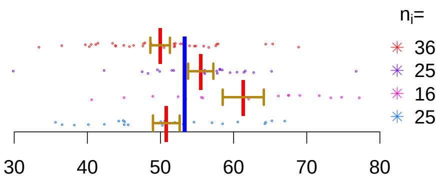
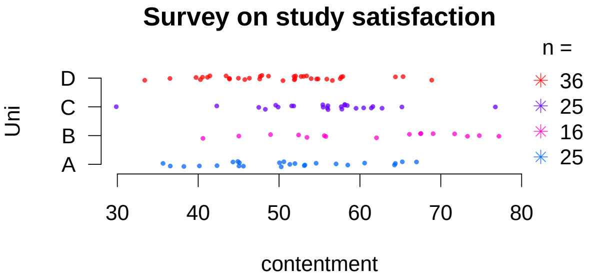
Analysis of variance and multiple testing



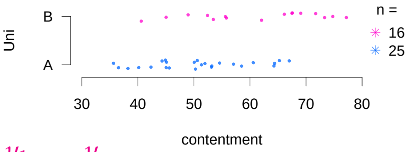
For comments please email: M. Messer, michael.messer@tuwien.ac.at

Reminder

At four universities students of a certain study program were interviewed regarding the level of their satisfaction with the study situation. An extensive survey had to be filled out. Subsequently, for every respondent a global value of 'contentment' was evaluated.

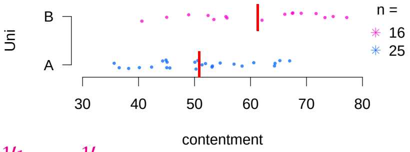


Reminder: two-sample situation



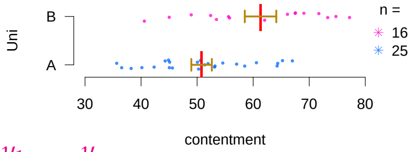
- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$

Reminder: two-sample situation



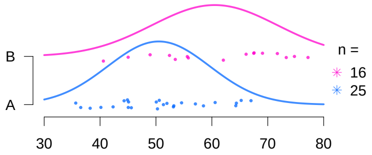
- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?

Reminder: two-sample situation



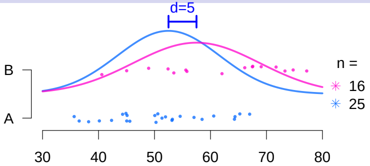
- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x/\sqrt{n_1}$ and $sem_y = s_y/\sqrt{n_2}$

Reminder: two-sample situation



- Data $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x / \sqrt{n_1}$ and $sem_y = s_y / \sqrt{n_2}$
- In the model $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \dots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \dots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$

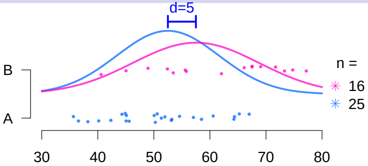
Reminder: two-sample situation

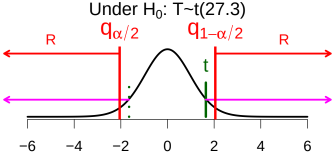


- Data $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x / \sqrt{n_1}$ and $sem_y = s_y / \sqrt{n_2}$
- In the model $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \dots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \dots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$

  it holds under $H_0 : \mu_y - \mu_x = d_0$ that

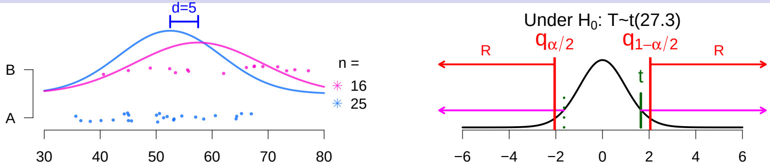
Reminder: two-sample situation





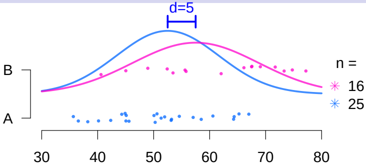
- Data $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x / \sqrt{n_1}$ and $sem_y = s_y / \sqrt{n_2}$
- In the model $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \dots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \dots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
  it holds under $H_0 : \mu_y - \mu_x = d_0$ that $T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$ (R knows ν)

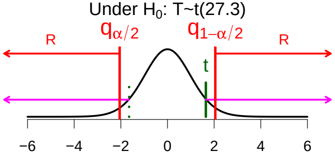
Reminder: two-sample situation



- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x / \sqrt{n_1}$ and $sem_y = s_y / \sqrt{n_2}$
- In the model $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
  it holds under $H_0 : \mu_y - \mu_x = d_0$ that $T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$ (R knows ν)
- Judge data via $t = \frac{(\bar{y} - \bar{x}) - d_0}{\sqrt{sem_y^2 + sem_x^2}}$ according to the $t(\nu)$-distribution

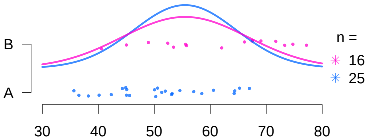
Reminder: two-sample situation

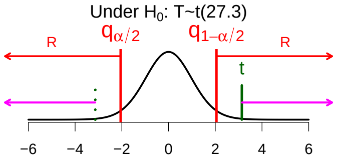




- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x/\sqrt{n_1}$ and $sem_y = s_y/\sqrt{n_2}$
- In the model $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
  it holds under $H_0 : \mu_y - \mu_x = d_0$ that $T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$ (R knows ν)
- Judge data via $t = \frac{(\bar{y} - \bar{x}) - d_0}{\sqrt{sem_y^2 + sem_x^2}}$ according to the $t(\nu)$-distribution
- Example: for $d_0 = 5$ it is $p > \alpha = 5\%$ resp. $t \notin R$, thus do not reject H_0 on the α-level

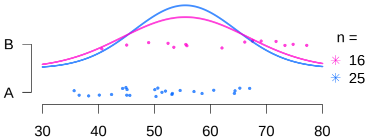
Reminder: two-sample situation

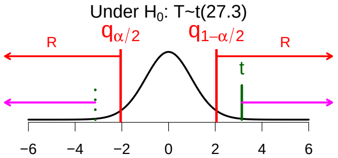




- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x/\sqrt{n_1}$ and $sem_y = s_y/\sqrt{n_2}$
- In the model $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
  it holds under $H_0 : \mu_y - \mu_x = d_0$ that $T := \frac{(\bar{Y}-\bar{X})-d_0}{\sqrt{SEM_y^2+SEM_x^2}} \sim t(\nu)$ (R knows ν)
- Judge data via $t = \frac{(\bar{y}-\bar{x})-d_0}{\sqrt{sem_y^2+sem_x^2}}$ according to the $t(\nu)$-distribution
- Example: for $d_0 = 5$ it is $p > \alpha = 5\%$ resp. $t \notin R$, thus do <u>not</u> reject H_0 on the α-level
- In practice often $d_0 = 0$, i.e., 'no shift' $\rightarrow$ here rejected

Reminder: two-sample situation

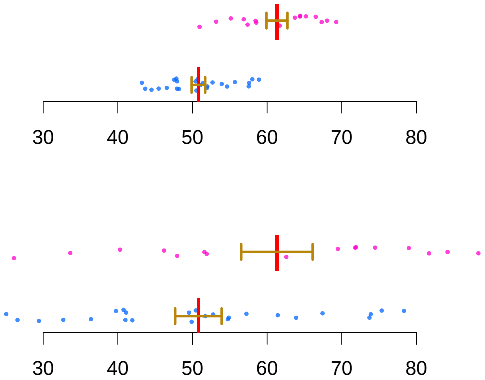




- Data $x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}$
- Mean $\bar{x}$ and $\bar{y}$ far apart?
- Depends on the standard errors $sem_x = s_x / \sqrt{n_1}$ and $sem_y = s_y / \sqrt{n_2}$
- In the model $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ independent RVs with
  $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$ and
  $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$ with $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$
  it holds under $H_0 : \mu_y - \mu_x = d_0$ that $T := \frac{(\bar{Y} - \bar{X}) - d_0}{\sqrt{SEM_y^2 + SEM_x^2}} \sim t(\nu)$ (R knows ν)
- Judge data via $t = \frac{(\bar{y} - \bar{x}) - d_0}{\sqrt{sem_y^2 + sem_x^2}}$ according to the $t(\nu)$-distribution
- Example: for $d_0 = 5$ it is $p > \alpha = 5\%$ resp. $t \notin R$, thus do <u>not</u> reject H_0 on the α-level
- In practice often $d_0 = 0$, i.e., 'no shift' $\rightarrow$ here rejected

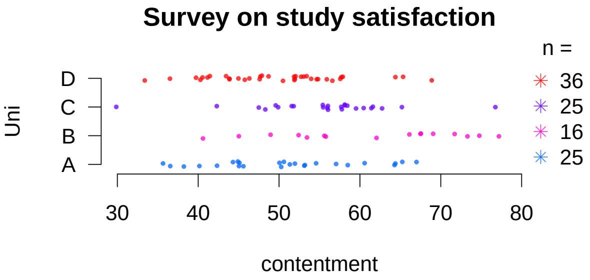
This is the Welch-test. Remember that there is also Student's t-test, were equal variances $\sigma_1^2 = \sigma_2^2$ are assumed, for which we found under H_0 that $T \sim t(n_1 + n_2 - 2)$ while a different scaling was used in T ('pooled variance')

Message



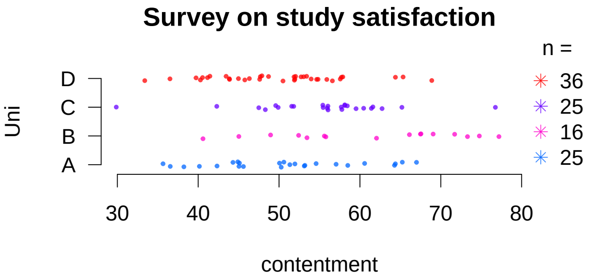
The *t*-statistic says:
the discrepancy between the groups is large,
if the variability within the groups is small

Today: Comparison of $k \geqslant 2$ groups



Here $k = 4$ groups

Today: Comparison of $k \geqslant 2$ groups

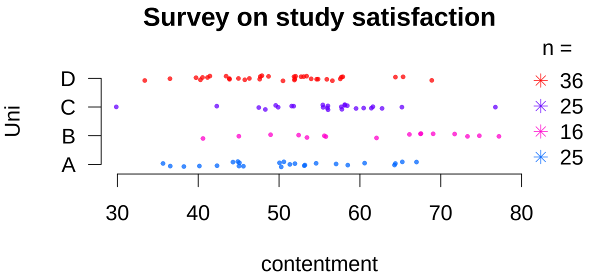


Here $k = 4$ groups

Two possibilities to deal with that:

- Multiple pairwise t-tests

Today: Comparison of $k \geqslant 2$ groups



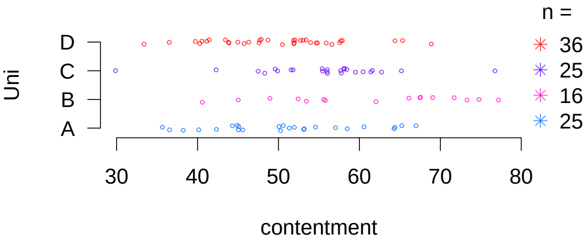
Here $k = 4$ groups

Two possibilities to deal with that:

- Multiple pairwise t-tests
- Analysis of variance (ANOVA)

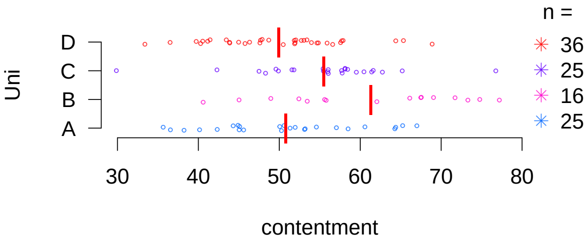
Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



Multiple testing

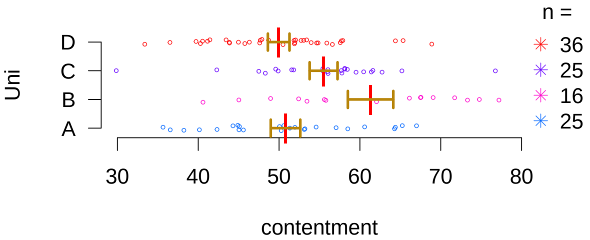
Multiple pairwise *t*-tests: for every pair a two-sample *t*-test



- means

Multiple testing

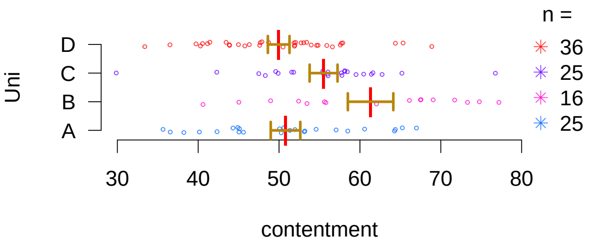
Multiple pairwise *t*-tests: for every pair a two-sample *t*-test



- means ± standard errors

Multiple testing

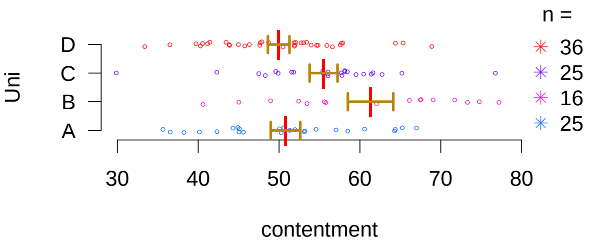
Multiple pairwise *t*-tests: for every pair a two-sample *t*-test



- means ± standard errors
- Naive: Discrepancy between *A* and *D* small, between *B* and *D* large

Multiple testing

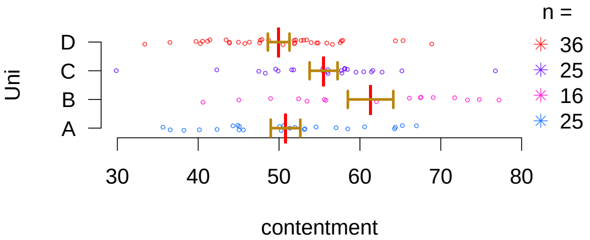
Multiple pairwise t-tests: for every pair a two-sample t-test



- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



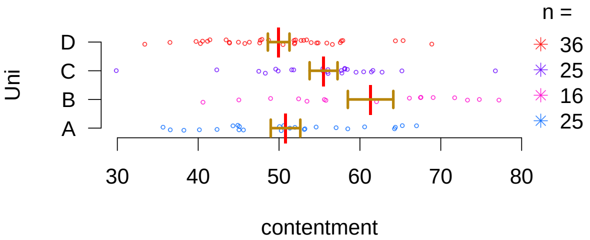
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	red	green	green	green	red	red

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



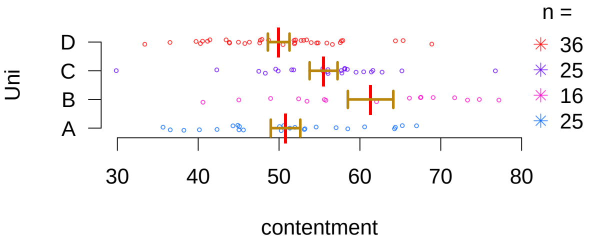
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004					

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



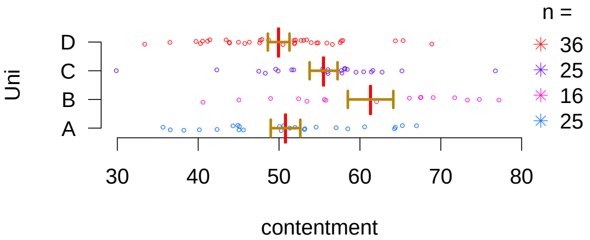
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004	0.067				

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



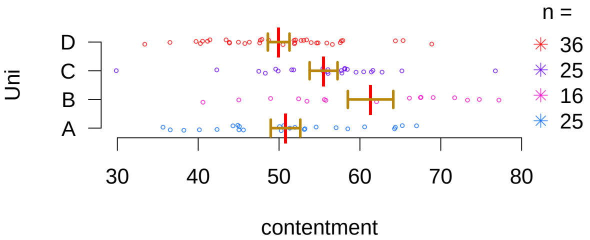
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004	0.067	0.710			

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



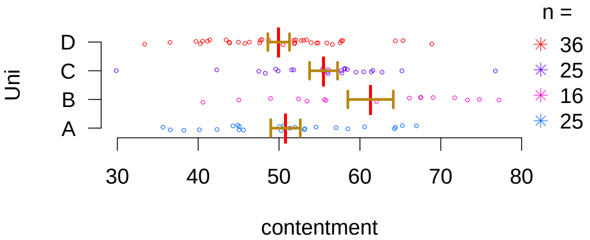
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004	0.067	0.710	0.090		

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



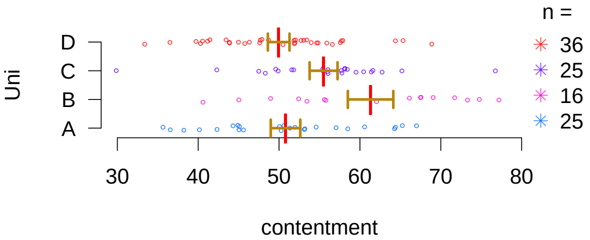
- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004	0.067	0.710	0.090	0.001	

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

Multiple pairwise t-tests: for every pair a two-sample t-test



- means $\pm$ standard errors
- Naive: Discrepancy between A and D small, between B and D large
- $k = 4$ groups $\rightarrow$ six t-tests ($d_0 = 0$, no shift)

groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value	0.004	0.067	0.710	0.090	0.001	0.014

red: $p < 5\%$, green: $p \geqslant 5\%$

Multiple testing

- With 4 groups the number of t-tests performed is 6

Multiple testing

- With 4 groups the number of t-tests performed is 6
- How many t-tests do we have to perform with k groups?

Multiple testing

- With 4 groups the number of t-tests performed is 6
- How many t-tests do we have to perform with k groups? Answer:

$$m = \binom{k}{2}$$

(the number of 2-element subsets of a k-element set)

Multiple testing

- With 4 groups the number of t-tests performed is 6
- How many t-tests do we have to perform with k groups? Answer:

$$m = \binom{k}{2}$$

(the number of 2-element subsets of a k-element set)

- We mention that

$$m = \frac{k(k-1)}{2}$$

grows quadratically in k

Multiple testing

- With 4 groups the number of t-tests performed is 6
- How many t-tests do we have to perform with k groups? Answer:

$$m = \binom{k}{2}$$

(the number of 2-element subsets of a k-element set)

- We mention that

$$m = \frac{k(k-1)}{2}$$

grows quadratically in k

- Meaning: the number of tests to perform increases rapidly with the number of groups

Multiple testing

- With 4 groups the number of t-tests performed is 6
- How many t-tests do we have to perform with k groups? Answer:

$$m = \binom{k}{2}$$

  (the number of 2-element subsets of a k-element set)
- We mention that

$$m = \frac{k(k-1)}{2}$$

  grows quadratically in k
- Meaning: the number of tests to perform increases rapidly with the number of groups
- … and the application of <u>many statistical tests</u> brings a <u>conceptional problem</u> …

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis $\underline{H_0 \text{ was true}}$

The multiple testing problem

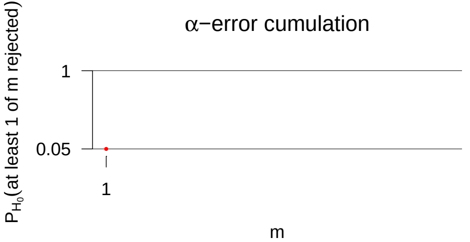
gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

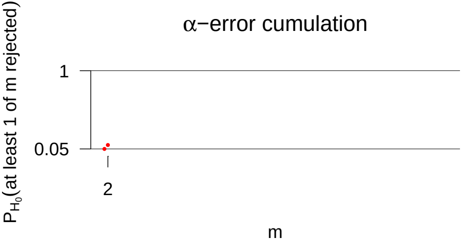


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

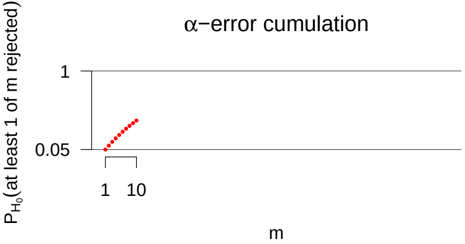


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

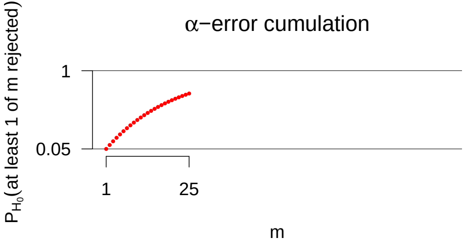


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

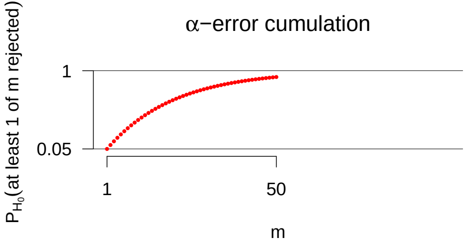


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

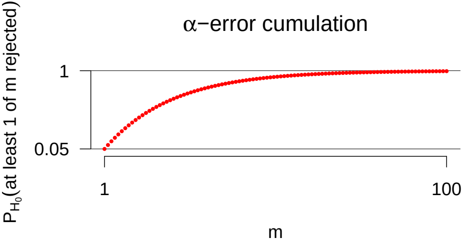


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

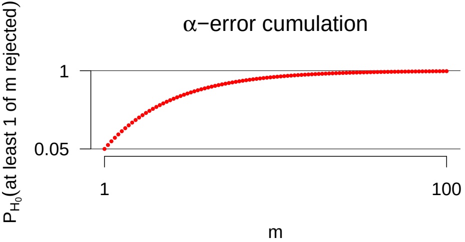


- In the picture $\alpha = 5\%$

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

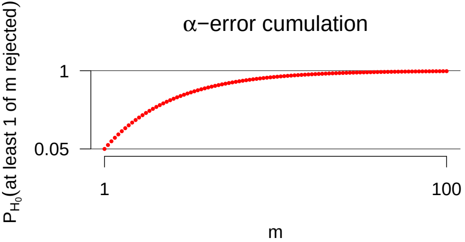


- In the picture $\alpha = 5\%$
- Message: The probability to commit the α-error at least once tended to 1, for $m \to \infty$ ('α-error cumulation')

The multiple testing problem

gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?

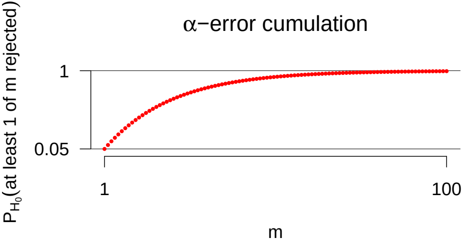


- In the picture $\alpha = 5\%$
- Message: The probability to commit the α-error at least once tended to 1, for $m \to \infty$ ('α-error cumulation')
- And this is bad news! If we only perform enough tests, then the null hypothesis will be rejected at some point (even if it was always true)

The multiple testing problem

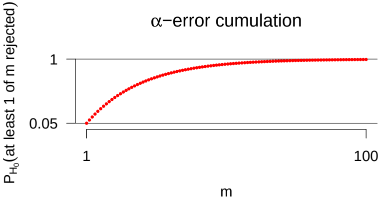
gedankenexperiment

- assume we perfomed m (independent) statistical tests at level α
- and we knew, that the null hypothesis H_0 was true
- What was the probability to commit the α-error (i.e., falsely reject) in at least one of the tests?



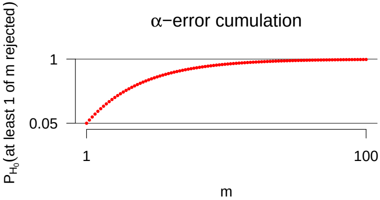
- In the picture $\alpha = 5\%$
- Message: The probability to commit the α-error at least once tended to 1, for $m \to \infty$ ('α-error cumulation')
- And this is bad news! If we only perform enough tests, then the null hypothesis will be rejected at some point (even if it was always true)
- Meaning for practice: do not perform unnecessary tests!

The multiple testing problem



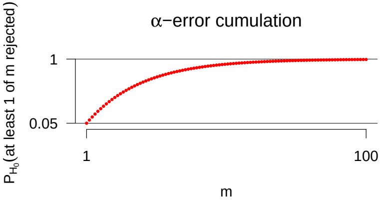
- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once

The multiple testing problem



- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once
- Let V_i be the event, that the i-th null hypothesis is rejected

The multiple testing problem

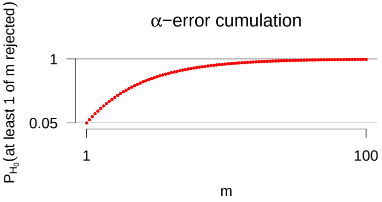


- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once
- Let V_i be the event, that the i-th null hypothesis is rejected
  - $m = 1$

$$\mathbb{P}_{H_0}(V_1) = \alpha$$

by construction of the test

The multiple testing problem



- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once
- Let V_i be the event, that the i-th null hypothesis is rejected
  - $m = 1$

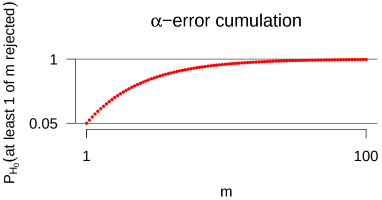
    $$\mathbb{P}_{H_0}(V_1) = \alpha$$

    by construction of the test
  - $m = 2$:

    $$\mathbb{P}_{H_0}(V_1 \cup V_2) = 1 - \mathbb{P}_{H_0}(V_1^c \cap V_2^c) = 1 - \mathbb{P}_{H_0}(V_1^c)\mathbb{P}_{H_0}(V_2^c) = 1 - (1 - \alpha)^2$$

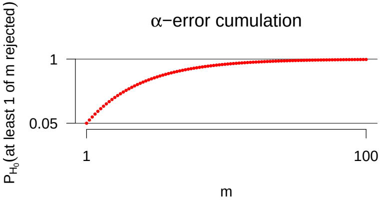
    used: $\mathbb{P}$('at least one') $= 1 - \mathbb{P}$('none') as well as indpendence

The multiple testing problem



- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once
- Let V_i be the event, that the i-th null hypothesis is rejected
  - $m = 1$
  $$\mathbb{P}_{H_0}(V_1) = \alpha$$
  by construction of the test
  - $m = 2$:
  $$\mathbb{P}_{H_0}(V_1 \cup V_2) = 1 - \mathbb{P}_{H_0}(V_1^c \cap V_2^c) = 1 - \mathbb{P}_{H_0}(V_1^c)\mathbb{P}_{H_0}(V_2^c) = 1 - (1-\alpha)^2$$
  used: $\mathbb{P}(\text{'at least one'}) = 1 - \mathbb{P}(\text{'none'})$ as well as indpendence
  - general m:
  $$\mathbb{P}_{H_0}(\bigcup_{i=1}^{m} V_i) = 1 - \mathbb{P}_{H_0}(\bigcap_{i=1}^{m} V_i^c) = 1 - \prod_{i=1}^{m} \mathbb{P}_{H_0}(V_i^c) = \boxed{1 - (1-\alpha)^m}$$

The multiple testing problem



- m (independent) tests at level α
- Assume H_0 always holds true
- Looking for the probability to commit the α-error at least once
- Let V_i be the event, that the i-th null hypothesis is rejected
  - $m = 1$
  $$\mathbb{P}_{H_0}(V_1) = \alpha$$
  by construction of the test
  - $m = 2$:
  $$\mathbb{P}_{H_0}(V_1 \cup V_2) = 1 - \mathbb{P}_{H_0}(V_1^c \cap V_2^c) = 1 - \mathbb{P}_{H_0}(V_1^c)\mathbb{P}_{H_0}(V_2^c) = 1 - (1 - \alpha)^2$$
  used: $\mathbb{P}$('at least one') $= 1 - \mathbb{P}$('none') as well as indpendence
  - general m:
  $$\mathbb{P}_{H_0}(\bigcup_{i=1}^{m} V_i) = 1 - \mathbb{P}_{H_0}(\bigcap_{i=1}^{m} V_i^c) = 1 - \prod_{i=1}^{m} \mathbb{P}_{H_0}(V_i^c) = \boxed{1 - (1 - \alpha)^m} \to 1 \; (m \to \infty)$$

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests
- More precisely: Correct the original significance level α to

$$\boxed{\alpha^* = \frac{\alpha}{m}}$$

known as the *Bonferroni-correction* ('divide by number of tests')

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests
- More precisely: Correct the original significance level α to

$$\boxed{\alpha^* = \frac{\alpha}{m}}$$

  known as the *Bonferroni-correction* ('divide by number of tests')
- Consequence: the probability to commit the α-alpha error at least once is restricted by α

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests
- More precisely: Correct the original significance level α to

$$\boxed{\alpha^* = \frac{\alpha}{m}}$$

  known as the *Bonferroni-correction* ('divide by number of tests')
- Consequence: the probability to commit the α-alpha error at least once is restricted by α
- Quantitatively:

$$\mathbb{P}_{H_0}\Big(\bigcup_{i=1}^{m} V_i\Big) \leqslant \sum_{i=1}^{m} \mathbb{P}_{H_0}(V_i) = m \cdot \alpha^* = \alpha$$

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests
- More precisely: Correct the original significance level α to

$$\boxed{\alpha^* = \frac{\alpha}{m}}$$

  known as the *Bonferroni-correction* ('divide by number of tests')
- Consequence: the probability to commit the α-alpha error at least once is restricted by α
- Quantitatively:

$$\mathbb{P}_{H_0}\left(\bigcup_{i=1}^{m} V_i\right) \leqslant \sum_{i=1}^{m} \mathbb{P}_{H_0}(V_i) = m \cdot \alpha^* = \alpha$$

- Alternatively: Use the original level α, but correct every p-value p to

$$\boxed{p^* = m \cdot p}$$

('multiplication with the number of tests')

Way out: Bonferroni correction

- m tests at level α
- Problem: the global α-error is increased
- A possible way out: be more 'strict' with each of the m tests
- More precisely: Correct the original significance level α to

$$\boxed{\alpha^* = \frac{\alpha}{m}}$$

known as the *Bonferroni-correction* ('divide by number of tests')

- Consequence: the probability to commit the α-alpha error at least once is restricted by α
- Quantitatively:

$$\mathbb{P}_{H_0}\Big(\bigcup_{i=1}^{m} V_i\Big) \leqslant \sum_{i=1}^{m} \mathbb{P}_{H_0}(V_i) = m \cdot \alpha^* = \alpha$$

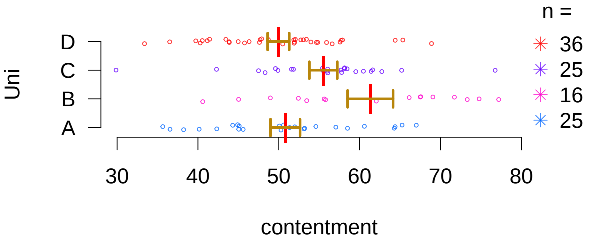
- Alternatively: Use the original level α, but correct every p-value p to

$$\boxed{p^* = m \cdot p}$$

('multiplication with the number of tests')

- Equivalent, as $p \leqslant \alpha^* \Leftrightarrow p^* \leqslant \alpha$

Multiple testing and Bonferroni-correction

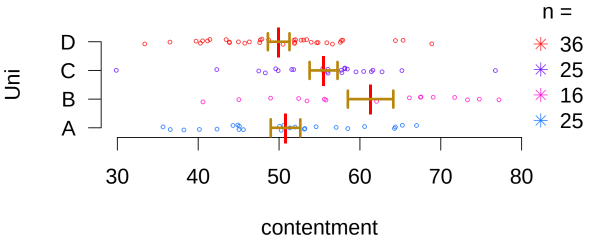


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)						

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

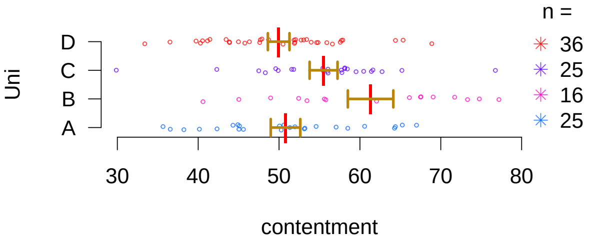


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024					

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

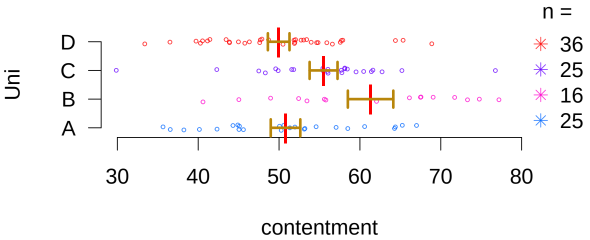


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268				

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

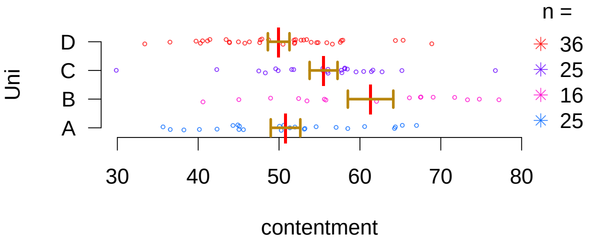


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1			

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

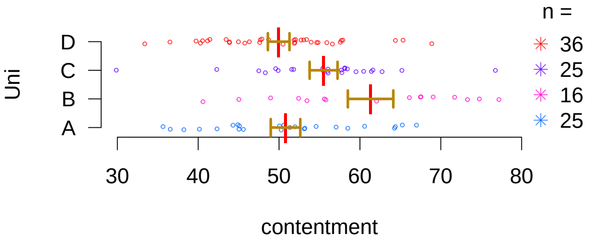


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1	0.540		

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

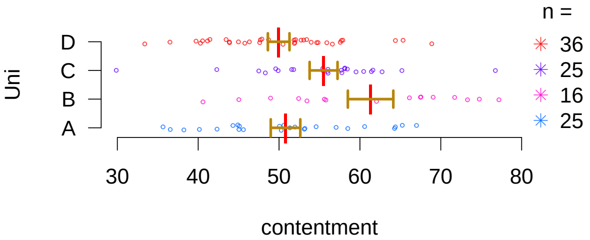


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1	0.540	0.006	

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

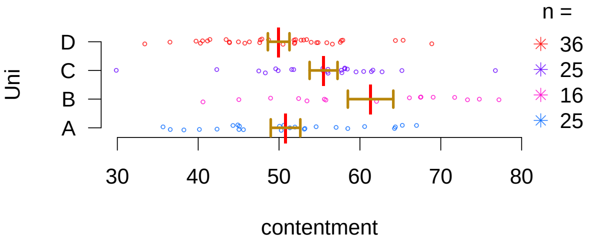


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1	0.540	0.006	0.084

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$

Multiple testing and Bonferroni-correction

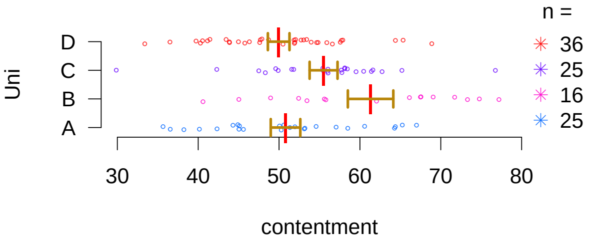


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1	0.540	0.006	0.084

red: $< 5\%$, green: $\geqslant 5\%$

- Bonferroni: multiply the original p-value with $m = 6$
- One additional null hypothesis cannot be rejected after the Boferroni correction.

Multiple testing and Bonferroni-correction

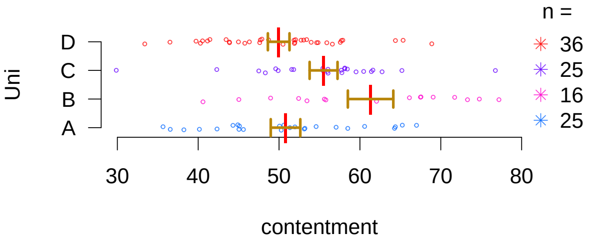


groups	A/B	A/C	A/D	B/C	B/D	C/D
p-value p (prior Bonferroni)	0.004	0.067	0.710	0.090	0.001	0.014
p-value p^* (after Bonferroni)	0.024	0.268	1	0.540	0.006	0.084

red: $< 5\%$, green: $\geqslant 5\%$

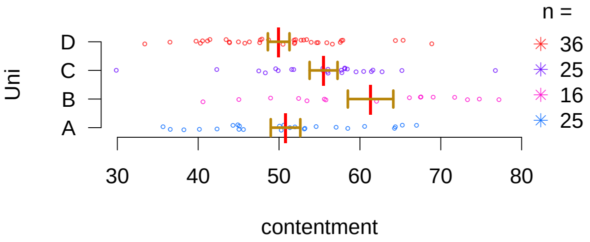
- Bonferroni: multiply the original p-value with $m = 6$
- One additional null hypothesis cannot be rejected after the Boferroni correction. We are more strict!

Analysis of variance



- $k = 4$ groups
- $m = 6$ pairwise t-tests $\rightarrow$ the global α-error is increased

Analysis of variance

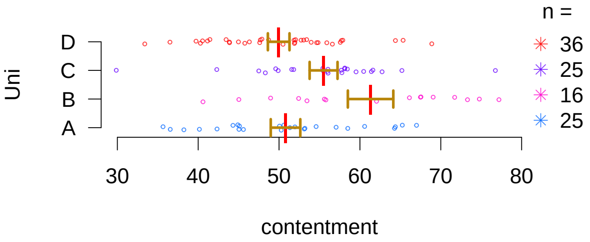


- $k = 4$ groups
- $m = 6$ pairwise t-tests $\rightarrow$ the global α-error is increased
- Another possible 'way out':

  The *analysis of variance*

  short ANOVA

Analysis of variance

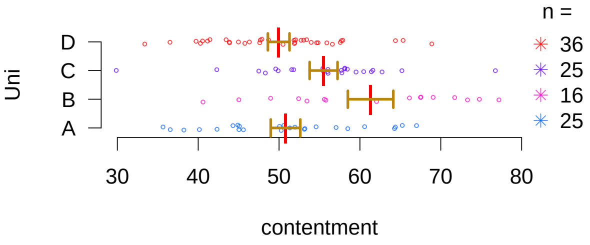


- $k = 4$ groups
- $m = 6$ pairwise t-tests $\rightarrow$ the global α-error is increased
- Another possible 'way out':

  The *analysis of variance*

  short ANOVA
- Basic idea: test all k groups simultaneously

Analysis of variance

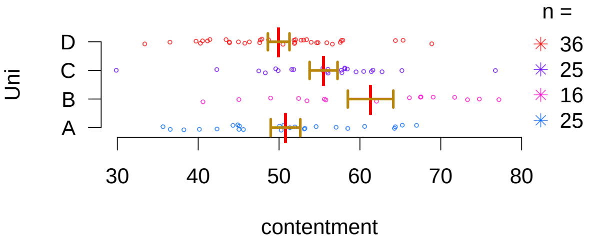


- $k = 4$ groups
- $m = 6$ pairwise t-tests $\rightarrow$ the global α-error is increased
- Another possible 'way out':

  The *analysis of variance*

  short ANOVA
- Basic idea: test all k groups simultaneously
- Null hypothesis: the data all derive from the same normal distribution ($\rightarrow$ particularly no shift)

Analysis of variance



- $k = 4$ groups
- $m = 6$ pairwise t-tests $\rightarrow$ the global α-error is increased
- Another possible 'way out':

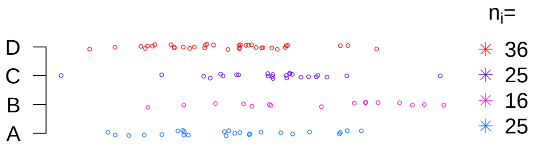
  The *analysis of variance*

  short ANOVA
- Basic idea: test all k groups simultaneously
- Null hypothesis: the data all derive from the same normal distribution ($\rightarrow$ particularly no shift)
- Intuitively in the graphic: this is not very plausible, because the group means are partly far away (multiple sem) from each other

ANOVA - Model

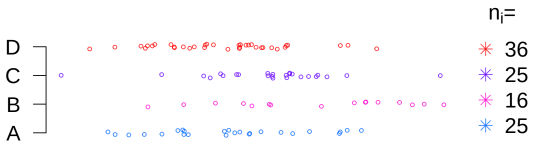
Model:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$



ANOVA - Model

<u>Model</u>:
Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

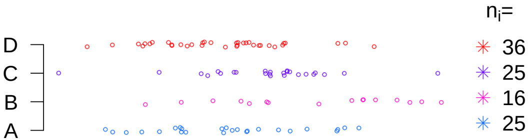


- Notation
  - k groups

ANOVA - Model

Model:

Let $X_{1,1},\ldots,X_{1,n_1},\quad X_{2,1},\ldots,X_{2,n_2},\quad \ldots,\quad X_{k,1},\ldots,X_{k,n_k}$ be independent RVs and for $i=1,\ldots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\ldots,n_i$, with $(\mu_1,\ldots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$

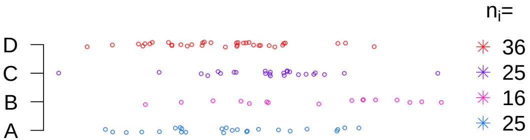


- Notation
  - k groups
  - n_i observations in the i-th group

ANOVA - Model

Model:

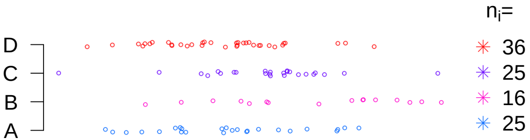
Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$



- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$

<u>Model</u>:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

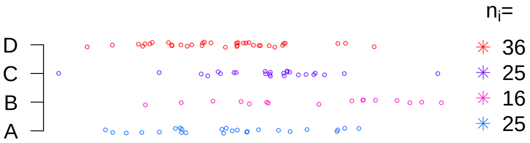


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence

ANOVA - Model

Model:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

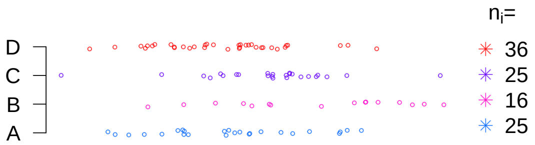


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution

ANOVA - Model

Model:

Let $X_{1,1}, \ldots, X_{1,n_1}, \;\; X_{2,1}, \ldots, X_{2,n_2}, \;\; \ldots, \;\; X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

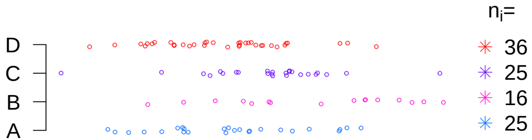


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)

ANOVA - Model

Model:

Let $X_{1,1}, \dots, X_{1,n_1}, \quad X_{2,1}, \dots, X_{2,n_2}, \quad \dots, \quad X_{k,1}, \dots, X_{k,n_k}$ be independent RVs and for $i = 1, \dots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \dots, n_i$, with $(\mu_1, \dots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

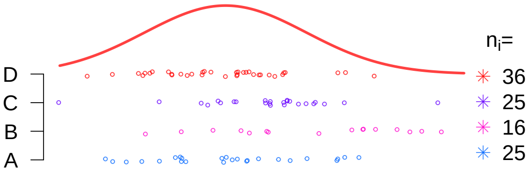


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)
  - the variance σ^2 is constant

ANOVA - Model

Model:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

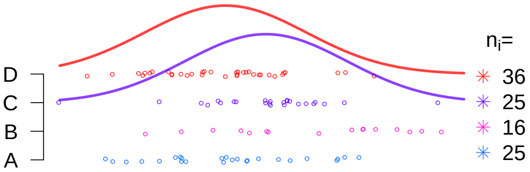


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)
  - the variance σ^2 is constant

ANOVA - Model

<u>Model</u>:

Let $X_{1,1},\ldots,X_{1,n_1},\ \ X_{2,1},\ldots,X_{2,n_2},\ \ \ldots,\ \ X_{k,1},\ldots,X_{k,n_k}$ be independent RVs and for $i=1,\ldots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\ldots,n_i$, with $(\mu_1,\ldots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$

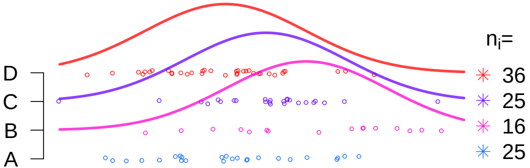


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible ’shift’ between groups)
  - the variance σ^2 is constant

ANOVA - Model

<u>Model</u>:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

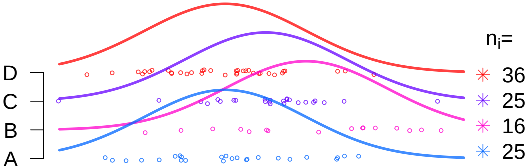


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)
  - the variance σ^2 is constant

ANOVA - Model

<u>Model</u>:

Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

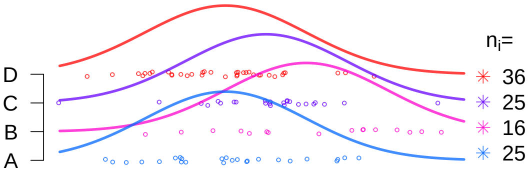


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)
  - the variance σ^2 is constant

ANOVA - Model

<u>Model</u>:

Let $X_{1,1}, \ldots, X_{1,n_1}$, $X_{2,1}, \ldots, X_{2,n_2}$, $\ldots$, $X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$

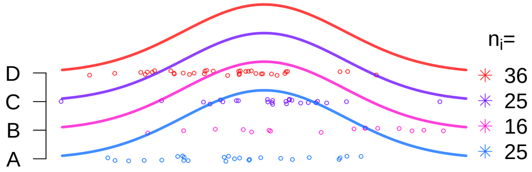


- Notation
  - k groups
  - n_i observations in the i-th group
  - the j-th observation in the i-th group is $X_{i,j}$
- Assumptions
  - independence
  - normal distribution
  - own (unknown) expectation μ_i in the i-th group (i.e., possible 'shift' between groups)
  - the variance σ^2 is constant

Note that for $k = 2$ groups, this model equals the model in Student's two-sample t-test.

ANOVA - Null hypothesis

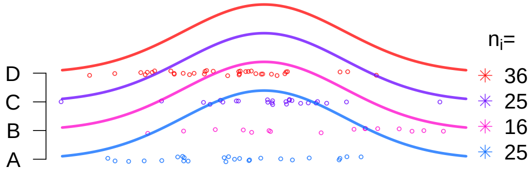
Let $X_{1,1},\dots,X_{1,n_1},\;\; X_{2,1},\dots,X_{2,n_2},\;\; \dots,\;\; X_{k,1},\dots,X_{k,n_k}$ be independent RVs and for $i=1,\dots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\dots,n_i$, with $(\mu_1,\dots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$



- Null hypothesis: $H_0:\mu_1=\mu_2=\cdots=\mu_k$
  the distributions are equal, in particular no shift

ANOVA - Null hypothesis

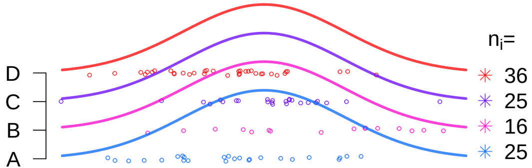
Let $X_{1,1},\ldots,X_{1,n_1},\;\; X_{2,1},\ldots,X_{2,n_2},\;\; \ldots,\;\; X_{k,1},\ldots,X_{k,n_k}$ be independent RVs and for $i = 1,\ldots,k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1,\ldots,n_i$, with $(\mu_1,\ldots,\mu_k,\sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$



- Null hypothesis: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$
  the distributions are equal, in particular no shift
- Alternative: at least two expectations are not equal, $\mu_i \neq \mu_\ell$

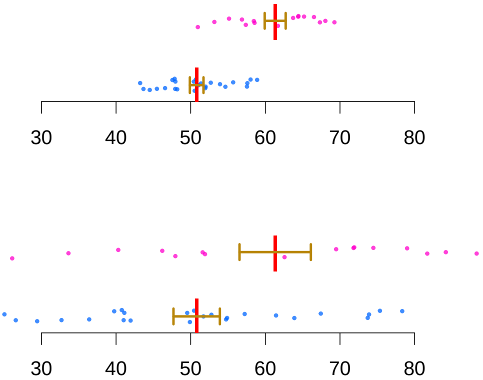
ANOVA - Null hypothesis

Let $X_{1,1}, \dots, X_{1,n_1}, \quad X_{2,1}, \dots, X_{2,n_2}, \quad \dots, \quad X_{k,1}, \dots, X_{k,n_k}$ be independent RVs and for $i = 1, \dots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \dots, n_i$, with $(\mu_1, \dots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$



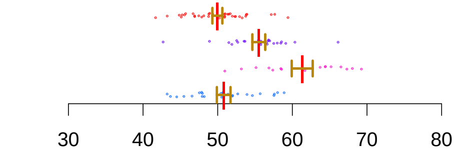
- Null hypothesis: $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$
  the distributions are equal, in particular no shift
- Alternative: at least two expectations are not equal, $\mu_i \neq \mu_\ell$
- How can we test the null hypothesis?

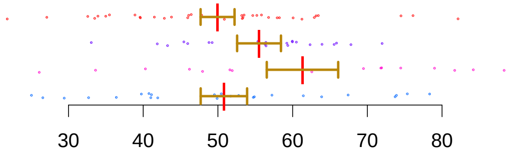
Reminder: *t*-Statistic



The *t*-statistic says:
the discrepancy <u>between</u> the groups is large,
if the variability <u>within</u> the groups is small

Idea of the ANOVA analog



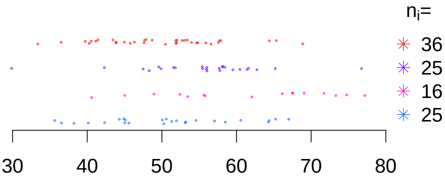


The f-statistic will tell us:
the variability <u>between</u> the groups is large,
if the variability <u>within</u> the groups is small

F-Statistic

Naive:

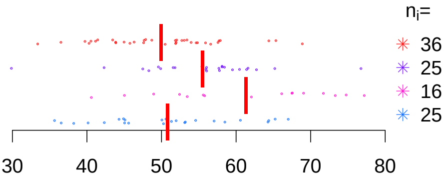
$$f \stackrel{!}{=} \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}$$



F-Statistic

Naive:

$$f \overset{!}{=} \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}$$



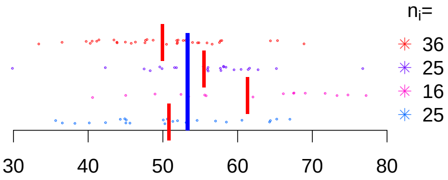
- Need 1.: group means, for $i = 1, \ldots, k$

$$\bar{X}_{i,\cdot} := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$$

F-Statistic

Naive:

$$f \stackrel{!}{=} \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}$$



- Need 1.: group means, for $i = 1, \dots, k$

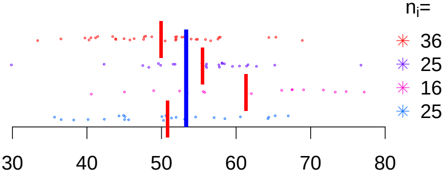
$$\bar{X}_{i,\cdot} := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$$

- Need 2.: global mean over all groups

$$\bar{X} := \frac{1}{n} \sum_{i,j} X_{i,j}$$

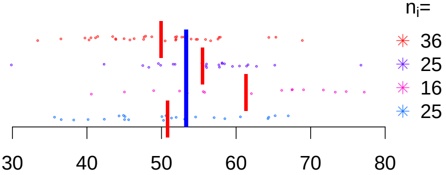
with $n = n_1 + \dots + n_k$ total number of observations

F-Statistic (google: R. A. Fisher)



- $\bar{X}_{i,\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n} \sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$

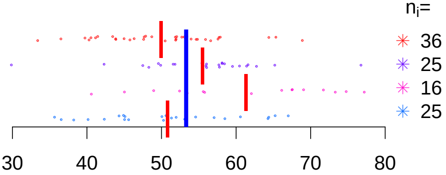
F-Statistic (google: R. A. Fisher)



- $\bar{X}_{i,\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n} \sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1} \sum_{i=1}^{k} n_i (\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

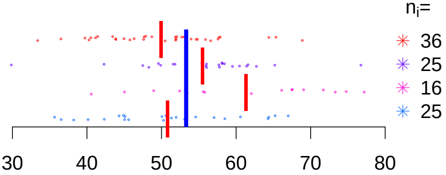
F-Statistic (google: R. A. Fisher)



- $\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n}\sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

F-Statistic (google: R. A. Fisher)

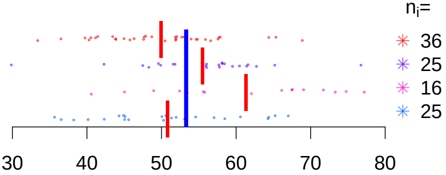


- $\bar{X}_{i,\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n} \sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1} \sum_{i=1}^{k} n_i (\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring

F-Statistic (google: R. A. Fisher)

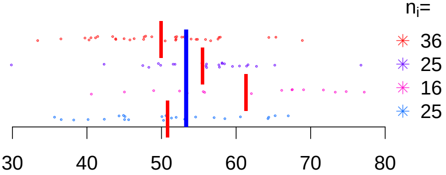


- $\textcolor{red}{\bar{X}_{i,\cdot}} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ and $\textcolor{blue}{\bar{X}} = \frac{1}{n} \sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1} \sum_{i=1}^{k} n_i (\textcolor{red}{\bar{X}_{i,\cdot}} - \textcolor{blue}{\bar{X}})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \textcolor{red}{\bar{X}_{i,\cdot}})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring
- If the evaluation of the data f is very large, this speaks against H_0

F-Statistic (google: R. A. Fisher)

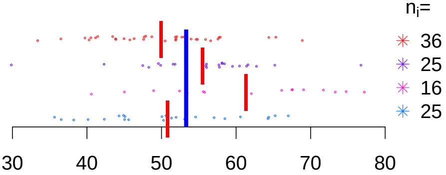


- $\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n}\sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring
- If the evaluation of the data f is very large, this speaks against H_0
  - Because then the numerator is large compared to the denominator

F-Statistic (google: R. A. Fisher)

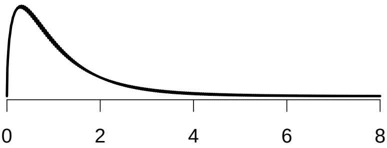


- $\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n}\sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring
- If the evaluation of the data f is very large, this speaks against H_0
  - Because then the numerator is large compared to the denominator
  - The test is always right-sided (only large values speak against H_0)

F-Statistic (google: R. A. Fisher)

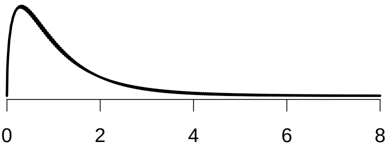


- $\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n}\sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring
- If the evaluation of the data f is very large, this speaks against H_0
  - Because then the numerator is large compared to the denominator
  - The test is always right-sided (only large values speak against H_0)
- The distribution of F is known under H_0: the Fisher-distribution with $(k-1)$-numerator- and $(n-k)$-denominator- degrees of freedom, short $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$

F-Statistic (google: R. A. Fisher)



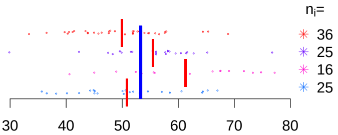
- $\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ and $\bar{X} = \frac{1}{n}\sum_{i,j} X_{i,j}$ with $n = \sum_{i=1}^{k} n_i$
- The F-Statistic is

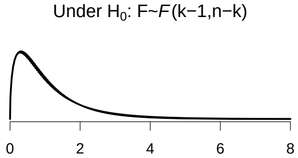
$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2} = \frac{\text{'variability between the groups'}}{\text{'variability within the groups'}}}$$

- $F \geqslant 0$, due to squaring
- If the evaluation of the data f is very large, this speaks against H_0
  - Because then the numerator is large compared to the denominator
  - The test is always right-sided (only large values speak against H_0)
- The distribution of F is known under H_0: the Fisher-distribution with $(k-1)$-numerator- and $(n-k)$-denominator- degrees of freedom, short $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$ (R knows it well, `rf(...)`, `pf(...)` etc.)

For more details on the $\mathcal{F}(k-1, n-k)$-distribution and the derivation of the distribution of F see e.g., Messer, M. and Schneider, G. *Statistik: Theorie und Praxis im Dialog*, Springer Berlin

F-test



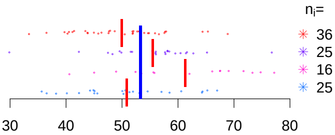


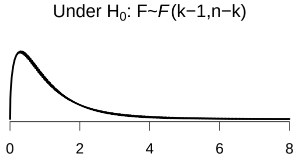
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\color{red}{\bar{X}_{i,\cdot}} - \color{blue}{\bar{X}})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \color{red}{\bar{X}_{i,\cdot}})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$

F-test



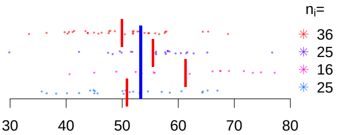


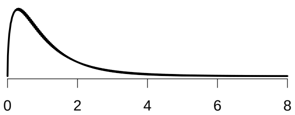
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$

F-test



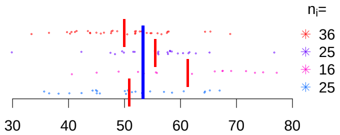


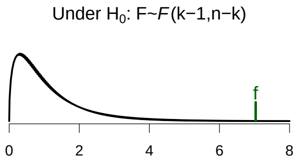
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
  - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$

F-test



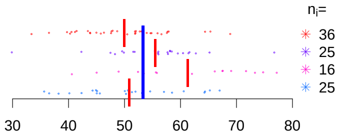


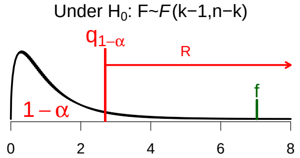
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
  - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$
  - thus: $f \approx 7.0$

F-test



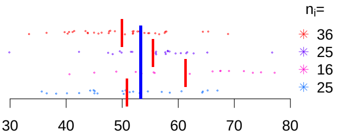


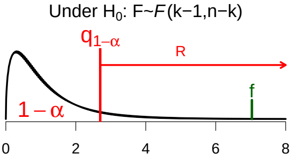
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1} \sum_{i=1}^{k} n_i (\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k} \sum_{i,j} (X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
  - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$
  - thus: $f \approx 7.0$
- For $\alpha = 5\%$ it is $q_{1-\alpha} \approx 2.69$ (95%-quantile of $\mathscr{F}(3, 98)$), thus $R \approx [2.69, \infty)$

F-test



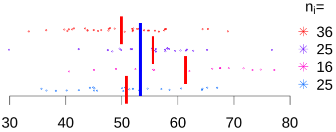


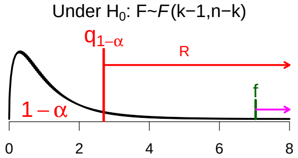
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
  - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$
  - thus: $f \approx 7.0$
- For $\alpha = 5\%$ it is $q_{1-\alpha} \approx 2.69$ (95%-quantile of $\mathscr{F}(3, 98)$), thus $R \approx [2.69, \infty)$
- $f \in R$, and hence reject H_0 on the 5%-level

F-test



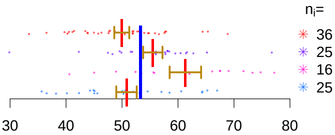


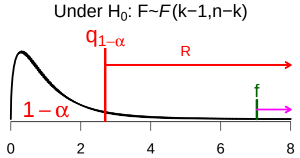
- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
  - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
  - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$
  - thus: $f \approx 7.0$
- For $\alpha = 5\%$ it is $q_{1-\alpha} \approx 2.69$ (95%-quantile of $\mathscr{F}(3, 98)$), thus $R \approx [2.69, \infty)$
- $f \in R$, and hence reject H_0 on the 5%-level
- $p = \mathbb{P}_{H_0}(F \geqslant f) \approx 0.00025 = 1/4000$. Only in about one of 4000 cases we observe a discrepancy, which is at least as extreme as in the data, if the null hypothesis holds true.

F-test





- Null hypothesis $H_0 : \mu_1 = \cdots = \mu_k$
- F-statistic

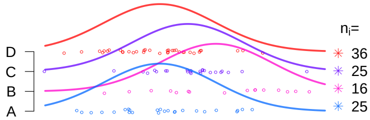
$$\boxed{F := \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j} - \bar{X}_{i,\cdot})^2}}$$

- $F \overset{H_0}{\sim} \mathscr{F}(k-1, n-k)$
- Evaluation of the data yields f
    - $k = 4$ groups, $n_1 = 25$, $n_2 = 16$, $n_3 = 25$, $n_4 = 36$, thus $n = 102$
    - means: $\bar{x}_{1,\cdot} \approx 50.8$, $\bar{x}_{2,\cdot} \approx 61.3$, $\bar{x}_{3,\cdot} \approx 55.5$, $\bar{x}_{4,\cdot} \approx 50.0$ and $\bar{x} \approx 53.3$
    - thus: $f \approx 7.0$
- For $\alpha = 5\%$ it is $q_{1-\alpha} \approx 2.69$ (95%-quantile of $\mathscr{F}(3, 98)$), thus $R \approx [2.69, \infty)$
- $f \in R$, and hence reject H_0 on the 5%-level
- $p = \mathbb{P}_{H_0}(F \geqslant f) \approx 0.00025 = 1/4000$. Only in about one of 4000 cases we observe a discrepancy, which is at least as extreme as in the data, if the null hypothesis holds true.
- Result plausible, given the variability (sem) of the means

ANOVA - Summary

Model:

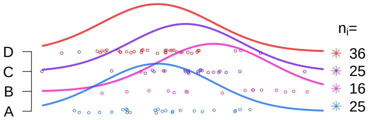
- Let $X_{1,1},\ldots,X_{1,n_1},\ \ X_{2,1},\ldots,X_{2,n_2},\ \ \ldots,\ \ X_{k,1},\ldots,X_{k,n_k}$ be independent RVs and for $i=1,\ldots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\ldots,n_i$, with $(\mu_1,\ldots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$



ANOVA - Summary

Model:

- Let $X_{1,1}, \ldots, X_{1,n_1}, \quad X_{2,1}, \ldots, X_{2,n_2}, \quad \ldots, \quad X_{k,1}, \ldots, X_{k,n_k}$ be independent RVs and for $i = 1, \ldots, k$ let $X_{i,j} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$, with $(\mu_1, \ldots, \mu_k, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}^+$
- Thus, k groups and $n = \sum_{i=1}^k n_i$ observations in total



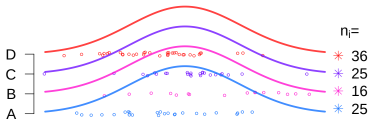
ANOVA - Summary

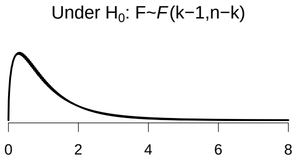
Model:

- Let $X_{1,1},\dots,X_{1,n_1},\ \ X_{2,1},\dots,X_{2,n_2},\ \ \dots,\ \ X_{k,1},\dots,X_{k,n_k}$ be independent RVs and for $i=1,\dots,k$ let $X_{i,j}\sim N(\mu_i,\sigma^2)$ for $j=1,\dots,n_i$, with $(\mu_1,\dots,\mu_k,\sigma^2)\in\mathbb{R}^k\times\mathbb{R}^+$
- Thus, k groups and $n=\sum_{i=1}^k n_i$ observations in total

Under $H_0:\mu_1=\cdots=\mu_k$ it holds

$$F := \frac{\frac{1}{k-1}\sum_{i=1}^k n_i(\bar{X}_{i,\cdot}-\bar{X})^2}{\frac{1}{n-k}\sum_{i,j}(X_{i,j}-\bar{X}_{i,\cdot})^2}\sim\mathscr{F}(k-1,n-k)$$





ANOVA in R

```
# Sizes of groups as vector (k groups)
ni       <- c(n1,n2,...,nk)
# Data as vector, all from group 1, then all from group 2 etc..
x        <- c(...)
# Groups as factor variable
gr       <- factor(rep(1:k,ni))
# Perform ANOVA (F-test)
anova(aov(x~gr))
# Output
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value   Pr(>F)
gr         3 1709.8  569.92  7.0332 0.000248 ***
Residuals 98 7941.2   81.03
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
```

- 1. column: degrees of freedom (first row numerator, second row denominator)
- 3. column: numerator and denominator of the f-statistic
- 4. column: f-statistic
- 5. column: p-value   (2.column $=$ 1.column $\cdot$ 3.column)

In practice

Procedure in practice often:

- First apply the ANOVA...

In practice

Procedure in practice often:

- First apply the ANOVA...
- ...if the null hypothesis is rejected, then perform post-hoc *t*-tests (and correct the *p*-values e.g., using Bonferroni correction)

In practice

Procedure in practice often:

- First apply the ANOVA...
- ...if the null hypothesis is rejected, then perform post-hoc t-tests (and correct the p-values e.g., using Bonferroni correction)

Rethink of whether the data comply with the model assumptions of the ANOVA:

- Independence and own expectation per group
  $\rightarrow$ makes sense due to the practical setup

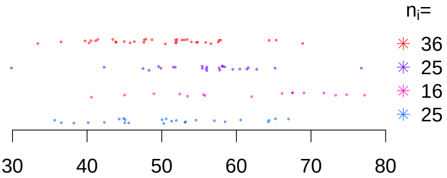
In practice

Procedure in practice often:

- First apply the ANOVA...
- ...if the null hypothesis is rejected, then perform post-hoc *t*-tests (and correct the *p*-values e.g., using Bonferroni correction)

Rethink of whether the data comply with the model assumptions of the ANOVA:

- Independence and own expectation per group
  $\rightarrow$ makes sense due to the practical setup
- normal distribution and constant variance (between the groups)
  $\rightarrow$ plausible when looking at the data



Thank you!