 9
```

	variable	n	mean	sd	p25	p50	p75	skewness	kurtosis
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	dif_thickness	8	-4	10.7	-13	-3	4	0.0295	-1.37
## 2	affected_eye	8	455	27.7	436.	459	478.	-0.493	-0.985
## 3	unaffected_eye	8	459	31.3	442	470	480.	-1.11	0.794

3.4.3 Shapiro-Wilk test for normality

We can apply Shapiro-Wilk test to check for normality. However, here, normality test is not helpful because of the small sample (the test is underpowered).

```
eyes %>%
  shapiro_test(dif_thickness)
```

variable	statistic	p
dif_thickness	0.944	0.651

3.5 Run the Wilcoxon Signed-Rank test

The differences between the two measurements can be tested using a rank test such as Wilcoxon Signed-Rank test.

```
eyes %>%  
  wilcox_test(dif_thickness ~ 1)
```

.y.	group1	group2	n	statistic	p
dif_thickness	1	null model	8	7.5	0.309

We can also use the wilcox.test form base R:

```
wilcox.test(eyes$dif_thickness)
```

or we can use the initial variables, affected_eye and unaffected_eye :

```
wilcox.test(eyes$affected_eye, eyes$unaffected_eye, paired = TRUE)
```

The result is not significant ($p = 0.309 > 0.05$). However, we can't be certain that there is not difference in corneal thickness in affected and unaffected eyes because the sample size is very small.

Note Some summary statistics need to be presented alongside the p-value, to allow the result of the test to be interpreted. We usually present the median, the interquartile range, or the full range.

3.6 Present the results in a summary table

Characteristic	affected_eye, N = 8 ¹	unaffected_eye, N = 8 ¹	p-value ²
thickness (microns)	459.0 (436.5, 478.5)	470.0 (442.0, 479.5)	0.3

¹Median (IQR)

²Wilcoxon signed rank test with continuity correction

There is not evidence from this small study with patients of glaucoma that the thickness of the cornea in affected eyes, median (IQR) = 459 (436.5, 478.5) μm , differs from unaffected eyes 470 (442, 479.5) μm , ($p=0.309 >0.05$).