P-values and statistical tests
2. Contingency tables

Marek Gierliński
Division of Computational Biology

Hand-outs available at http://is.gd/statlec

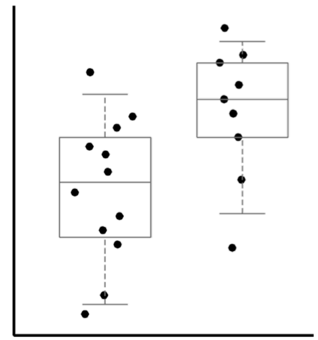
Statistical testing

Null hypothesis
H_0: no effect

All other assumptions

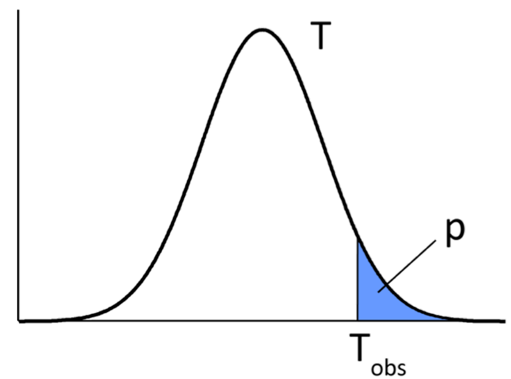
Significance level
$\alpha = 0.05$

Data



Test statistic T_{obs}

Statistical test against H_0



p-value: probability that the observed effect is random

$p < \alpha$
Reject H_0
Effect is real

$p \geq \alpha$
Insufficient evidence

Contingency tables

Drug treatment

	No treatment	Drug X
No improvement	57	32
Improvement	13	46

Enrichment

	In cluster	Outside cluster
With GO-term	6	1
Without GO-term	38	623

Cell counting

	WT	KO
G1	50	61
S	172	175

Cell counting

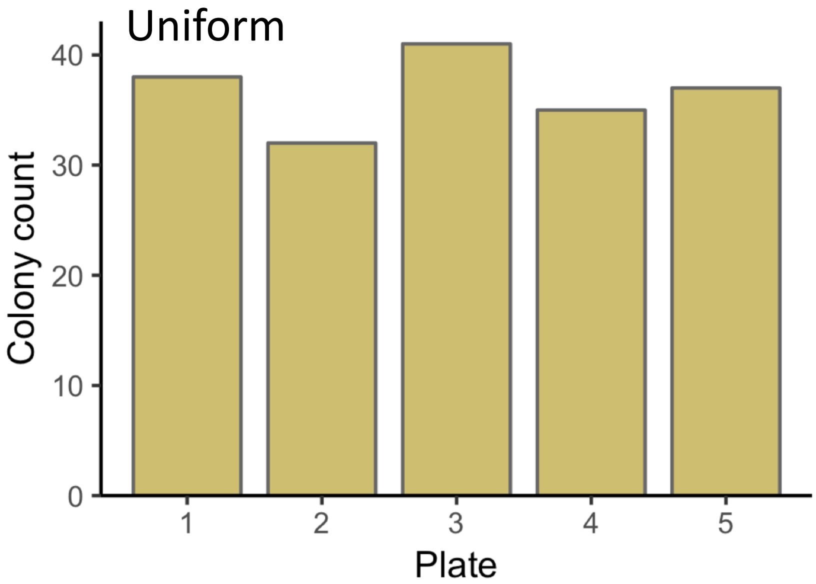
	WT	KO1	KO2	KO3
G1	50	61	78	43
S	172	175	162	178
G2	55	45	47	59

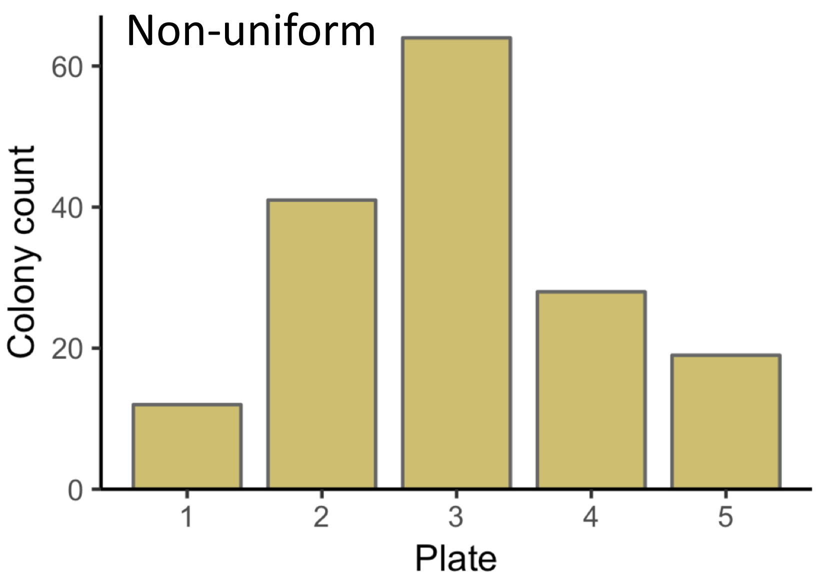
Chi-square test

Goodness-of-fit test

Pipetting experiment

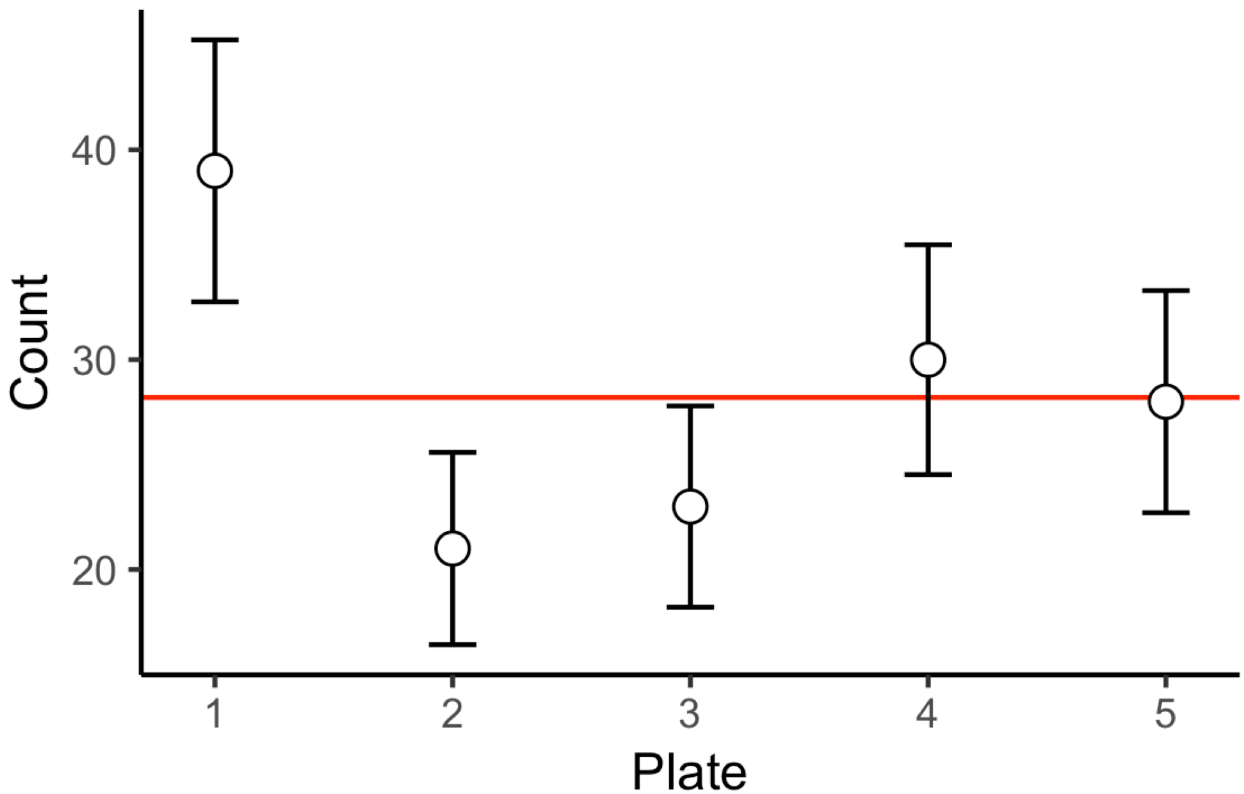
- Dilution plating over five plates
- Aliquots taken from the same culture
- Good pipetting: uniform distribution of counts





Chi-square goodness-of-fit test

	Plate				
	1	2	3	4	5
Observed	39	21	23	30	28
Expected	28.2	28.2	28.2	28.2	28.2



Test statistic

	Plate				
	1	2	3	4	5
Observed	39	21	23	30	28
Expected	28.2	28.2	28.2	28.2	28.2
$\chi_i = \frac{O_i - E_i}{\sqrt{E_i}}$	2.03	-1.36	-0.98	0.34	-0.04

- We have observed (O_i) and expected (E_i) counts, $i = 1, 2, \dots, n$
- Test statistic is

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

$$\chi^2 = \frac{(39 - 28.2)^2}{28.2} + \frac{(21 - 28.2)^2}{28.2} + \frac{(23 - 28.2)^2}{28.2} + \frac{(30 - 28.2)^2}{28.2} + \frac{(28 - 28.2)^2}{28.2}$$
$$= 7.05$$

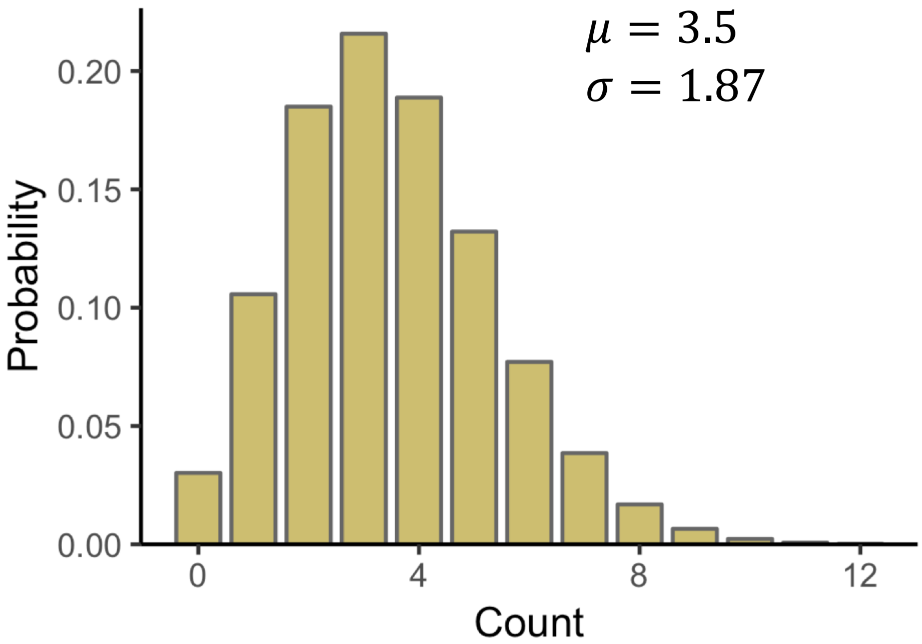
Note: Poisson distribution

- Distribution of counts
- E.g. colony count

- Standard deviation is the root of the mean

$$\sigma = \sqrt{\mu}$$

- For large counts Poisson is similar to normal distribution



χ_i are approximately normal

	Plate				
	1	2	3	4	5
Observed	39	21	23	30	28
Expected	28.2	28.2	28.2	28.2	28.2
$\chi_i = \frac{O_i - E_i}{\sqrt{E_i}}$	2.03	-1.36	-0.98	0.34	-0.04

- Observations O_i have Poisson distribution with mean E_i and standard deviation $\sqrt{E_i}$

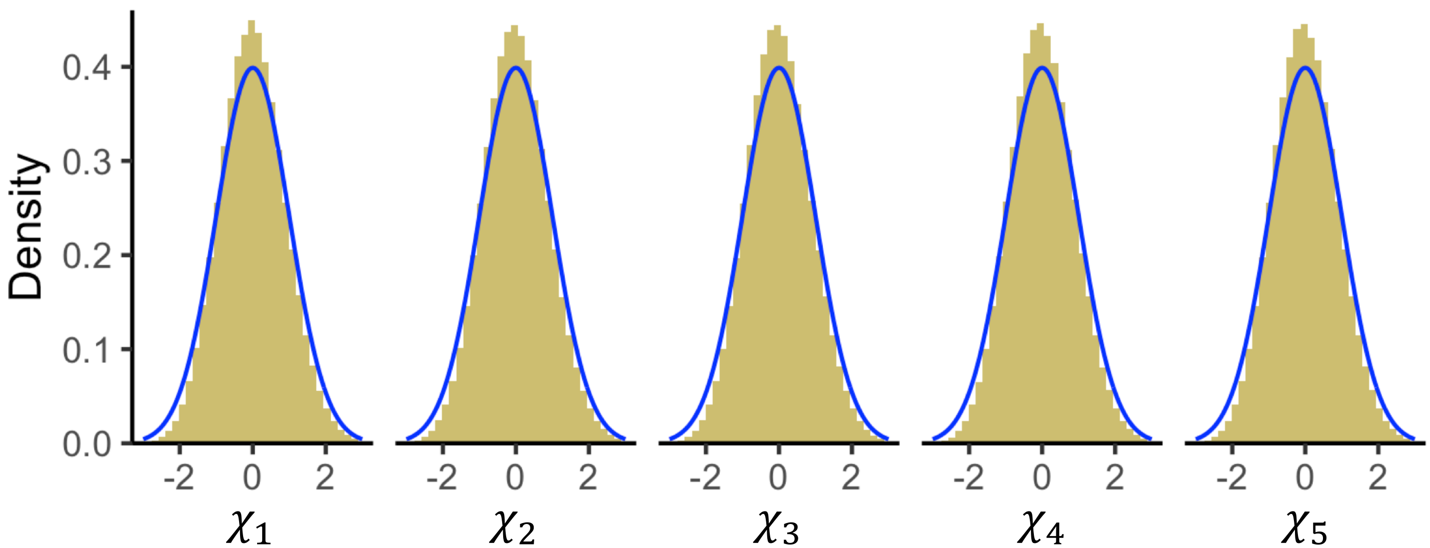
$$\chi_i = \frac{O_i - E_i}{\sqrt{E_i}} \quad \text{looks very much like} \quad Z = \frac{X - \mu}{\sigma}$$

- Then χ_i roughly follows standardized normal distribution (i.e., centred at 0 and with standard deviation of 1)

Gedankenexperiment

- Simulate dilution plating experiment 1 million times
- Generate random counts with the same total count (141) as the original data
- Uniform distribution between plates: null hypothesis

	Plate				
	1	2	3	4	5
Observed	**39**	**21**	**23**	**30**	**28**
Sim 1	33	23	32	22	31
Sim 2	30	22	25	28	36
Sim 3	29	30	32	18	32
...	...	...	...	...	...

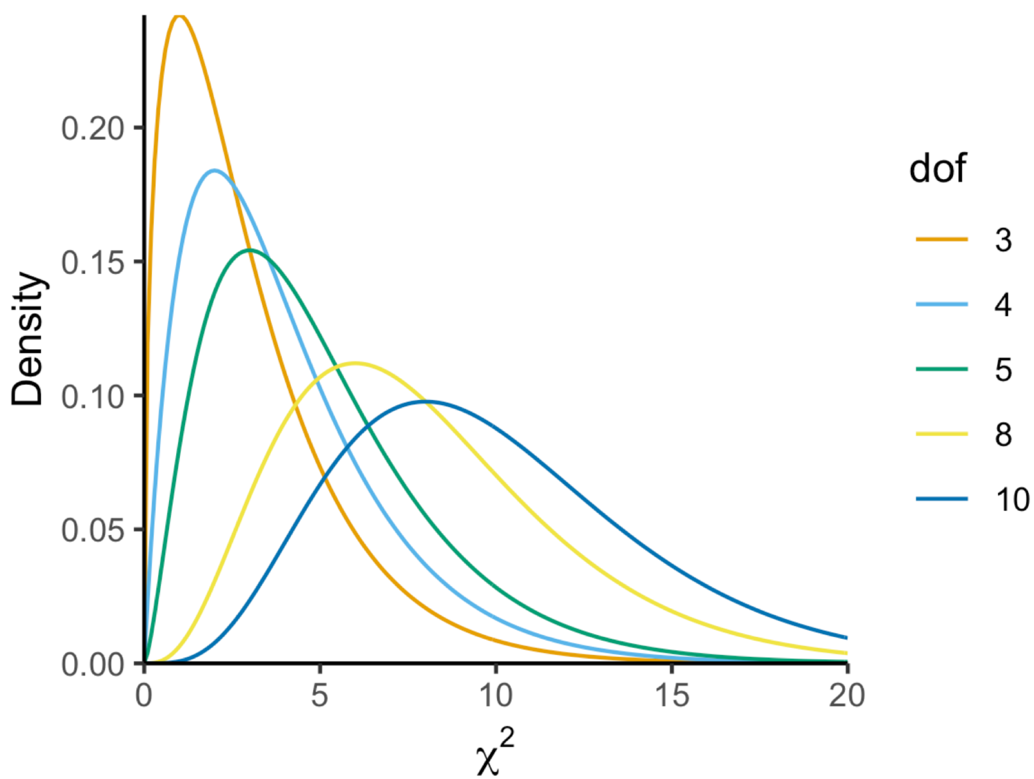


Note: chi-square distribution

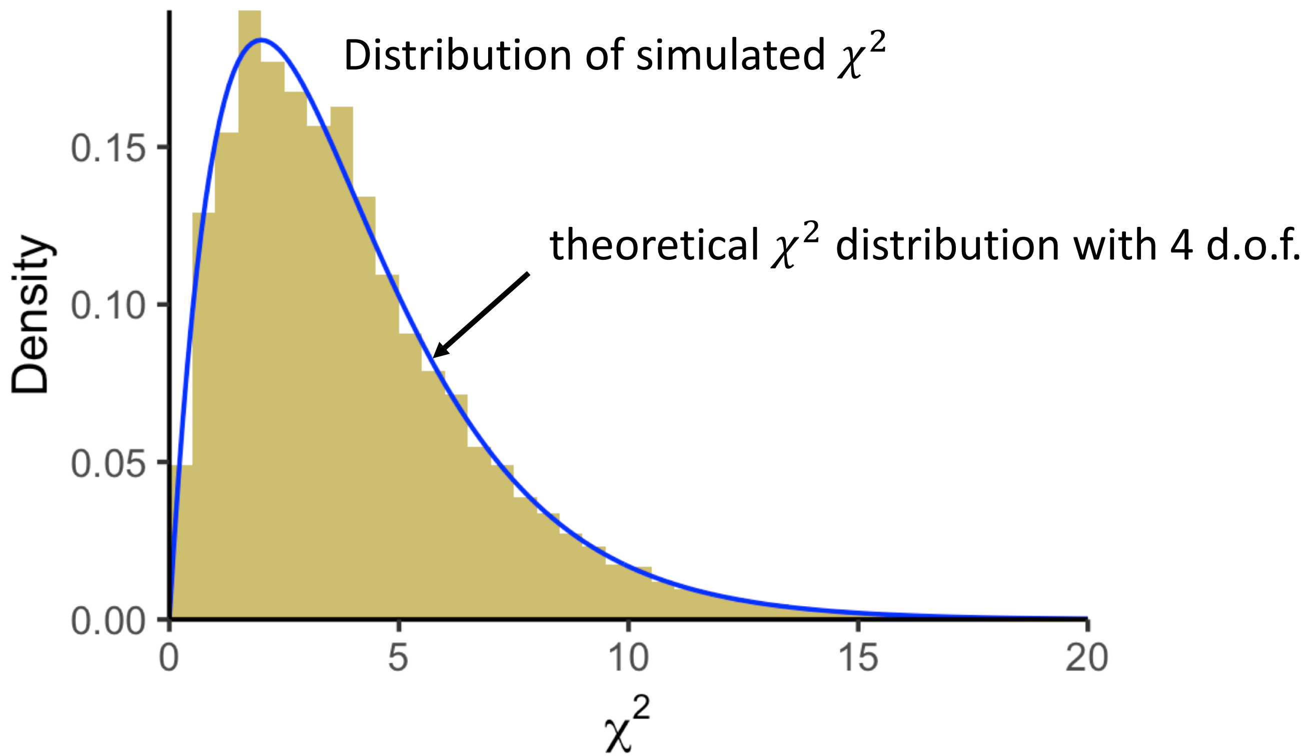
- Definition: a sum of squares of independent standard normal variables

$$\chi^2 = \sum_{i=1}^{n} \chi_i^2$$

- is distributed with χ^2 distribution with $n - 1$ degrees of freedom



Null distribution

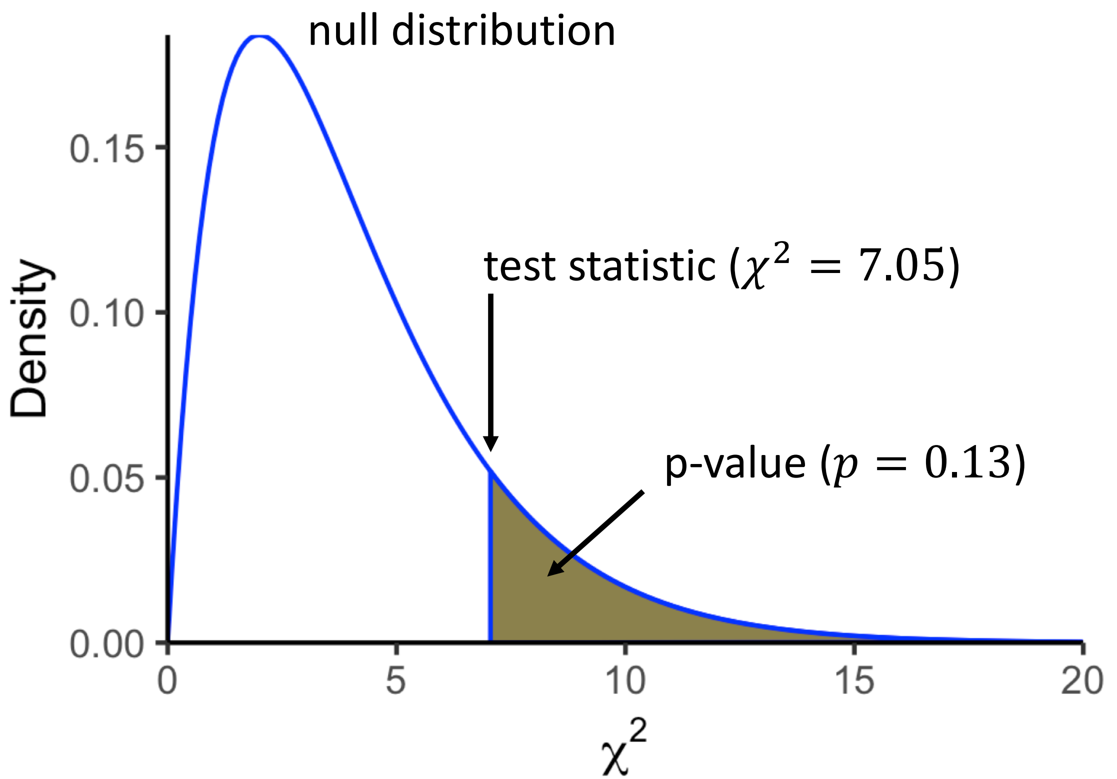


Null distribution represents all random samples when the null hypothesis is true

Chi-square test

- Test statistic (observed): $\chi^2 = 7.05$
- P-value: probability of observing this, or more extreme, effect by chance, if H_0 is true

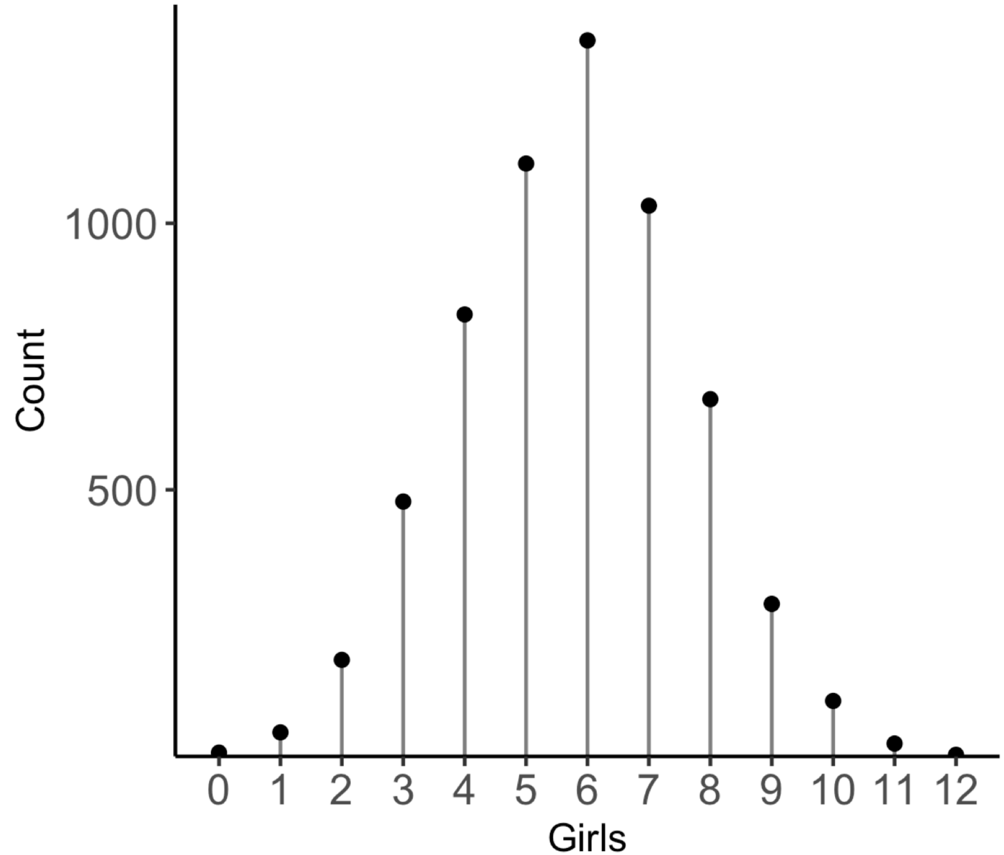
```
> Obs <- c(39, 21, 23, 30, 28)
> Exp <- mean(Obs)
> chi2 <- sum((Obs - Exp)^2 / Exp)
> chi2
[1] 7.049645
> 1 - pchisq(chi2, length(Obs) - 1)
[1] 0.1332878
```



Geissler (1889)

- Birth data from a hospital in Saxony, 1876-1885
- Includes 6115 sibships of 12 children
- Girl/boy ratio $\hat{p} = 0.481 \pm 0.004$ (95% CI)
- Does it follow binomial distribution?

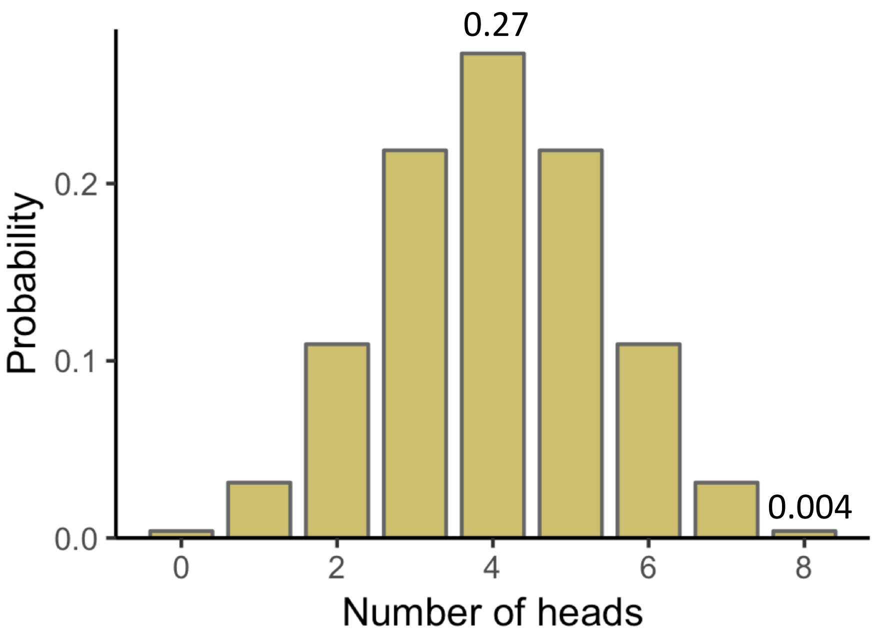
No. girls	Observed
0	7
1	45
2	181
3	478
4	829
5	1112
6	1343
7	1033
8	670
9	286
10	104
11	24
12	3



Note: binomial distribution

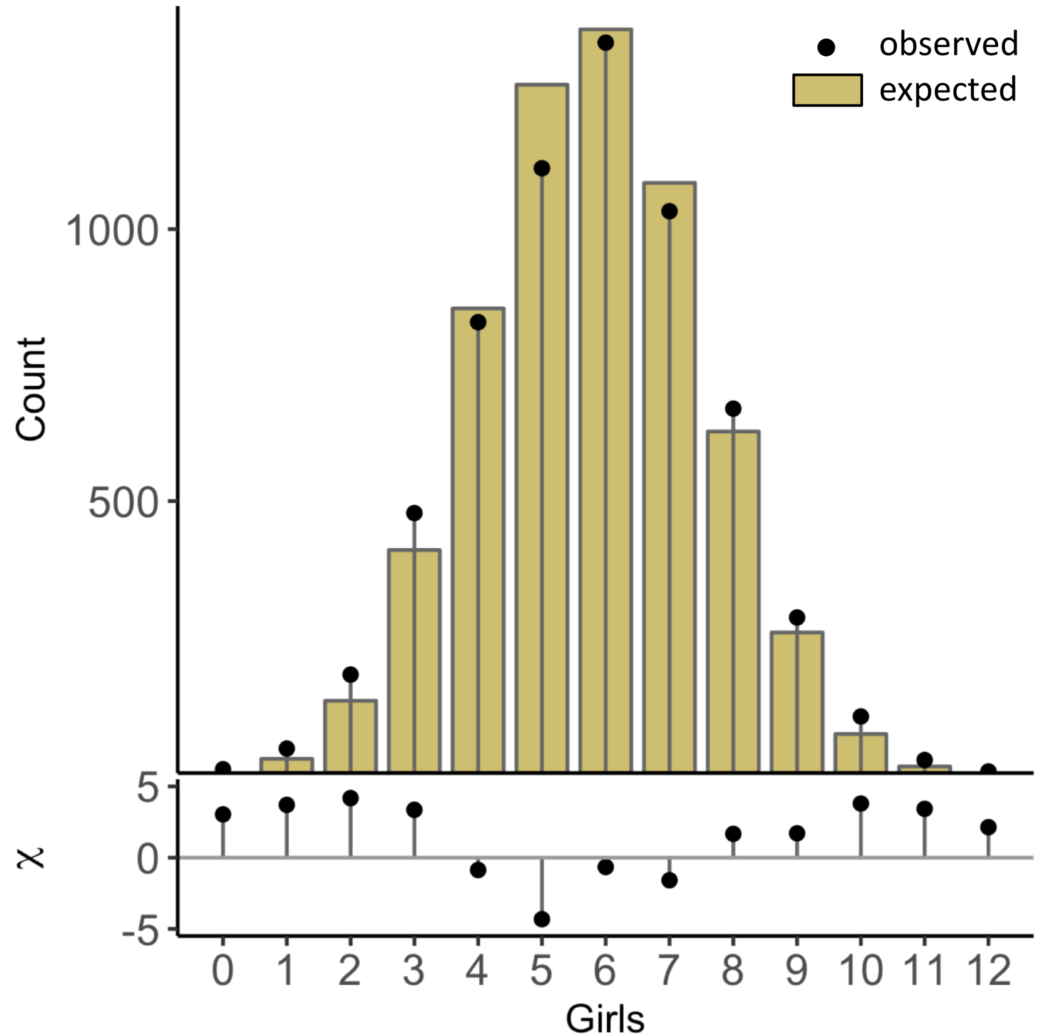
- n repeated trials
- Two possible outcomes, probability p and $1 - p$

- Example: toss a coin ($p = 0.5$)



Geissler (1889)

# girls	O_i	E_i	χ_i	χ_i^2
0	7	2.3	3.04	9.2
1	45	26.1	3.70	13.7
2	181	132.8	4.18	17.5
3	478	410	3.36	11.3
4	829	854.2	-0.86	0.8
5	1112	1265.6	-4.32	18.7
6	1343	1367.3	-0.66	0.4
7	1033	1085.2	-1.58	2.5
8	670	628.1	1.67	2.8
9	286	258.5	1.71	2.9
10	104	71.8	3.80	14.4
11	24	12.1	3.43	11.7
12	3	0.9	2.14	4.6
				110.5



$$\chi^2 = 110.5$$
$$\text{d.o.f.} = 11$$
$$p = 0$$

Degrees of freedom

- Degrees of freedom = independent pieces of available information

- Input data: $N = 141$ objects segregated into $n = 5$ categories
- The total number of counts, N, is fixed

33	23	32	22	31	**141**

- We lose one degree of freedom, $n - 1$ left

- Lose one degree of freedom per each model parameter found
- Binomial proportion from the input data ($\hat{p} = 0.481$)
- We use up more information and have $n - 2$ degrees of freedom

Chi-square goodness-of-fit d.o.f. = $n - 1 - m$
n is size of data
m is the number of model parameters

Chi-square test: how it works

Observations	O_i	Poisson distribution
Errors *chi*	$\chi_i = \frac{O_i - E_i}{\sqrt{E_i}}$	Normal distribution
Test statistic *chi-square*	$\chi^2 = \sum_{i=1}^{n} \chi_i^2$	Chi-square distribution $n - 1 - m$ degrees of freedom
P-value	p	Probability of obtaining observed or more extreme result by chance

Chi-square goodness-of-fit test: summary

Input	Counts from n categories
Assumptions	Observations are random and independent Mutual exclusivity (no overlap between categories) Errors are normal
Usage	Compare the observed counts with a theoretical distribution
Null hypothesis	Number of observations in each category is equal to that predicted by the theoretical distribution
Comments	Approximate test Breaks down for small numbers (total count < 100) For small numbers use the exact multinomial (or binomial) test Be careful with the number of degrees of freedom!

Chi-square test

Test of independence

Chi-square test of independence

- Comparing observed (O_{ij}) with expected (E_{ij}) values
- Expected values are

$$E_{ij} = Np_i p_j$$

  - p_i – proportion in row i
  - p_j – proportion in column j
  - N – total number

	Drug A	Drug B	Total	Proportion
Improvement	12 18.6	30 23.3	42	23.3%
No improvement	68 61.3	70 76.8	138	76.7%
Total	80	100	180	
Proportion	44.4%	55.6%		

Observed

Estimated
$180 \times 0.767 \times 0.556 = 76.8$

- Expected values = null hypothesis
- Proportions in columns (rows) are equal
  - $\frac{18.6}{61.3} = 0.30$
  - $\frac{23.3}{76.8} = 0.30$
- Improvement proportion is independent of the drug choice

Chi-square test of independence

- Comparing observed (O_{ij}) with expected (E_{ij}) values
- Expected values are

$$E_{ij} = N p_i p_j$$

  - p_i – proportion in row i
  - p_j – proportion in column j
  - N – total number

	Drug A	Drug B	Total	Proportion
Improvement	12 18.6	30 23.3	42	23.3%
No improvement	68 61.3	70 76.8	138	76.7%
Total	80	100	180	
Proportion	44.4%	55.6%		

- Test statistic

$$\chi^2 = \sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\chi^2 = \frac{(12-18.6)^2}{18.6} + \frac{(30-23.3)^2}{23.3} + \frac{(68-61.3)^2}{61.3} + \frac{(70-76.8)^2}{76.8} = 5.59$$

$$p_{\mathrm{chi2}} = 0.018$$
$$p_{\mathrm{Fisher}} = 0.013$$

- $\nu = (n_{\mathrm{rows}} - 1)(n_{\mathrm{columns}} - 1)$
- P-value is from χ^2 distribution with 1 d.o.f.
- Corresponds to two-sided Fisher's test

Chi-square test for independence

- Flow cytometry experiment
- WT and three KOs
- Take about 280 cells in each condition
- Establish cell cycle stage

	WT	KO1	KO2	KO3
G1	50	61	78	43
S	172	175	162	178
G2	55	45	47	59

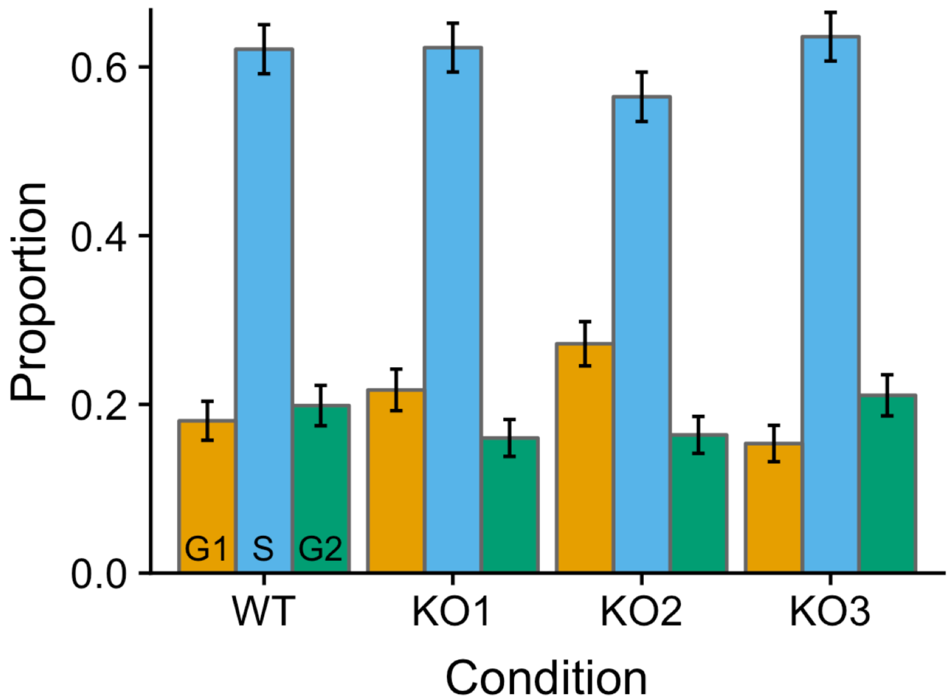
- Are the any differences between the WT and KOs?

$$\chi^2 = 15.1$$

$$\nu = (4-1)(3-1) = 6$$

$$p = 0.02$$

- But what does it mean?



Independence of proportions

- Like in Fisher's test

	C1	C2	C3	C4
G1	10	20	30	40
G2	20	40	60	80

- Rows and columns are independent
- Proportions between rows do not depend on the choice of column
- Proportions between columns do not depend on the choice of row

- Proportions in each row are 1:2:3:4
- Proportions in each column are 1:2

- This contingency table is consistent with the null hypothesis

Pairwise comparison

- Null hypothesis: proportions of cells in G1-S-G2 stages are the same for each condition
- $p = 0.02$, reject the null hypothesis

- Pairwise comparison
- WT vs. KO1

	WT	KO1	KO2	KO3
G1	50	61	78	43
S	172	175	162	178
G2	55	45	47	59

	WT	KO1
G1	50	61
S	172	175
G2	55	45

$\chi^2 = 2.09$
$\nu = 2$
$p = 0.35$

Comparison	p-value	Adj. p-value
WT vs. KO1	0.35	1
WT vs. KO2	0.03	0.19
WT vs. KO3	0.69	1
KO1 vs. KO2	0.28	1
KO1 vs. KO3	0.08	0.49
KO2 vs. KO3	0.002	0.01

One versus others

- Compare each column vs. the sum of others
- WT vs. others

	WT	others
G1	50	182
S	172	515
G2	55	151

$$\chi^2 = 1.72$$
$$\nu = 2$$
$$p = 0.42$$

	WT	KO1	KO2	KO3
G1	50	61	78	43
S	172	175	162	178
G2	55	45	47	59

Comparison	p-value	Adj. p-value
WT	0.42	1
KO1	0.50	1
KO2	0.006	0.02
KO3	0.03	0.12

Chi-square test of independence: summary

Input	$n_r \times n_c$ contingency table table contains counts
Assumptions	Observations are random and independent (no before-after) Mutual exclusivity (no overlap between categories) Errors are normal
Usage	Examine if there is an association (contingency) between two variables; whether the proportions in “groups” depend on the “condition” (and vice versa)
Null hypothesis	The proportions between rows do not depend on the choice of column
Comments	Approximate test Use when you have large numbers For small numbers use Fisher’s test For before-after data use McNemar’s test

How to do it in R?

```
# Colony count test (goodness-of-fit test)
> chisq.test(c(39, 21, 23, 30, 28), p=rep(1/5, 5))

        Chi-squared test for given probabilities

data:  c(39, 21, 23, 30, 28)
X-squared = 7.0496, df = 4, p-value = 0.1333

# Drug comparison
> chisq.test(rbind(c(12, 30), c(68, 70)), correct=FALSE)

        Pearson's Chi-squared test

data:  rbind(c(12, 30), c(68, 70))
X-squared = 5.5901, df = 1, p-value = 0.01806
# Flow cytometry experiment
> cells <- rbind(c(50, 61, 78, 43), c(172, 175, 162, 178), c(55, 45, 47, 59))
> chisq.test(cells)

        Pearson's Chi-squared test

data:  cells
X-squared = 15.122, df = 6, p-value = 0.01933
```

G test

G-test

- Similar to chi-square test
- Based on log-likelihood ratio
- Test statistic

$$G = 2 \sum_i O_i \ln \frac{O_i}{E_i}$$

- G is chi-square distributed with $n - 1$ degrees of freedom (n categories)

- For large numbers chi-square test and G-test give very similar results

Reminder: chi-square statistic

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

G test is like chi-square test

- You can use G test just like chi-square test:
  - Goodness-of-fit test
  - Test of independence

	Plate				
	1	2	3	4	5
Obs	39	21	23	30	28
Exp	28.2	28.2	28.2	28.2	28.2

$\chi^2 = 7.05$
$p = 0.13$

$G = 6.85$
$p = 0.14$

- Results are very similar
- Chi-square test is an approximation of the G test

- G is additive, chi-square is not

	WT	KO1	KO2	KO3
G1	1	61	78	43
S	172	175	162	178
G2	55	45	47	59

$\chi^2 = 15.1$
$p = 0.02$

$G = 15.0$
$p = 0.02$

G test for replicated experiments

Replicate 1

	WT	KO1	KO2	KO3
G1	50	61	78	43
S	172	175	162	178
G2	55	45	47	59

$G = 15.0 \quad p = 0.02$

Replicate 2

	WT	KO1	KO2	KO3
G1	54	75	77	34
S	180	168	167	180
G2	50	41	49	50

$G = 21.1 \quad p = 0.002$

Replicate 3

	WT	KO1	KO2	KO3
G1	48	69	80	49
S	172	166	180	168
G2	63	38	43	45

$G = 16.5 \quad p = 0.01$

Pooled data

	WT	KO1	KO2	KO3
G1	152	205	235	126
S	524	509	509	519
G2	168	124	139	154

$G = 44.9 \quad p = 5{\times}10^{-8}$

G test for replicated experiments

- Perform G test for each replicate
- Find the total G

$$G_{\text{tot}} = G_1 + G_2 + \cdots + G_n$$

$$\nu_{\text{tot}} = \nu_1 + \nu_2 + \cdots \nu_n$$

- Find G_{pool} and ν_{pool} from pooled data
- Find heterogeneity G

$$G_{\text{het}} = G_{\text{tot}} - G_{\text{pool}}$$

$$\nu_{\text{het}} = \nu_{\text{tot}} - \nu_{\text{pool}}$$

- Find p-value for $G_{\text{tot}}, \nu_{\text{tot}}$ and $G_{\text{het}}, \nu_{\text{het}}$ from χ^2 distribution

	G	d.o.f	p-value
Replicate 1	15.0	6	0.02
Replicate 2	21.1	6	0.002
Replicate 3	16.5	6	0.01
Total	52.6	18	3×10^{-5}
Pooled	44.9	6	5×10^{-8}
Heterogeneity	7.7	12	0.8

G test for replicated experiments

- G represents deviation from the null hypothesis
- We can split total G into

$$G_{\text{tot}} = G_{\text{het}} + G_{\text{pool}}$$

G_{het}: variation among replicates

G_{pool}: deviation of the pooled data from H_0

	G	d.o.f	p-value
Replicate 1	15.0	6	0.02
Replicate 2	21.1	6	0.002
Replicate 3	16.5	6	0.01
Total	52.6	18	3×10^{-5}
Pooled	44.9	6	5×10^{-8}
Heterogeneity	7.7	12	0.8

- Use G_{tot} to test the null hypothesis
- However, if G_{het} is large (and p_{het} significant), the deviation from H_0 is due to variation between replicates

G test: summary

Input	$n_r \times n_c$ contingency table table contains counts possible replicates in cells
Assumptions	Observations are random and independent Mutual exclusivity (no overlap between categories) Errors are normal
Usage	Examine if there is an association (contingency) between two variables; whether the proportions in “groups” depend on the “condition” (and vice versa)
Null hypothesis	The proportions between rows do not depend on the choice of column
Comments	Very similar to chi-square test G and d.o.f. are additive Can be used for replicated experiments Not to be confused with ANOVA!

How to do it in R?

```
> install.packages("DescTools")
> library(DescTools)

# Flow cytometry experiment, first replicate
> flcyt1 <- rbind(c(50,61,78,43), c(172,175,162,178), c(55,45,47,59))
> GTest(flcyt1)

        Log likelihood ratio (G-test) test of independence without correction

data:  flcyt
G = 14.994, X-squared df = 6, p-value = 0.0203

# The remaining replicates and the pooled value are found in the same fashion

# Finding p-value for total and heterogeneity G
> 1 - pchisq(52.6, 18)
[1] 3.024812e-05

> 1 - pchisq(7.7, 12)
[1] 0.8081131
```

McNemar’s test

Within-subjects test

Before and after example

Infections before and after treatment (the same patients)

ID	Before	After
1	0	0
2	0	0
3	1	0
4	1	1
…	…	…
500	1	0
501	0	0
502	1	1
Sum	121	34
Proportion	24%	6.8%

		After: no	After: yes
Before	no	321	34
	yes	121	26

```
> mcnemar.test(rbind(c(321, 34), c(121, 26)))

McNemar's Chi-squared test with continuity correction

data:  rbind(c(321, 34), c(121, 26))
McNemar's chi-squared = 47.716, df = 1, p-value = 4.926e-12
```

Contingency table tests

Test	Table	To test if...	Comments
Fisher's exact	2×2	rows and columns are independent; proportions are equal	Works for small numbers, some consider it too conservative
Chi-square goodness-of-fit	$1 \times n$	observed counts follow a theoretical distribution	Requires categorical data, doesn't work for continuous distributions
Chi-square test of independence	$n_r \times n_c$	rows and columns are independent; proportions are equal	Similar to Fisher's works better with large numbers
G-test of independence	$n_r \times n_c$	rows and columns are independent; proportions are equal	Similar to chi-square test, more powerful, can take replicates into account
McNemar's test	2×2	symmetry of rows and columns	Appropriate for paired data, e.g., before-after data on the same subjects

Hand-outs available at http://tiny.cc/statlec