

Statistics for LMU Bio Master's programs  
**Wilcoxon test applied to biological example data**

Dirk Metzler

May 17, 2020

Wilcoxon's rank sum test

is also called

Mann-Whitney U test  
(up to some minor technical differences)

## References

- [1] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* **1**:8083.
- [2] Mann, H. B., Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**:5060.

## Contents

## Contents

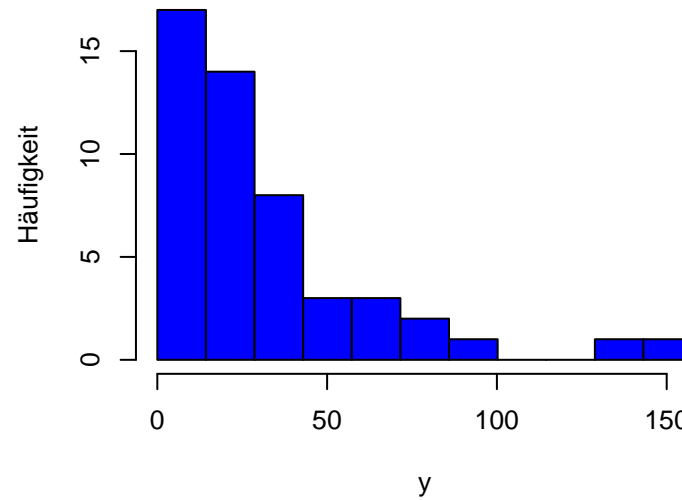
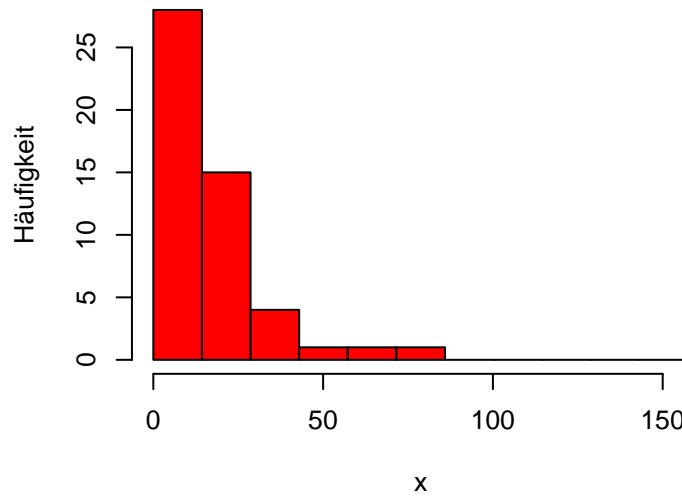
<b>1</b>	<b>Motivation</b>	<b>1</b>
<b>2</b>	<b>Application to Hipparion tooth measurements</b>	<b>2</b>
<b>3</b>	<b>Wilcoxon signed-rank test</b>	<b>3</b>

## 1 Motivation

For (more or less) bell-shaped symmetrically distributed observations  
(or for sufficiently large sample sizes)  
we can apply the  $t$ -Test for the null hypothesis  $\mu_1 = \mu_2$ :  
The  $t$ -statistic is then nearly Student-t-distributed.

This can be different especially for very asymmetric long-tailed distributions

Assume that we have to compare these distributions:



### Examples

- Waiting times
- Distances of dispersal
- Frequencies of cell types

Wanted: a test that does not assume a distribution

Such tests are called *non-parametric*

## 2 Application to Hipparion tooth measurements

Remember Welch's *t*-test

We examine one trait in two populations:

Population	1	2
mean	$\mu_1$	$\mu_2$

Null hypothesis:  $\mu_1 = \mu_2$

We draw samples from the two populations with means  $\bar{x}_1$   $\bar{x}_2$

To check the null hypothesis  $H_0$ , we compute the *t-Statistic*[1ex]

$$t = \frac{\bar{x}_1 - \bar{x}_2}{f} \quad \text{with} \quad f^2 = \left( \frac{s_1}{\sqrt{n_1}} \right)^2 + \left( \frac{s_2}{\sqrt{n_2}} \right)^2$$

*p*-value under  $H_0$ :  $p \approx \mathbb{P}(|T_g| \geq |t|)$  ( $g$ =(estimated) degrees of freedom depends on  $n_1, n_2, s_1, s_2$ )

If the distributions are not close to normal, *Wilcoxon's rank-sum Statistic* can be used.

```
> t.test(md[Art=="africanum"],md[Art=="libycum"])

Welch Two Sample t-test

data: md[Art == "africanum"] and md[Art == "libycum"]
t = -3.2043, df = 54.975, p-value = 0.002255
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.1025338 -0.9453745
sample estimates:
mean of x mean of y
 25.91026  28.43421
```

```
> wilcox.test(md[Art=="africanum"],md[Art=="libycum"])

Wilcoxon rank sum test with continuity correction

data: md[Art == "africanum"] and md[Art == "libycum"]
W = 492, p-value = 0.01104
alternative hypothesis: true location shift is not equal to 0
```

```
Warning message:
In wilcox.test.default(md[Art == "africanum"], md[Art == "libycum"]) :
  kann bei Bindungen keinen exakten p-Wert Berechnen
```

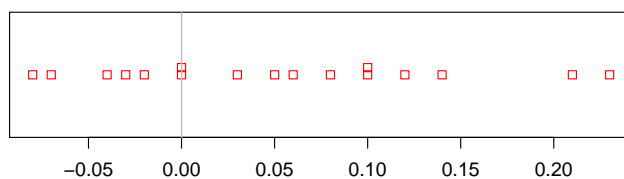
### IMPORTANT!

Is the Wilcoxon test really appropriate here? Not clear because its null hypothesis is that the data come from the same distribution, not just that the means are equal. If we want to test whether the means are different but allow the standard deviations to be different (like in the assumptions of Welch's t-test), the Wilcoxon test cannot be applied!

## 3 Wilcoxon signed-rank test

Remember the pied flycatchers?

$$x := \text{"green length"} - \text{"blue length"}$$

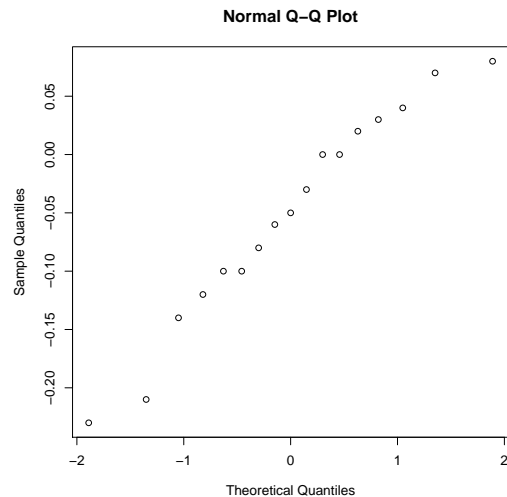


t-Test with R

```
> x <- length$green-length$blue
> t.test(x)
```

### One Sample t-test

```
data: x
t = 2.3405, df = 16, p-value = 0.03254
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.004879627 0.098649784
sample estimates:
mean of x
0.05176471
```



### Wilcoxon test for paired samples

Wilcoxon signed rank test:

given independent samples  $X_1, X_2, \dots, X_n$  from an unknown distribution.

null hypothesis: distribution of  $X_i$  is symmetric around 0.0.

Application: paired samples  $(Y_1, Z_1), \dots, (Y_n, Z_n)$ . null hypothesis:  $Y_i$  and  $Z_i$  have same distribution.

Apply Wilcoxon signed rank test to

$$X_1 = Y_1 - Z_1, X_2 = Y_2 - Z_2, \dots, X_n = Y_n - Z_n.$$

`wilcox.test(flycatchers$blue, flycatchers$green, paired=TRUE)` and  
`wilcox.test(flycatchers$blue-flycatchers$green)` both give:

Wilcoxon signed rank test with continuity correction

```
data: flycatchers$blue and flycatchers$green
V = 22.5, p-value = 0.03553
alternative hypothesis: true location shift is not equal to 0
```

Warning messages:

- 1: In `wilcox.test.default(flycatchers$blue, flycatchers$green, paired = TRUE)` :  
kann bei Bindungen keinen exakten p-Wert Berechnen
- 2: In `wilcox.test.default(flycatchers$blue, flycatchers$green, paired = TRUE)` :  
kann den exakten p-Wert bei Nullen nicht berechnen

Warning message: calculation imprecise as some of the input values are 0  
statistic: sum of signed ranks.

signed rank: take rank of absolute value, equipped with sign of original value

example:

values:	-2.6	-2.5	-2.3	-1.3	-0.6	1.6	2.2	6.1
absolute values:	2.6	2.5	2.3	1.3	0.6	1.6	2.2	6.1
ranks:	7	6	5	2	1	3	4	8
signed ranks:	-7	-6	-5	-2	-1	3	4	8

Test statistic:  $3 + 4 + 8 = 15$

### Some of what you should be able to explain

- When should you apply non-parametric tests?
- Wilcoxon test: requirements and null hypothesis.
- The statistics of the Wilcoxon test and the signed-rank test.
- How to calculate the  $p$  value for the Wilcoxon test.