

Swiss Institute of  
Bioinformatics

## Introduction to Statistics

Joao Lourenço ([joao.lourenco@sib.swiss](mailto:joao.lourenco@sib.swiss)) and Rachel Marcone ([rachel.marcone@sib.swiss](mailto:rachel.marcone@sib.swiss))

January 2024

# Parametric and non parametric tests

## *T-tests: summary*

T-test in general  
Used to compare means

One-sample t-test  
Compare the mean of a sample to a given number

Two-sample t-test  
Compare the means of two samples

Paired t-test  
Compare the difference between pairs of related data points

## *Assumptions for the t-test*

- This test is one of the most widely used tests. However, it can be used only if the background assumptions are satisfied.
  - Data values must be **independent**.
  - Data in each group must be obtained via a **random sample** from the population.
  - The variances for the two independent groups are **equal**.
  - Data in each group are **normally** distributed.
  - Data values are **continuous**.

## *T-test in practice*

- Analysts generally don't use the standard (Student) t-test.
- The Welch t-test takes into account different sample variances.
- If the variances are actually the same, it provides the same results as the standard Student t-test.

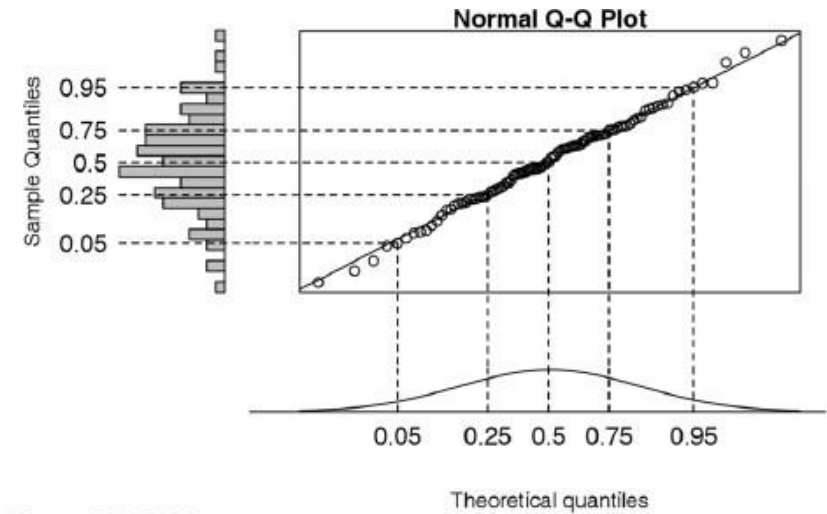
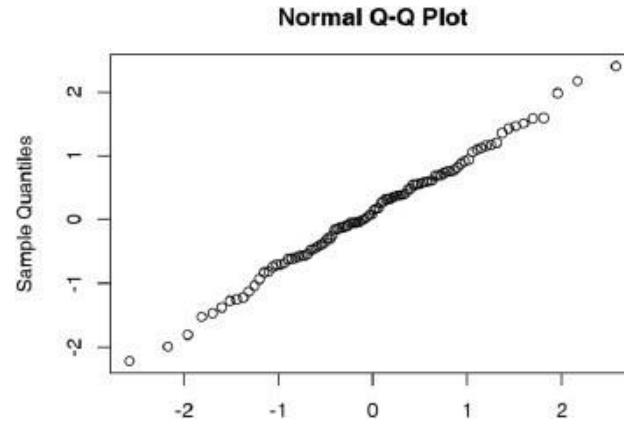
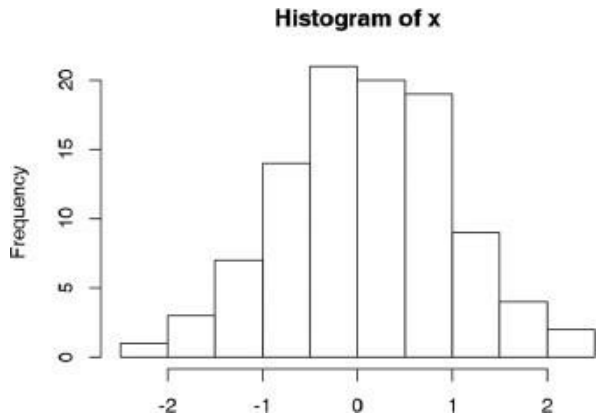
```
> t.test(KO_WT$weight ~ KO_WT$genotype)
```

### **Welch Two Sample t-test**

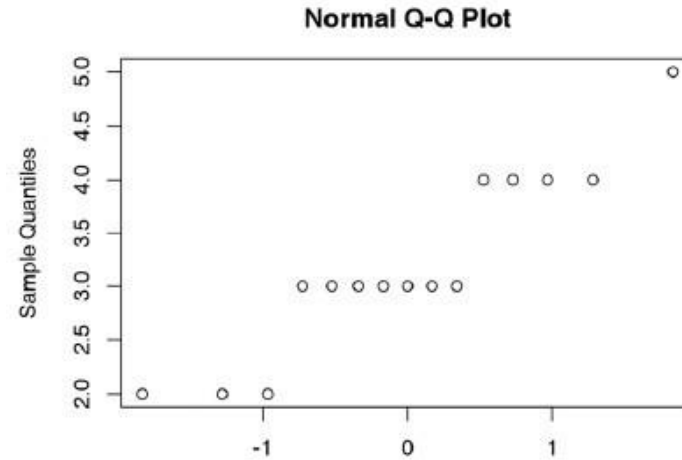
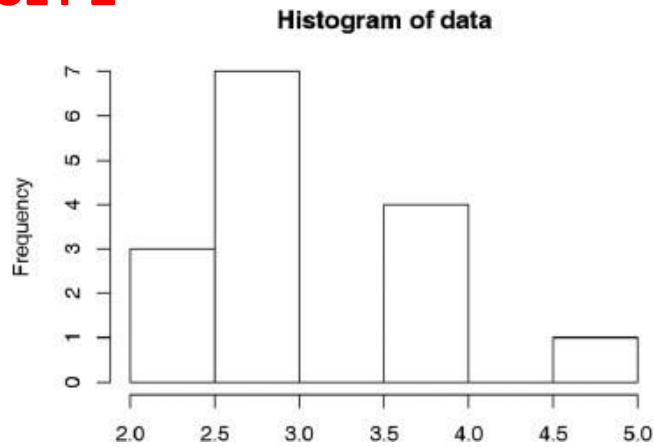
```
data: KO_WT$weight by KO_WT$genotype
t = -1.4261, df = 18.905, p-value = 0.1702
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0078465  0.5705536
sample estimates:
mean in group KO mean in group WT
      30.32366      31.54231
```

# Graphical tools to assess normality

## DATASET 1



## DATASET 2



## Normality tests

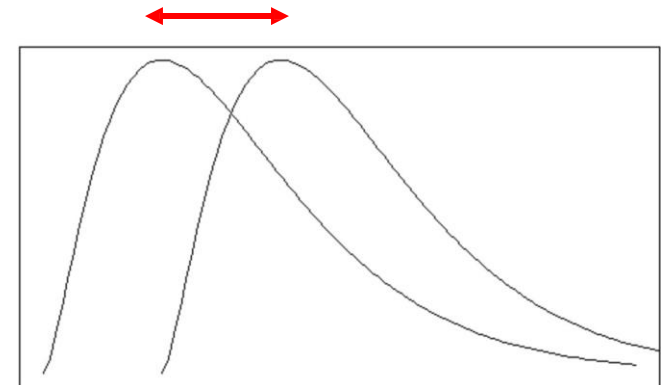
- Examples: Shapiro-Wilk, Kolmogorov-Smirnov.

```
> shapiro_test(weight$weight)
# A tibble: 1 x 3
  variable      statistic p.value
  <chr>          <dbl>   <dbl>
1 weight$weight  0.902   0.166
```

- Normality tests are statistical tests
  - They test **against** normality (and can not show/prove normality)
  - They may not have enough power
  - Conversely, if  $n$  is large, even a very small departure from normality may be significant.

## *What happens if the data is not normally distributed ?*

- Transform the data using the log function
  - Examples of data that should be logged:
    - Highly asymmetrical data
    - Data spanning several orders of magnitude
    - Data originating from ratios
- Another possible solution: non-parametric tests
  - Wilcoxon rank-sum test (WRS)
  - Mann–Whitney U test
  - Mann–Whitney–Wilcoxon (MWW)





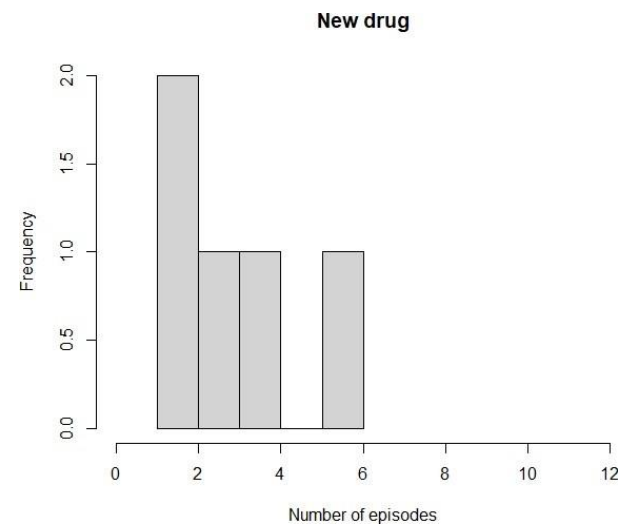
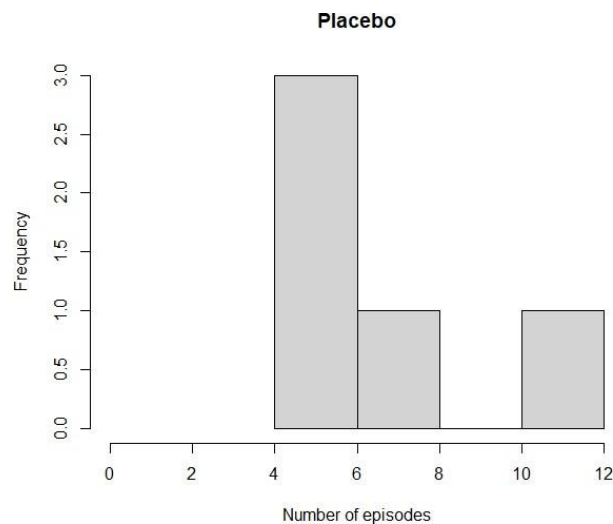
## *Mann–Whitney U test*

- General formulation:
  - All the observations from both groups are independent of each other
  - The variable is ordinal or numerical
  - $H_0$ : the distributions of both populations are equal
  - $H_1$ : the distributions are not equal

## *Mann–Whitney U test: an example*

Consider a Phase II clinical trial designed to investigate the effectiveness of a new drug to reduce symptoms of asthma in children. Participants are asked to record the number of episodes of shortness of breath over a 1 week period following receipt of either the new drug or a placebo. Is the difference significant ?

<b>Placebo</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>4</b>	<b>12</b>
New drug	3	6	4	2	1



## Mann–Whitney U test: an example

		Total sample (ordered)		Ranks	
Placebo	New drug	Placebo	New drug	Placebo	New drug
7	3		1		1
5	6		2		2
6	4		3		3
4	2	4	4	4.5	4.5
12	1	5		6	
		6	6	7.5	7.5
		7		9	
		12		10	

$$R_1 = 4.5 + 6 + 7.5 + 9 + 10 = 37$$

$$R_2 = 1 + 2 + 3 + 4.5 + 7.5 = 18$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 5(5) + \frac{5(6)}{2} - 37 = 3$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 5(5) + \frac{5(6)}{2} - 18 = 22$$

Test statistic for the Mann Whitney U Test is denoted  $U = \min(U_1, U_2)$

In R: `wilcox.test(data$placebo, data$drug)`

## Mann–Whitney U test: an example

Alpha = .05 (two-tailed)

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10
2							0	0	0
3				0	1	1	2	2	3
4			0	1	2	3	4	4	5
5		0	1	2	3	5	6	7	8
6		1	2	3	5	6	8	10	11
7		1	3	5	6	8	10	12	14
8	0	2	4	6	8	10	13	15	17
9	0	2	4	7	10	12	15	17	20
10	0	3	5	8	11	14	17	20	23

**Decision for the test: if  $U < U_{crit}$  we reject  $H_0$**

We do not reject  $H_0$  because  $3 > 2$ .

We do not have statistically significant evidence at  $\alpha = 0.05$ , to show that the two populations of numbers of episodes of shortness of breath are not equal.

Low power because of small sample sizes ?

```
> t.test(placebo,newdrug)
```

```
Welch Two Sample t-test
```

```
data: placebo and newdrug
```

```
t = 2.199, df = 6.6639, p-value = 0.06574
```

```
alternative hypothesis: true difference in means is not equal to 0
```

## *Take home message*

- How to handle (non-)normality
  - Start by making sure that your data is of the right type
  - If the data spans several orders of magnitude, is made of ratios or is very asymmetrical, consider taking the log.
  - If  $n$  is large ( $>20$  or  $30$ ), then you should not have to worry.
  - As long as your data is not "too" far from normality (and you are careful with borderline  $p$ -values), the  $t$ -test will return a meaningful value.
- How to handle (non-)normality: alternatively
  - Perform both a  $t$ -test and a Wilcoxon rank-sum test
  - If they return a coherent result, it is stronger than if you performed only one test.
  - If the results are incoherent, look at the data to find out what produced this.