P-values and statistical tests
5. Non-parametric methods

Marek Gierliński
Division of Computational Biology

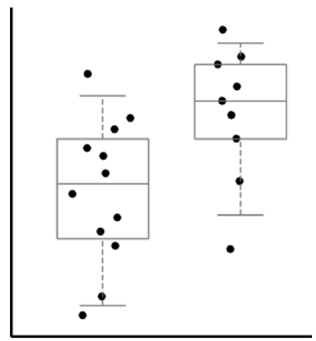
Hand-outs available at http://is.gd/statlec

Statistical test

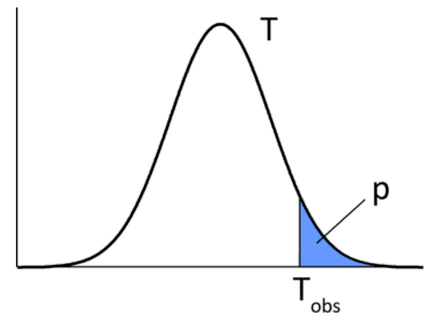
Null hypothesis
H_0: no effect

Significance level
$\alpha = 0.05$

Data



Statistic T



p-value

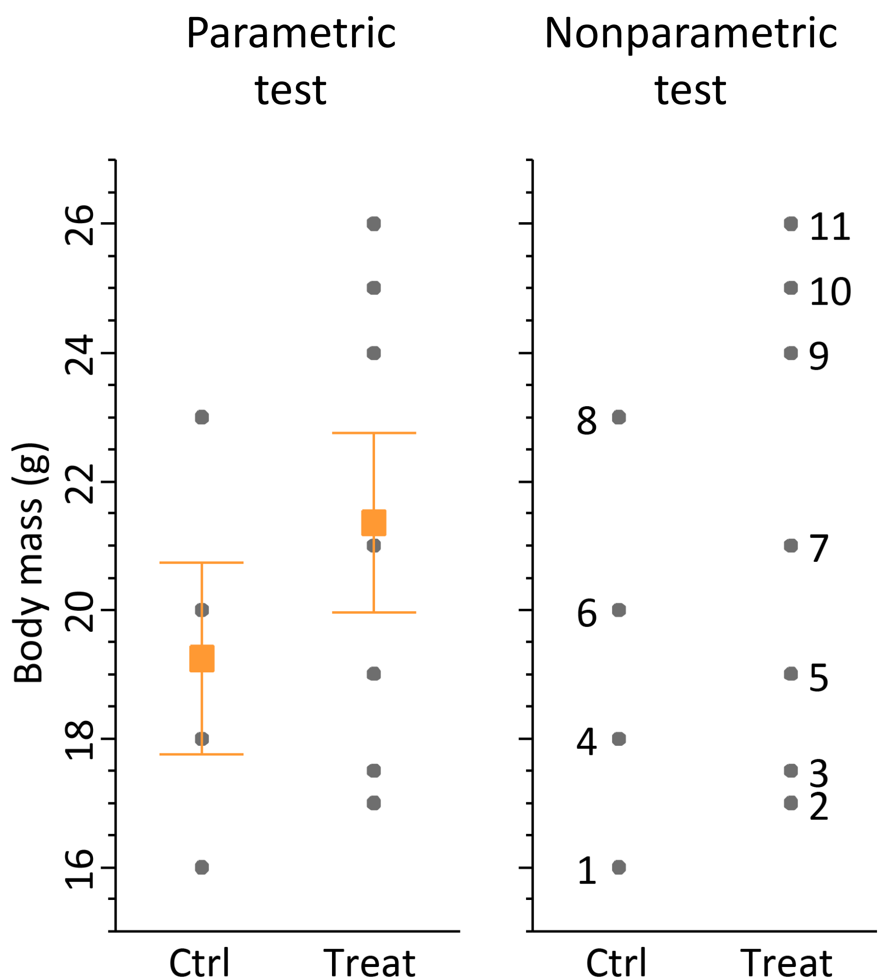
$p < \alpha$
Reject H_0

$p \geq \alpha$
Insufficient evidence

Nonparametric methods

- Parametric methods:
  - require finding parameters (e.g. mean)
  - sensitive to distributions
  - don’t work in some cases
  - more powerful

- Nonparametric methods:
  - based on ranks
  - distribution-free
  - wider application
  - less powerful



Mann-Whitney test
(Wilcoxon rank-sum test)

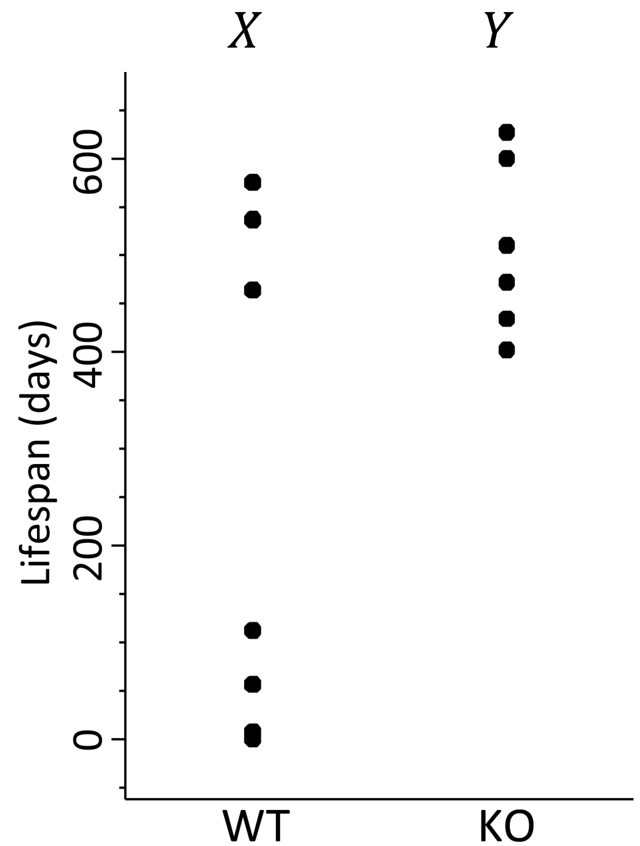
a nonparametric alternative to t-test

Mann-Whitney test

- Two samples representing random variables X and Y
- Null hypothesis: there is no shift in location (and/or change in shape)

$$H_0: \; P(X > Y) = P(Y > X)$$

- Only ranks matter, not actual values



Mann-Whitney test

- Two samples:

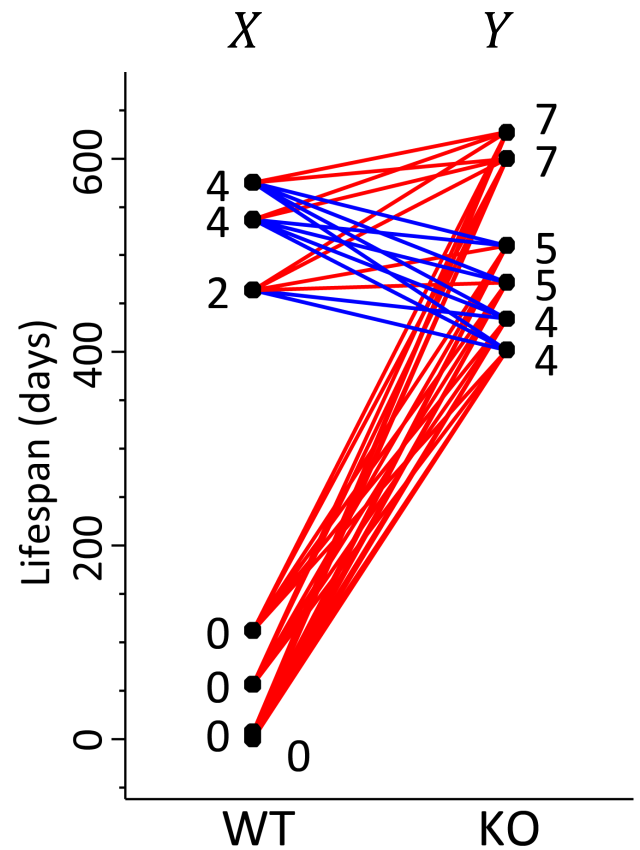
  $x_1, x_2, \dots, x_{n_x}$
  $y_1, y_2, \dots, y_{n_y}$

- For each x_i count the number of y_j, such that $x_i > y_j$
- The sum of these counts over all x_i is U_x

- Do the same for y_j and find U_y

- Test statistic

$$U = \min(U_x, U_y)$$



$U_x = 10$ $U_y = 32$

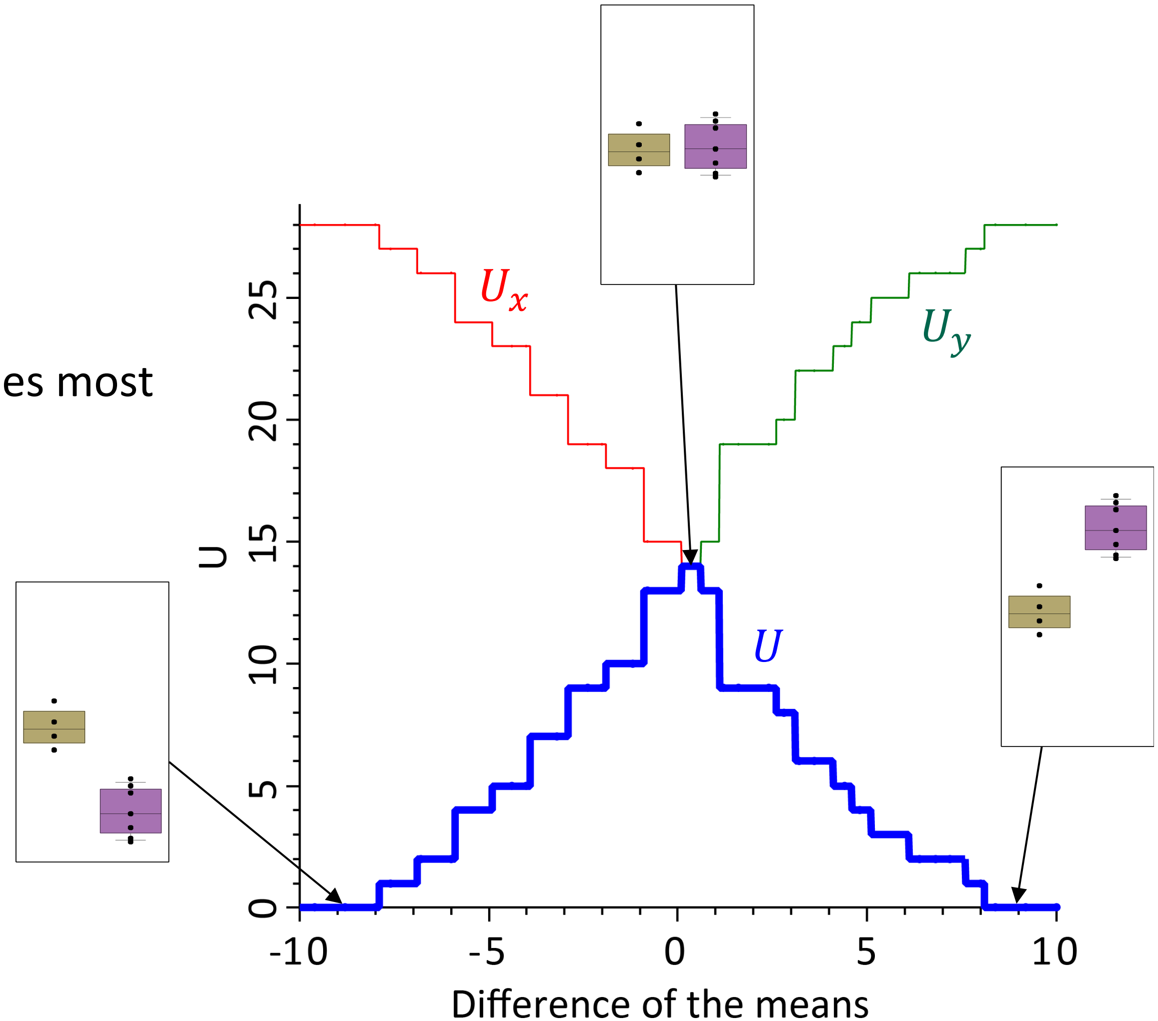
$U = 10$

Mann-Whitney test

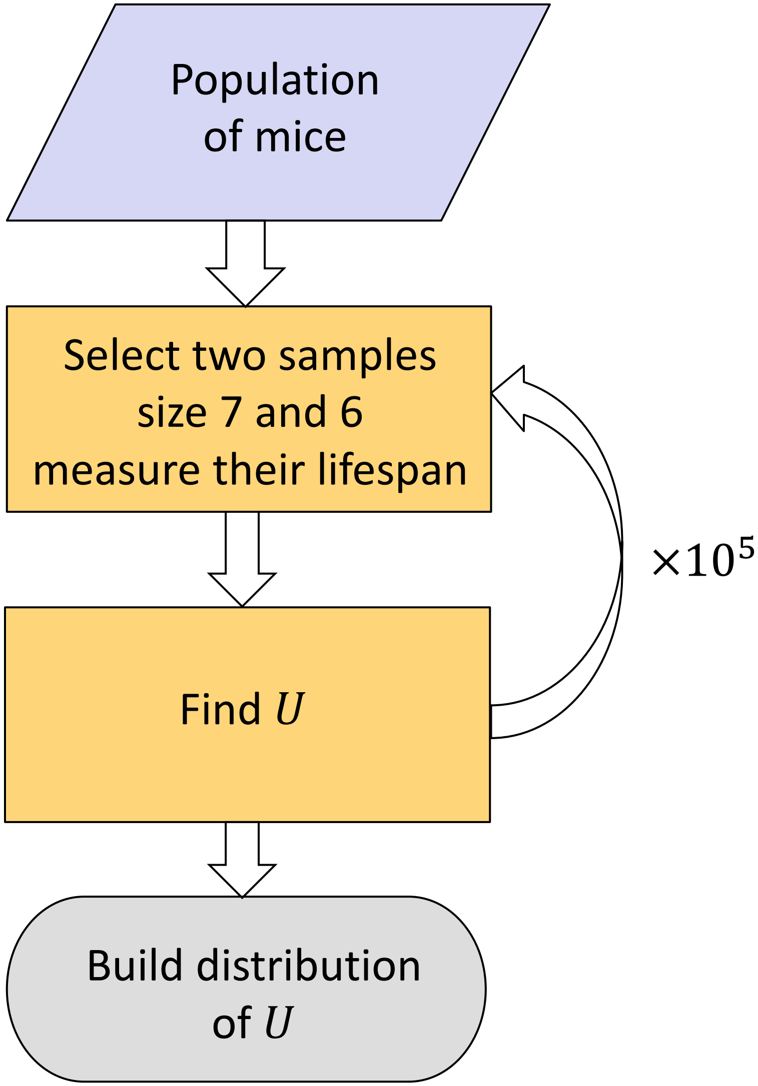
- U measures difference in location between the samples

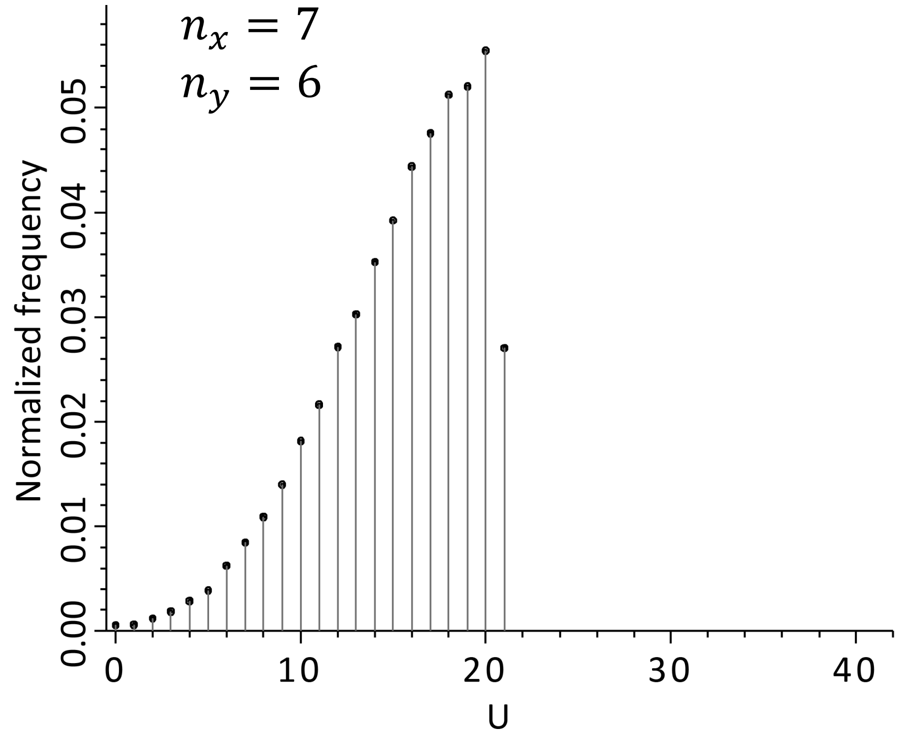
- With no overlap $U = 0$
- Direction not important

- $U = \max = \left\lfloor \frac{n_x n_y}{2} \right\rfloor$ when samples most similar



Null distribution





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

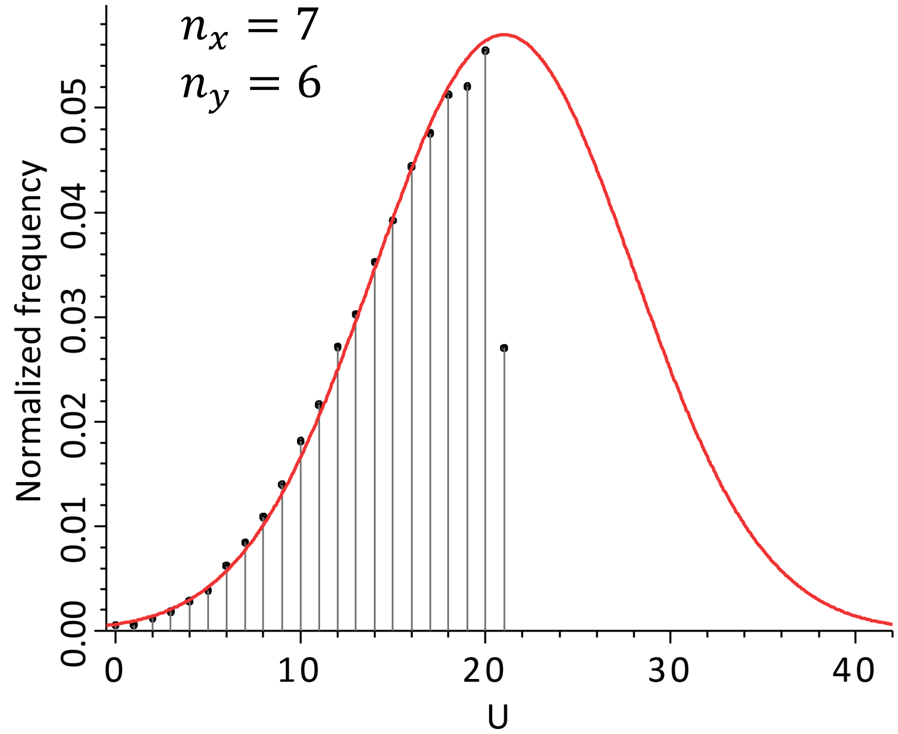
Null distribution

- For large samples U is approximately normally distributed (half of it) with

$$\mu_U = \frac{n_x n_y}{2}$$

$$\sigma_U = \sqrt{\frac{n_x n_y (n_x + n_y + 1)}{12}}$$

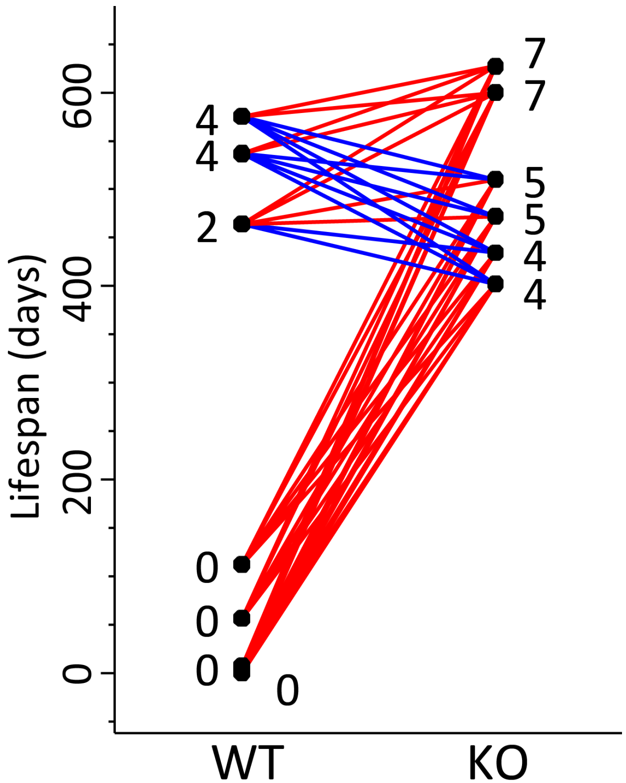
- For smaller samples exact solutions are available (tables or software)

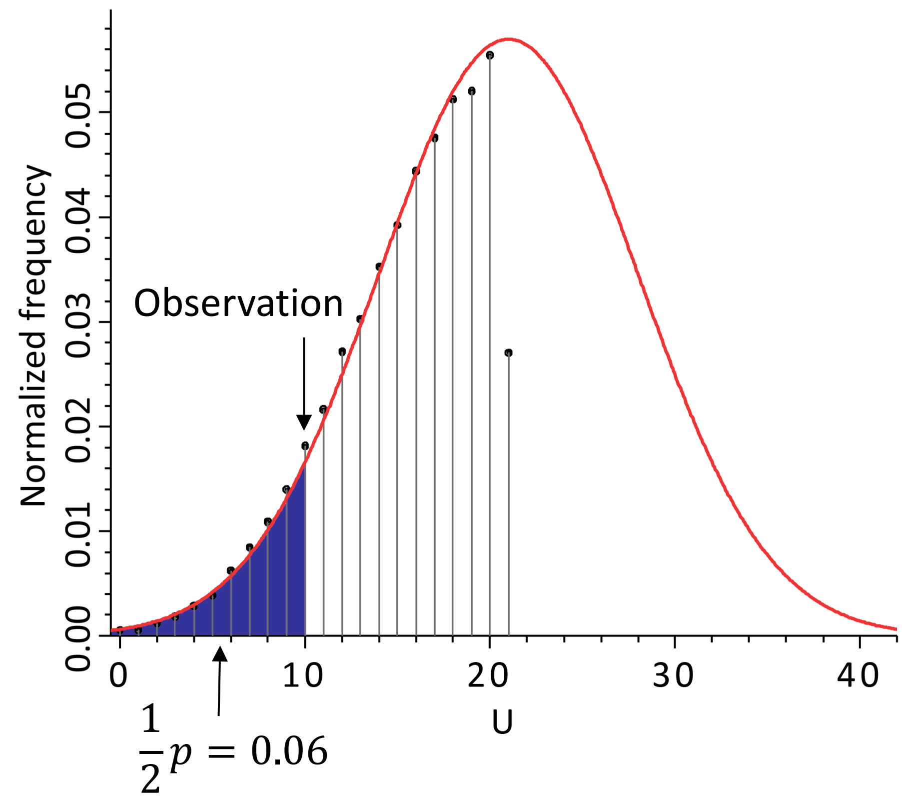


$$\mu_U = \frac{7\times 6}{2} = 21$$

$$\sigma_U = \sqrt{\frac{7\times 6\times(7+6+1)}{12}} = 7$$

P-value





$U_x = 10 \qquad U_y = 32$

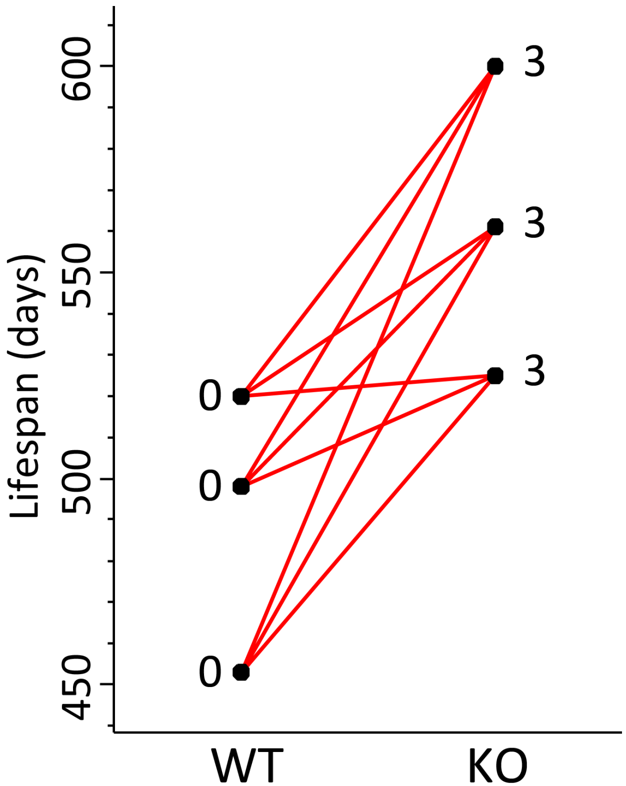
$U = 10$

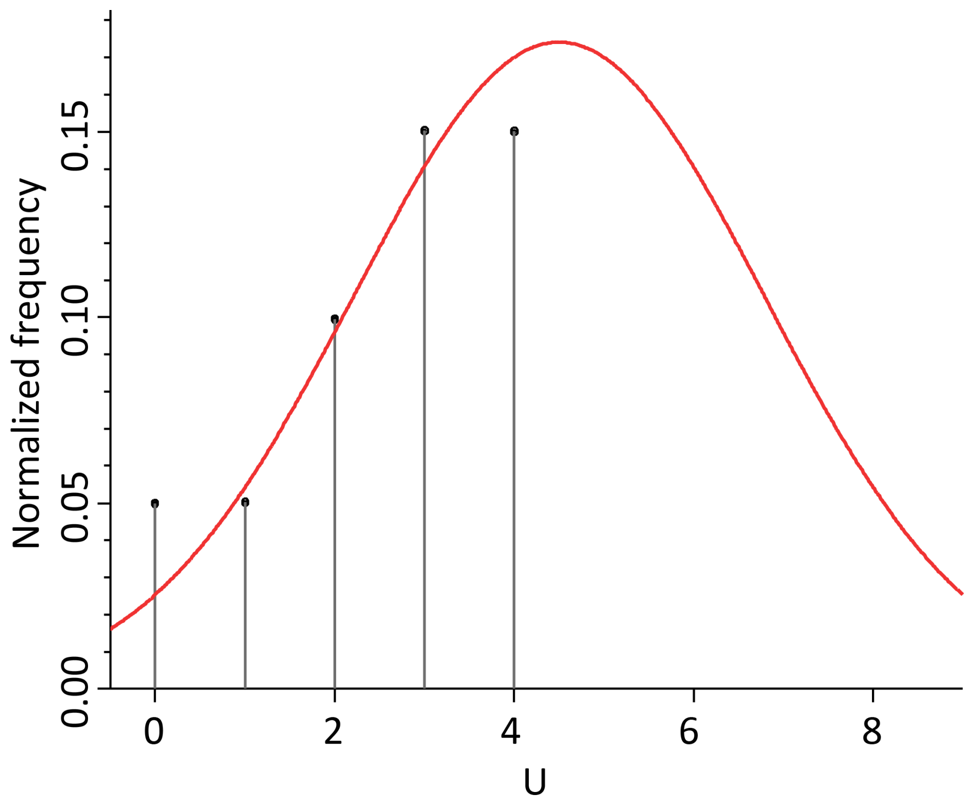
$\mu_U = 21$ g
$\sigma_U = 7$ g
$Z = \frac{U - \mu_U}{\sigma_U} = -1.57$
$p = 0.12$
Exact solution: $p = 0.14$

Limited usage for small samples

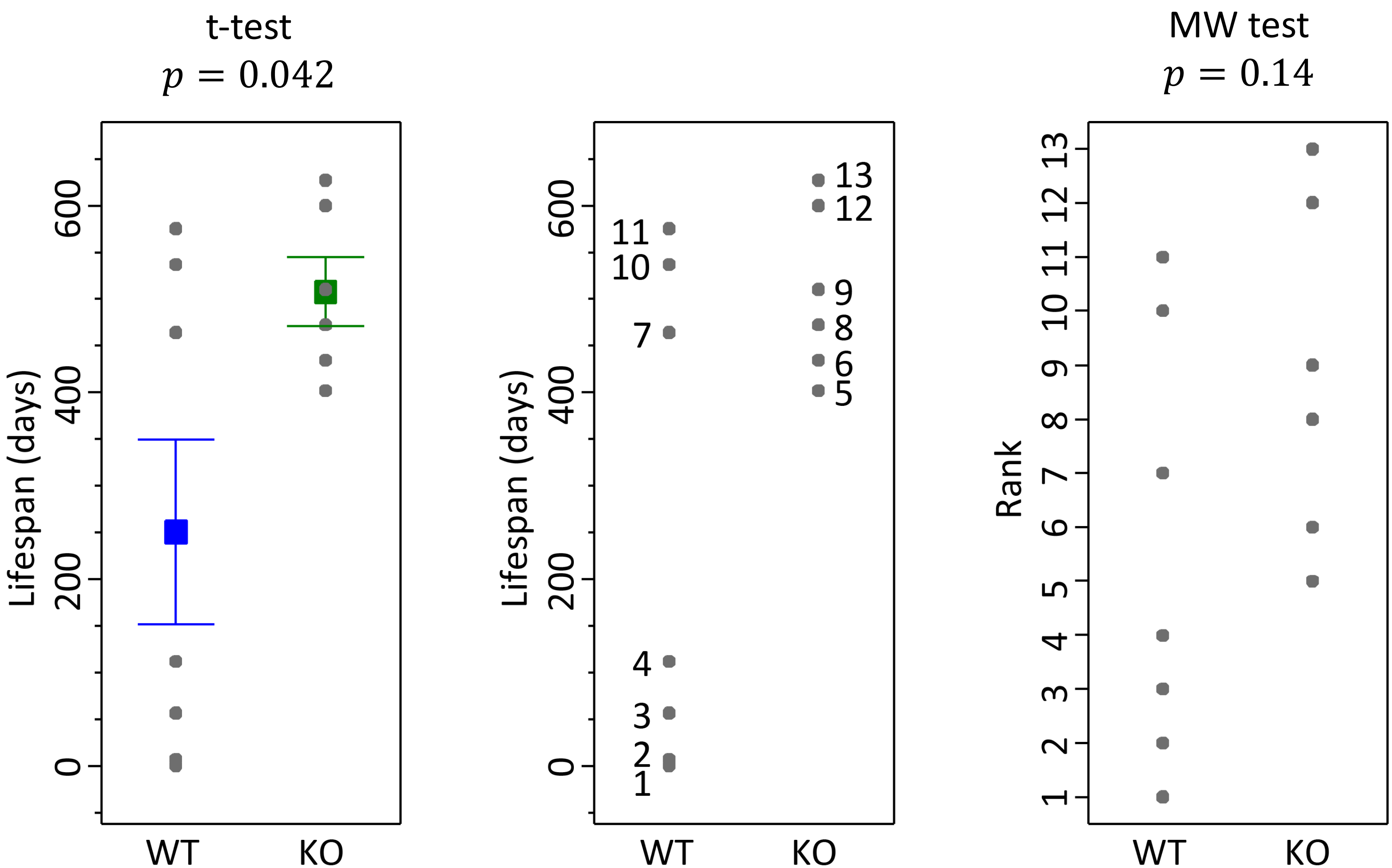
$p = 0.1$

Only 5 possible p-values:
0.1, 0.2, 0.4, 0.7, 1





Comparison to t-test



Mann-Whitney can compare medians, but…

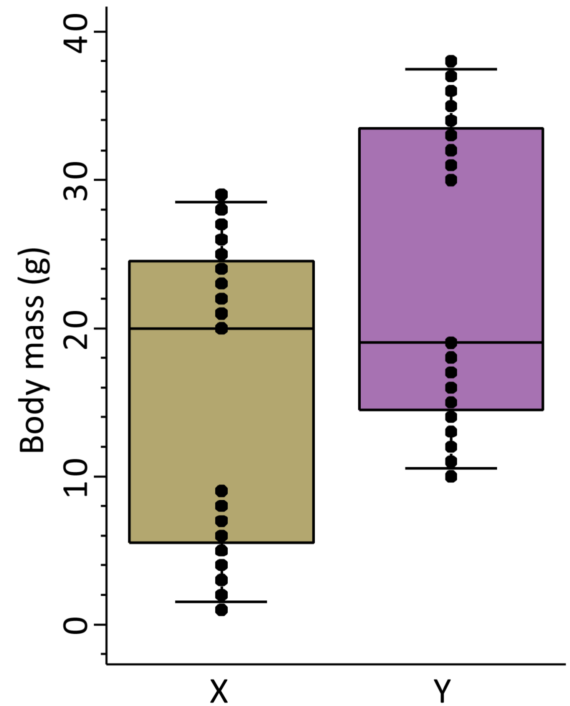
- Consider two samples in the figure
- Yes, I know they are contrived

- Medians are similar, but $\text{med}\, X > \text{med}\, Y$

- Mann-Whitney test gives $U = 100$ and one-sided $p = 0.02$
- Y exceeds X!

- Mann-Whitney test is sensitive to change in location (median) and/or shape
- If shapes are the same, then MW test can be a test of medians

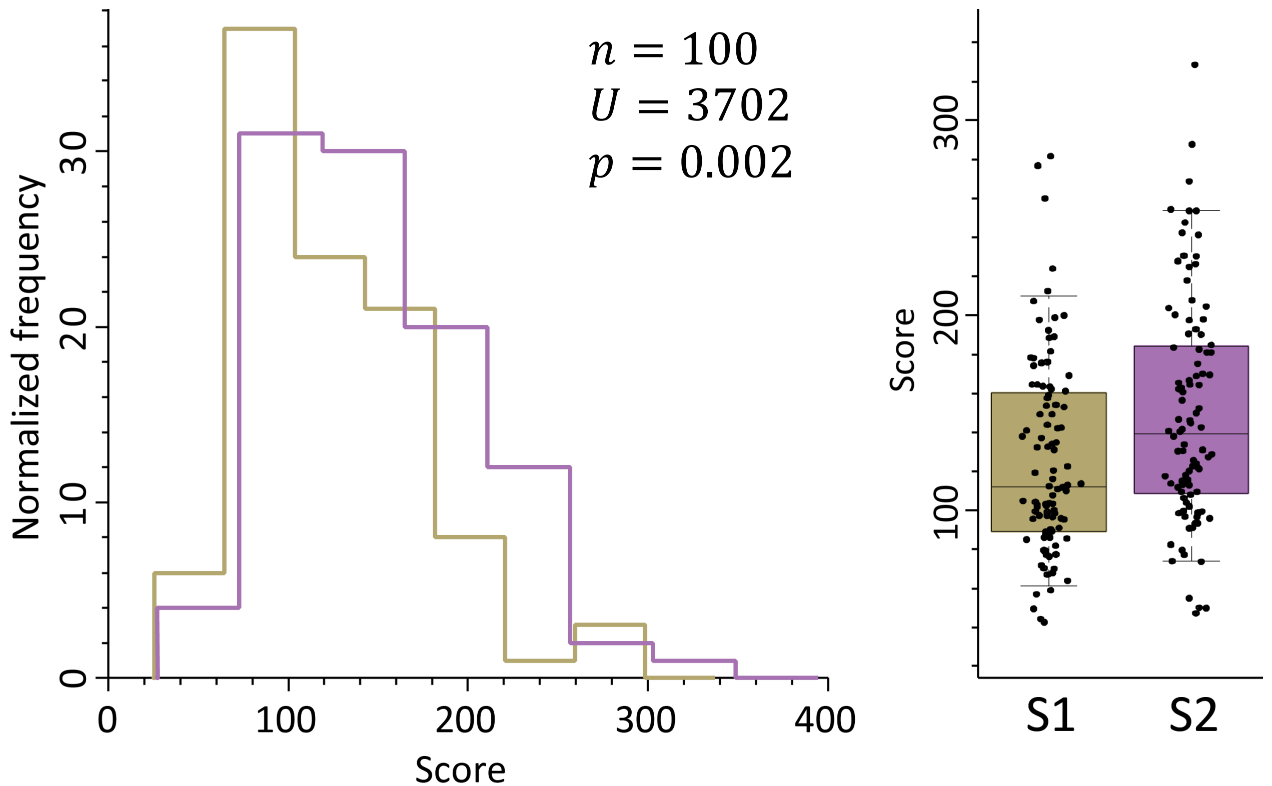
- Otherwise, use Mood's test for medians



What is Mann-Whitney test really for?

- If data are distributed (roughly) normally, use t-test

- MW test is good for weird distributions, e.g. ‘scores’
- Ordinal variables, e.g., manual score 1-5



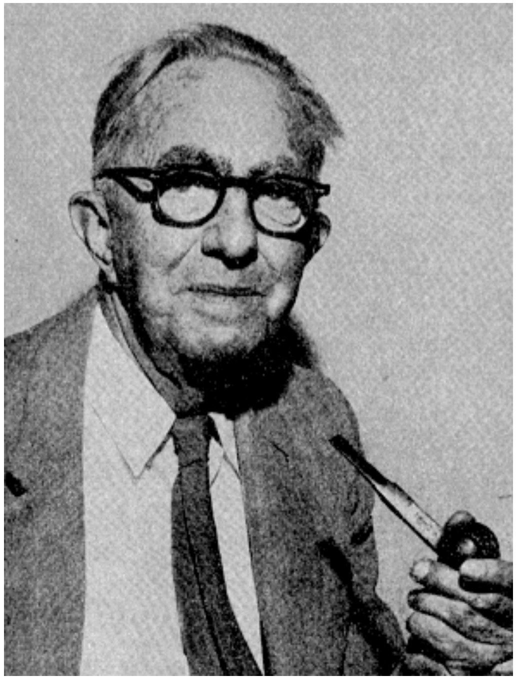
How to do it in R?

```
> x = c(0, 7, 56, 112, 464, 537, 575)
> y = c(402, 434, 472, 510, 600, 627)
# Mann-Whitney test
> wilcox.test(x, y)

        Wilcoxon rank sum test

data:  x and y
W = 10, p-value = 0.1375
alternative hypothesis: true location shift is not equal to 0

# Mood's test for medians
> mood.test(x,y)

        Mood two-sample test of scale

data:  x and y
Z = 0.55995, p-value = 0.5755
alternative hypothesis: two.sided
```

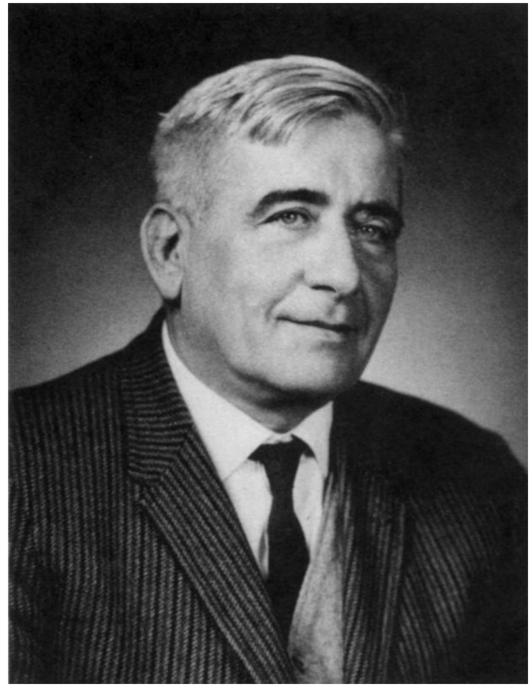
Mann-Whitney test: summary

Input	two samples of n_1 and n_2 values values can be ordinal
Assumptions	Samples are random and independent (no before/after tests) If used for medians, both distributions must be the same
Usage	Compare location and shape of two samples
Null hypothesis	There is no shift in location and/or change in shape Stronger version: both samples are from the same distribution
Comments	Non-parametric counterpart of t-test Less powerful than t-test (use t-test if distributions symmetric) Not very useful for small samples Doesn’t really give the effect size

Mann-Whitney-Wilcoxon



Frank Wilcoxon
(1892-1965)



Henry Berthold Mann
(1905-2000)

Donald Ransom Whitney
(1915-2007)

Wilcoxon, F. (1945) ”Individual Comparisons by Ranking Methods” *Biometrics Bulletin* **1**, 80–83

Mann, H. B.; Whitney, D. R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other" *Annals of Mathematical Statistics* **18**, 50–60

Wilcoxon signed-rank test

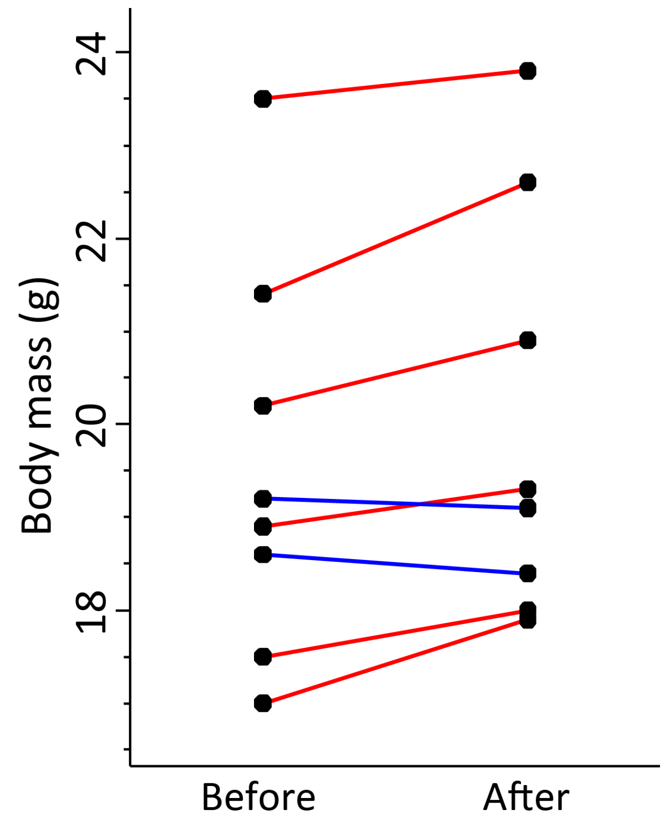
a nonparametric alternative to paired t-test

Paired data

- Samples are paired
- For example: mouse weight before and after obesity treatment

- Null hypothesis: difference between pairs follows a symmetric distribution around zero

- Example: mouse body mass (g)



Before:	21.4	20.2	23.5	17.5	18.6	17.0	18.9	19.2
After:	22.6	20.9	23.8	18.0	18.4	17.9	19.3	19.1

Wilcoxon signed-rank test

- Find the differences:

  $$\Delta_i = |y_i - x_i|$$

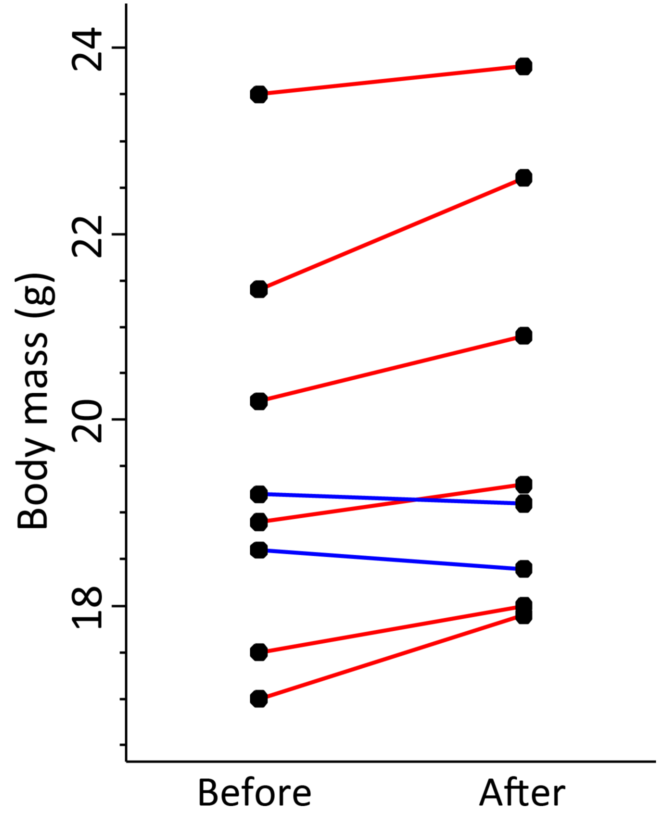
  $$s_i = \mathrm{sgn}(y_i - x_i)$$

- Order and rank the pairs according to Δ_i

  R_i - rank of the *i*-the pair

- Test statistic:

  $$W = \sum_{i=1}^{n} s_i R_i$$

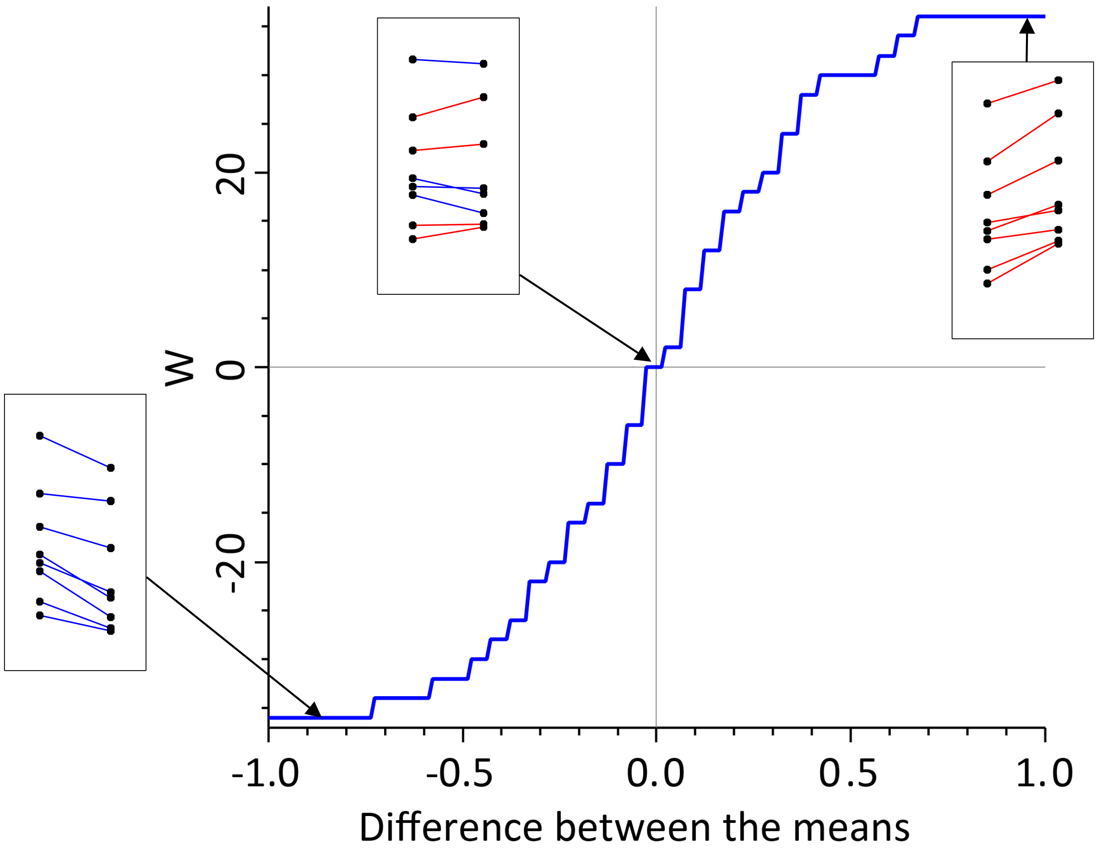


Wilcoxon signed-rank test

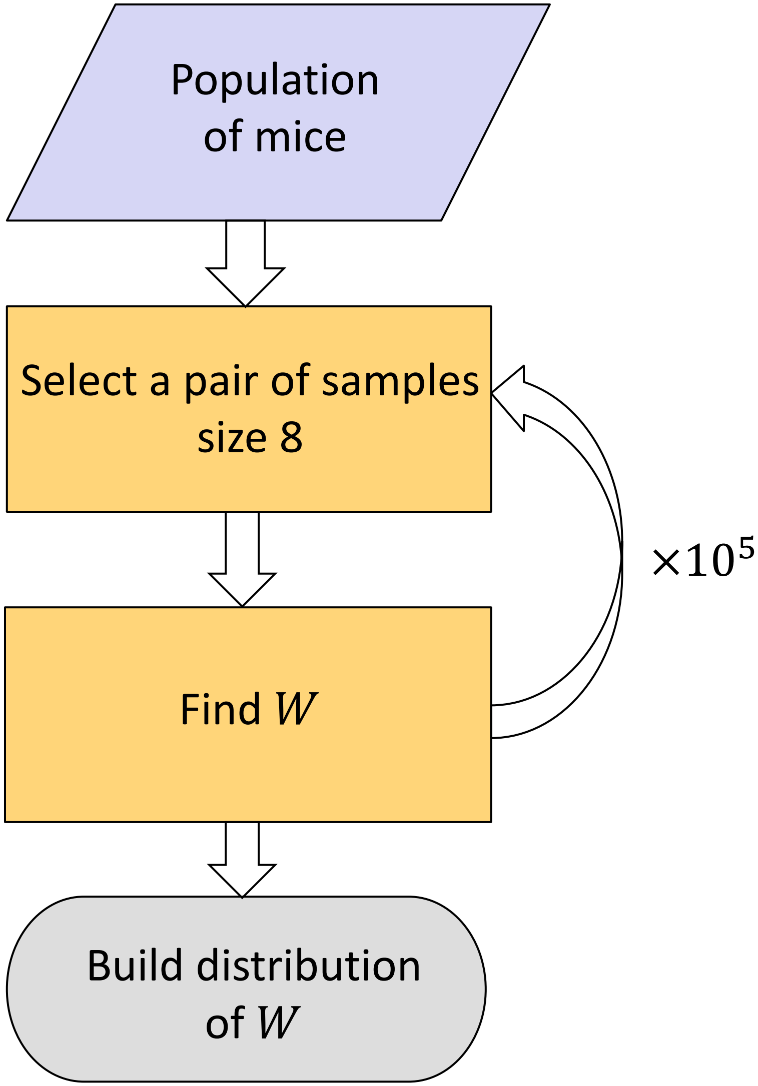
- W measures difference in location between pairs of points

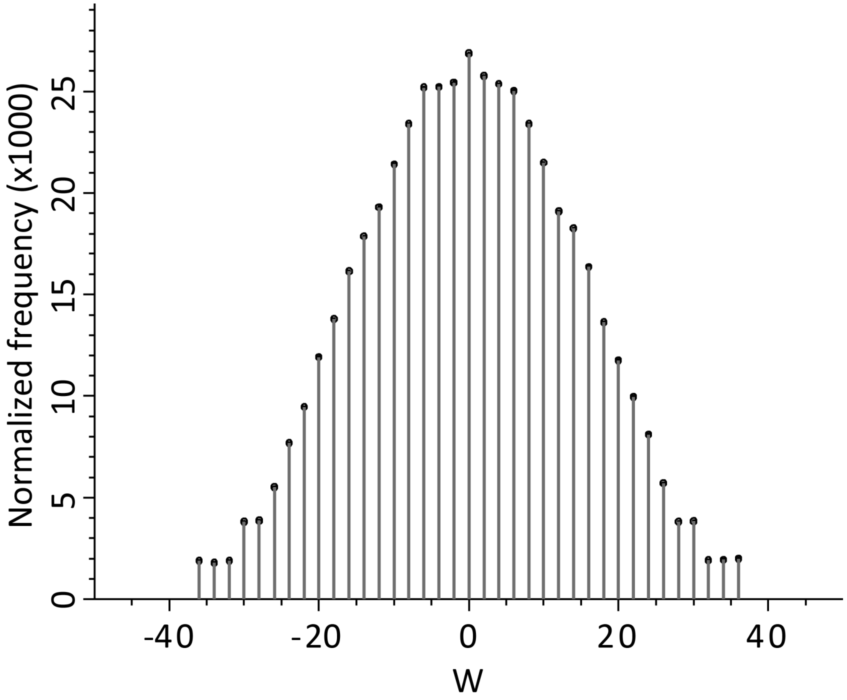
- Direction is important

- $W = 0$ when samples most similar



Null distribution





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

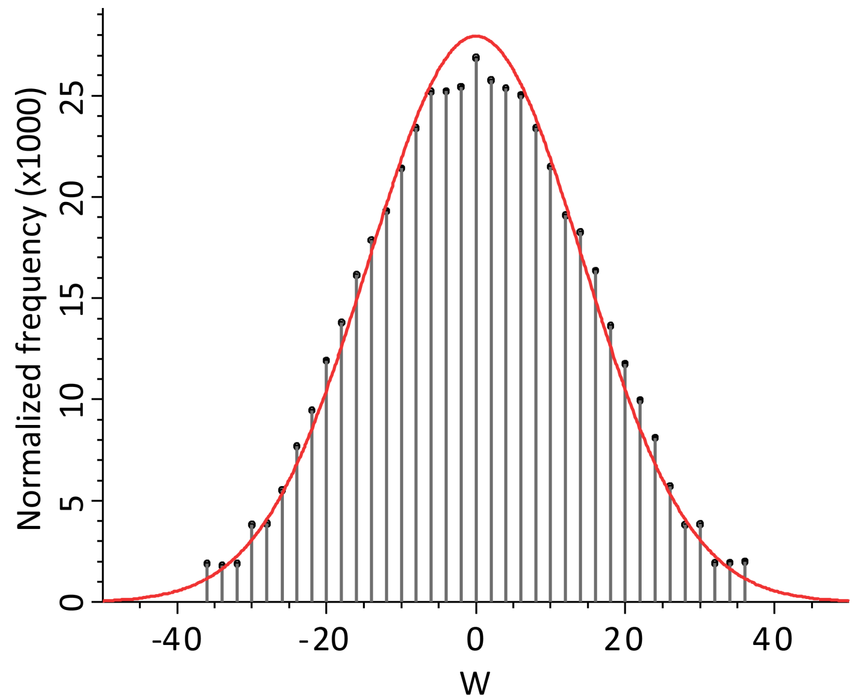
Null distribution

- For large samples W is approximately normally distributed with

$$\mu_W = 0$$

$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

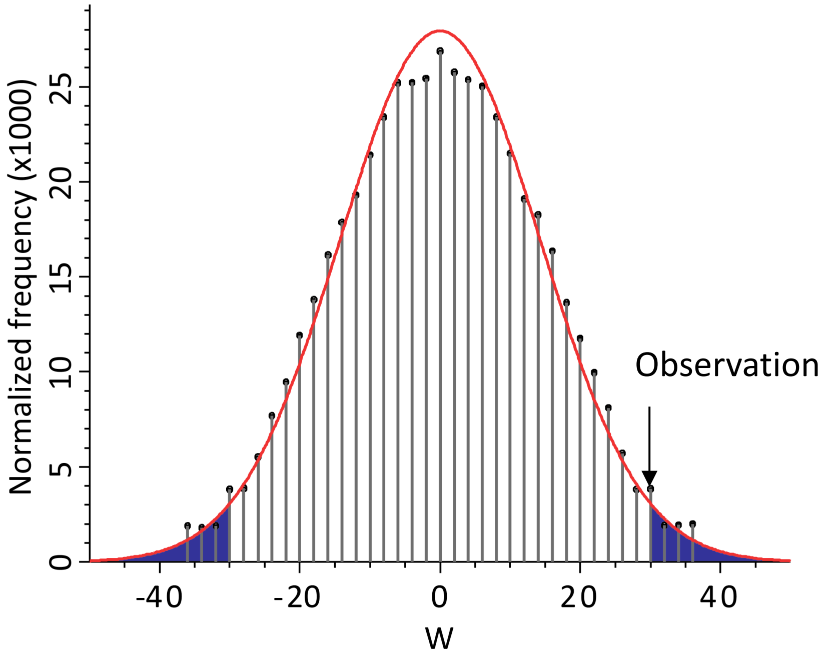
- For smaller samples exact solutions are available (tables or software)



$$n = 8$$

$$\sigma_W = \sqrt{\frac{8 \times 9 \times 17}{6}} = \sqrt{204} \approx 14.3$$

P-value



x_i	y_i	Δ_i	R_i	s_i	$s_i R_i$
19.2	19.1	0.1	1	-1	-1
18.6	18.4	0.2	2	-1	-2
23.5	23.8	0.3	3	1	3
18.9	19.3	0.4	4	1	4
17.5	18.0	0.5	5	1	5
20.2	20.9	0.7	6	1	6
17.0	17.9	0.9	7	1	7
21.4	22.6	1.2	8	1	8
					30

$n = 8$

$W = 30$

$\sigma_W = 14.3$

$Z = W/\sigma_W = 2.10$

$p = 0.036$

$p_{\text{exact}} = 0.039$

How to do it in R?

```
# Paired t-test
> before = c(21.4, 20.2, 23.5, 17.5, 18.6, 17.0, 18.9, 19.2)
> after = c(22.6, 20.9, 23.8, 18.0, 18.4, 17.9, 19.3, 19.1)
> wilcox.test(before, after, paired=T)

        Wilcoxon signed rank test

data:  before and after
V = 3, p-value = 0.03906
alternative hypothesis: true location shift is not equal to 0
```

Wilcoxon signed-rank test: summary

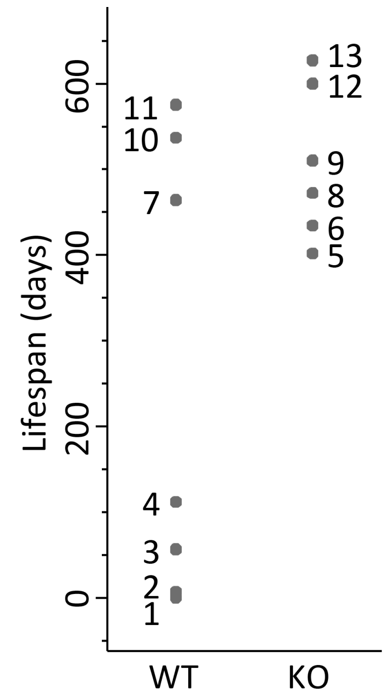
Input	Sample of n pairs of data (*before* and *after*) Values can be ordinal
Assumptions	Pairs should be random and independent
Usage	Discover change in individual points between *before* and *after*
Null hypothesis	There is no change between *before* and *after* is zero The difference between *before* and *after* follows a symmetric distribution around zero
Comments	Non-parametric counterpart of paired t-test Paired data only Doesn’t care about distributions Not very useful for small samples

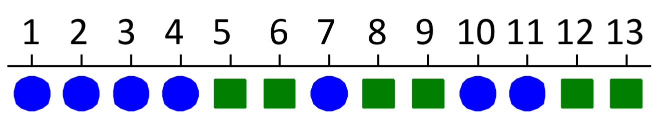
Kruskal-Wallis test

a nonparametric alternative to one-way ANOVA

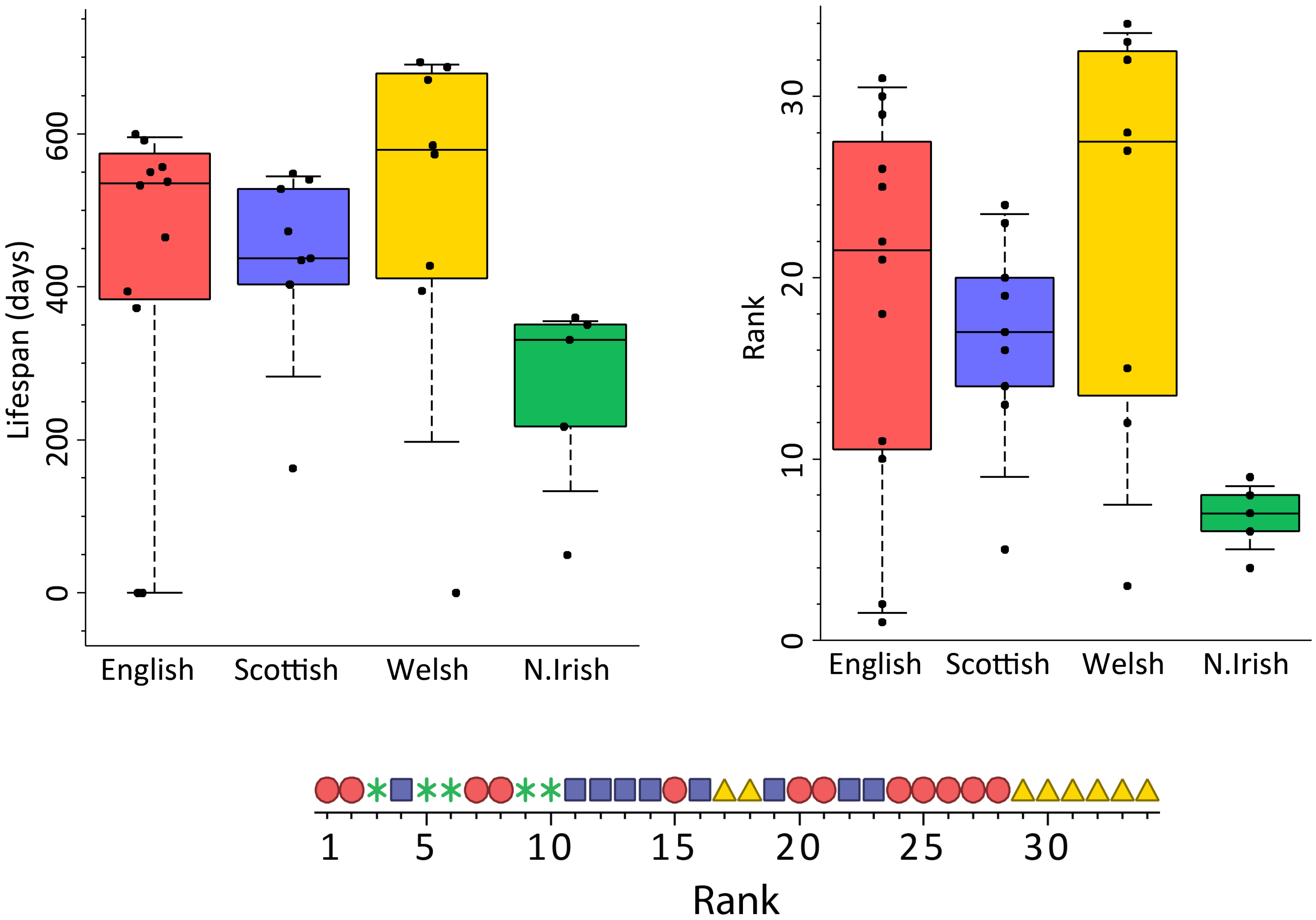
Alternative formulation of the Mann-Whitney test

- Rank pooled data from the smallest to the largest
- Null hypothesis: both samples are randomly distributed between available rank slots
- Can be extended to more than 2 samples

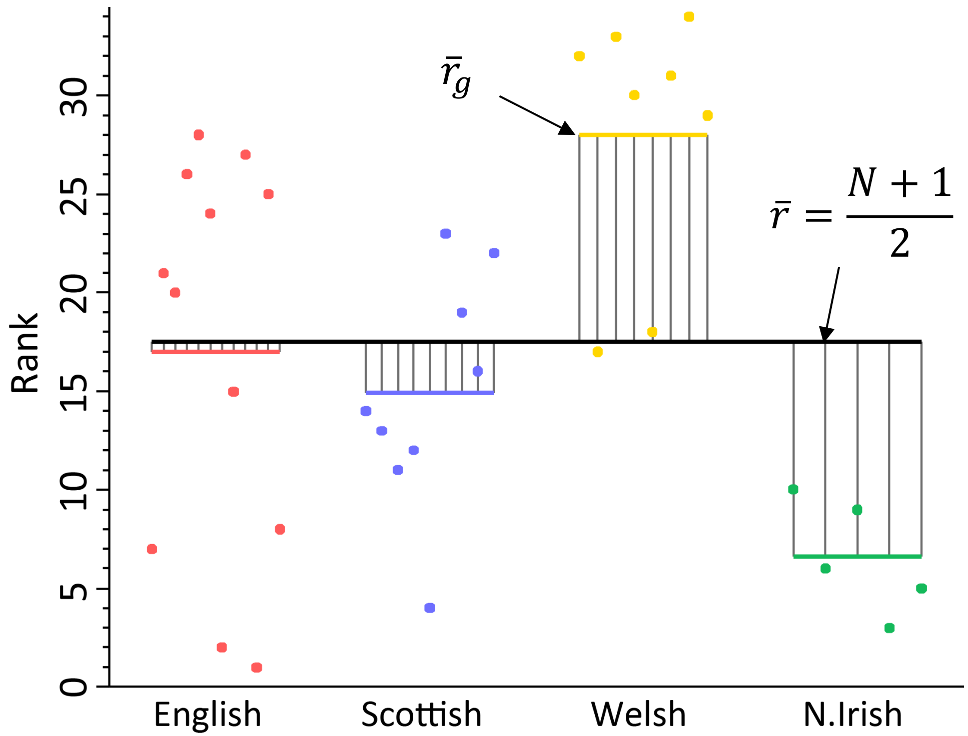




Ranked ANOVA



Variance between groups



Test statistic

- Sum of square residuals

$$SS_B = \sum_{g=1}^{n} n_g\left(\bar{r}_g - \bar{r}\right)^2$$

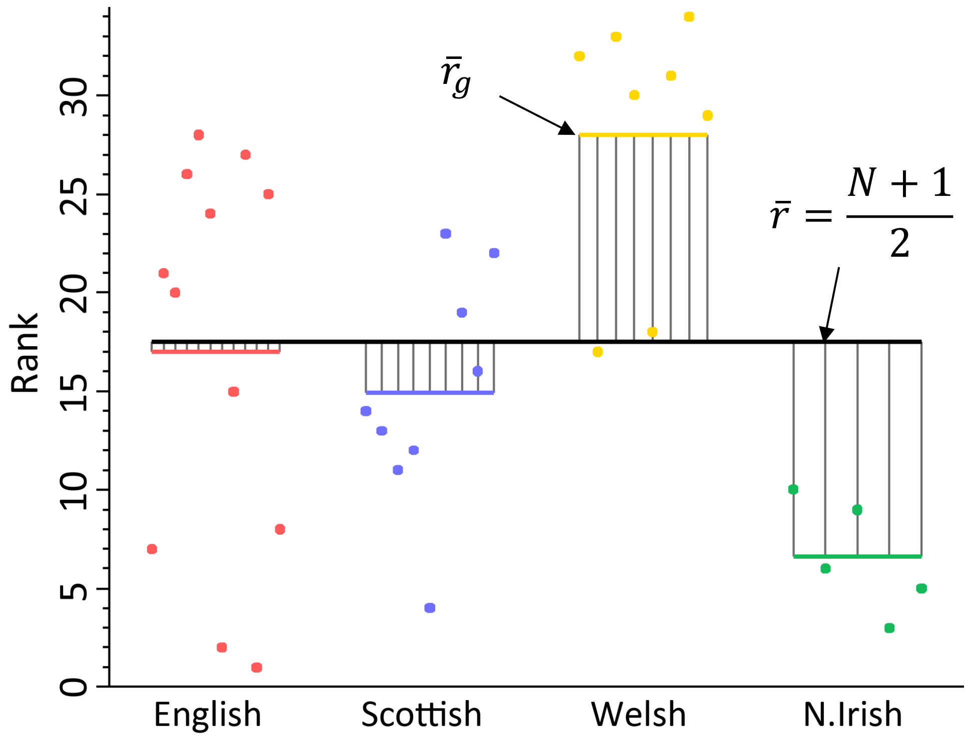
- Rank variance

$$\sigma^2 = \frac{1}{12}N(N+1)$$

- Test statistic

$$H = \frac{SS_B}{\sigma^2}$$

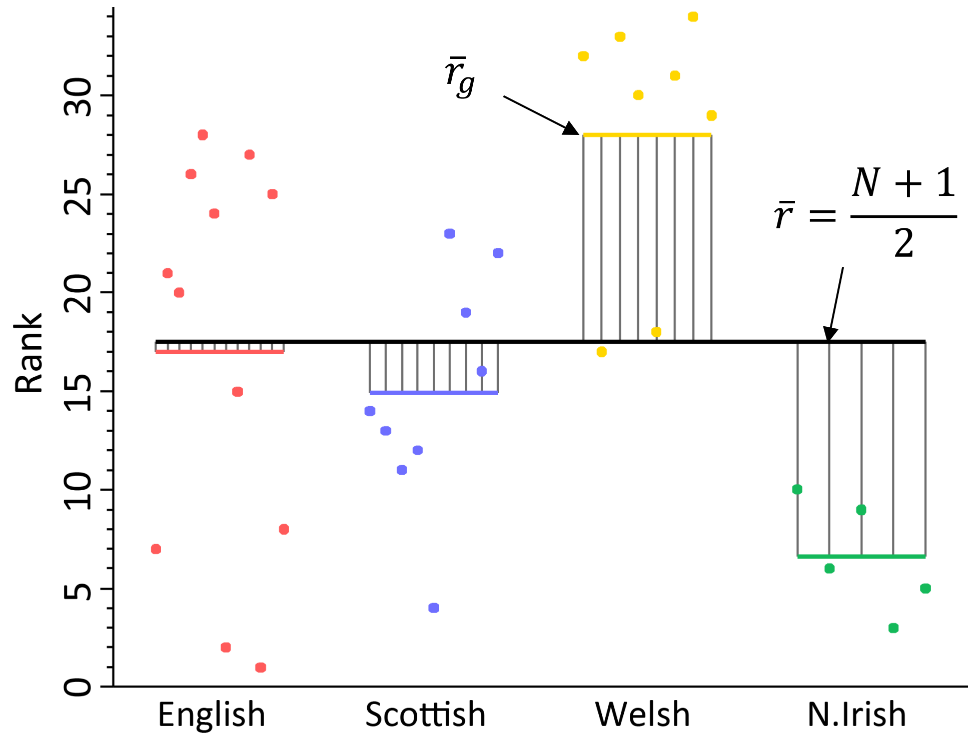
$$H = \frac{12}{N(N+1)}\sum_{g=1}^{n} n_g\left(\bar{r}_g - \frac{N+1}{2}\right)^2$$



Test statistic

$$H = \frac{12}{N(N+1)} \sum_{g=1}^{n} n_g \left(\bar{r}_g - \frac{N+1}{2} \right)^2$$

- where
  - n_g – number of points in group g
  - $\bar{r}_g$ – mean rank in group g
  - $\bar{r} = (N+1)/2$ – mean rank
  - N – number of all points
  - n – number of groups



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

- Null hypothesis: mean rank in each group is the same as total mean rank

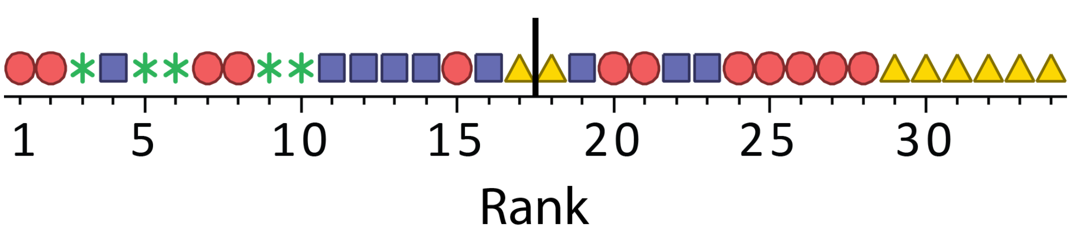
$$H_0 \colon \bar{r}_g = \frac{N+1}{2}$$

$$H = \frac{1}{\sigma^2} \sum_{g=1}^{n} n_g \left(\bar{r}_g - \bar{r} \right)^2$$

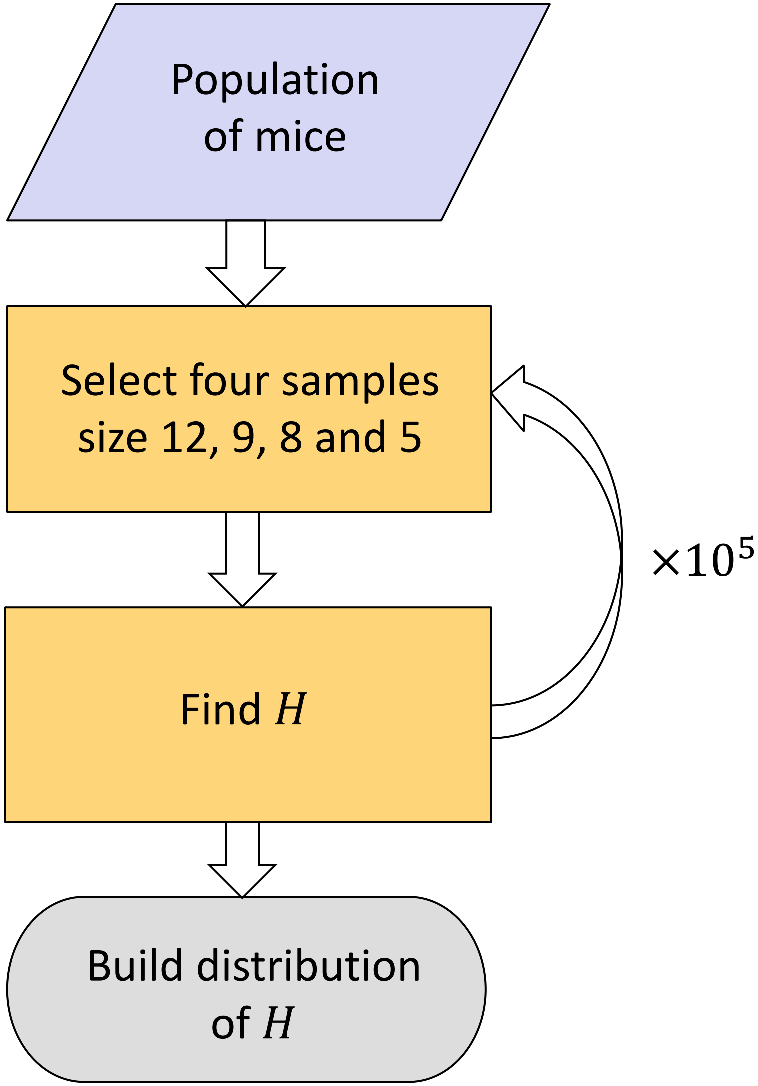
$$\bar{r} = \frac{N+1}{2} = 17.5 \qquad \sigma^2 = \frac{N(N+1)}{12} = 99.2$$

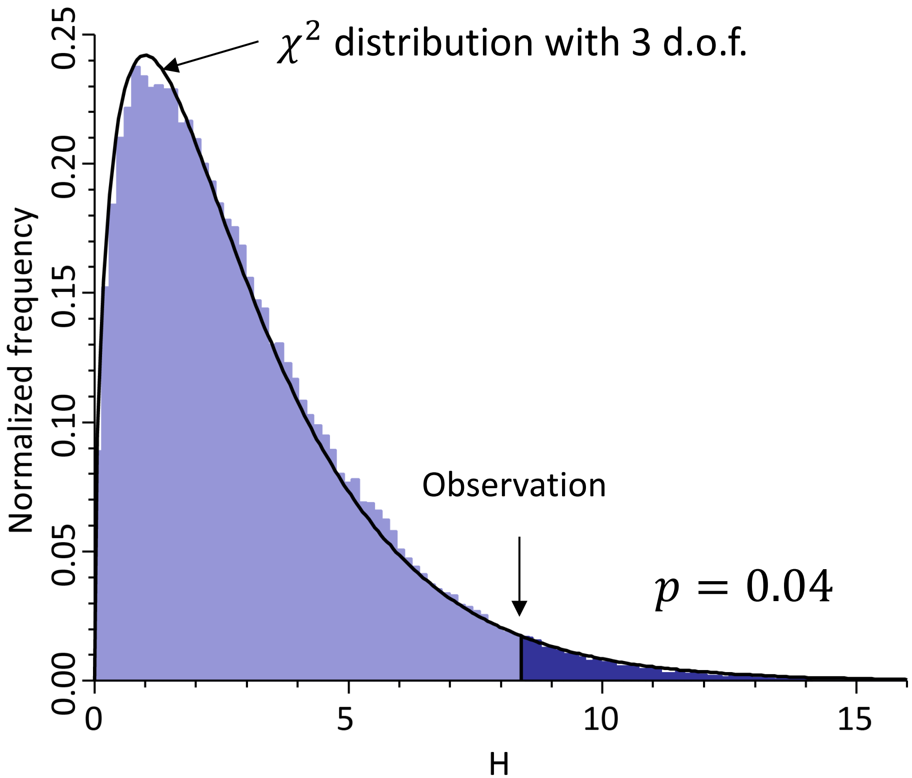
		English	Scottish	Welsh	N. Irish
Number	n_g	12	9	8	5
Mean rank	$\bar{r}_g$	18.96	16.78	22.81	$\boxed{6.80}$
Contribution to H	$\frac{n_g\left(\bar{r}_g - \bar{r}\right)^2}{\sigma^2}$	0.258	0.047	2.27	$\boxed{5.77}$

$$H = 8.36$$



Null distribution





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

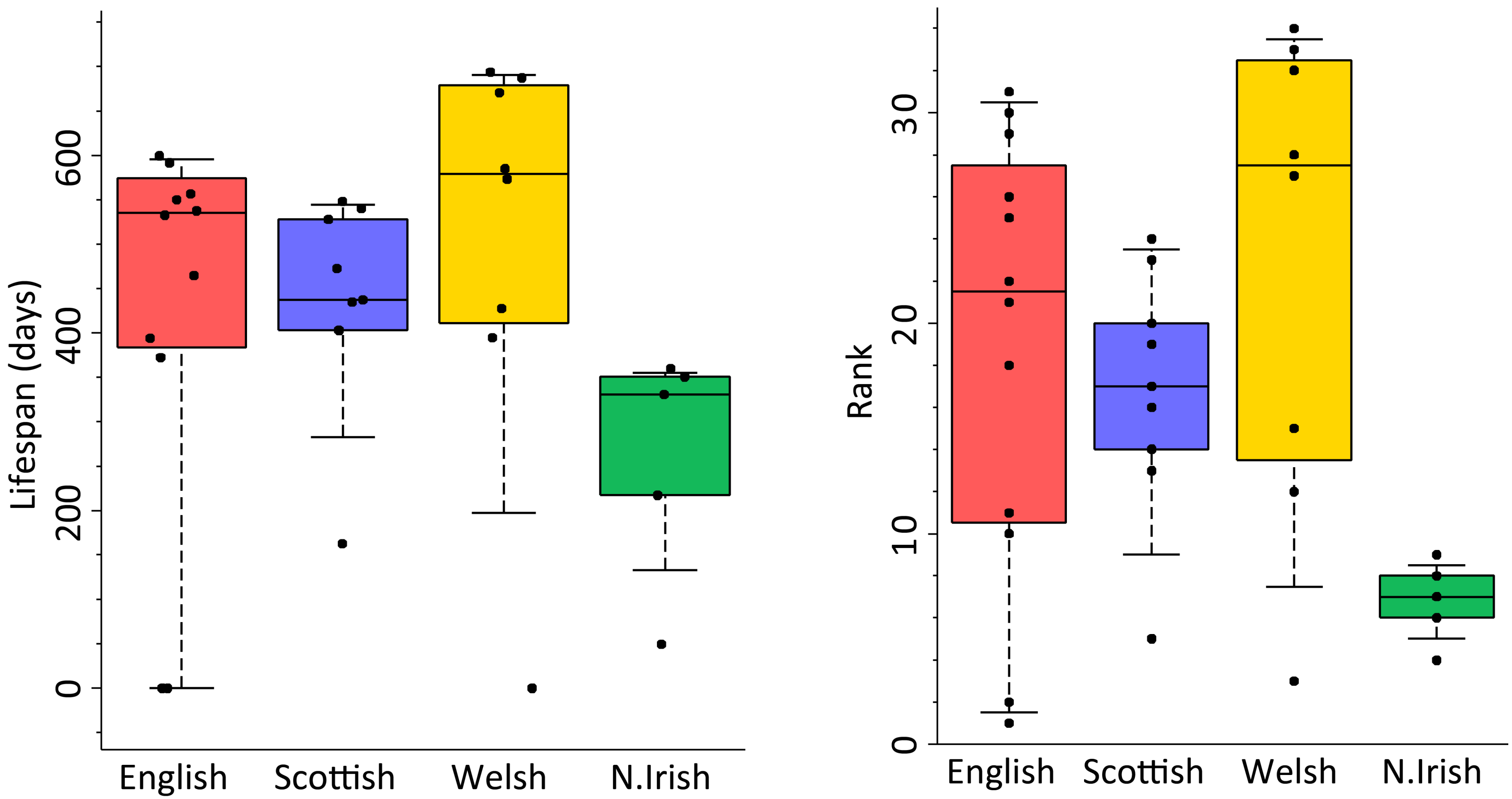
Comparison to ANOVA

ANOVA

$p = 0.18$

Kruskal-Wallis

$p = 0.04$



How to do it in R?

```
> mice = read.table('http://tiny.cc/mice_kruskal', header=T)
> kruskal.test(Lifespan ~ Country, data=mice)

        Kruskal-Wallis rank sum test

data:  Lifespan by Country
Kruskal-Wallis chi-squared = 8.3617, df = 3, p-value = 0.0391
```

What about two-way test?

- Scheirer-Ray-Hare extension to Kruskal-Wallis test
- Briefly: replace values with ranks and carry out two-way ANOVA

Scheirer C.J., Ray W.S. and Hare N (1976), The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs, *Biometrics,* **32**, 429-434

Kruskal-Wallis test: summary

Input	n samples of values N values divided into n groups
Assumptions	Samples are random and independent
Usage	Compare location and shape of n samples
Null hypothesis	Mean rank in each group is the same as total mean rank There is no change between groups
Comments	Doesn’t care about distributions

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