Statistical tests

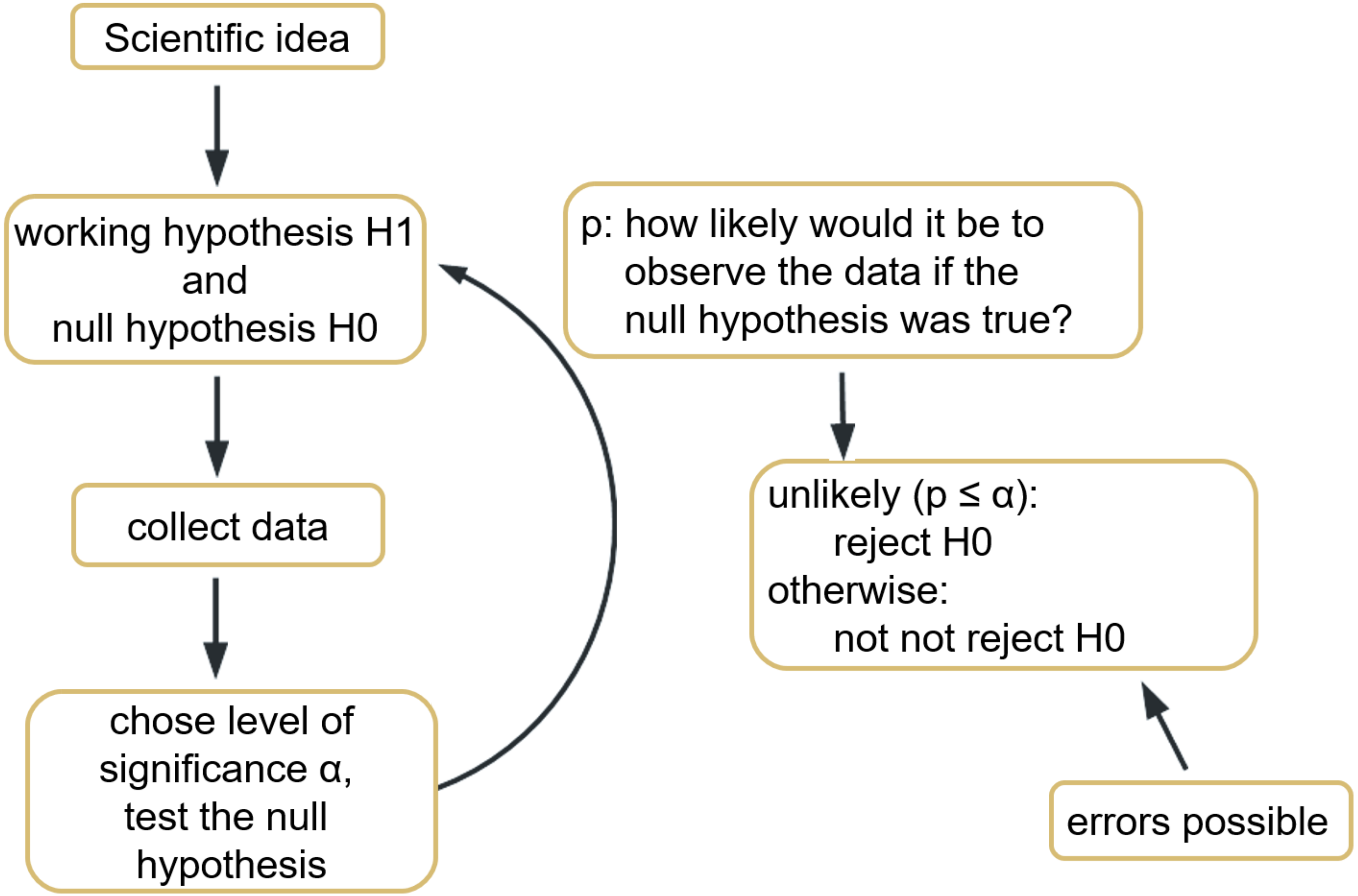
Introduction to R - Day 3

Instructor: Selina Baldauf

Freie Universität Berlin - Theoretical Ecology

2021-08-01 (updated: 2021-08-30)

Typical approach in statistics



Types of errors

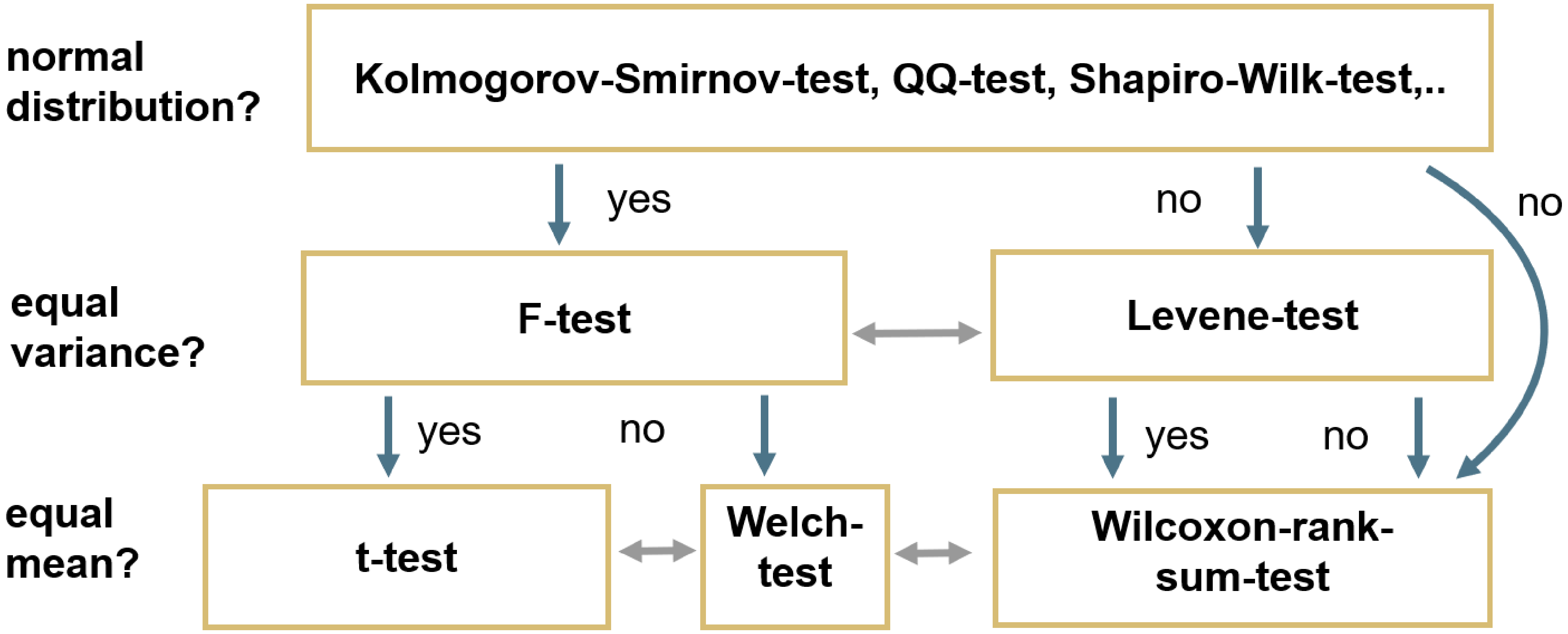
		Truth	
		H0 is true	H0 is false
Decision based on statistical test	Accept H0	Correct decision Probability: $1 - \alpha$	**Type II error** Probability: β
	Reject H0	**Type I error** Probability: α	Correct decision Probability: $1 – \beta$ (Power)

Keep error probability as low as possible

- Decrease type I errors: choose smaller α → but increases type II errors
- Decrease type II errors: more data or different test

→ There are trade-offs and errors can't be completely avoided, especially with multiple tests applied in a row

Overview of tests



Tests for normal distribution

Test for normal distribution

There are **various tests** and the outcome might differ!

Kolmogorov-Smirnov-Test (KS-Test)

- how much does the cumulative probability of observed data differs from normal distribution
- very general test (also works for other distributions)
- low power: data not normally distributed but H_0 not rejected

Shapiro-Wilk-Test

- how much does variance of observed data differ from normal distribution
- specific test only for normal distribution
- high power, also for few data points

Visual tests: QQ-Plot

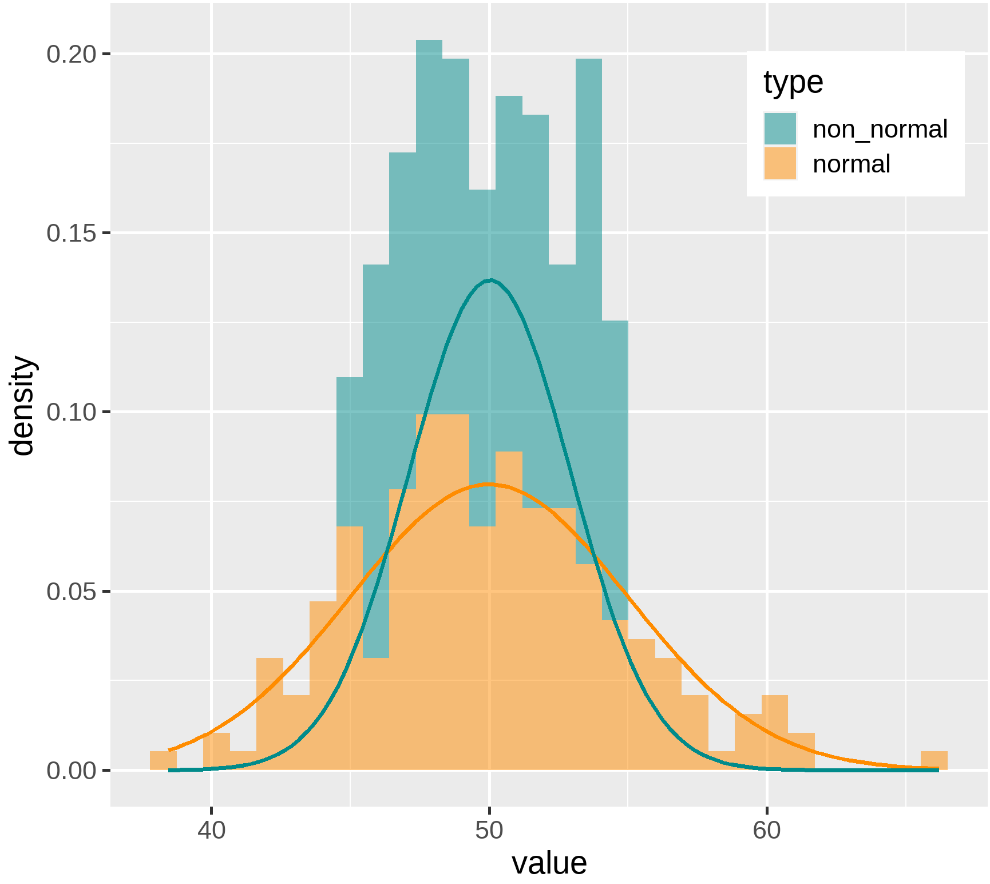
- quantiles of observed data plotted against quantiles of normal distribution
- scientist has to decide if normal or not

The data

Create a tibble with two variables

- `normal`: 200 normally distributed values with mean 50 and standard deviation 5
- `non_normal`: 200 uniformly distributed values between 45 and 55

```
set.seed(123)
mydata <- tibble(
  normal = rnorm(
    n = 200,
    mean = 50,
    sd = 5
  ),
  non_normal = runif(
    n = 200,
    min = 45,
    max = 55
  )
)
```



Kolmogorov-Smirnov-Test

H_0: Data does not differ from a normal distribution

```
# Tests need the data as a vector input
ks.test(mydata$normal,
  "pnorm",
  mean = mean(mydata$normal),
  sd = sd(mydata$normal)
)
##
##      One-sample Kolmogorov-Smirnov test
##
## data:  mydata$normal
## D = 0.054249, p-value = 0.5983
## alternative hypothesis: two-sided
```

- D: test statistic, largest difference between cumulative distributions
- p-value: probability to observe data if H_0 was true
- two-sided test: we don't know in which direction our data deviates from normal distribution

Result: The data does not deviate significantly from a normal distribution (KS-Test, D = 0.054, p > 0.05)

Kolmogorov-Smirnov-Test

H_0: Data does not differ from a normal distribution

```
ks.test(mydata$non_normal,
  "pnorm",
  mean = mean(mydata$non_normal),
  sd = sd(mydata$non_normal)
)
##
##      One-sample Kolmogorov-Smirnov test
##
## data:  mydata$non_normal
## D = 0.077198, p-value = 0.1843
## alternative hypothesis: two-sided
```

Result: The data does not deviate significantly from a normal distribution (KS-Test, D = 0.077, p > 0.05)

But: We know that our data is not normally distributed. Nevertheless, we would not reject H_0 → we would make a type II error here

Shapiro-Wilk-Test

H_0: Data does not differ from a normal distribution

```
shapiro.test(mydata$normal)
##
##      Shapiro-Wilk normality test
##
## data:  mydata$normal
## W = 0.99076, p-value = 0.2298
```

- W: test statistic
- p-value: probability to observe the data if H_0 was true

The data does not deviate significantly from a normal distribution (Shapiro-Wilk-Test, W = 0.991, p > 0.05)

Shapiro-Wilk-Test

H_0: Data does not differ from a normal distribution

```
shapiro.test(mydata$non_normal)
##
##      Shapiro-Wilk normality test
##
## data:  mydata$non_normal
## W = 0.95114, p-value = 2.435e-06
```

The data deviates significantly from a normal distribution (Shapiro-Wilk-Test, $W = 0.95$, $p < 0.05$)

Visual test with QQ-Plot

Points should match the straight line. Small deviations are okay.

```
ggplot(mydata, aes(sample = normal)) +
  stat_qq() +
  stat_qq_line()
```

```
ggplot(mydata, aes(sample = non_normal)) +
  stat_qq() +
  stat_qq_line()
```

Tests for equal variance

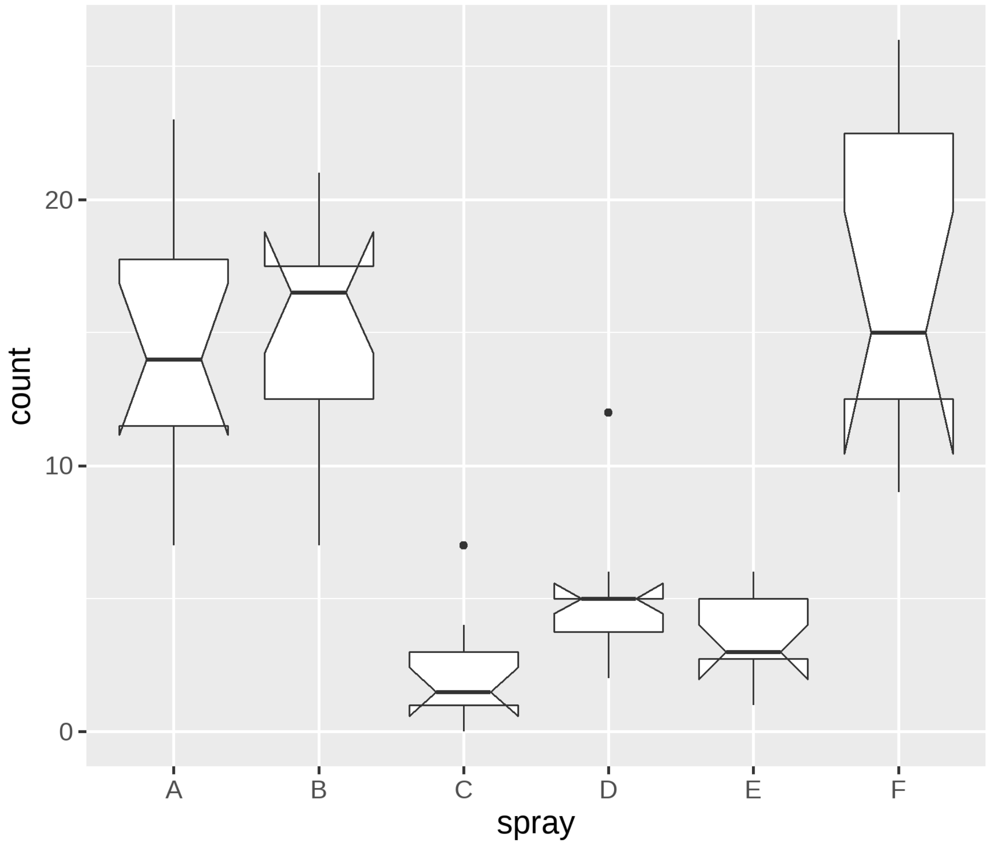
The data

Counts of insects in agricultural units treated with different insecticides.

Compare treatments A, B and E:

- Create subsets before: count variable for each treatment as a vector

```
TreatA <- filter(InsectSprays,
                 spray == "A")$count
TreatB <- filter(InsectSprays,
                 spray == "B")$count
TreatE <- filter(InsectSprays,
                 spray == "E")$count
```



Test for equal variance

First, test for normal distribution!

F-Test

- **normal distribution** of groups
- calculates ratio of variances (if equal, ratio = 1)
- p: How likely is ratio if variances were equal?
- only for normal distribution

Levene test

- **non-normal distribution** of groups
- compare difference between data sets with difference within data sets
- works with all distributions

Test for equal variances

If we want to compare variances between treatments A, B and E, we first test for normal distribution

```
shapiro.test(TreatA)
##
##      Shapiro-Wilk normality test
##
## data:  TreatA
## W = 0.95757, p-value = 0.7487
shapiro.test(TreatB)
##
##      Shapiro-Wilk normality test
##
## data:  TreatB
## W = 0.95031, p-value = 0.6415
shapiro.test(TreatE)
##
##      Shapiro-Wilk normality test
##
## data:  TreatE
## W = 0.92128, p-value = 0.2967
```

Result: All 3 treatments are normally distributed.

F-Test

H_0: Variances do not differ between groups

```
var.test(TreatA, TreatB)
##
##      F test to compare two variances
##
## data:  TreatA and TreatB
## F = 1.2209, num df = 11, denom df = 11, p-value = 0.7464
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3514784 4.2411442
## sample estimates:
## ratio of variances
##            1.22093
```

- F: test statistics, ratio of variances (if F = 1, variances are equal)
- df: degrees of freedom of both groups
- p-value: how likely is it to observe the data if H_0 was true?

Result: The variances of sprays A and B do not differ significantly (F-Test, $F_{11,11}$ = 1.22, p > 0.05)

F-Test

H_0: Variances do not differ between groups

```
var.test(TreatA, TreatE)
##
##      F test to compare two variances
##
## data:  TreatA and TreatE
## F = 7.4242, num df = 11, denom df = 11, p-value = 0.002435
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   2.137273 25.789584
## sample estimates:
## ratio of variances
##           7.424242
```

Result: The variances of sprays A and E differ significantly (F-Test, $F_{11,11}$ = 7.42, p < 0.05)

Test for equal means

Test for equal means

First test for normality and equal variances in the groups!

t-test

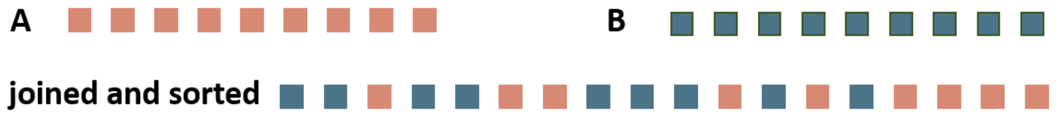
- **normal distribution** AND **equal variance**
- compares if mean values are within range or standard error of each other
- p: how likely is the difference if the means were equal

Welch-Test

- **normal distribution** but **unequal variance**
- corrected t-test

Wilcoxon rank sum test

- **non-normal distribution** and **unequal variance**
- compares rank sums of the data
- non-parametric



t-test

H_0: The samples do not differ in their mean

Treatment A and B: **normally distributed** and **equal variance**

```
t.test(TreatA, TreatB, var.equal = TRUE)
##
##      Two Sample t-test
##
## data:  TreatA and TreatB
## t = -0.45352, df = 22, p-value = 0.6546
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.643994  2.977327
## sample estimates:
## mean of x mean of y
##  14.50000  15.33333
```

- t: test statistics (t = 0 means equal means)
- df: degrees of freedom of t-statistics
- p-value: how likely is it to observe the data if H_0 was true?

Result: The means of spray A and B do not differ significantly (t = -0.45, df = 22, p > 0.05)

Welch-Test

H_0: The samples do not differ in their mean

Treatment A and E: **normally distributed** and **non-equal variance**

```
t.test(TreatA, TreatE, var.equal = FALSE)
##
##      Welch Two Sample t-test
##
## data:  TreatA and TreatE
## t = 7.5798, df = 13.91, p-value = 2.655e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   7.885546 14.114454
## sample estimates:
## mean of x mean of y
##      14.5       3.5
```

Result: The means of spray A and E do differ significantly ($t = 7.58$, $df = 13.9$, $p < 0.05$)

Wilcoxon-rank-sum Test

H_0: The samples do not differ in their mean

We don't need the Wilcoxon test to compare treatment A and B, but for the sake of an example:

```
wilcox.test(TreatA, TreatB)
##
##     Wilcoxon rank sum test with continuity correction
##
## data:  TreatA and TreatB
## W = 62, p-value = 0.5812
## alternative hypothesis: true location shift is not equal to 0
```

Result: The means of spray A and E do not differ significantly (W = 62, p > 0.05)

Paired values

Are there pairs of data points?

Example: samples of invertebrates across various rivers before and after sewage plants.

- for each plant, there is a pair of data points (before and after the plant)
- Question: Is the change (before-after) significant

Use `paired = TRUE` in the test.

```
t.test(TreatA, TreatB, var.equal = TRUE, paired = TRUE)
t.test(TreatA, TreatB, var.equal = FALSE, paired = TRUE)
wilcox.test(TreatA, TreatB, paired = TRUE)
```

Careful: your treatment vector both have to have the same order

Now you

Task 1: Statistical tests

Find the task description here