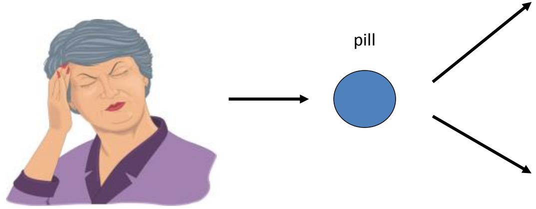
Chi-square test
Analyzing frequency data

Example: headache

Effective:
no headache



Not effective:
remaining headache

Experiment

1. group: patients taking the ***medicine***
   - headache (*a*)
   - no headache (*b*)
2. group: patients taking the ***placebo***
   - headache (c)
   - no headache (d)

(a,b,c,d are frequency data)

Contingency table

	headache	no headache	Total
1. group	a	b	a+b
2. group	c	d	c+d
total	a+c	b+d	n

So-called 2 x 2 table.

Ideal case

Condition: there is no placebo effect!

medicine is effective

	headache	no headache	Total
1. group	0	a+b	a+b
2. group	c+d	0	c+d
total	c+d	a+b	n

medicine is not effective

	headache	no headache	Total
1. group	a+b	0	a+b
2. group	c+d	0	c+d
total	n	0	n

Nullhypothesis

	headache	no headache	Total
1. group	a	b	a+b
2. group	c	d	c+d
total	a+c	b+d	n

If the medicine is

similar to the placebo we expect:

$$\frac{a}{b}=\frac{c}{d} \longrightarrow a\times d=b\times c$$

Nullhypothesis: the medicine is same as the placebo.

Chi-Square test for independence.

Independent case

Remember: P(AB) = P(A) x P(B) if A and B are independent from each other. (P(AB), P(A) and P(B) may be estimated by relative frequencies.)

$$\frac{a}{n}\approx\frac{a+b}{n}\times\frac{a+c}{n}$$

	headache	no headache	Total
1. group	a	b	a+b
2. group	c	d	c+d
total	a+c	b+d	n

Observed proportion: a/n

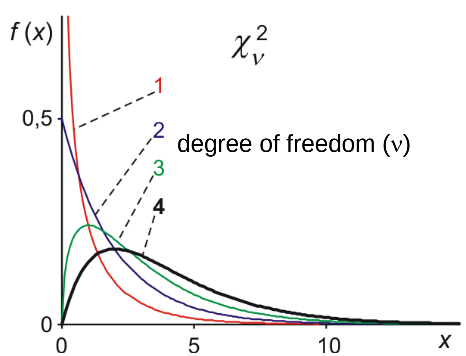
Expected proportion: $\frac{a+b}{n}\times\frac{a+c}{n}$

transformation

$$\chi^2=\sum\frac{(O-E)^2}{E}$$

a/n ~ P(AB) – (no effect in the 1. group)
(a+b)/n ~ P(A) – (belongs to the 1. group)
(a+c)/n ~ P(B) - (no effect)

χ^2-distribution



χ^2-distribution

Shortcut formula
for 2 x2 tables:

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Nullhypothesis: χ^2-value is equal to 0.
The difference is due to the **sampling error**.

χ^2-distribution describes the random deviations of the χ^2-value.

Decision

Same, then in the case of *t*-distribution. We use χ^2-distribution.

Expected value is 0, if the null hypothesis is true.

p ≤ α - reject the null hypothesis else accept.

degree of freedom: in this special case = 1.

In general:

d.f.=(r-1)(c-1), where r – no. of rows
c – no. of columns

Small expected frequencies

May not be used if:

1. An expected frequency is 2 or less.
2. More than 20% of the expected frequencies are less than 5.

Fisher's exact test may be used.
Calculates the exact probability for the given table.

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Remember!
n! = multiplying the integers from 1 to n.

Decision is based on the P.

The Chi-Square test Goodness-of-Fit test

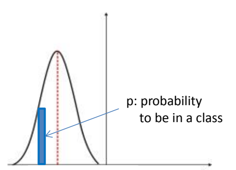
Example: testing normality of the larger diameter of the frog red blood cells.

Observed frequencies:

size (μm)

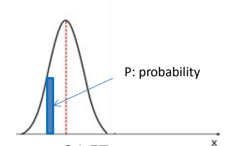
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	n
4	10	9	20	26	27	37	42	48	53	45	39	35	17	18	10	7	5	450

Nullhypothesis (H_0):
Data has normal distribution. Calculate the average and the s_d from the sample!
Calculate expected frequency from the normal distribution!
in a class = np (see figure)



Chi-Square test

avg = 24.57 μm;
s_d = 3.62 μm



μ = 24.57 μm

$n_i \sim 450 \times P$

Expected frequencies:

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
2.8	5.1	8.9	14.2	21	29	37	44	48	49	46.4	41	33	25	18	11	6.9	7.2

Observed frequencies:

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	n
4	10	9	20	26	27	37	42	48	53	45	39	35	17	18	10	7	5	450

Degree of freedom = m – b – 1
m: no. of classes. (in example = 18)
b: no. of parameters (in example = 2)

Calculation:
p = 0.96

We accept the nullhypothesis.

Chi-Square test
Test for homogenity

Example: Is there difference between girls and boys wearing glasses?

H_0: There is no difference.
(independent!)

P(**W**ith) ~ 76/200; P(**B**oys) ~ 97/200
Independent case:
P(W and B) = P(W) x P(B)
expected freq. ~ n x (PW and B)
= 200 * 76/200 * 97/200 ~ 36.9

Observed frequencies

	with	without	
boys	48	49	97
girls	28	75	103
	76	124	200

Expected frequencies

	with	without	
boys	36.9	60.1	97
girls	39.1	63.9	103
	76	124	200

Calculation

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{200\cdot(48\cdot 75-49\cdot 28)^2}{76\cdot 124\cdot 97\cdot 103}$$

$\chi^2 \approx 10.5$ d.f. = 1 $p \approx 0.001$

Decision:
We reject the nullhypothesis. There is significant difference between boys and girls.

Hypothesis test?

- Set up the **nullhypothesis!**
- Look for a **variable with known distribution**.
- Calculate the **probability of the random deviation** on the base of the distribution.
- If it is smaller than the significance level **reject**, in opposite case **accept the nullhypothesis**.
- That's all!

Conditions for tests

test	condition
One-sample t-test	One group, one variable, normal distribution
Two-sample t-test	Two independent groups, one variable, normal distribution, the standard deviations may be the same in the groups
ANOVA	3 or more independent groups, one variable, normal distribution
Sign test	One group, numerical or ordinal quantity
Wilcoxon's signed rank-test	One group, numerical or ordinal quantity
Mann-Whitney U-test	Two independent groups, numerical or ordinal quantity
Kruskall-Wallis test	3 or more groups, numerical quantity
Pearson's correlation test	One group, two variables, normal distribution
Spearman's correlation test	One group, two variables, numerical or ordinal quantity
Chi-Square test (independency)	Two or more groups, frequency data
Chi-Square test (homogenity)	Two or more groups, frequency data
Chi-Square test (fit)	One group, known distribution, frequency data