Biological Data Analysis

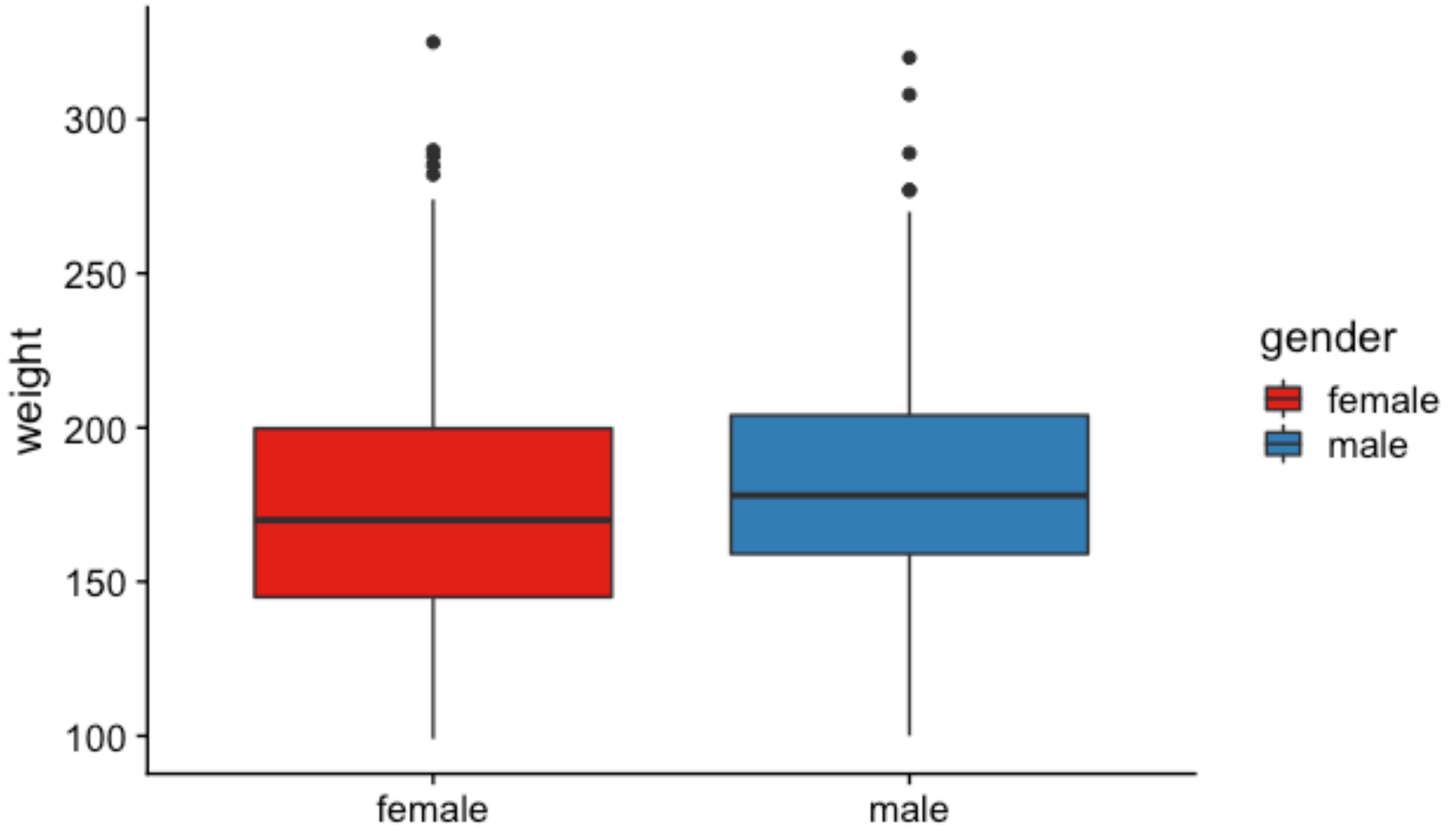
Carl Herrmann
IPMB - Universität Heidelberg

Institut für Pharmazie und Molekulare Biotechnologie

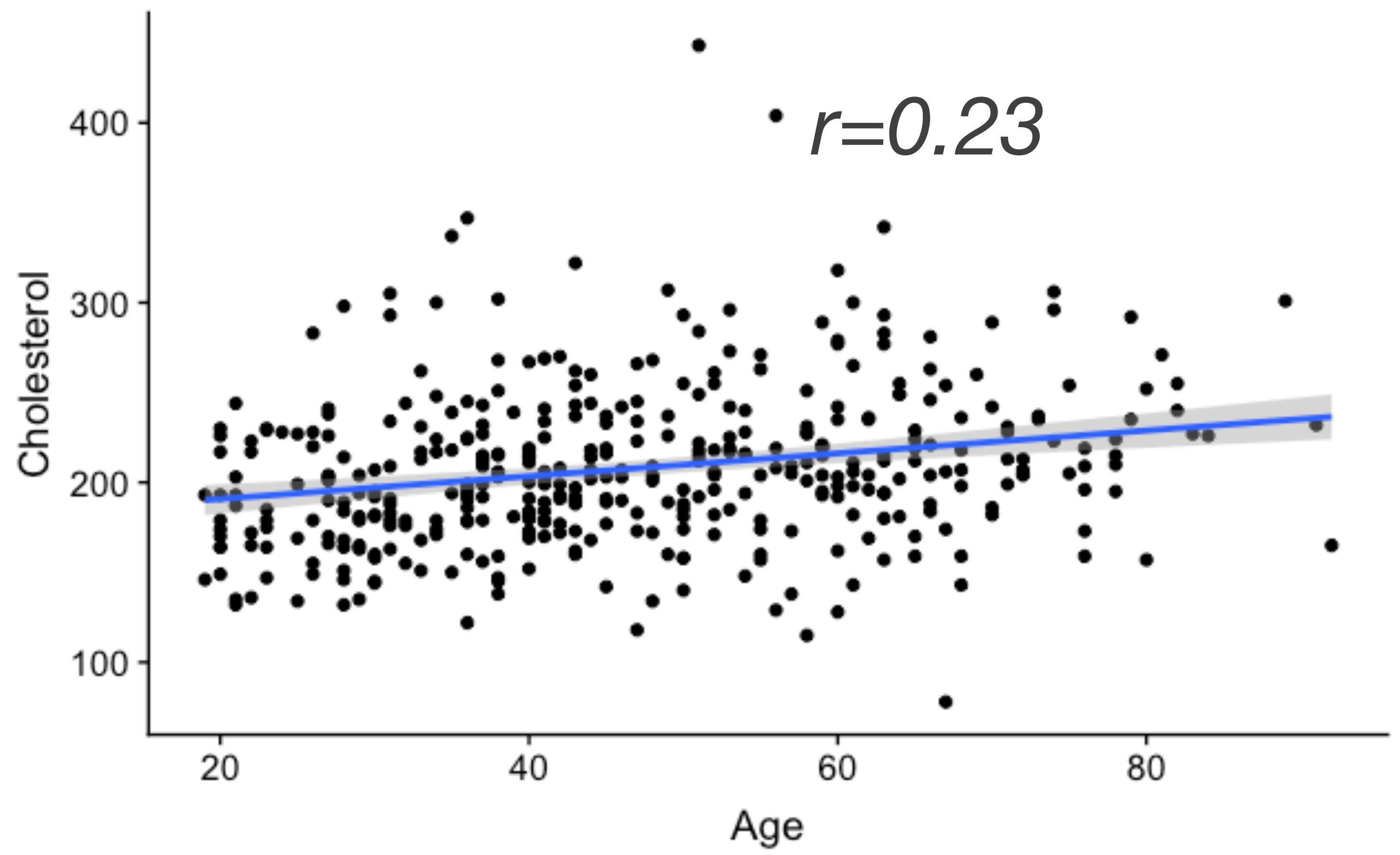
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

7. Hypothesis testing

Are observations significant?



170 178



- For the cohort, we observe:
  - a difference in man/women weights
  - a non-zero correlation between age and cholesterol
- But:
  - would we observe this in another cohort??
  - Does this hold for the entire (unknown) population?

    → ***is this difference/correlation significant?***

Hypothesis testing: what do we need?

- **Question** that we want to investigate:
  - *is there a **GENERAL** weight difference between men/women?*
  - *is there a **GENERAL** non-zero correlation between age/cholesterol?*
- ***Null-hypothesis (H_0)***: this is the “no-effect” Hypothesis
  - no difference between the **expectations of the random variables** X_m=weight of Men and X_w = weights of Women
  - no correlation between the random variables X_{chol} and X_{age}: cor(Xchol,Xage)=0
- ***Alternative hypothesis (H_1):***
  - $E(X_m) \neq E(X_w)$
  - $cor(X_{age}, X_{chol}) \neq 0$
- **Test-statistics**: numerical value that can be computed from the data, with known distribution under H_0

Example

- **Study**: effect of fertilizer F1 on plant growth
  - no-fertilizer: h = 1.5m
  - fertilizer on n = 10 samples:

    x = `{1.47,1.62,1.51,1.61,1.27,1.51,1.55,1.49,1.44,1.5}`
- **Random variable:** plant height X after treatment with F1
- **Question**: *does the treatment with fertilizer* ***enhance*** *plant growth?*
- **Hypothesis**:
  - H0 : no $\quad E(X) \leq h = 1.5m$
  - H1: yes $\quad E(X) > h = 1.5m$
- Effect size: $\bar{x} - h = -0.003$
- size of random effect: $s/\sqrt{n} = 0.031$

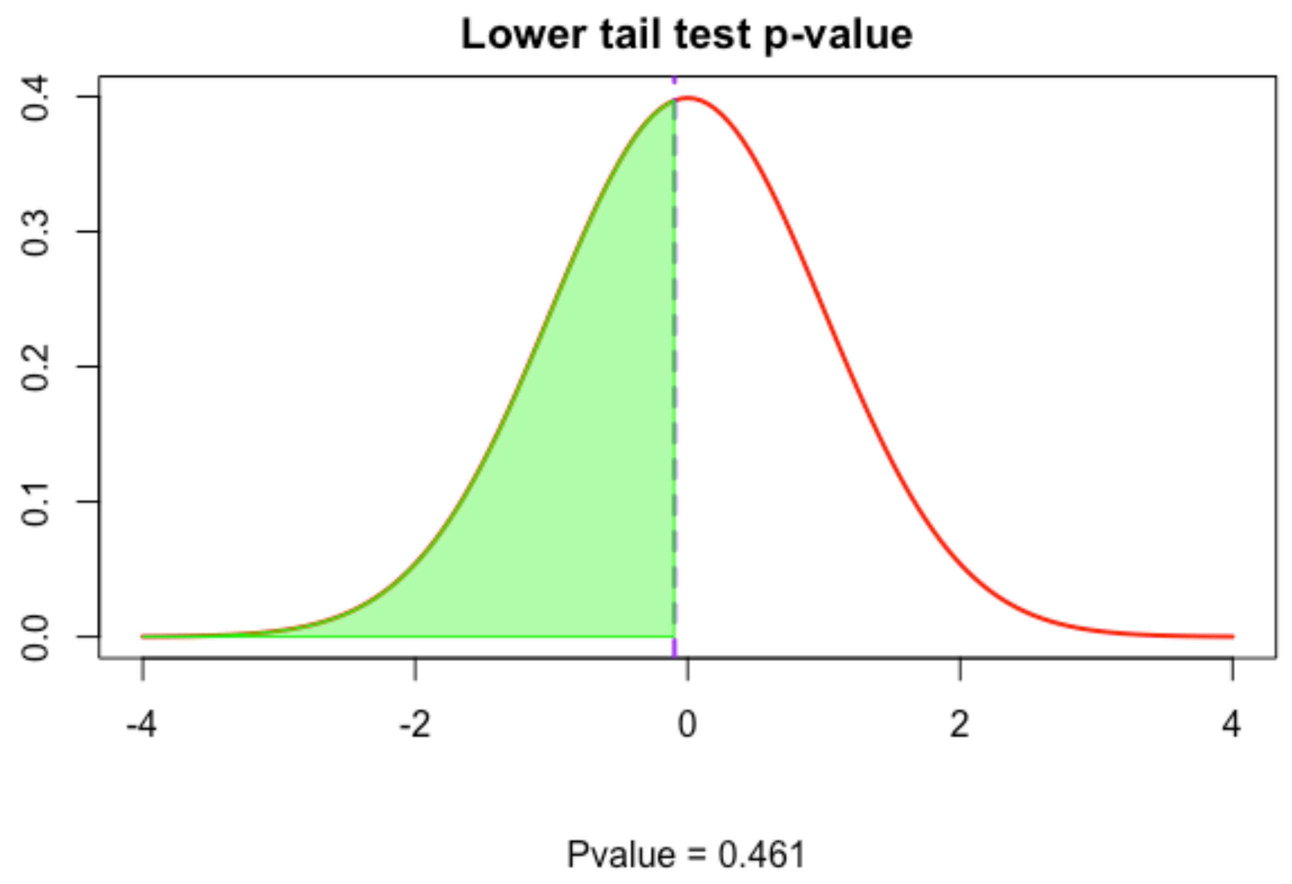
$$t = \frac{\bar{x} - h}{s/\sqrt{n}} = -0.09$$

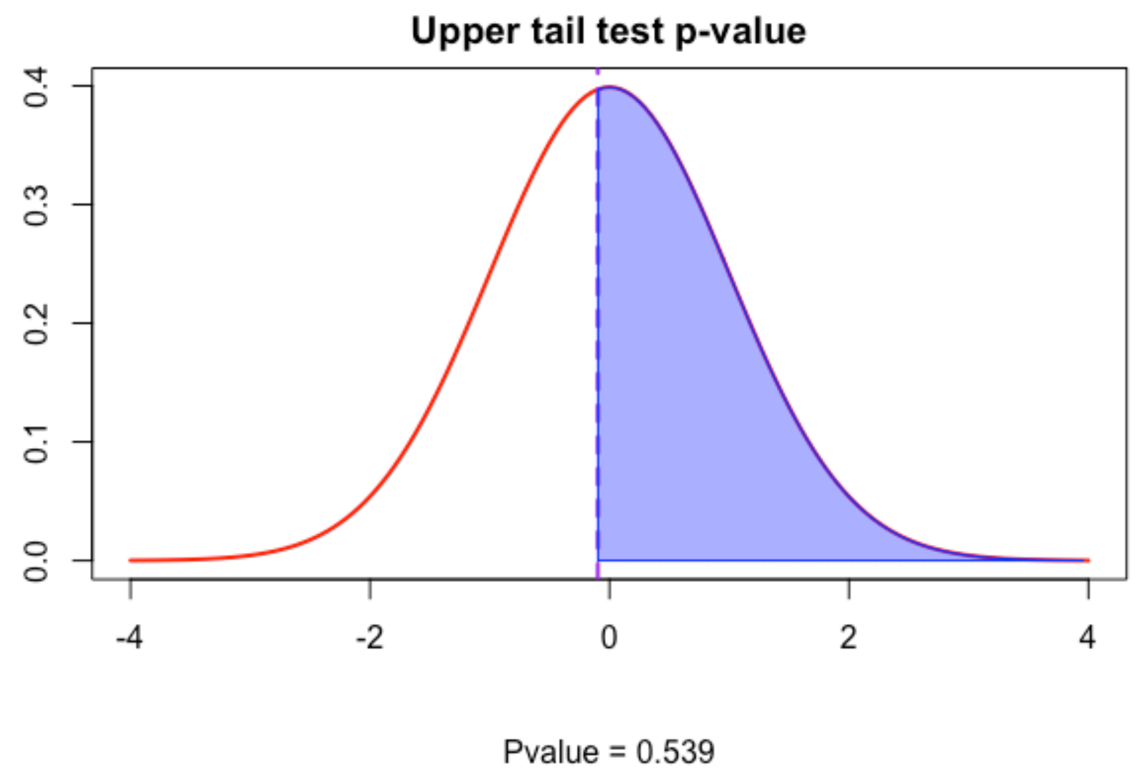
s = standard deviation of sample

- What are typical values of t **under the H_0 hypothesis**?

Example

- Distribution of t under the H_0 hypothesis
- Vertical line = observed value of test statistics t
- Green = probability to observe under H_0 a lower value of t
- Blue = probability to observe under H_0 a larger value of t
- Here: Blue = 53.9% of total area





Conclusion: if H0 (= no effect) is true, there is a **53.9% probability** to observe a value of t larger or equal to the one observed

→ not unlikely, hence no reason to distrust H0 (= no effect)

Example

- **Study**: effect of fertilizer F2 on plant growth
  - no-fertilizer: h = 1.5m
  - fertilizer on n = 10 samples:

    `x = {1.47,1.62,1.61,1.61,1.47,1.51,1.55,1.59,1.64,1.5}`
- **Random variable:** plant height X after treatment with F2
- **Question**: *does the treatment with fertilizer **enhance** plant growth?*
- **Hypothesis**:
  - H0 : no $E(X) \leq h = 1.5m$
  - H1: yes $E(X) > h = 1.5m$
- Effect size: $\bar{x} - h = 0.057$
- size of random effect: $s/\sqrt{n} = 0.02$

$$\left.\begin{array}{l} \bar{x} - h = 0.057 \\ s/\sqrt{n} = 0.02 \end{array}\right\} \quad t = \frac{\bar{x} - h}{s/\sqrt{n}} = 2.77$$

s = standard deviation of sample

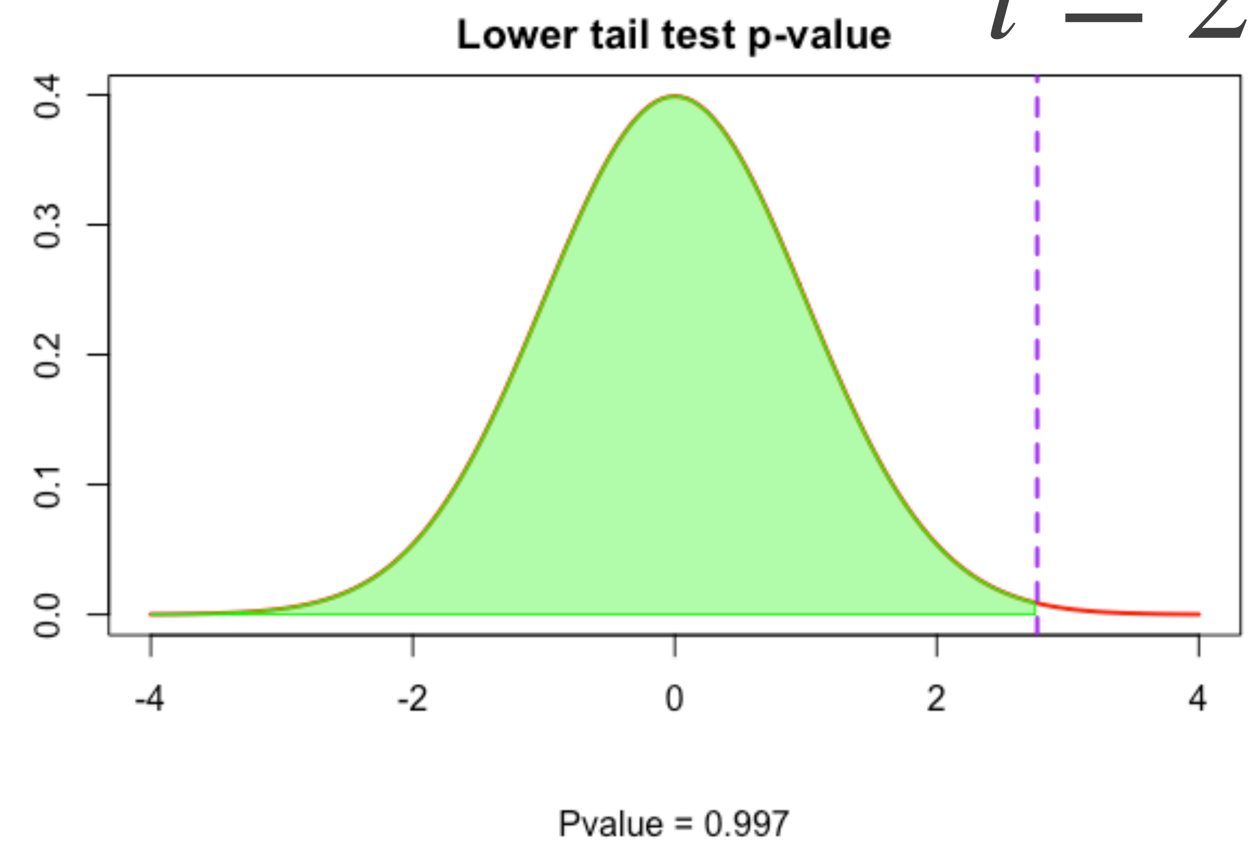
- What are typical values of t **under the H_0 hypothesis**?

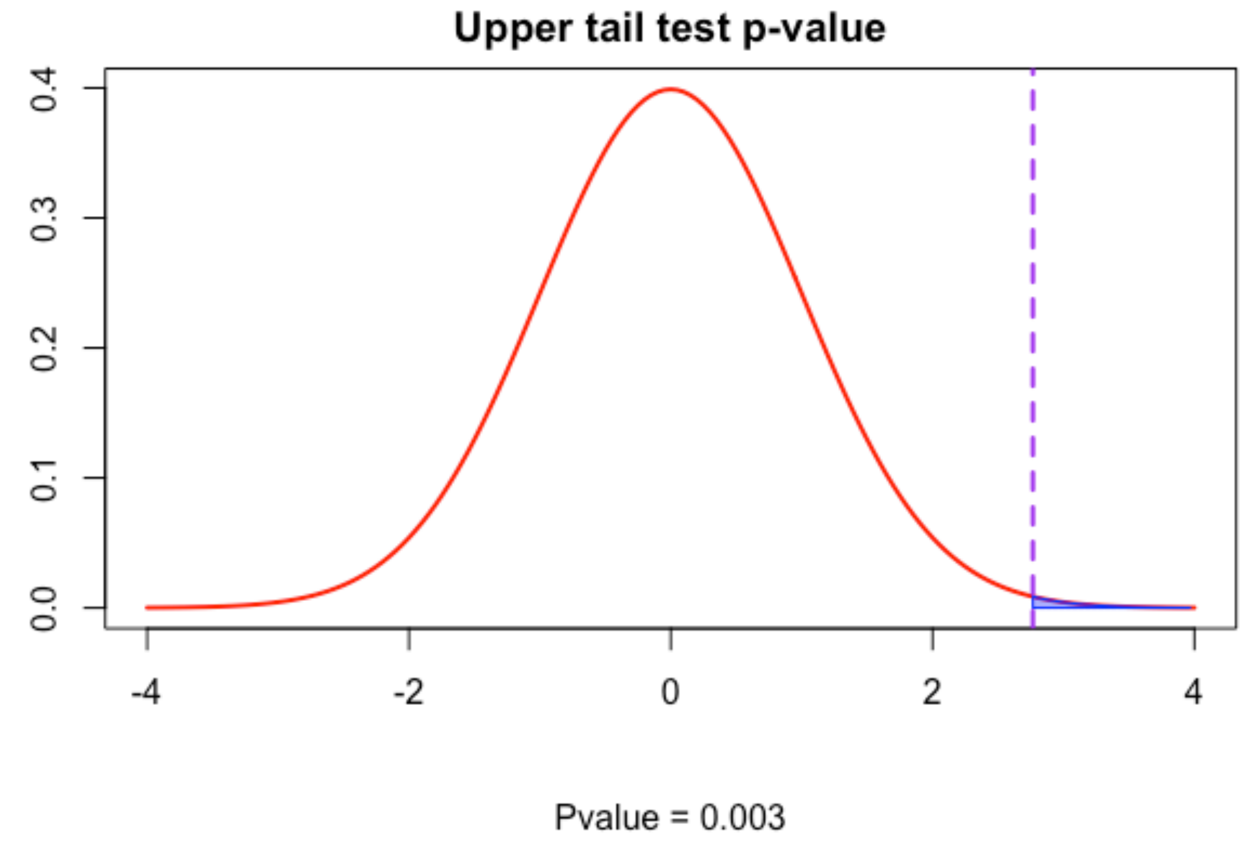
Example

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Distribution of t under the H_0 hypothesis
- Vertical line = observed value of test statistics t
- Green = probability to observe under H_0 a lower value of
- Blue = probability to observe under H_0 a larger value of t
- Here: Blue = 0.3% of total area

$$t = 2.77$$





Conclusion: if H0 (= no effect) is true, there is a **0.3% probability** to observe a value of t larger or equal to the one observed
→ ***very unlikely, H0 is probably not true and should be rejected***

Example

- **Study**: effect of fertilizer F3 on plant growth
  - no-fertilizer: $h = 1.5$m
  - fertilizer on $n = 10$ samples:

    `x = {1.47,1.45,1.31,1.41,1.47,1.51,1.55,1.39,1.44,1.5}`
- **Question**: *does the treatment with fertilizer* ***enhance*** *plant growth?*
- **Hypothesis**:
  - H0 : no, $E(X) \leq h = 1.5m$
  - H1: yes $E(X) > h = 1.5m$
- Effect size: $\bar{x} - h$
- size of random effect: $s/\sqrt{n}$

$$t = \frac{\bar{x} - h}{s/\sqrt{n}} = -2.32$$

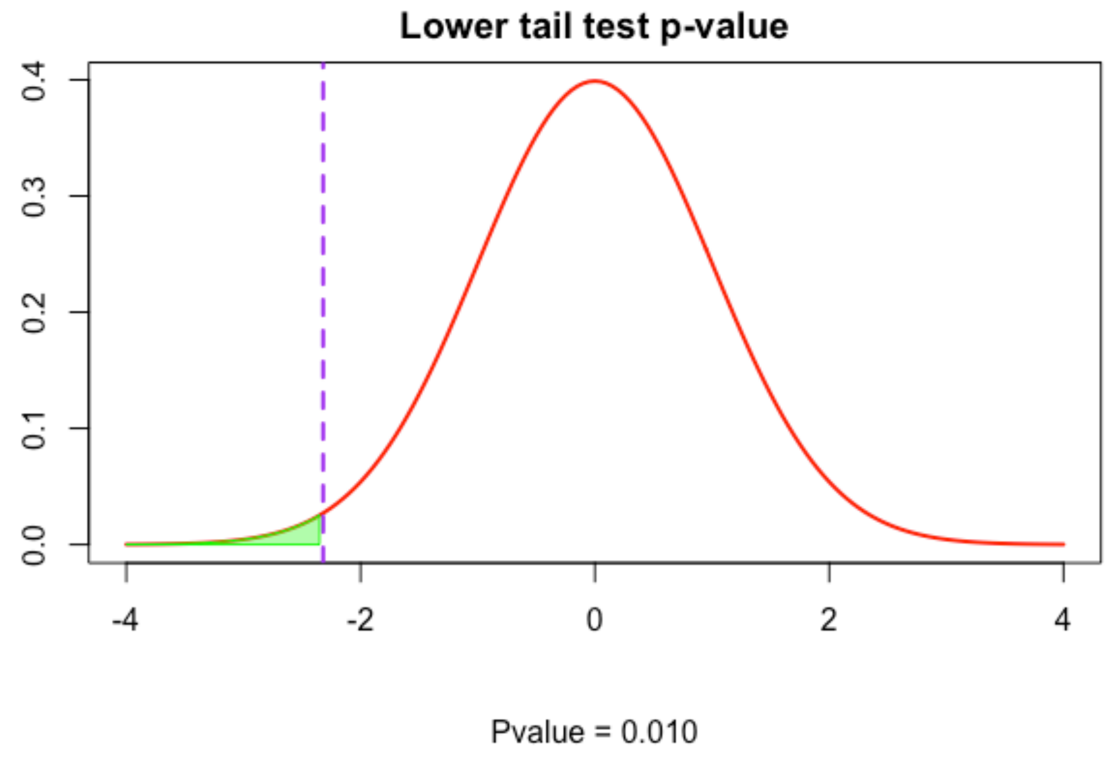
s = standard deviation of sample

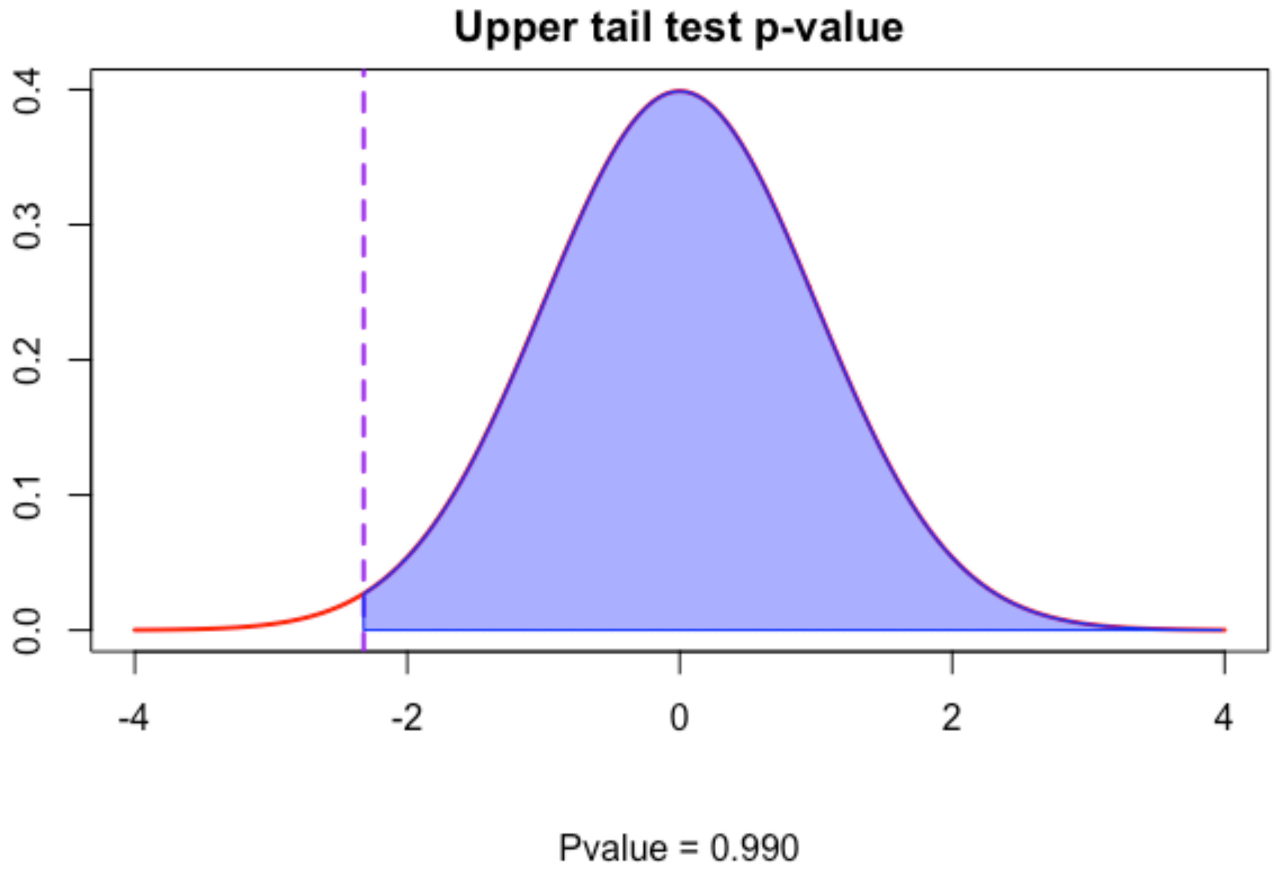
- What are typical values of t **under the H_0 hypothesis**?

Example

$$t = -2.32$$

- Distribution of t under the H_0 hypothesis
- Vertical line = observed value of test statistics t
- Green = probability to observe under H_0 a lower value of t
- Blue = probability to observe under H_0 a larger value of t
- Here: Blue = 99% of total area



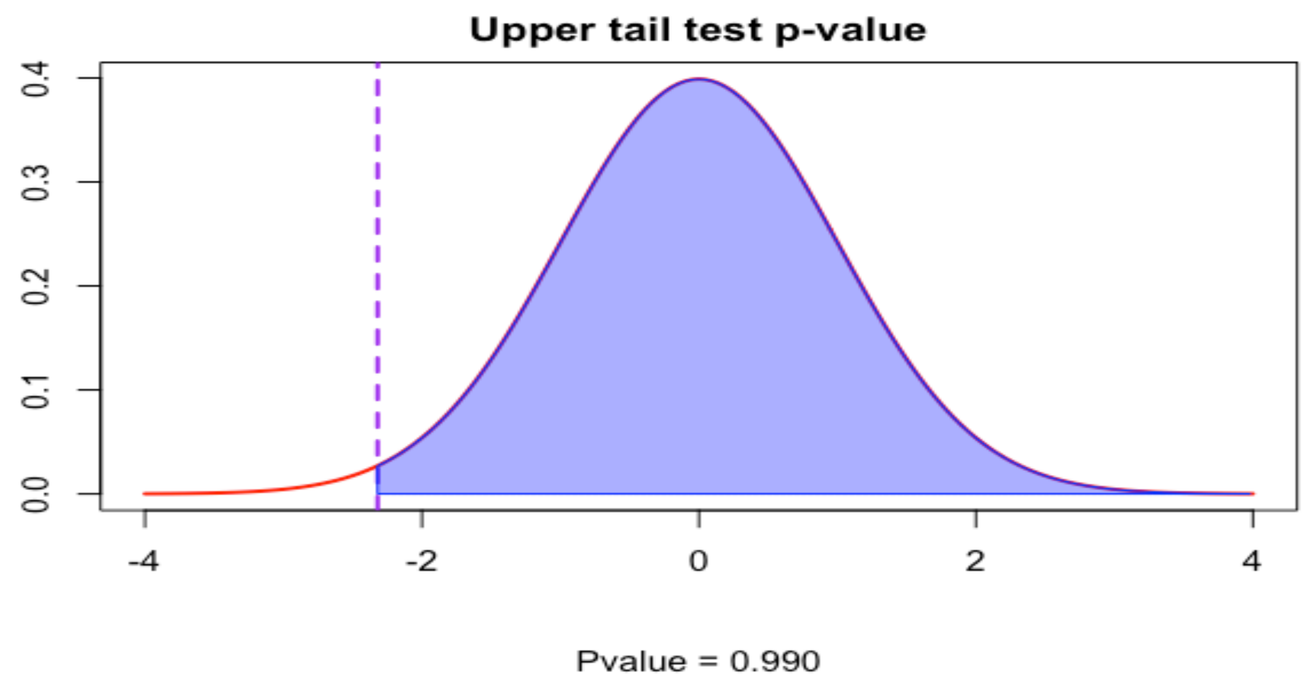


Conclusion: if H_0 (= no effect) is true, there is a **99% probability** to observe a value of t larger or equal to the one observed
→ ***very likely, H_0 cannot be rejected…***

What was the question again?

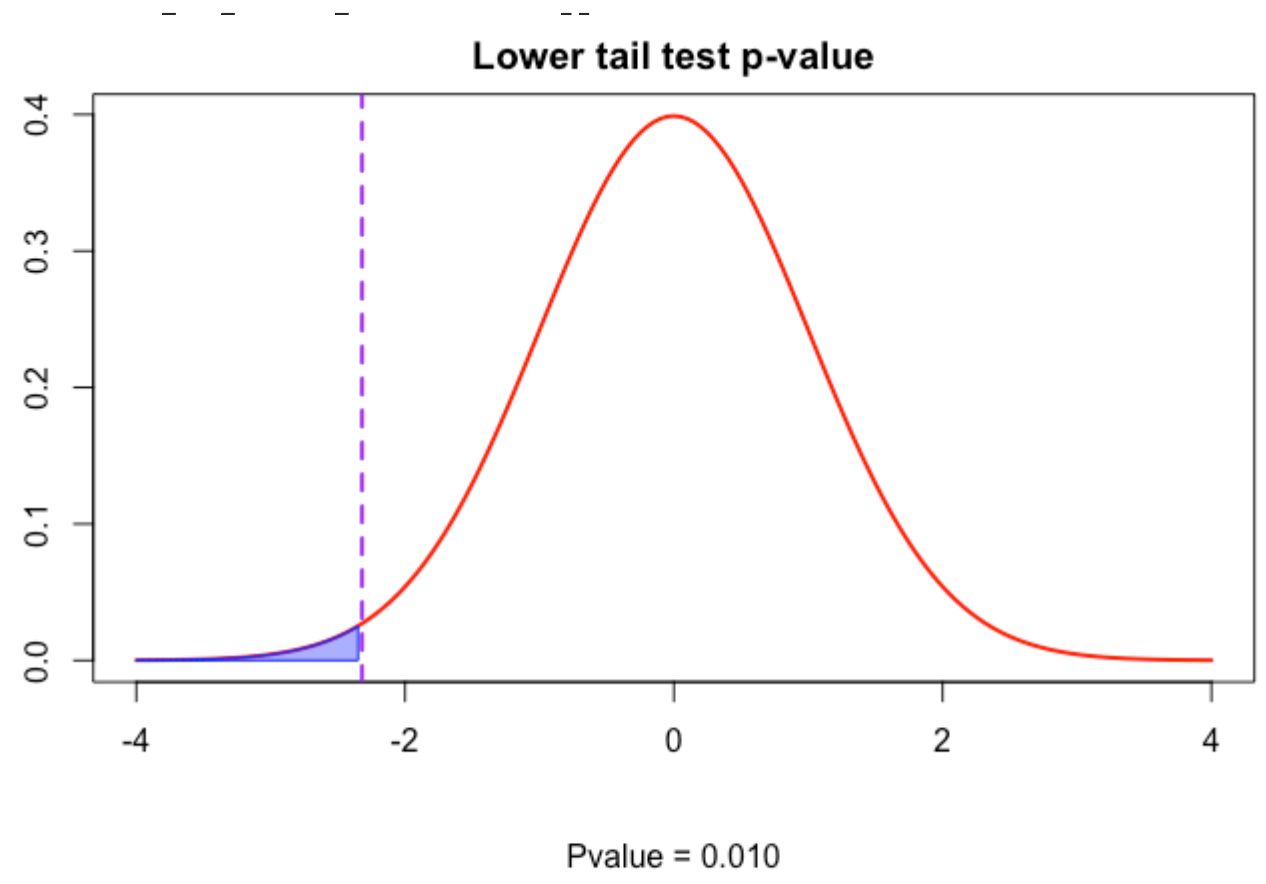
IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

- Question 1:
  does the treatment with fertilizer ***enhance*** *plant growt*
  (→ expected direction of effect is implicit: ***"upper tail***
  H0: no! H1: yes!



blue area = 99%: H0 cannot be rejected!

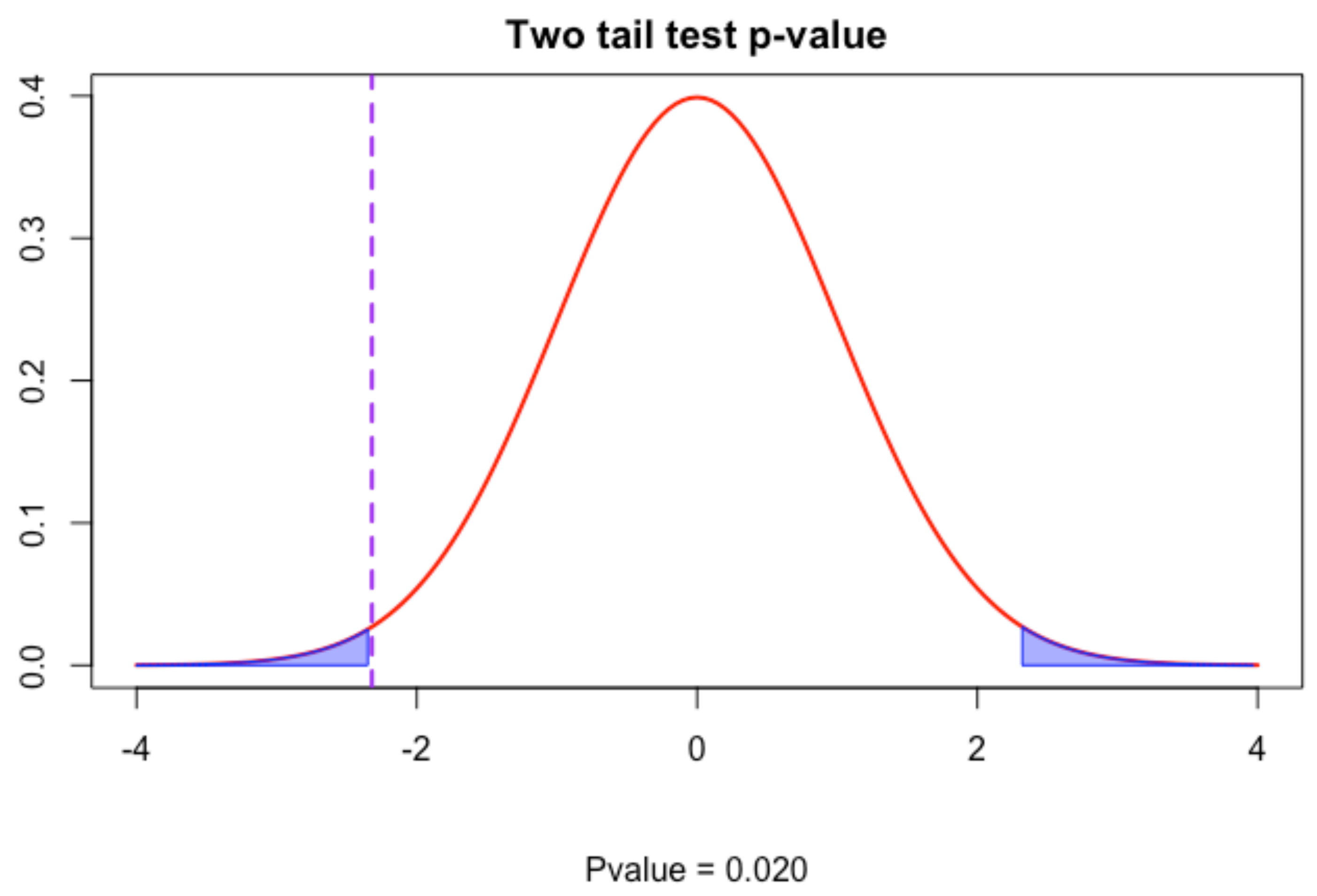
- Question 2:
  does the treatment with fertilizer ***reduce*** *plant growth?*
  (→ expected direction of effect is implicit: ***"lower tail o***
  H0: no! H1: yes!



blue area = 1%: H0 very unlikely

What was the question again?

- Question 3:
  *does the treatment with fertilizer **has an effect** on plai*
  *(→ no direction implicit: **“two-sided test”**)*
  H0: no! H1: yes!



blue area = 2%: H0 very unlikely

P-value

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

the p-value is the **probability** of obtaining a

- larger (one-sided upper tail)
- smaller (one-sided lower tail)
- more extreme (two-sided or two tailed)

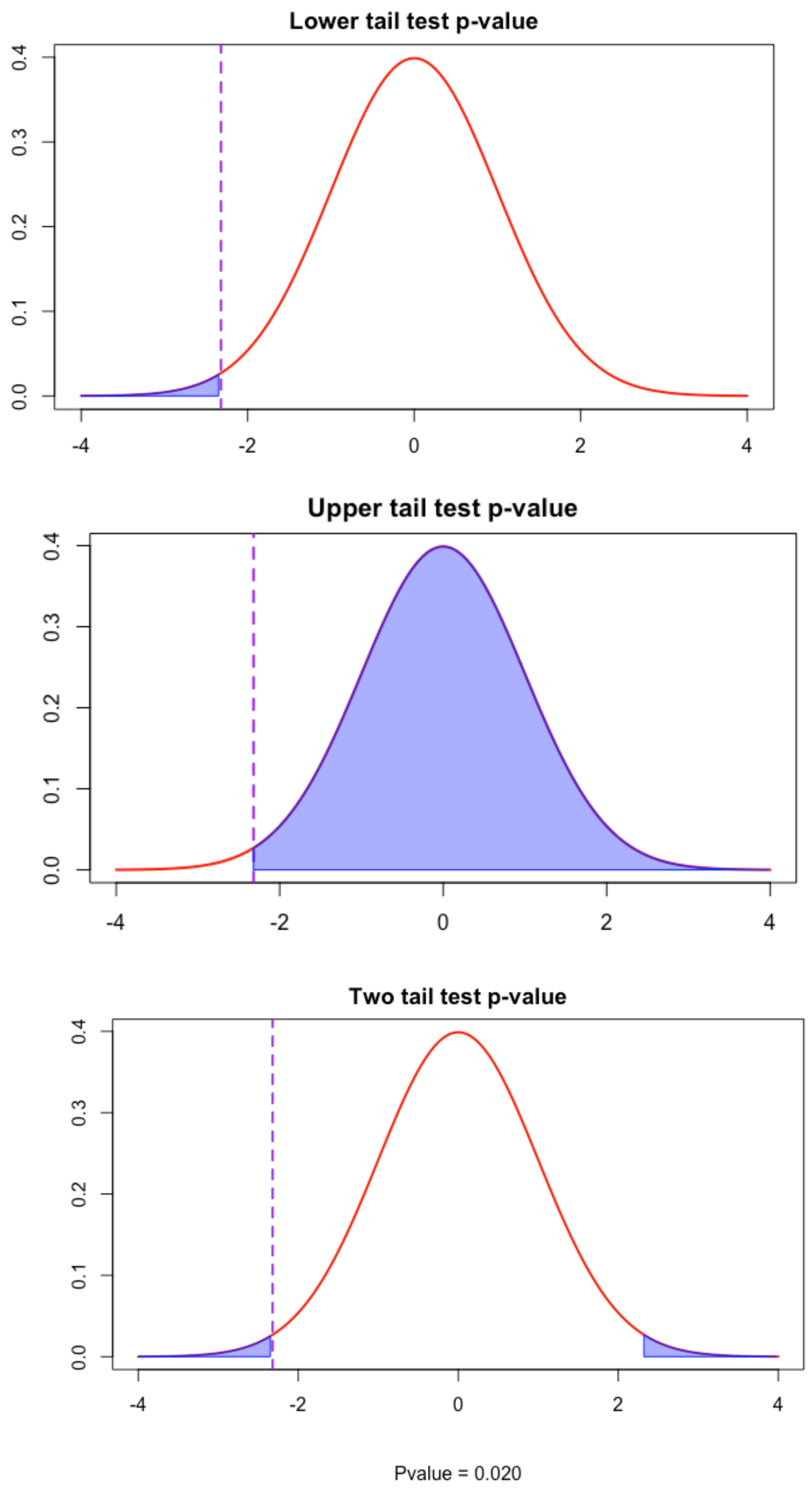
value of the test statistics **if H_0 is valid!**

The p-value represents the area under the H0 curve

- above observed value (one-sided upper tail)
- below observed value (one-sided lower tail)
- more extreme than pbserved value (two-sided or two tailed)

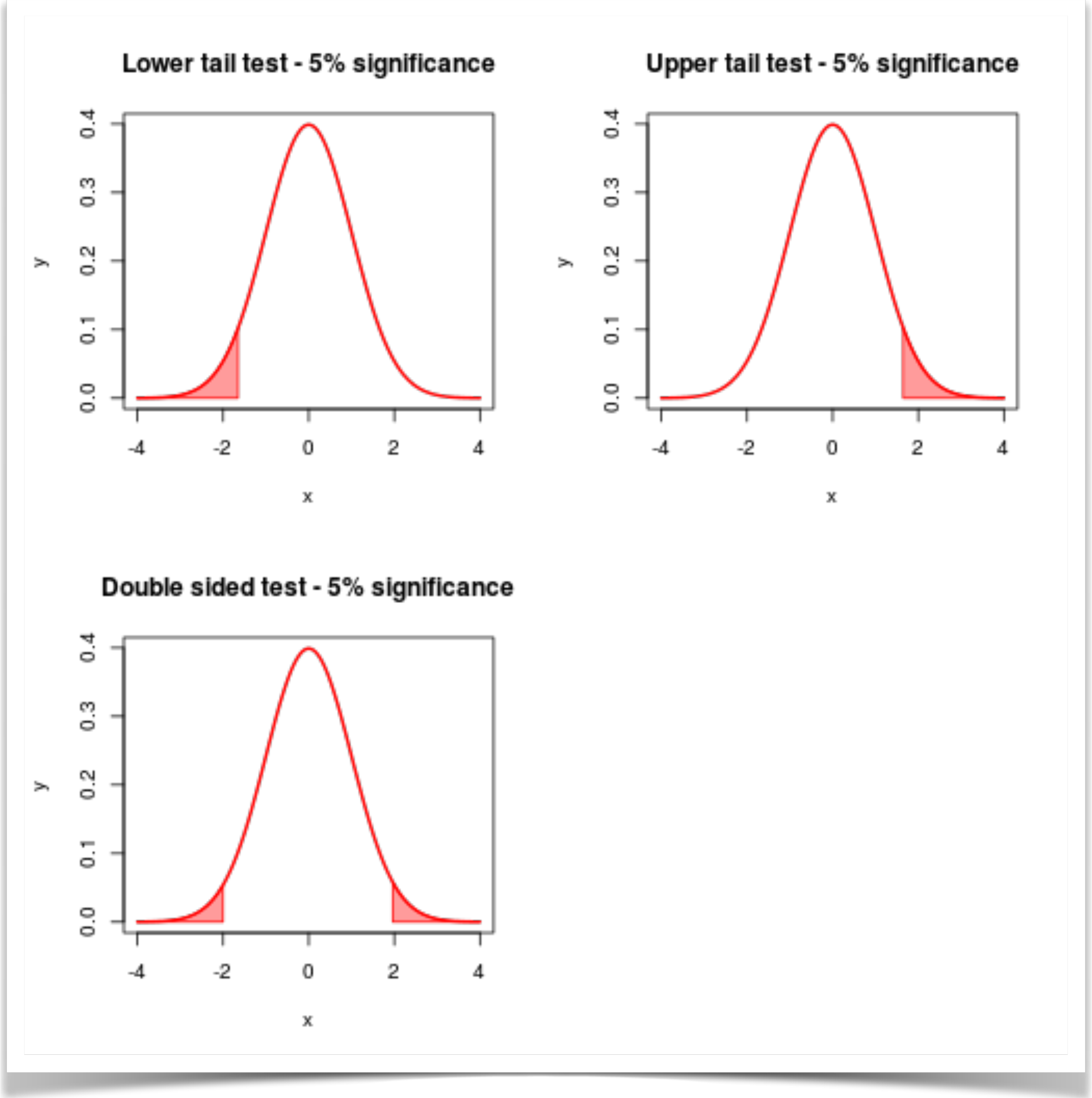
The probability of the two sided test is **twice** the smallest probability of the upper-tail or lower-tail test

$$p_{2sided} = 2\ min(p_{lower-tail}, p_{upper-tail})$$



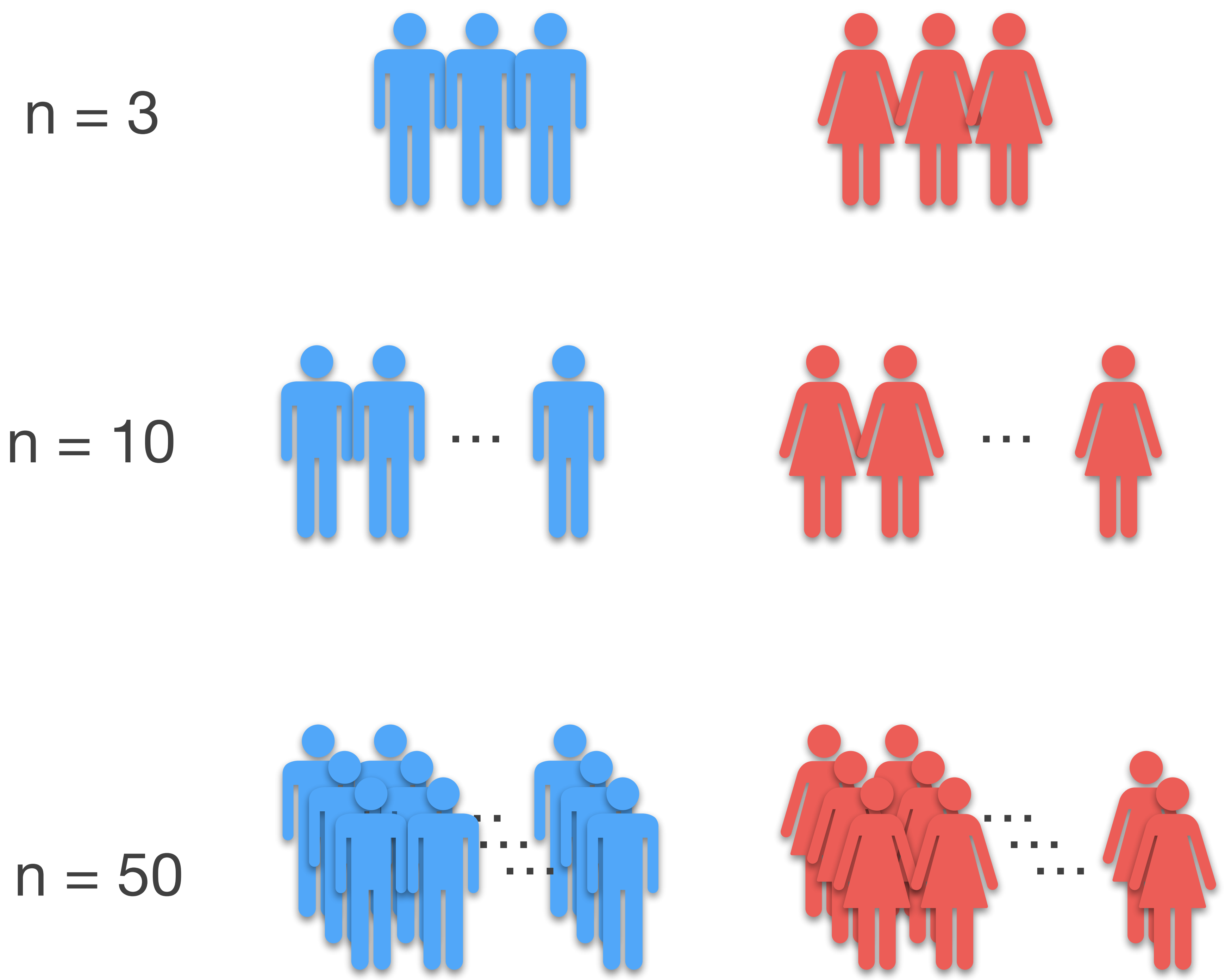
Significance

- When is a probability low, very low, or high?
- Define a **significance level α**
- $p < \alpha$:
  - H_0 hypothesis can be rejected
  - the observed effect is significant
  - H_1 is statistically proven
- $p > \alpha$:
  - effect is not sufficient to reject H_0
  - observed effect is compatible with statistical fluctuations
  - H_0 is not proven, maybe with a larger sample, the effect could become significant
- $\alpha = 0.05$ has become a standard value (but no golden rule!)



Effect size vs. significance

comparing mean weights



n	w.m	w.f	Difference	p	-log(p)
3	69.51	66.48	3.03	3.76E-02	1.425
10	70.08	66.98	3.10	1.42E-08	7.846
50	70.17	67.24	2.93	1.11E-24	23.957

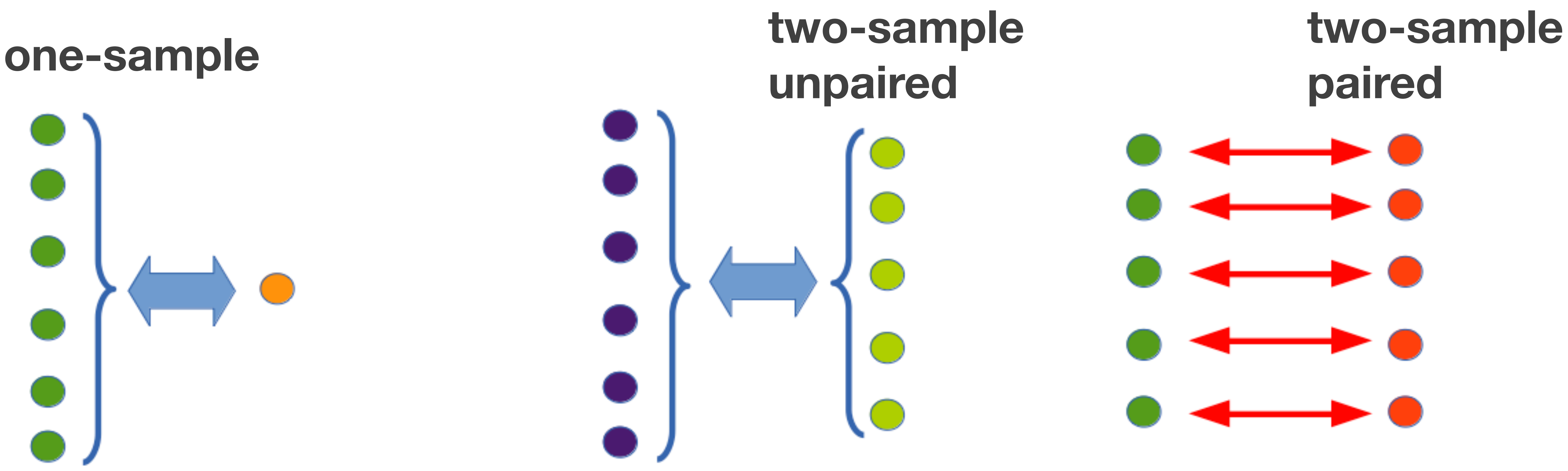
- A small effect size can become significant for large n
- A large effect size can be none-significant if n is low

7. Hypothesis testing

7.2 Testing the mean - t-tests

Test on mean values

- Hypothesis on mean values can be investigated using a ***t-test***
- Family of tests with different version:
  - **one-sample test**: *is the mean body temperature 37.7 C?*
  - **two-sample test, unpaired**: *do men and women have different mean cholesterol levels?*
  - **two-sample test, paired**: *is there a change in cholesterol level after a one-month egg rich diet?*



(do both samples have equal variance?)

t-test test statistics

Type	test statistics	degrees of freedom	note
one sample	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	n-1	
two-sample unpaired (same variance)	*(Student t-test)* $t = \frac{\bar{x}_1 - \bar{x}_2}{s_{12}\sqrt{1/n_1 + 1/n_2}}$	n_1+n_2-2	$s_{12} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$
two-sample unpaired (diff. variance)	*(Welch t-test)* $t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$	(*)	$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
two-sample paired	$t = \frac{\bar{x}_D - \mu}{s_D/\sqrt{n}}$	n-1	x_D = difference between pairs *mu* = expected difference

(*) $$\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}.$$

Distribution under H0

- The test statistics of the t-tests under H_0 are distributed according to a **t-distribution** with the corresponding number of degrees of freedom
- for large sample sizes, the H_0 distribution is the standard normal distribution *N(0,1)*

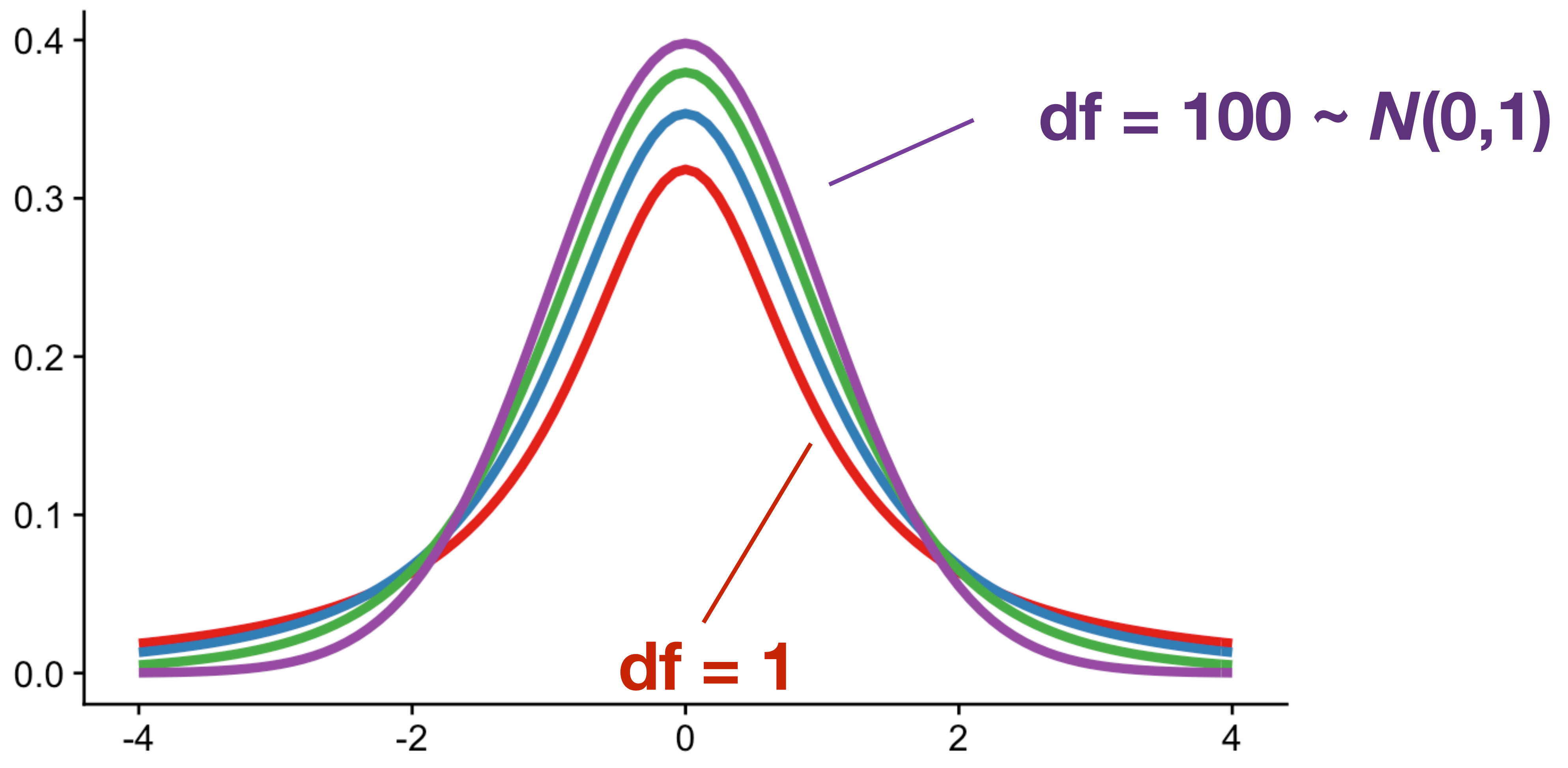
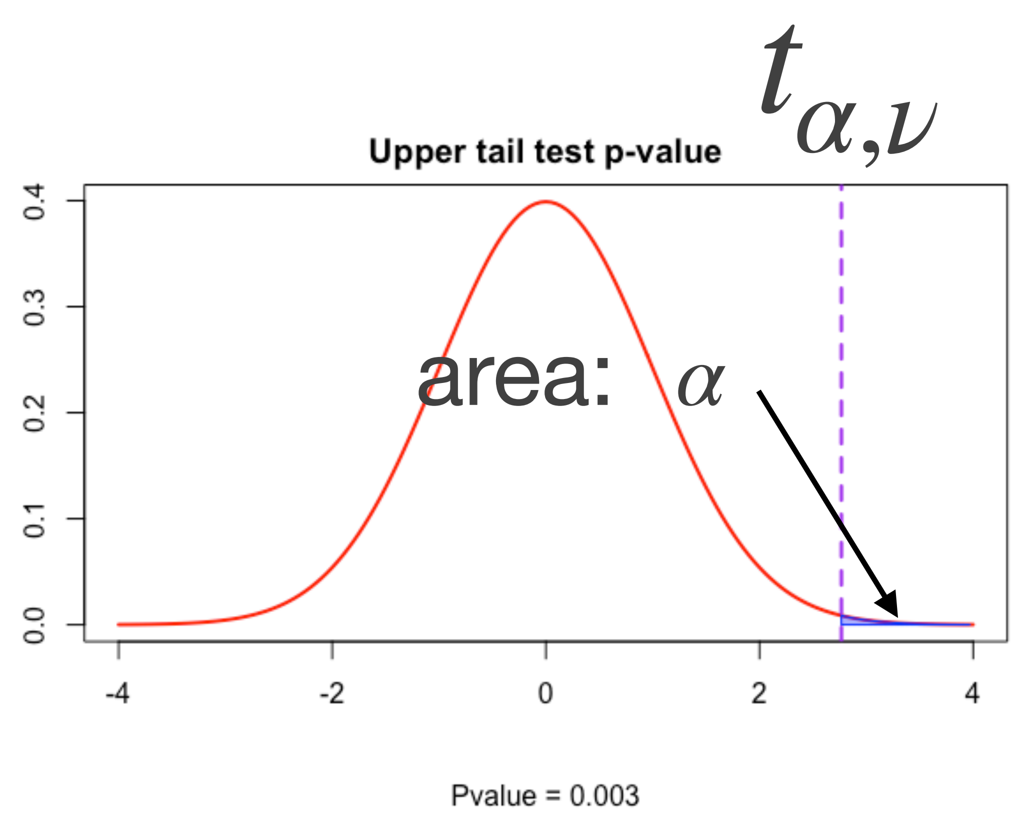


Table of critical values

two-sample t-test, unpaired, one-sided

ν	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781



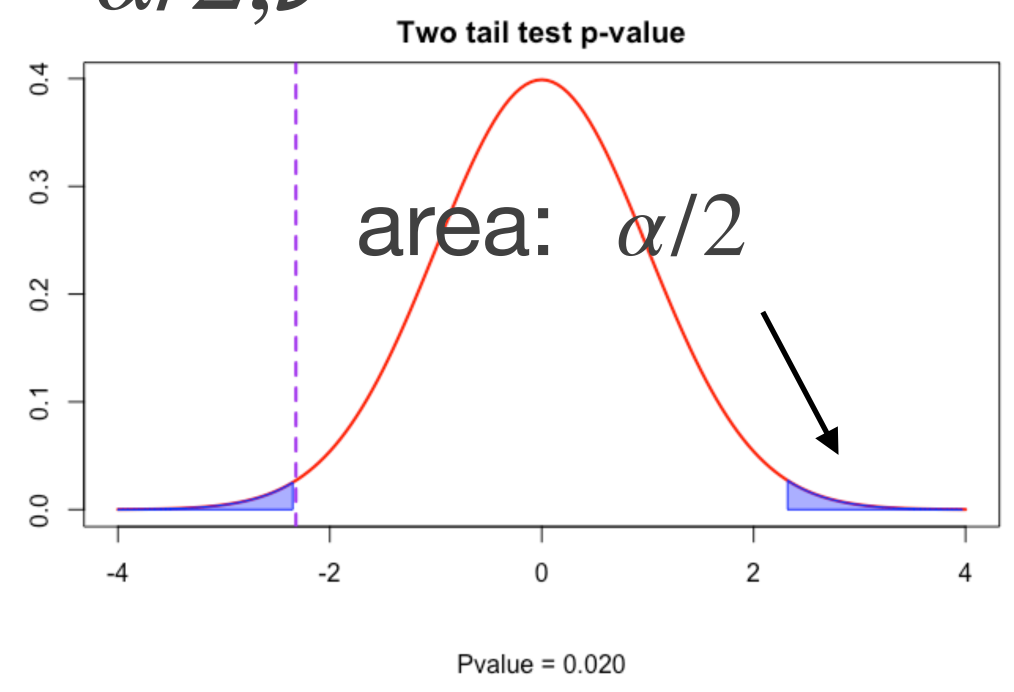
- Example (1-sample t-test)
  - alpha = 0.05
  - t = 2.01
  - sample size n = 8 → $\nu = n - 1 = 7$
- **one-sided t-test**
  - critical value $t_{0.05,7}$ = 1.895
  - t > $t_{0.05,7}$: test is significant for alpha = 0.05
  - ***H_0 can be rejected: result is significant!***

Table of critical values

two-sample t-test, unpaired, two-sided

ν	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781

$t_{\alpha/2,\nu}$



- Example (1-sample t-test)
  - alpha = 0.05
  - t = 2.01
  - sample size n = 8 → $\nu = n - 1 = 7$
- **two-sided** t-test
  - critical value $t_{0.025,7}$ = 2.365
  - t < $t_{0.025,7}$: test is NOT significant for alpha = 0.05
  - ***H_0 cannot be rejected: test is NOT significant***

Running a t-test in R

two-sample unpaired, two-sided

t = test statistics
df = degrees of freedom

confidence interval differences of the means

```
> t.test(weight.m,weight.f,var.equal=TRUE)

        Two Sample t-test
data:  weight.m and weight.f

t = 1.8265, df = 400, p-value = 0.06852

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

95 percent confidence interval:
 -0.5669448 15.4259192

sample estimates:
mean of x mean of y
 181.9167  174.4872
```

Running a t-test in R

Institut für Pharmazie und Molekulare Biotechnologie

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

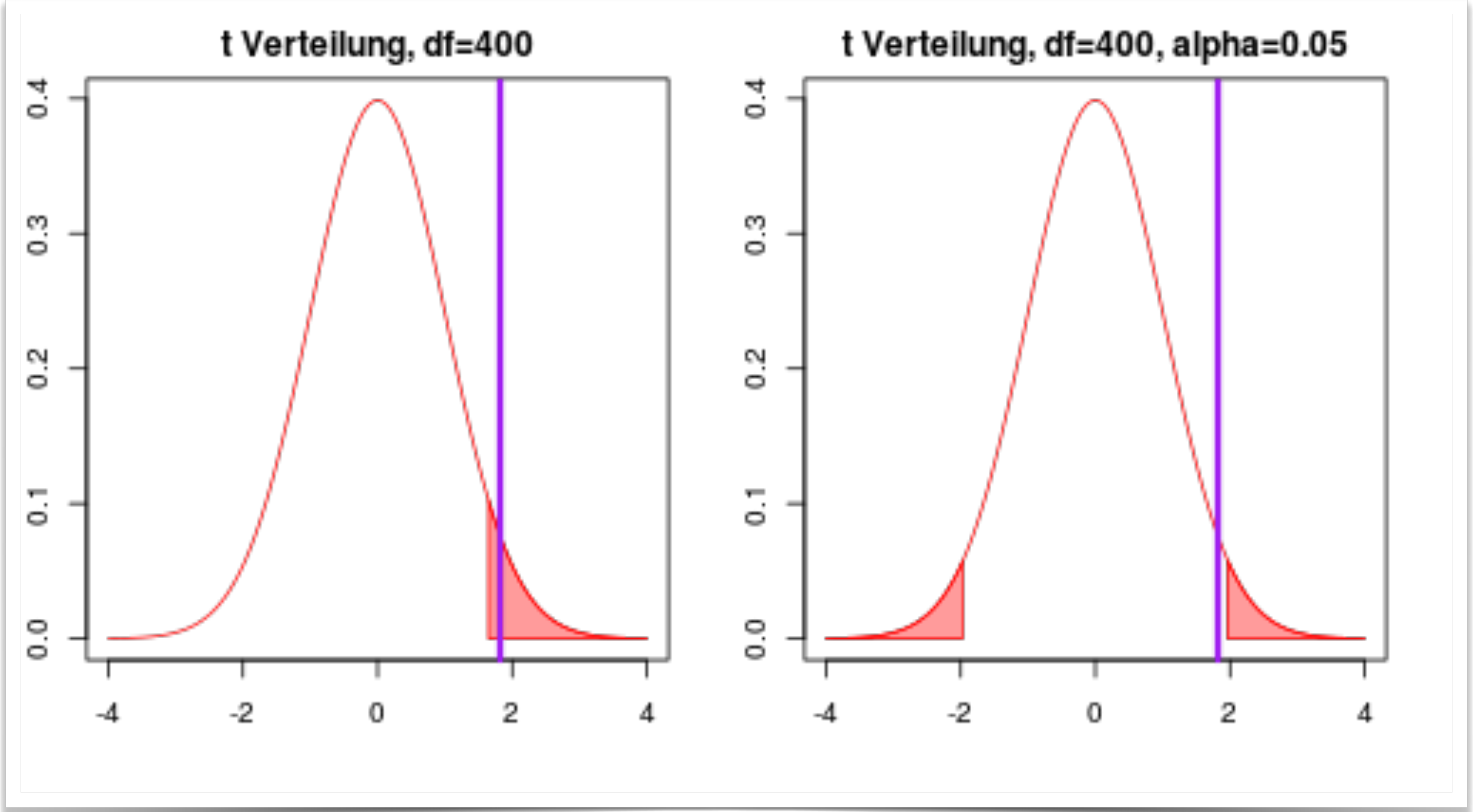
two-sample unpaired, one-sided

t = test statistics
df = degrees of freedom

confidence interval differences of the means

```
>t.test(weight.m,weight.f,alternative="greater",var.equal=TRUE)

        Two Sample t-test
data:  weight.m and weight.f

t = 1.8265, df = 400, p-value = 0.03426

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:
 0.723444      Inf
sample estimates:
mean of x mean of y
 181.9167  174.4872
```

Running a t-test in R



one-sided t-test
→ significant

two-sided t-test
→ non significant

Running a t-test in R

two-sample Welch unpaired, one-sided

t = test statistics
df = degrees of freedom

confidence interval differences of the means

```
>t.test(weight.m,weight.f,alternative=“greater")

        Welch Two Sample t-test
data:  weight.m and weight.f

t = 1.8453, df = 372.446, p-value = 0.0329

alternative hypothesis: true difference in means
is greater than 0

95 percent confidence interval:
 0.7903498       Inf

sample estimates:
mean of x mean of y
 181.9167  174.4872
```

Paired t-test

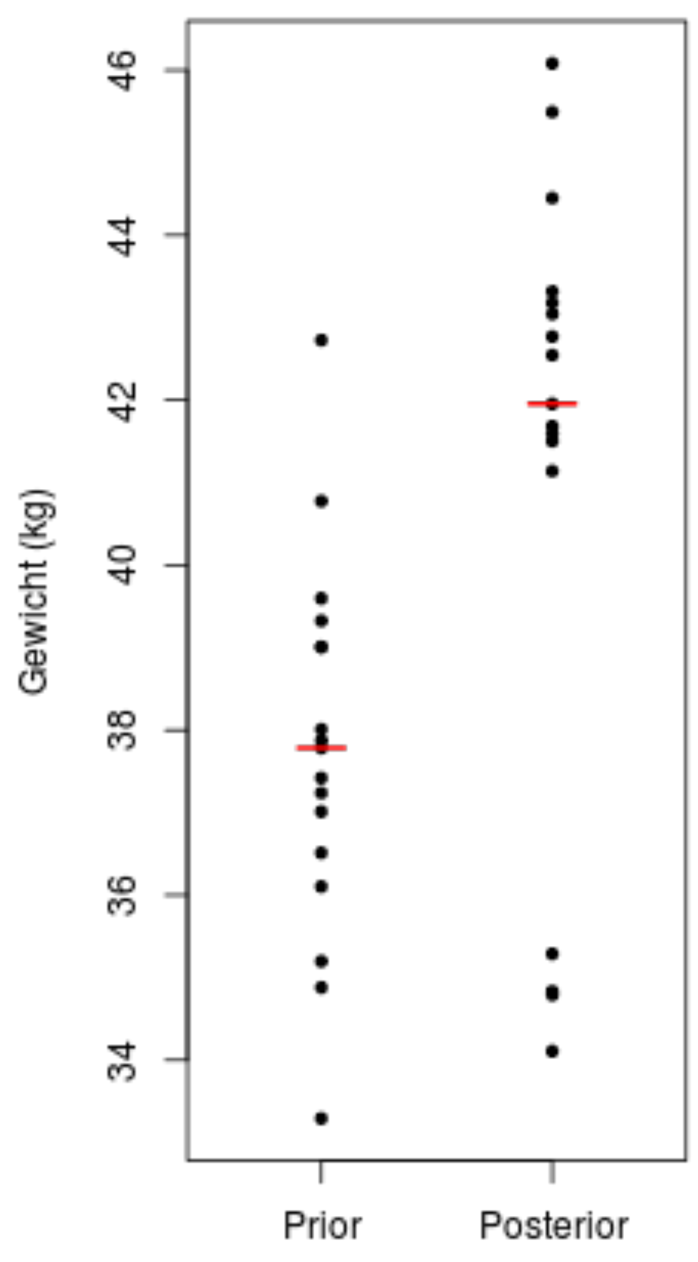
- 2 samples with equal number of elements
- each element of sample A can be associated to one element of sample B
    - patients before (A) and after (B) treatment
    - technical replicates

$$t = \frac{\bar{x}_D - \mu}{s_D / \sqrt{n}}$$

$\bar{x}_D$ = mean of differences

μ = expected difference

Treatment against anorexia
Weight before/after treatment



unpaired: $p = 5 \cdot 10^{-3}$

When can we apply t-test?

- There are several conditions that must be fulfilled to apply a t-test
- **Normality**: data must be (approximately) normaly distributed
  → check using
  - QQ-plot
  - statistical tests: Shapiro-Wilks / Kolmogorov-Smirnov
  - if not, apply non-parametrical test

- **Variance** of samples must be equal
  - if so: Student t-test
  - if not: Welch t-test

- **Independance**: independent samples: values in one sample should not be influenced by those in the second sample

7. Hypothesis testing

7.3 Non-parametric tests

Non-parametric tests

- If the condition of normality of the data is not met, use **non-parametric tests**
- These do not require any specific distribution of the data
- Values of the data are **converted to ranks** *(remember the Spearman correlation!)*

- **Wilcoxon Rank Tests**
  - unpaired: ***Wilcoxon Rank Sum Test*** (a.k.a, Mann-Whitney U test)

  - paired : ***Wilcoxon signed rank test***

Wilcoxon Rank Sum Test
Mann-Whitney U Test

- 2 samples with numerical values

$$X = \{x_1, x_2, \ldots, x_{n_1}\} \quad Y = \{y_1, y_2, \ldots, y_{n_2}\}$$

- Values are merged and ranked in increasing order

$$Z = X \cup Y$$

- R_1 is the sum of the ranks of the first probe (first probe is the one giving the smallest *U*)
- **Test statistics**

$$U = R_1 - \frac{n_1(n_1+1)}{2}$$

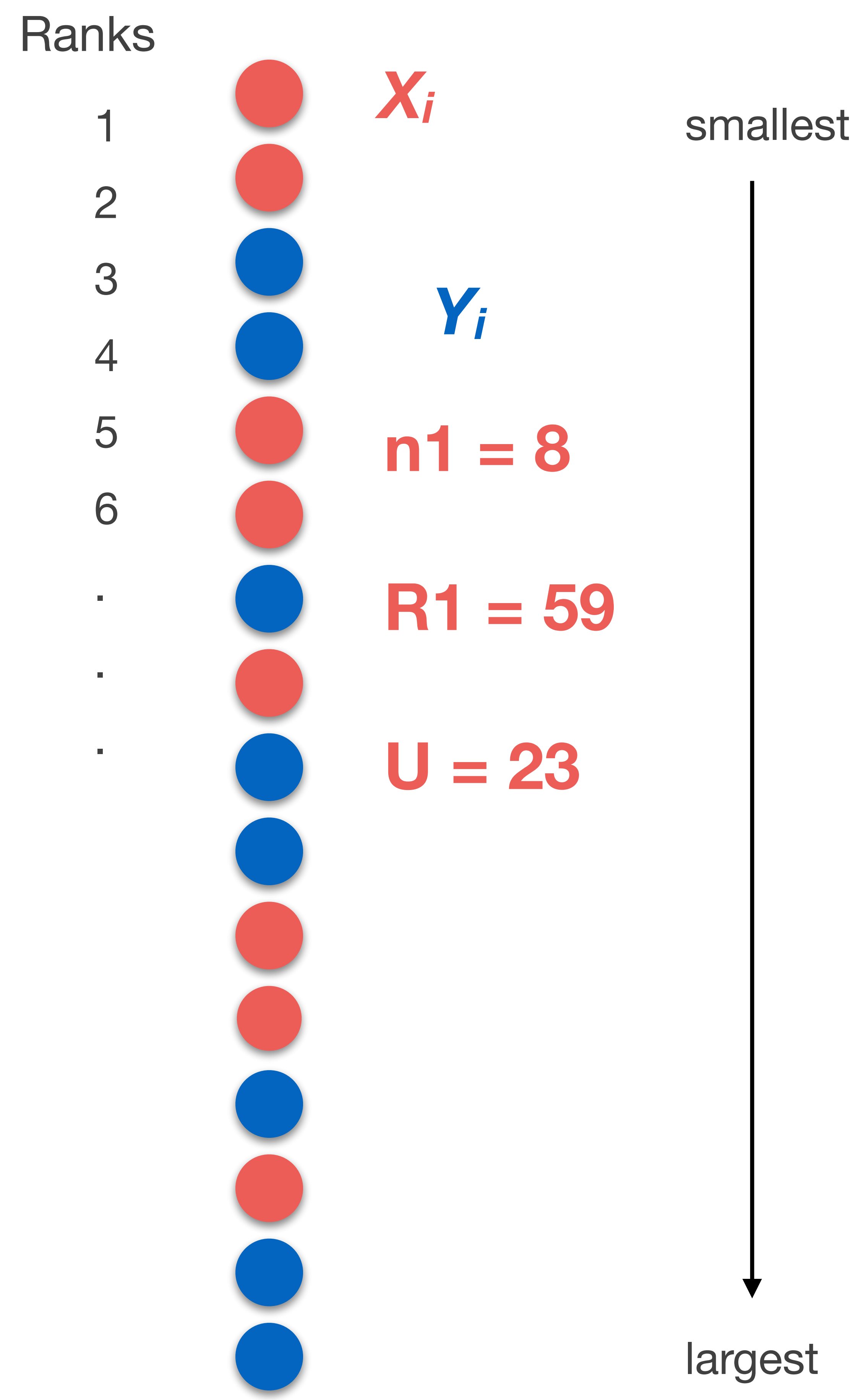
- **H_0 :** $E(X) = E(Y)$ (2-sided test)

  $E(X) > E(Y)$

  (1-sided test)

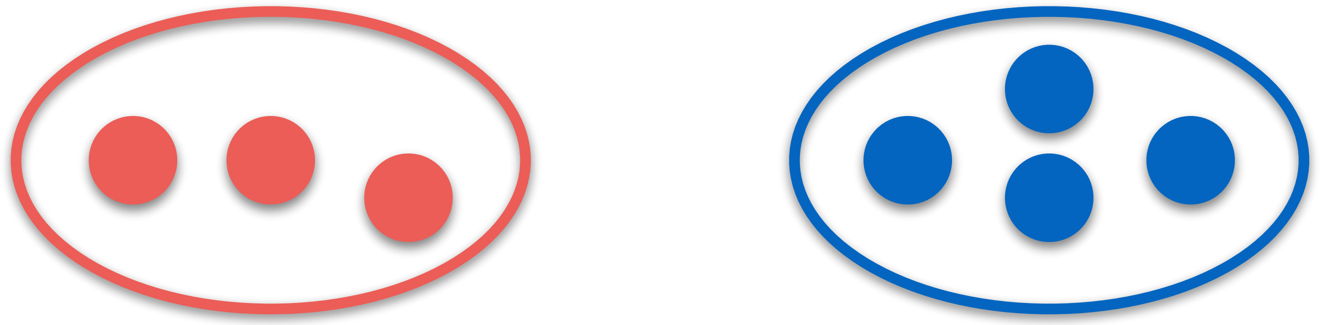
  $E(X) < E(Y)$

What is U if all reds are smaller than the blues?



Wilcoxon Rank Sum Test
Mann-Whitney U Test

Remember: the smaller U, the more significant

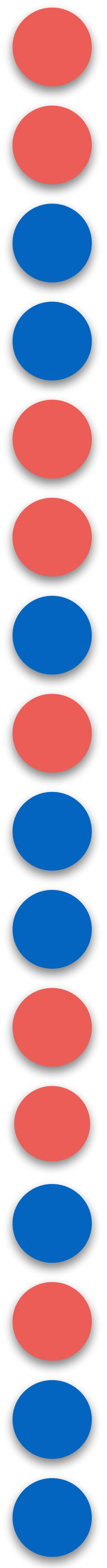


!! Values of α are for two-sided test !!

n_2	α	n_1: 3	4	5	6	7	8	9	10	11	12	13	14	15
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16
8	.05	2	4	6	8	10	$\boxed{13}$	15	17	19	22	24	26	29
	.01	--	1	2	4	6	$\boxed{7}$	9	11	13	15	17	18	20

X_i

Y_i



smallest

largest

n1 = 8

R1 = 59

U = 23

U is larger than the critical values for α=0.05 or 0.01

→ ***H_0* cannot be rejected**

→ **test non significant**

Wilcoxon Signed Rank Test
(2 paired probes)

- 2 samples with numerical values

$$X = \{x_1, x_2, \ldots, x_n\} \qquad Y = \{y_1, y_2, \ldots, y_n\}$$

- D_i = differences between pairs
- R_i = ranks of the differences $|D_i|$
- *Test statistics:*

$$W_+ = \sum_{i=1}^{N_+} R_{i,D_i>0} \qquad W_- = \sum_{i=1}^{N_-} R_{i,D_i<0}$$

$$W = min(W_+, W_-)$$

- Question: do the positive/negative differences have different ranks?
  → H_0: no!

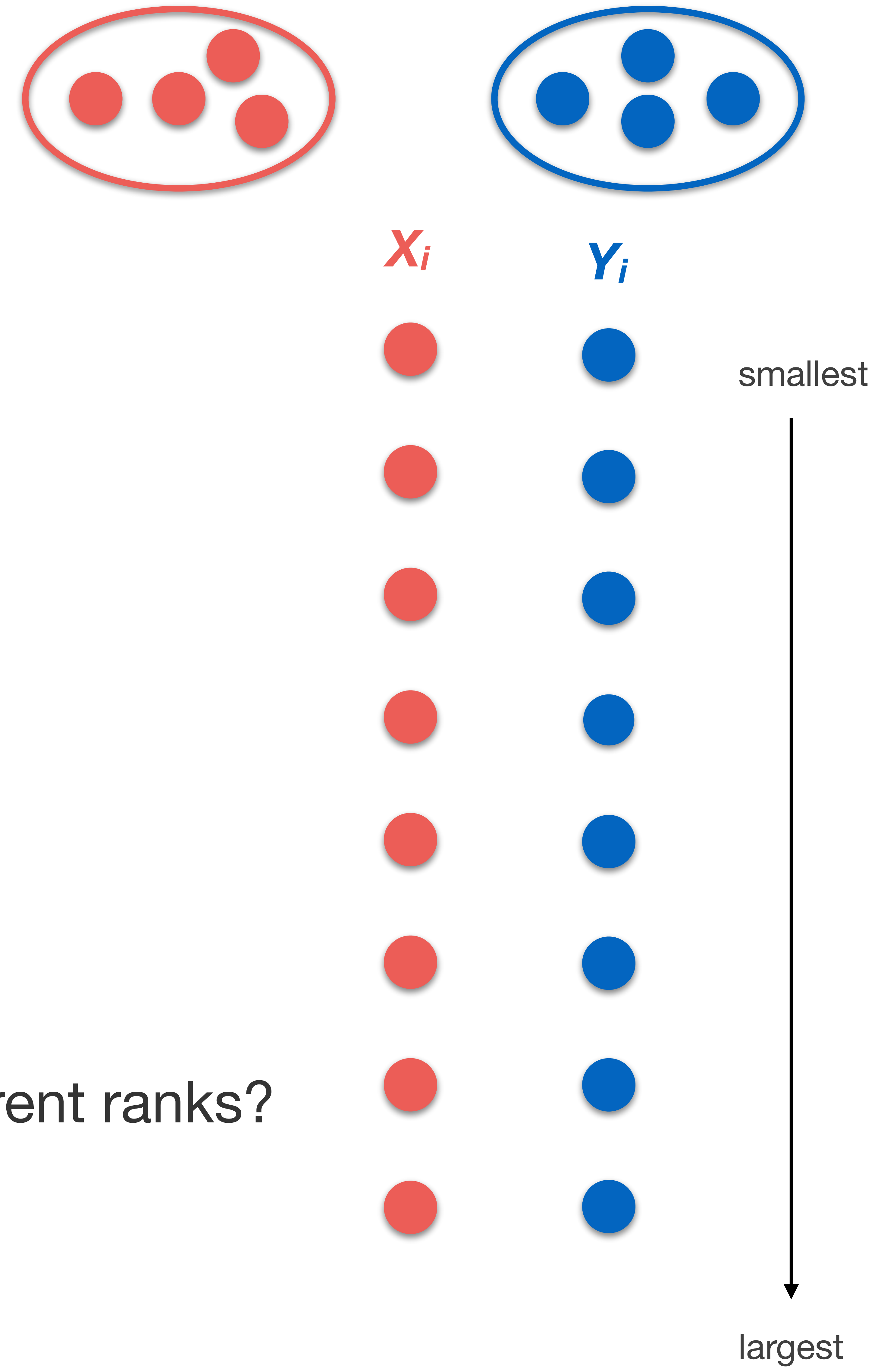


Table of critical values for Wilcoxon signed-rank test

- The smaller, the more significant
- Example:
  - n=14
  - W = 22
  - Non-significant for 2-tailed test and α =0.05
  - Significant for 1-tailed test and α =0.05
  - Non significant for 1-tailed test and α = 0.01

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

Wilcoxon Signed Rank Test
(2 paired probes)

```
> X
   Prior  Post Diff AbsDiff ranks SignedRanks
1   76.9  76.8 -0.1     0.1     1          -1
2   79.6  76.7 -2.9     2.9     2          -2
3   81.6  77.8 -3.8     3.8     3          -3
4   89.9  93.8  3.9     3.9     4           4
5   80.5  75.2 -5.3     5.3     5          -5
6   86.0  91.5  5.5     5.5     6           6
7   86.0  91.7  5.7     5.7     7           7
8   94.2 101.6  7.4     7.4     8           8
9   83.5  92.5  9.0     9.0     9           9
10  82.5  91.9  9.4     9.4    10          10
11  87.3  98.0 10.7    10.7    11          11
12  83.3  94.3 11.0    11.0    12          12
13  83.8  95.2 11.4    11.4    13          13
14  77.6  90.7 13.1    13.1    14          14
15  82.1  95.5 13.4    13.4    15          15
16  86.7 100.3 13.6    13.6    16          16
17  73.4  94.9 21.5    21.5    17          17

> W.p <- sum(X[X$Diff>0,'ranks'])
> W.m <- sum(X[X$Diff<0,'ranks'])

> W.p
[1] 142
> W.m
[1] 11
```

```
> wilcox.test(X$Prior,X$Post,paired=TRUE)

        Wilcoxon signed rank test

data:  X$Prior and X$Post
V = 11, p-value = 0.0008392

alternative hypothesis: true location shift
is not equal to 0
```

(two-sided test)

7. Hypothesis testing

7.4 Proportion tests

Proportion tests

- This class of tests can be used when searching for
  - **relation between different categorical variables**
    Is there a relation between social background and school grades?
  - comparison of **observed** vs. **expected** counts
    Is there a significant gender bias in the math department if 4 professors out of 10 are women?
- Two tests are generally used
  - **Fisher-Exact test** (FET): gives an exact p-value, used for small samples
  - **chi-square test**: for larger samples (n>5 in each category)
  - both tests are equivalent for large n

Fisher Exact Test

- Tests for a significant relationship between 2 variables
- Starting point: contingency table

	iPhone	other	Total
Men	**4**	**1**	5
Women	**2**	**3**	5
Total	6	4	10

Proportion iPhone/other:
- Men : 4/1 = 4
- Women: 2/3 = 0.66

Odds-Ratio:

OR = (4/1)/(2/3) = 6

If we would <u>randomly</u> distribute 6 iPhone and 4 other smartphones to 5 men and 5 women, how often would we get a larger/smaller*/more extreme

*smaller: < 1/6

**More extreme: > 6 or < 1/6

What is H0?

	iPhone	other	Total
Men	**3**	**2**	5
Women	**3**	**2**	5
Total	6	4	10

H_0: The proportion of men with iPhone is **equal** to the proportion of women with iPhones (2-sided)

$$OR = 1$$

H_0:The proportion of men with iPhones is **not higher** that the proportion of women with iPhones (1-sided)

$$OR \leq 1$$

H_0:The proportion of men with iPhones is **not lower** that the proportion of women with iPhones (1-sided)

$$OR \geq 1$$

Random permutations

If I randomly distribute 6 iPhones and 4 other phones
to 5 women and 5 men, how likely it is to obtain this table?

	iPhone	other	Total
Men	**4**	**1**	5
Women	**2**	**3**	5
Total	6	4	10

Random permutations

	iPhone	other	Total
Men	**4**	**1**	5
Women	**2**	**3**	5
Total	6	4	10

$$p = \frac{\binom{6}{4}\cdot\binom{5}{4}\cdot 4!\cdot\binom{5}{2}\cdot 2!\cdot\binom{4}{1}\cdot 3!}{10!} = 0.238 \qquad OR = 6$$

	iPhone	other	Total
Men	**5**	**0**	5
Women	**1**	**4**	5
Total	6	4	10

$$p = 0.023; \quad OR = \frac{5/0}{1/4} = +\infty$$

	iPhone	other	Total
Men	**3**	**2**	5
Women	**3**	**2**	5
Total	6	4	10

$$p = 0.4761; \quad OR = \frac{3/2}{3/2} = 1$$

Random permutations

	iPhone	other	Total
Men	**2**	**3**	5
Women	**4**	**1**	5
Total	6	4	10

$$p = 0.238; \quad OR = \frac{2/3}{4/1} = 1/6$$

	iPhone	other	Total
Men	**1**	**4**	5
Women	**5**	**0**	5
Total	6	4	10

$$p = 0.023; \quad OR = \frac{1/4}{5/0} = 0$$

$$p_{1-sided} = 0.238 + 0.0238 = 0.2619 \quad (OR \geq 6)$$

$$p_{2-sided} = 0.238 + 0.0238 + 0.238 + 0.0238 = 0.5238$$

$$(OR \leq \frac{1}{6} \quad or \quad OR \geq 6)$$

MoBi students

	iPhone	other	Total
Men	**8**	**19**	27
Women	**16**	**16**	32
Total	24	35	59

```
Fisher's Exact Test for Count Data

data:  X
p-value = 0.1831
alternative hypothesis: true odds
ratio is not equal to 1
95 percent confidence interval:
 0.1230632 1.3943512
sample estimates:
odds ratio
 0.4273899
```

chi-square test

- The chi-square test compares **observed** and **expected** counts
- Starting point is a **contingency table**
- Test statistics

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

- H_0: expected and observed proportions are equal
- H_0 distribution: chi2 distribution with *n-1* degrees of freedom for *n* observations
- Application possible when $O_i > 2$ and $O_i > 5$ in 80% of observations
- *Note: the chi-square test is always a 1-sided upper tail test!*

Observed

	iPhone	other	Total
Men	**14**	**30**	44
Women	**5**	**20**	25
Total	19	50	69

Observed proportions

	iPhone	other	Total
Men	**31.8%**	**68.2%**	100%
Women	**20%**	**80%**	100%
Total	27.5%	72.5%	100%

H0 proportions

	iPhone	other	Total
Men	**27.5%**	**72.5%**	100%
Women	**27.5%**	**72.5%**	100%
Total	27.5%	72.5%	100%

Expected counts under H0

	iPhone	other	Total
Men	**12.1**	**31.9**	44
Women	**6.9**	**18.1**	25
Total	19	50	69

$$\chi^2 = \frac{(14-12.1)^2}{12.1} + \frac{(30-31.9)^2}{31.9} + \frac{(5-6.9)^2}{6.9} + \frac{(20-18.1)^2}{18.1}$$
$$= 0.6022$$

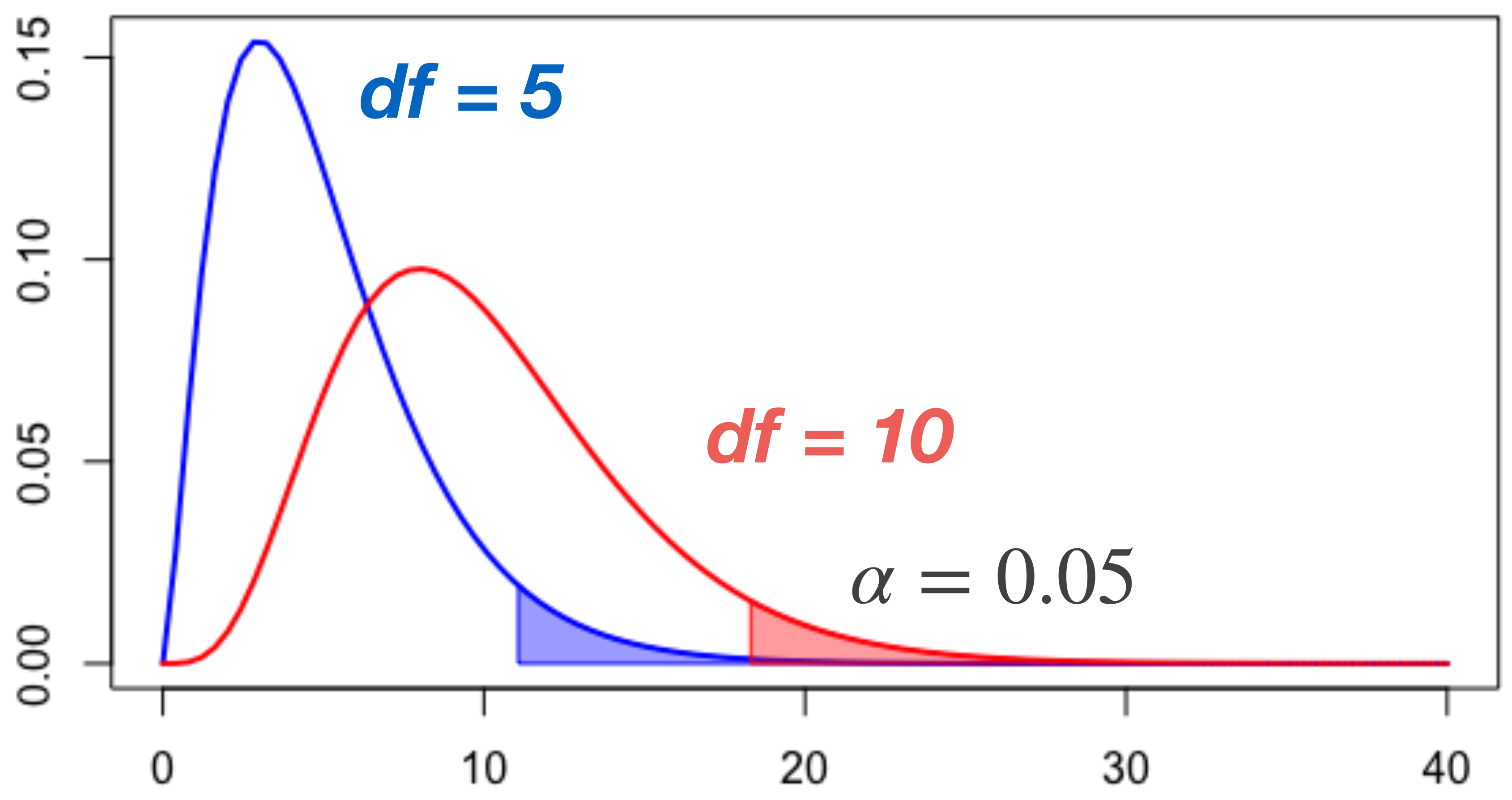
degrees of freedom = (rows-1) x (columns-1)

chi-square distribution

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Critical values

	0.025	0.05	0.1
df = 1	5.02	3.84	2.71
df = 2	7.38	5.99	4.61
df = 3	9.35	7.81	6.25
df = 4	11.14	9.49	7.78
df = 5	12.83	11.07	9.24
df = 6	14.45	12.59	10.64
df = 7	16.01	14.07	12.02
df = 8	17.53	15.51	13.36
df = 9	19.02	16.92	14.68
df = 10	20.48	18.31	15.99



$\alpha = 0.05$

$\chi^2 = 0.6022$ **not significant...**

$df = 1$

More than 2 categories

Institut für Pharmazie und Molekulare Biotechnologie

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Side effects

	weak	medium	strong	Total
Drug A	**25**	**11**	**13**	49
Drug B	**9**	**14**	**11**	34
Total	34	25	24	83

	weak	medium	strong	Total
Drug A	**51%**	**22.5%**	**26.5%**	100%
Drug B	**26.5%**	**41.2%**	**32.3%**	100%
Total	41%	30.1%	28.9%	100%

```
> table(sideeffect)
    SideEffect
Drug weak medium strong
   A     25      11      13
   B      9      14      11

> chisq.test(table(sideeffect))
        Pearson's Chi-squared test
data:  table(sideeffect)
X-squared = 5.5257, df = 2, p-value = 0.06311


> fisher.test(table(sideeffect))
        Fisher's Exact Test for Count Data
data:  table(sideeffect)
p-value = 0.06375
alternative hypothesis: two.sided
```