P-values and statistical tests
4. ANOVA

Marek Gierliński
Division of Computational Biology

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

1. Introduction

Null hypothesis, statistical test, p-value
Fisher's test

2. Contingency tables

Chi-square test
G-test

3. T-test

One- and two-sample
Paired
One-sample variance test

4. ANOVA

One-way
Two-way

5. Non-parametric methods 1

Mann-Whitney
Wilcoxon signed-rank
Kruskal-Wallis

6. Non-parametric methods 2

Kolmogorov-Smirnov
Permutation
Bootstrap

7. Statistical power

Effect size
Power in t-test
Power in ANOVA

8. Multiple test corrections

Family-wise error rate
False discovery rate
Holm-Bonferroni limit
Benjamini-Hochberg limit
Storey method

9. What's wrong with p-values?

A lot

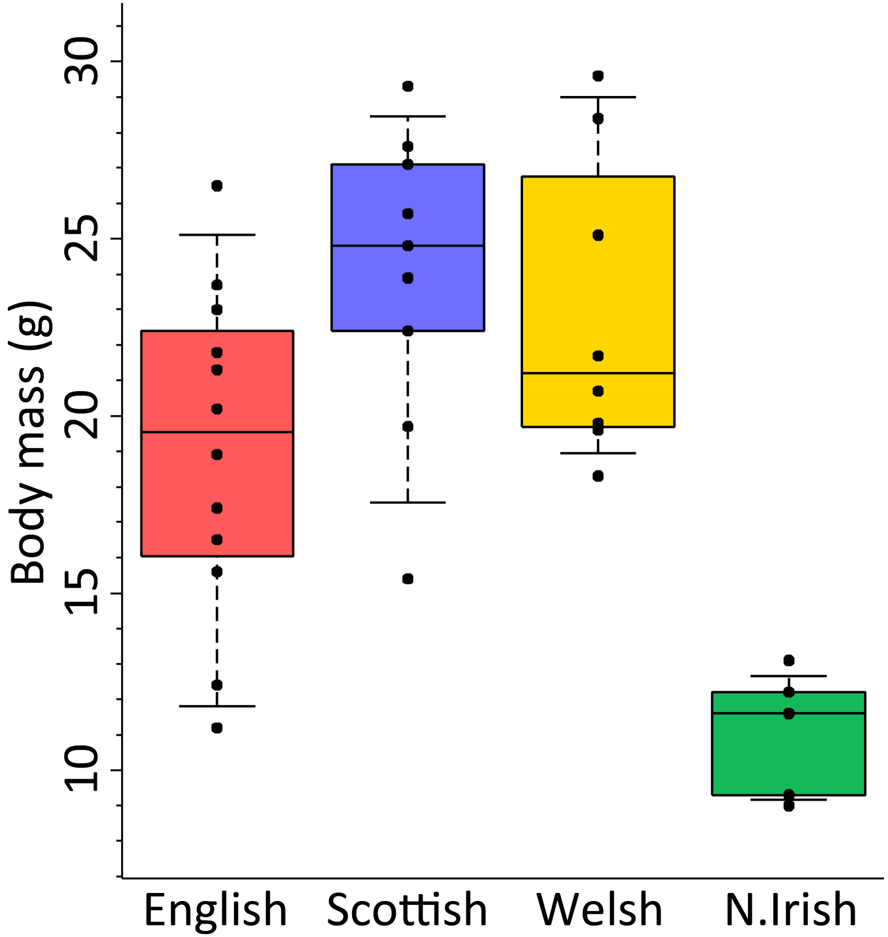
One-way ANOVA

One-way ANOVA

- Extension of the t-test to more than 2 groups

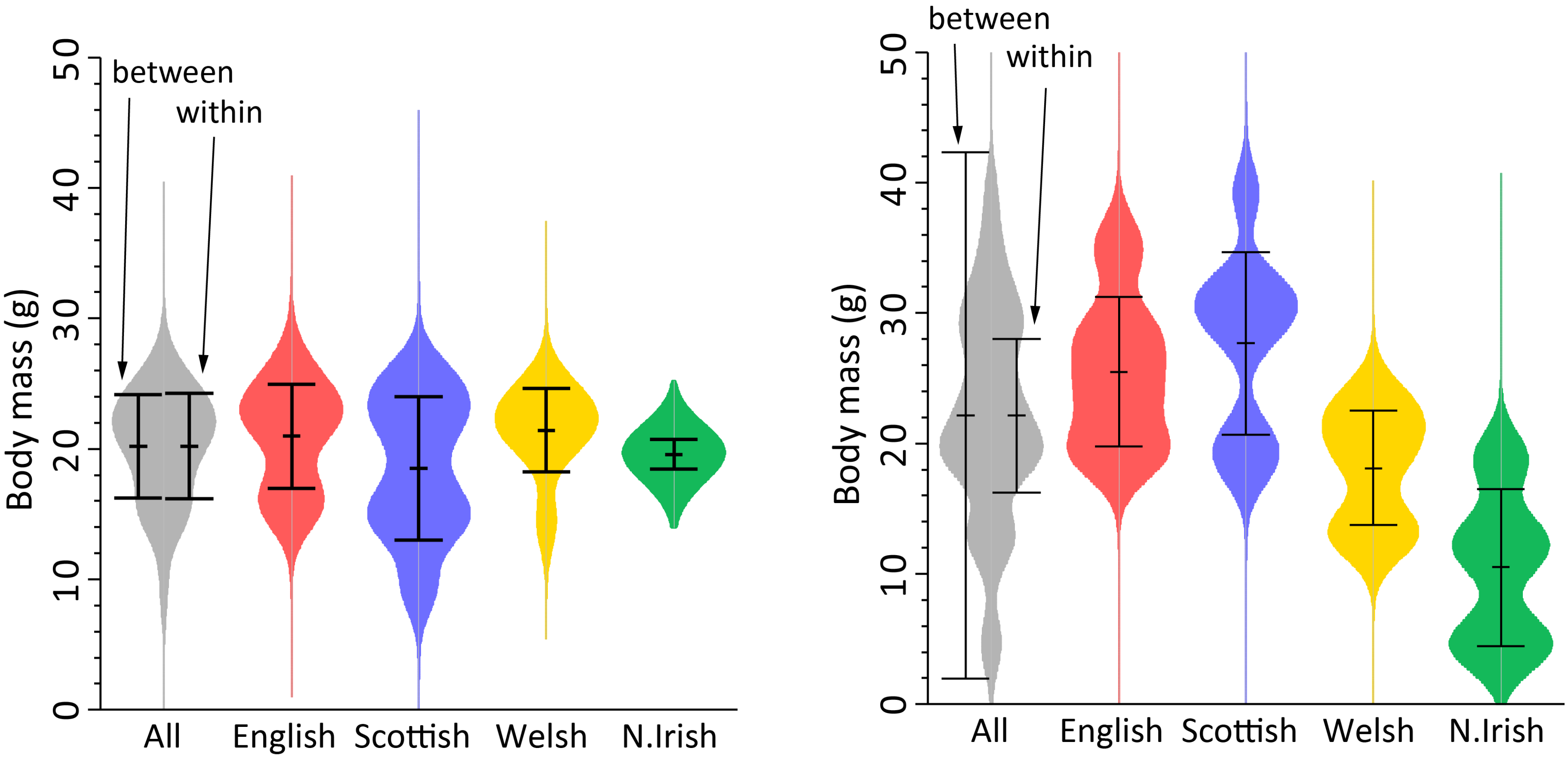
- Null hypothesis: all samples came from populations with the same mean
- H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$

- The null hypothesis is tested by comparing variances
- ANOVA – **AN**alysis **O**f **VA**riance



Variance between and within groups

- Variance within groups – typical variance in each group
- Variance between groups – how the sample mean varies from group to group



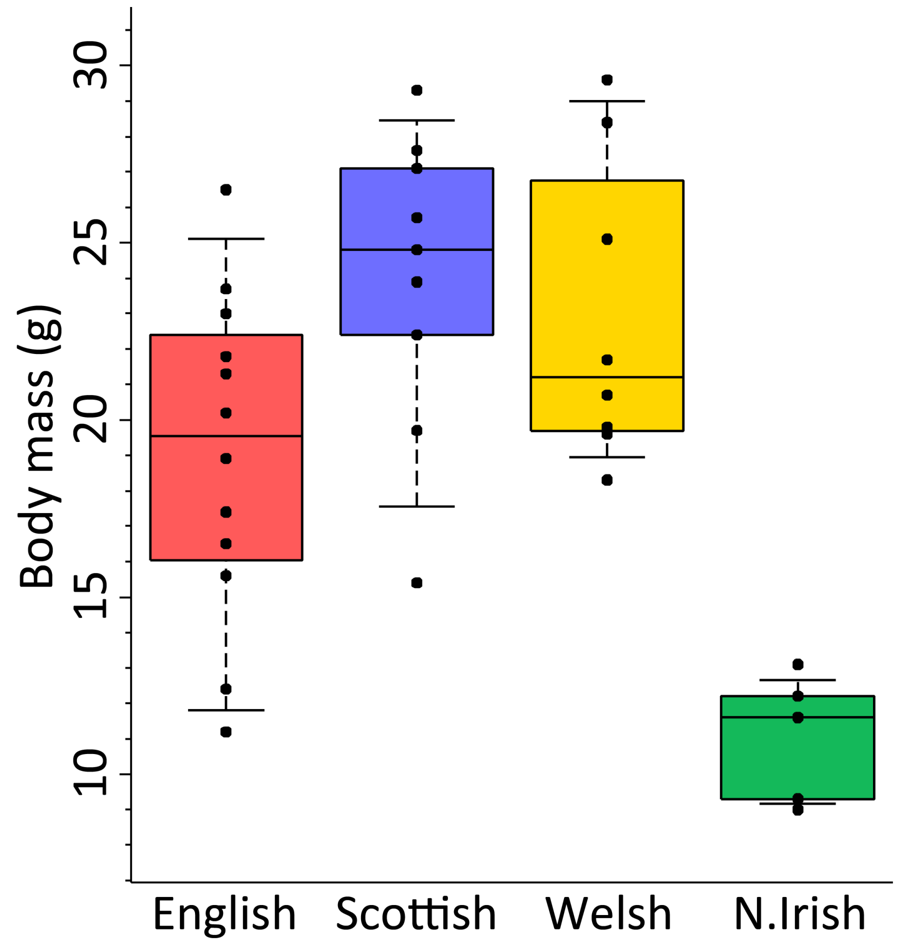
One-way ANOVA

- Null hypothesis: all samples came from populations with the same mean
- H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$

- Assumption: they all have common variance σ^2

- $n = 34$ data points
- $k = 4$ groups of data

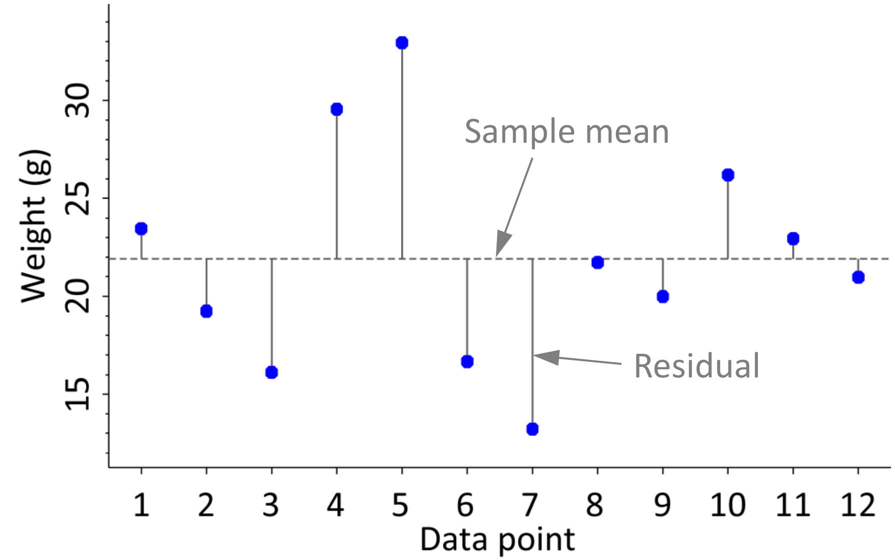
- n_g - number of points in group g
- x_{gi} - body mass, group g, mouse i
- $\bar{x}_g$ - mean in group g
- $\bar{x}$ - grand mean, across all data points



Variance

- One sample of size n
- Sample variance

$$SD_{n-1}^2 = \frac{1}{n-1}\sum_i (x_i - M)^2$$



- Generalized variance: mean square

$$MS = \frac{SS}{\nu}$$

- where
  - SS - sum of squared residuals
  - ν - number of degrees of freedom

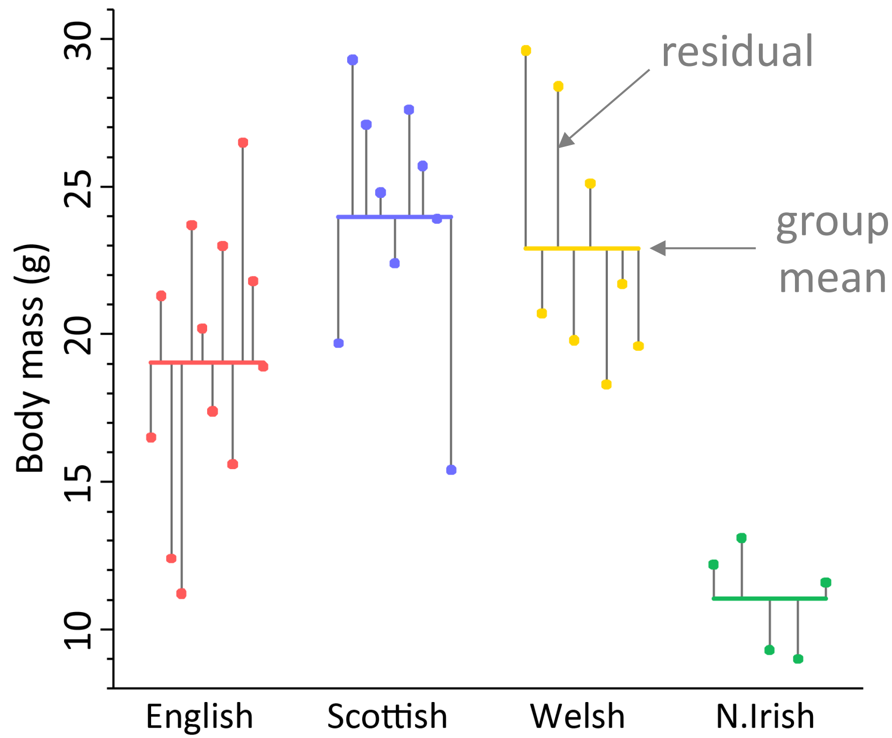
Variance within groups

- Variance within groups is

$$MS_W = \frac{SS_W}{\nu_W}$$

sum of squared residuals

number of degrees of freedom



- MS_W estimates the common variance, σ^2, regardless of the null hypothesis

$$SS_W = \sum_{g=1}^{k} \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g)^2 \qquad \nu_W = \sum_{g=1}^{k} (n_g - 1)$$

Variance within groups

- Variance within groups is

$$MS_W = \frac{SS_W}{\nu_W}$$

sum of squared residuals

number of degrees of freedom

SS_W	524
ν_W	30
MS_W	17.5

- MS_W estimates the common variance, σ^2, regardless of the null hypothesis

$$SS_W = \sum_{g=1}^{k} \sum_{i=1}^{n_g} \left(x_{gi} - \bar{x}_g\right)^2 \qquad \nu_W = \sum_{g=1}^{k} (n_g - 1)$$

Variance between groups

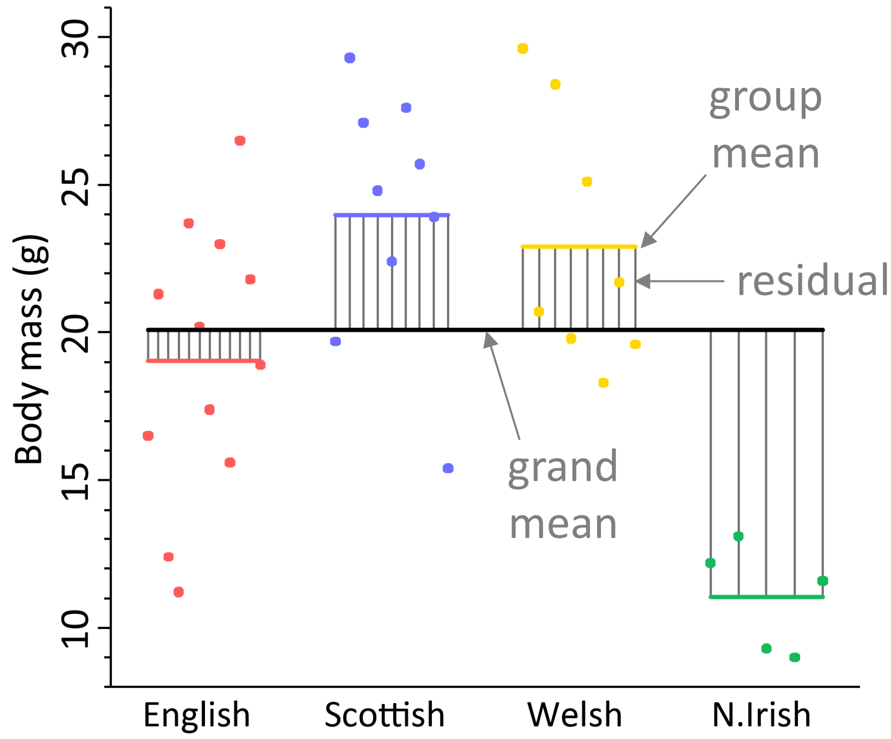
- Variance between groups is

$$MS_B = \frac{SS_B}{\nu_B}$$

sum of squared residuals

number of degrees of freedom

- MS_B estimates the common variance, σ^2, **only** when the null hypothesis is true



$$SS_B = \sum_{g=1}^{k} n_g \left(\bar{x}_g - \bar{x}\right)^2 \qquad \nu_B = k - 1$$

Variance between groups

- Variance between groups is

$$MS_B = \frac{SS_B}{\nu_B}$$

sum of squared residuals

number of degrees of freedom

SS_W	524
ν_W	30
MS_W	17.5
SS_B	623
ν_B	3
MS_B	208

- MS_B estimates the common variance, σ^2, **only** when the null hypothesis is true

$$SS_B = \sum_{g=1}^{k} n_g\left(\bar{x}_g - \bar{x}\right)^2 \qquad \nu_B = k - 1$$

F test

- MS_W estimates the common variance, σ^2, regardless of the null hypothesis

- MS_B estimates the common variance, σ^2, **only** when the null hypothesis is true

- Test for equality of variances: F-test

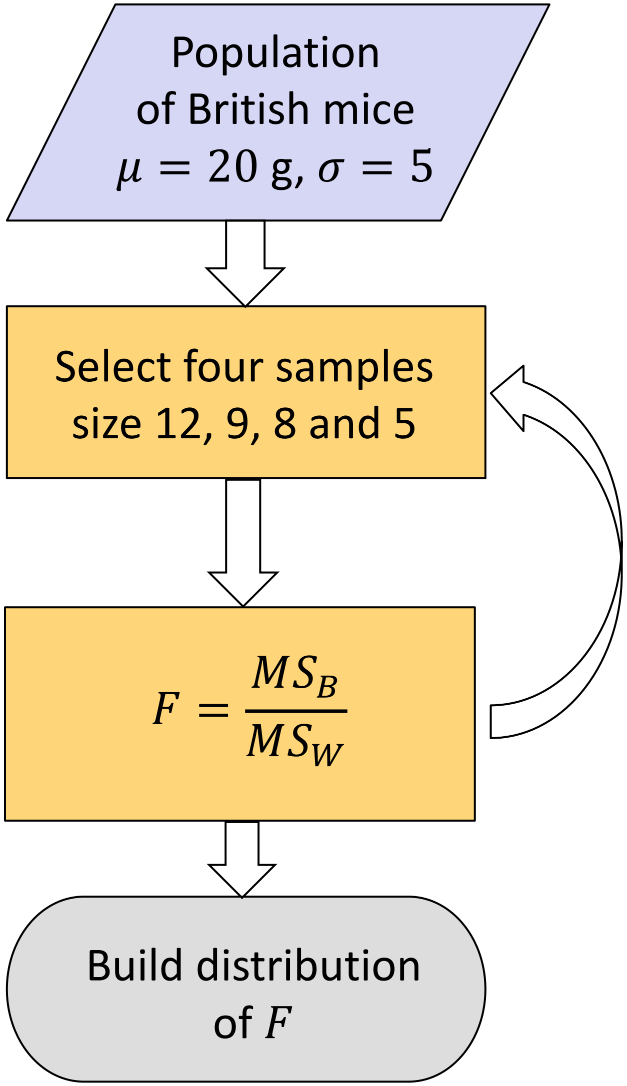
$$F = \frac{MS_B}{MS_W}$$

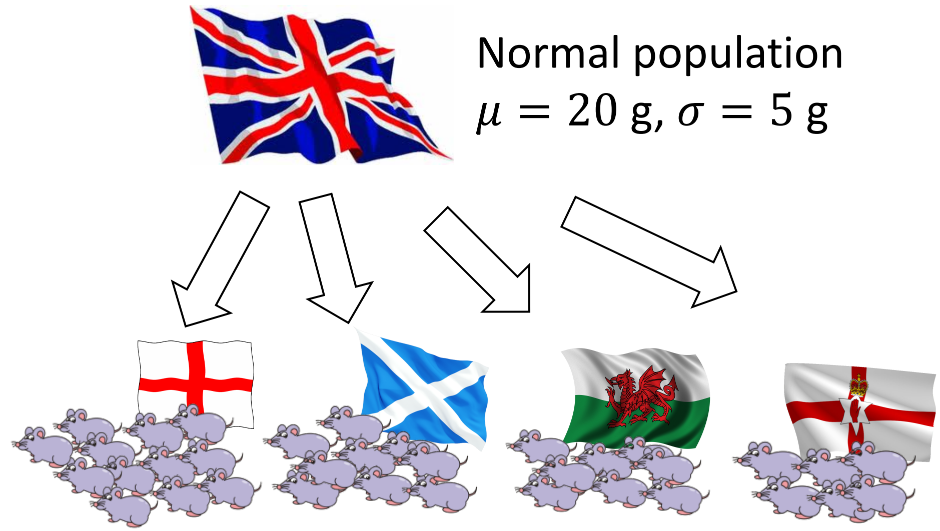
- Degrees of freedom: ν_B, ν_W

- If H_0 is true, we expect $F \sim 1$

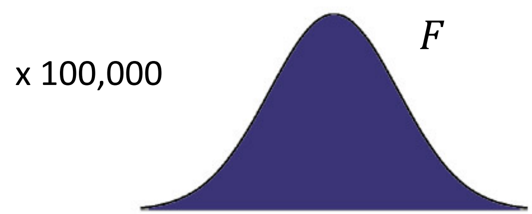
SS_W	524
ν_W	30
MS_W	17.5
SS_B	623
ν_B	3
MS_B	208
F	11.9

Null distribution

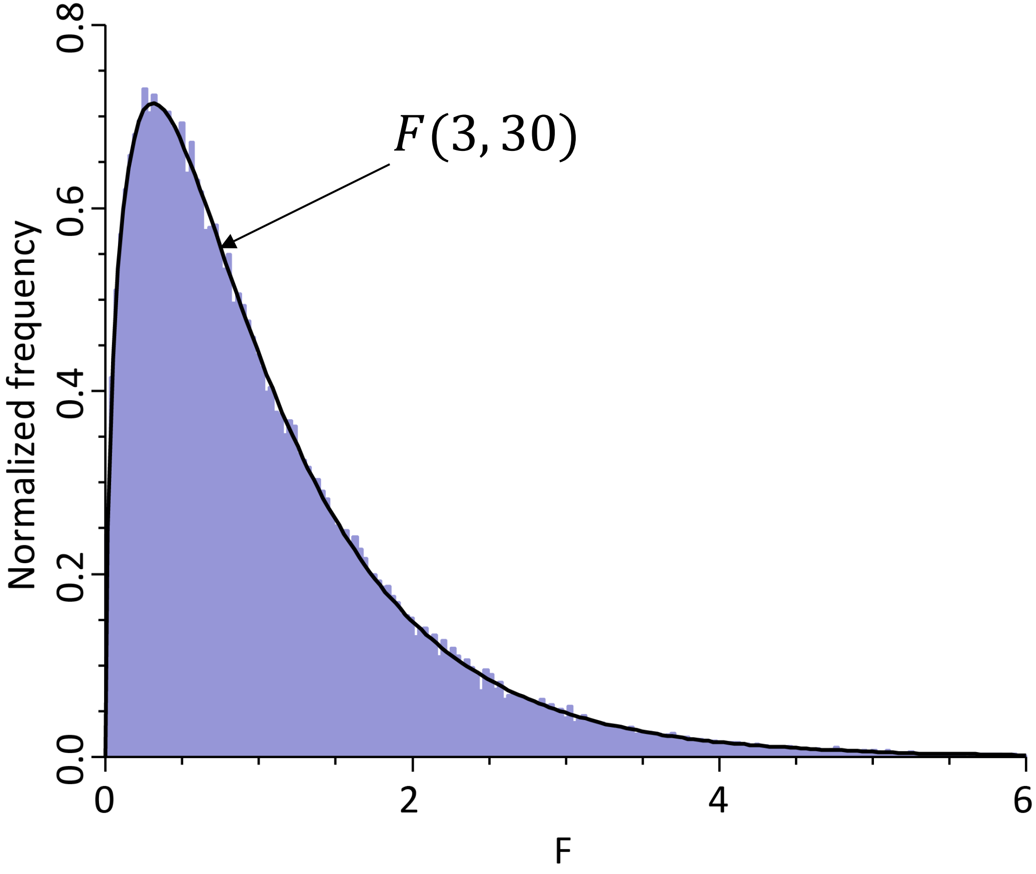




Normal population
$\mu = 20$ g, $\sigma = 5$ g



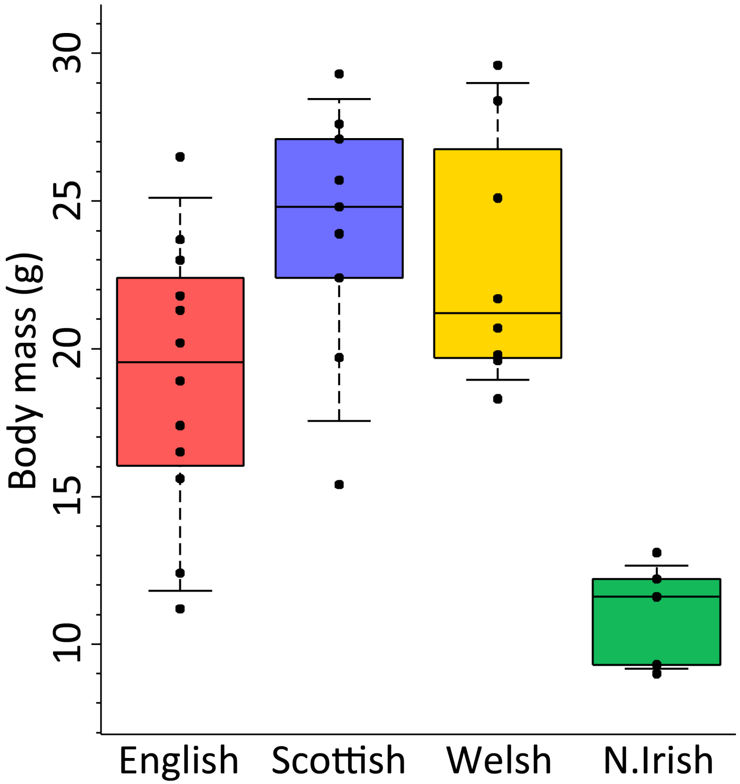
Null distribution = F-distribution

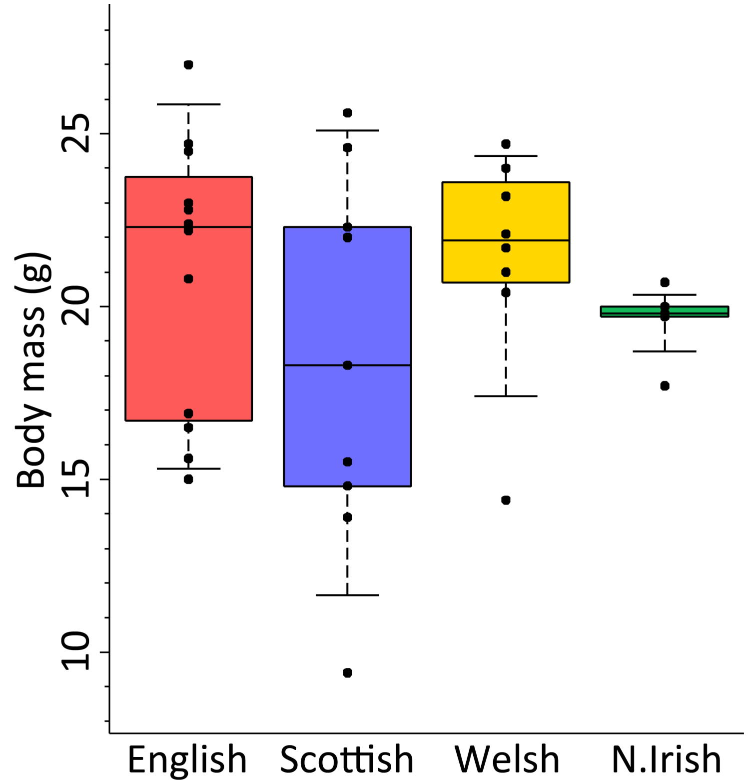


Effect vs. no effect

MS_W	17.5 g^2
MS_B	208 g^2
F	11.9
p	3×10^{-5}

MS_W	16.4 g^2
MS_B	15.8 g^2
F	0.96
p	0.42





ANOVA assumptions

- Normality – data in each group are distributed normally
  - ANOVA is quite robust against non-normality
  - if strongly not normal (e.g. log-normal) – transform to normality
  - if this fails, use non-parametric Kruskal-Wallis test

- Independence – groups are independent
  - dependence: e.g., observations of the same subjects over time
  - if groups are not independent, ANOVA is not appropriate, use other methods

- Equality of variances – groups sampled from populations with the same variance
  - sometimes called homogeneity of variances, or homoscedasticity /ˌhoʊmoʊskəˈdæstɪsity/
  - if variances are not equal, use Welch’s approximated test

Test to compare variances

- Null hypothesis: samples come from populations with equal variances
- H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$

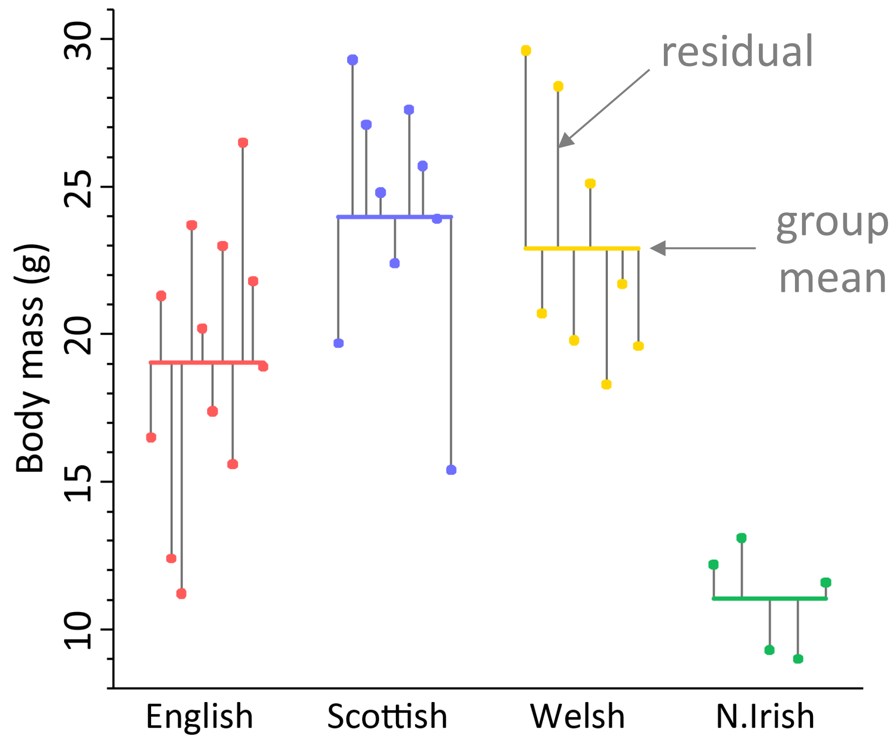
- Like ANOVA, except data x_{gi} are replaced by residuals R_{gi}:

  $R_{gi} = \left| x_{gi} - \bar{x}_g \right|$ - Levene's test

  $R_{gi} = \left| x_{gi} - \tilde{x}_g \right|$ - Brown-Forsythe test

- Test statistic:

$$W = \frac{MS_B}{MS_W}$$



Test to compare variances

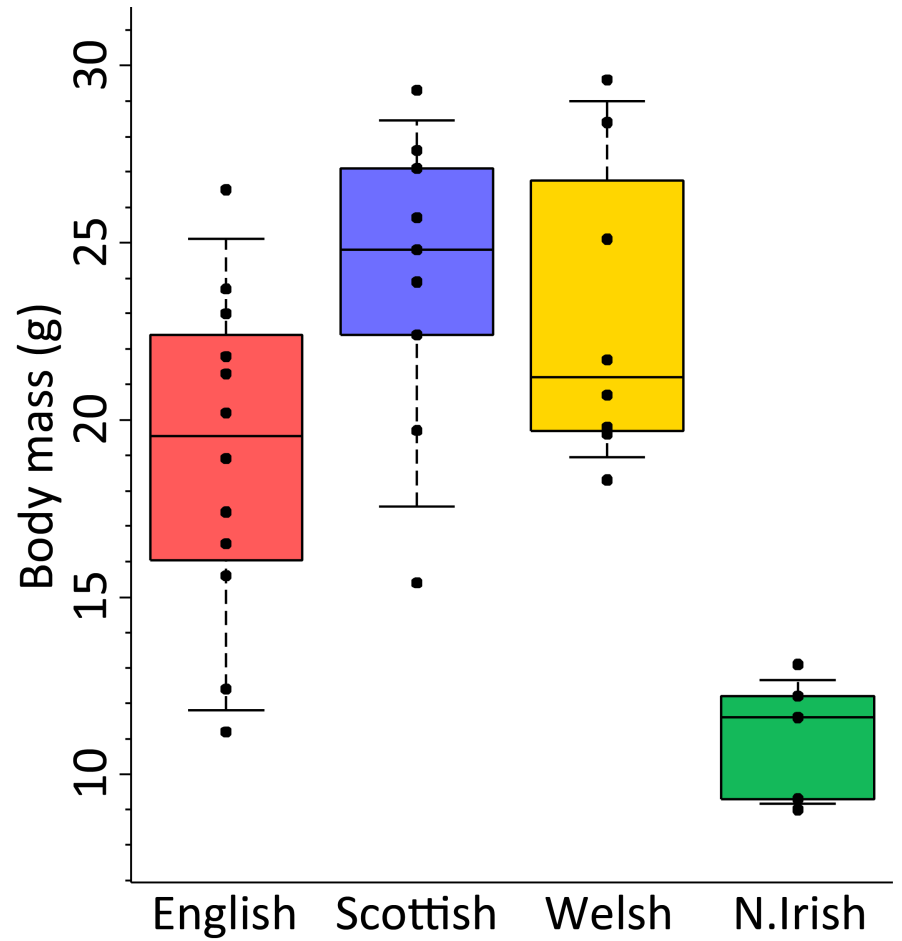
- Null hypothesis: samples come from populations with equal variances
- H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$
- Test statistic:

$$W = \frac{MS_B}{MS_W}$$

$MS_B = 6.40 \text{ g}^2$

$MS_W = 6.89 \text{ g}^2$

$W = 0.930$
$p = 0.44$



What if variances are not equal?

- B. L. Welch developed an approximated test
- Welch, B.L. (1951), "On the comparison of several mean values: an alternative approach", *Biometrika*, **38**, 330–336
- Skip the details...

- Mice data

	F	ν_1	ν_2	p
ANOVA	11.89	3	30	2.7×10^{-5}
Welch's test	28.95	3	15.96	10^{-6}

Post-hoc analysis: Tukey's test

- A multiple *t*-test
- Finds differences and p-values for each pair of categories
- Post-hoc test, you need ANOVA first
- Skip the details...

	Scottish	Welsh	N.Irish
Welsh	-1.1 0.95		
N.Irish	-12.9 0.00003*	-11.9 0.0001*	
English	-4.9 0.05	-3.9 0.20	8.0 0.006*

How to do it in R?

```
> mice
    Country Mass
1   English 16.5
2   English 21.3
3   English 12.4
4   English 11.2
5   English 23.7
6   English 20.2
7   English 17.4
8   English 23.0
9   English 15.6
10  English 26.5
11  English 21.8
12  English 18.9
13 Scottish 19.7
14 Scottish 29.3
15 Scottish 27.1
16 Scottish 24.8
17 Scottish 22.4
18 Scottish 27.6
19 Scottish 25.7
20 Scottish 23.9
21 Scottish 15.4
22    Welsh 29.6
23    Welsh 20.7
24    Welsh 28.4
25    Welsh 19.8
...
```

```
# ANOVA
> mice = read.table('http://tiny.cc/mice_1way', header=T)
> mice.aov = aov(Mass ~ Country, data=mice)
> summary(mice.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
country      3  622.7  207.56   11.89 2.67e-05 ***
Residuals   30  523.9   17.46
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Tukey’s Honest Significant Differences
> TukeyHSD(mice.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Mass ~ Country, data = mice)

$Country
                       diff         lwr       upr     p adj
N.Irish-English   -8.001667 -14.04998948 -1.953344 0.0059422
Scottish-English   4.947222  -0.06331043  9.957755 0.0539580
Welsh-English      3.858333  -1.32806069  9.044727 0.2023039
Scottish-N.Irish  12.948889   6.61101070 19.286767 0.0000277
Welsh-N.Irish     11.860000   5.38219594 18.337804 0.0001394
Welsh-Scottish    -1.088889  -6.61022696  4.432449 0.9494897
```

How to do it in R?

```
# Levene's test for equality of variances
> library(lawstat)
> levene.test(mice$Mass, mice$Country)

        modified robust Brown-Forsythe Levene-type test based on the absolute
deviations from the median

data:  mice$Mass
Test Statistic = 0.92948, p-value = 0.4386



# Welch's test for unequal variances
> oneway.test(Mass ~ Country, mice, var.equal=F)

        One-way analysis of means (not assuming equal variances)

data:  mass and country
F = 28.95, num df = 3.00, denom df = 15.96, p-value = 1.084e-06
```

Two-way ANOVA

ANOVA as a linear model (one-way)

measurement group mean noise

$$x_{ir} = \mu_i + \epsilon_{ir}$$

group replicate

grand mean group effect

$$x_{ir} = \mu + \alpha_i + \epsilon_{ir}$$

null hypothesis

$$\mathrm{H}_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

null hypothesis

$$\mathrm{H}_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

$$\forall i: \ \alpha_i = 0$$

ANOVA as a linear model (two-way)

measurement cell mean noise

$$x_{ijr} = \mu_{ij} + \epsilon_{ijr}$$

column row replicate

grand mean column effect row effect interaction effect

$$x_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$$

Column means are equal:

$$\mathrm{H}_0^{\mathrm{col}}\colon\ \mu_{1.} = \mu_{2.} = \cdots = \mu_{n_c.} \text{ or } \forall i\colon\ \alpha_i = 0$$

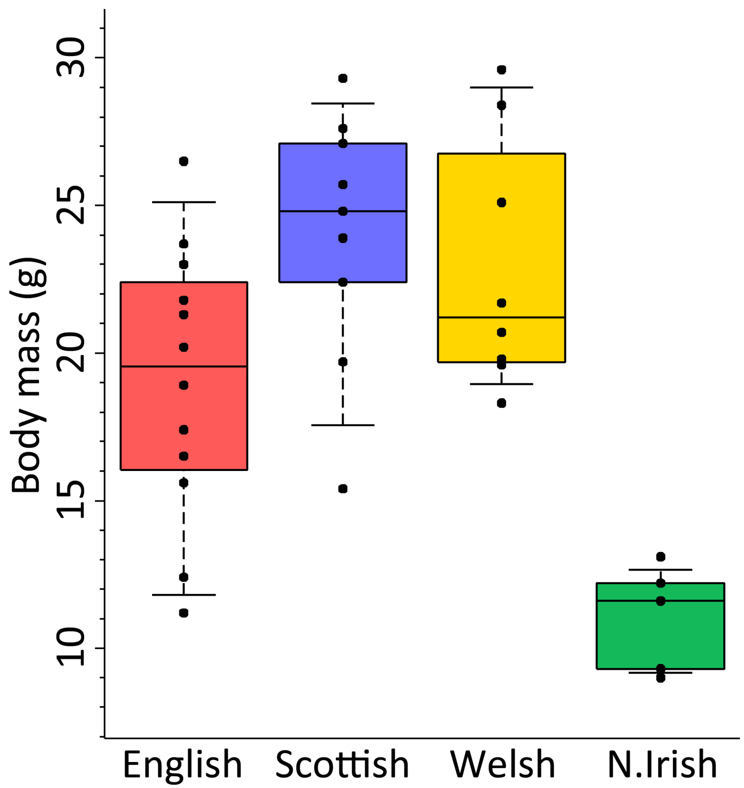
Row means are equal:

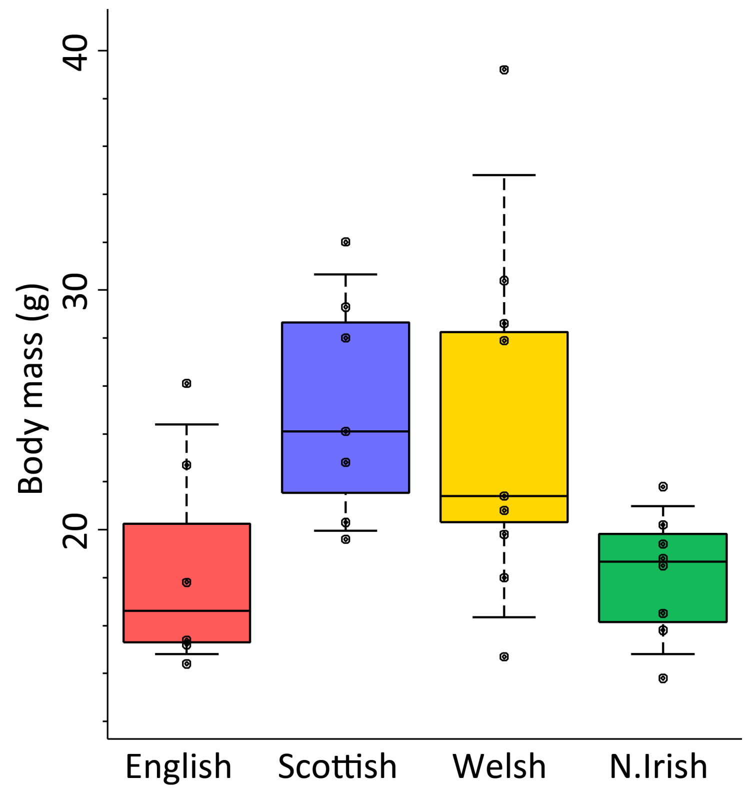
$$\mathrm{H}_0^{\mathrm{row}}\colon\ \mu_{.1} = \mu_{.2} = \cdots = \mu_{.n_r} \text{ or } \forall i\colon\ \beta_i = 0$$

There is no interaction between rows and columns:

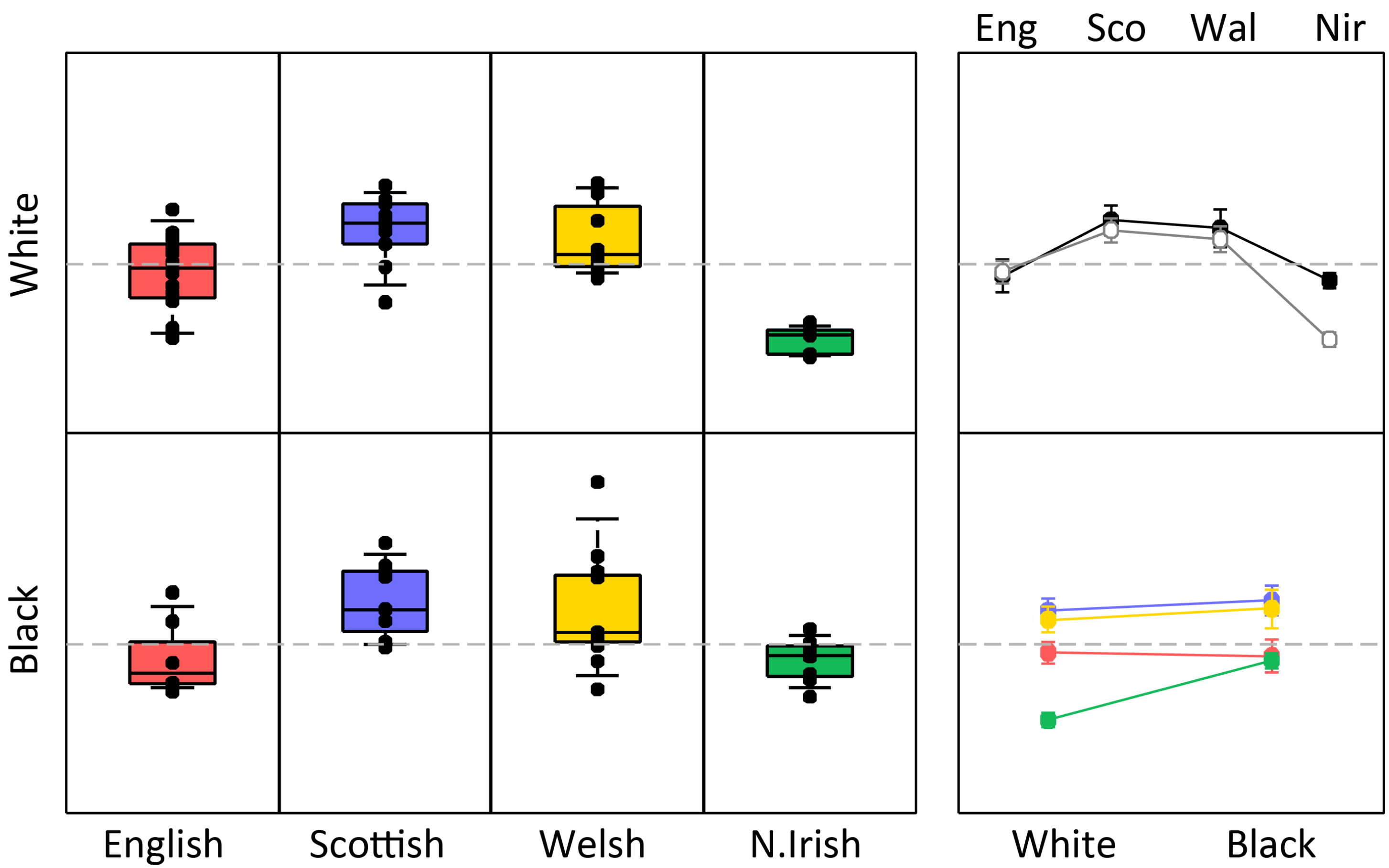
$$\mathrm{H}_0^{\mathrm{int}}\colon \forall i, j\colon\ \gamma_{ij} = 0$$

More mice!


Body mass (g)
10
15
20
25
30
English
Scottish
Welsh
N.Irish


Body mass (g)
20
30
40
English
Scottish
Welsh
N.Irish

Two-way ANOVA – two variables



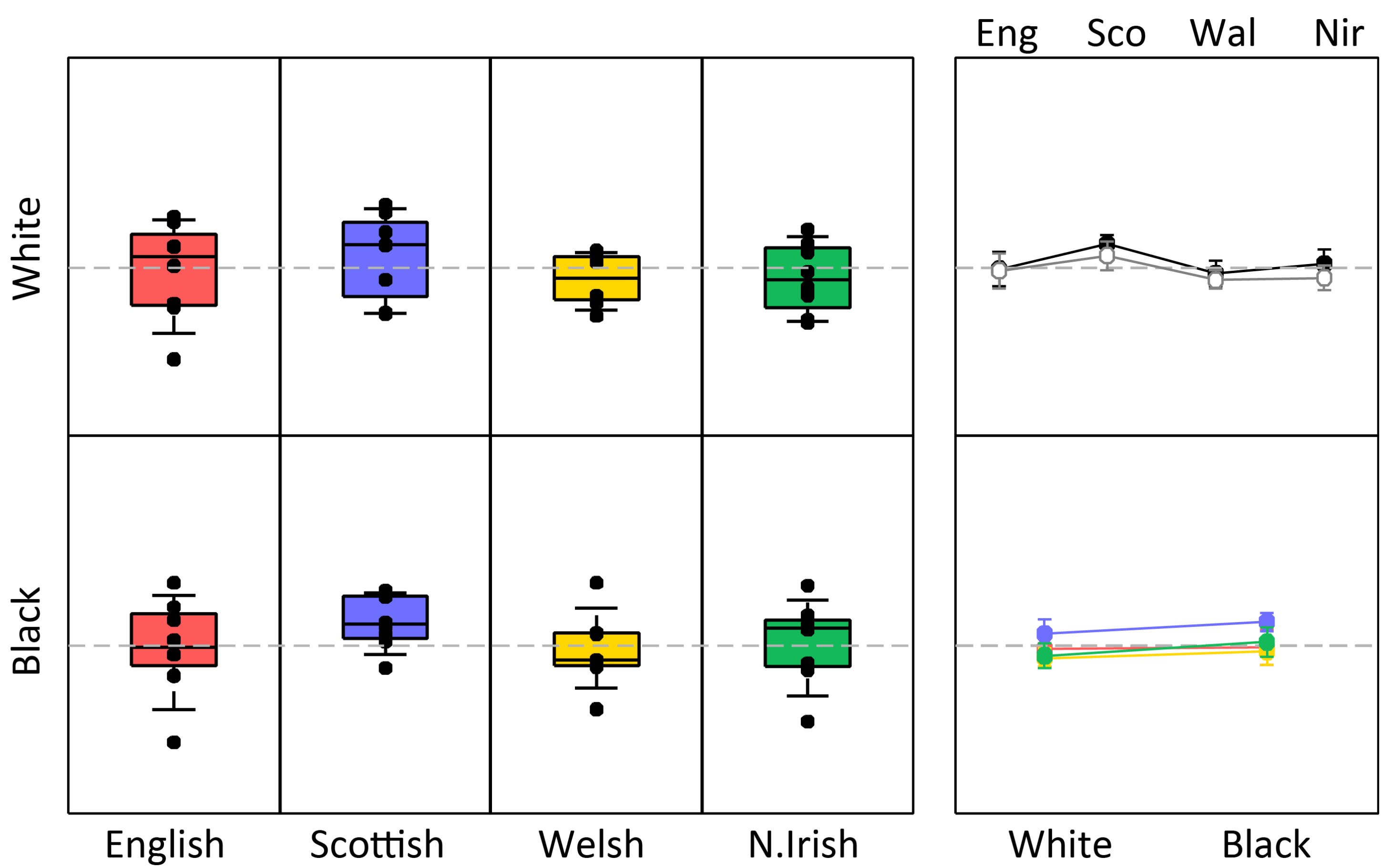
How to do it in R?

```
# 2-way ANOVA
> mice = read.table('http://tiny.cc/mice_2way', header=T)
> mice.lm = lm(Mass ~ Country + Colour + Country*Colour, mice)
> anova(mice.lm)
Analysis of Variance Table

Response:
Mass           Df  Sum Sq Mean Sq F value    Pr(>F)
Country         3  809.68 269.893 11.9366 3.598e-06 ***
Colour          1   59.87  59.873  2.6480    0.1092
Country:Colour  3  107.39  35.797  1.5832    0.2034
Residuals      57 1288.80  22.611
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1>
```

Null hypotheses: all three true

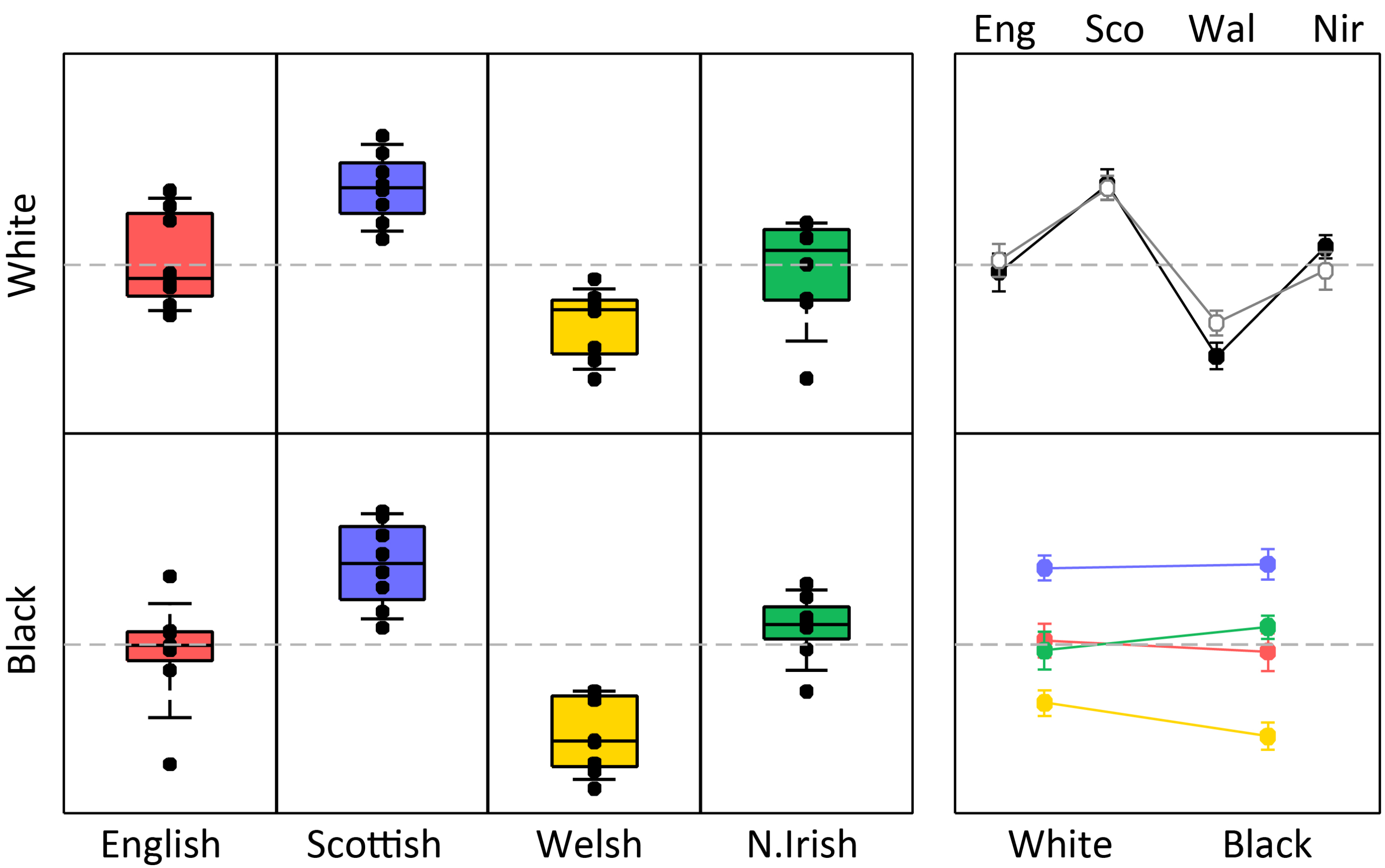
- $x_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$
- $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$



	p
columns	0.39
rows	0.24
interaction	0.97

Null hypotheses: columns not equal

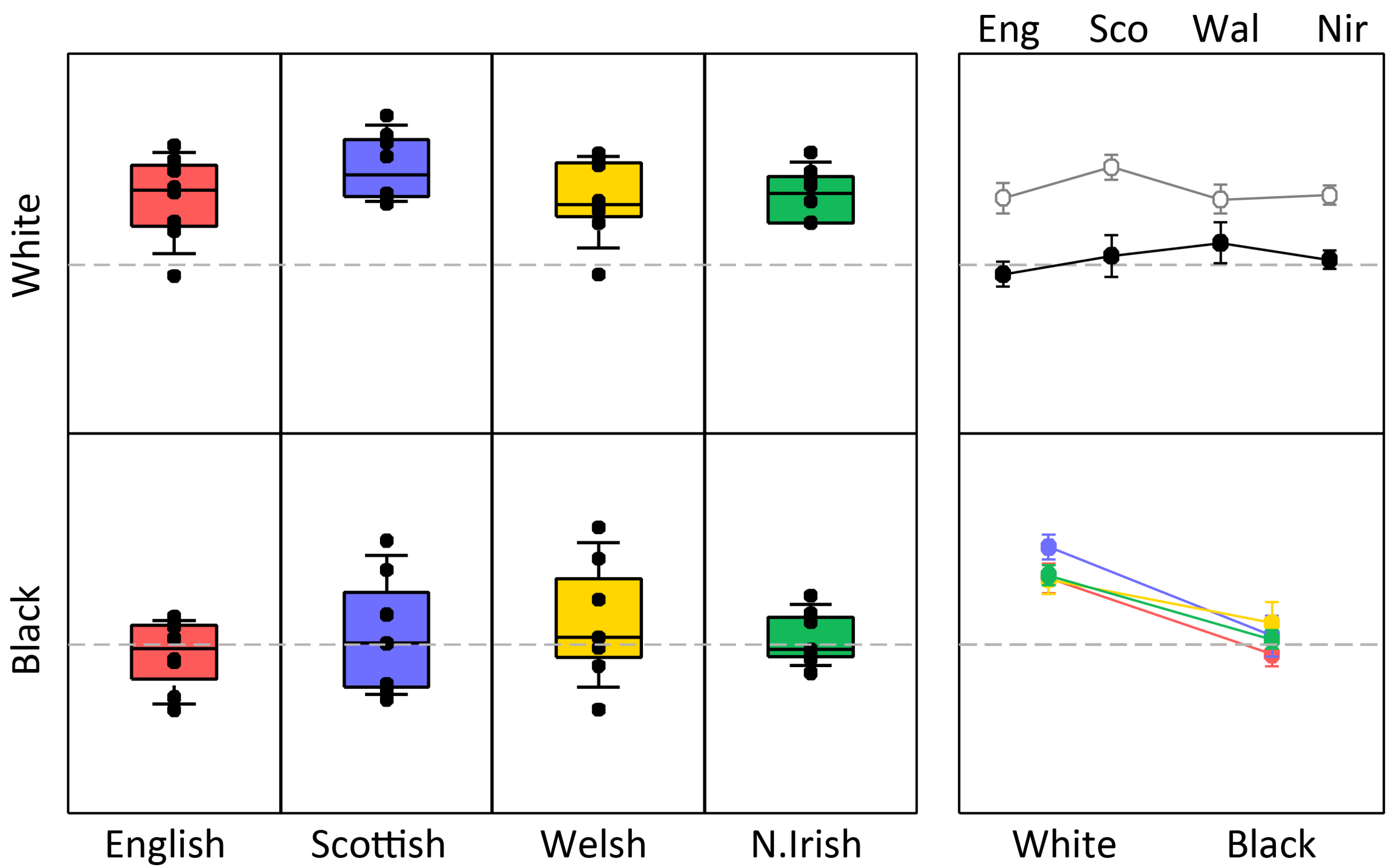
- $x_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$
- $\mathbf{A} = (0 \quad 10 \quad -10 \quad 0), \mathbf{B} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$



	p
columns	2×10^{-8}
rows	0.41
interaction	0.48

Null hypotheses: rows not equal

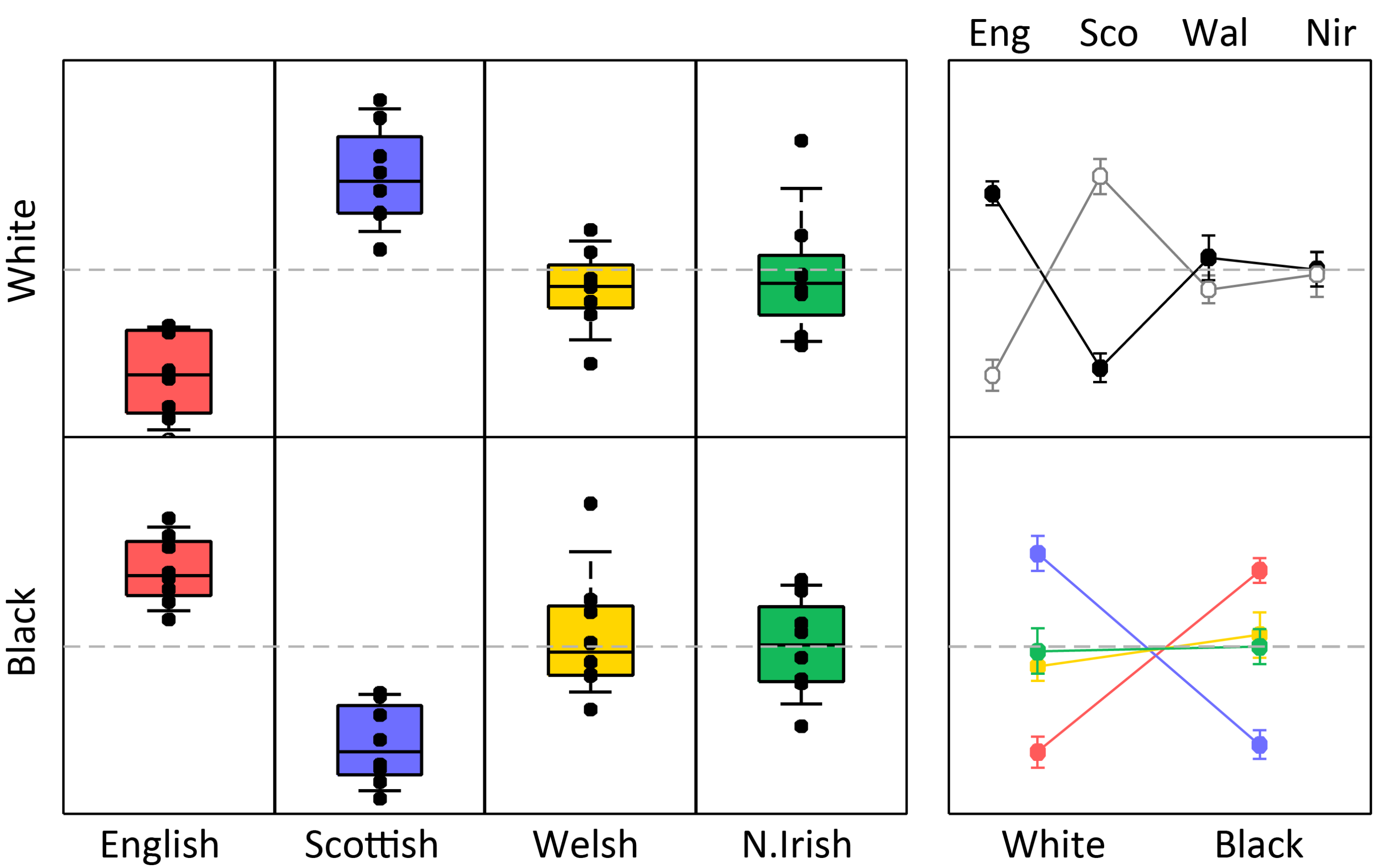
- $x_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$
- $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 10 \\ 0 \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$



	p
columns	0.69
rows	10^{-12}
interaction	0.30

Null hypotheses: interaction

- $x_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$
- $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} -10 & 10 & 0 & 0 \\ 10 & -10 & 0 & 0 \end{pmatrix}$

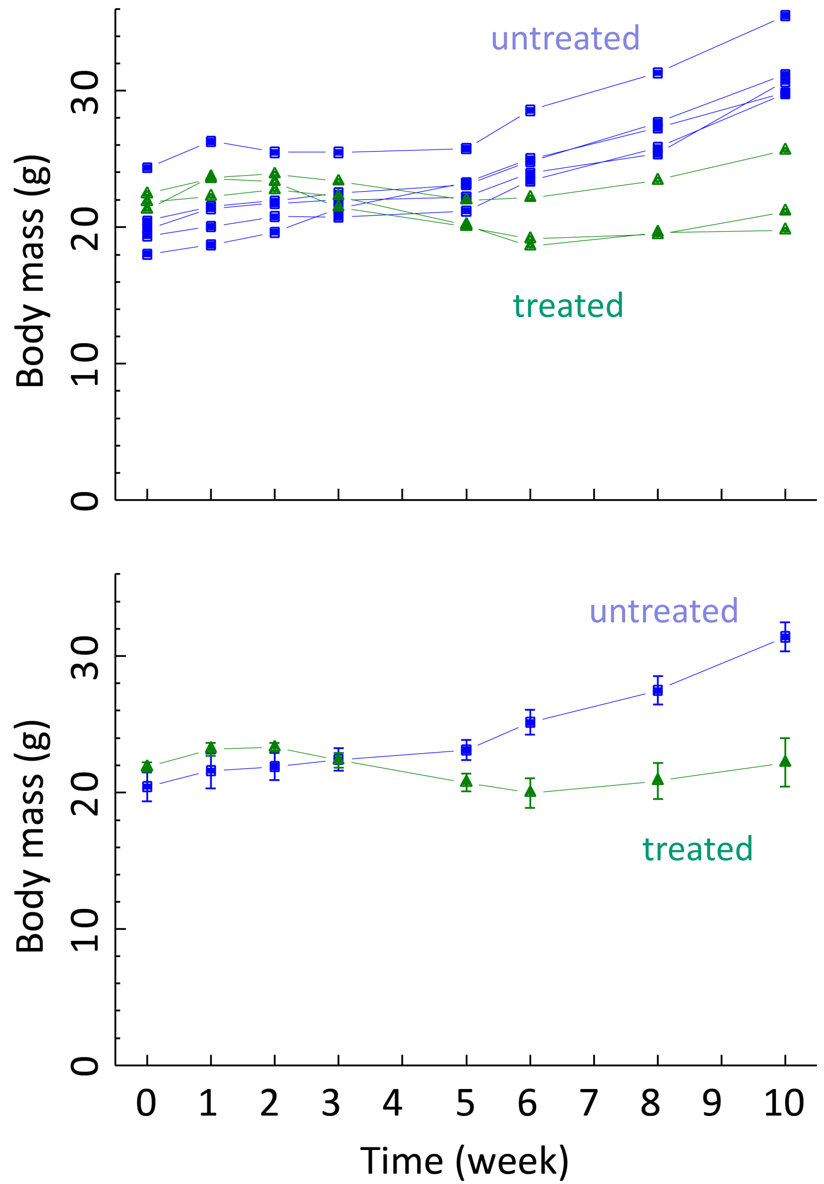


	p
columns	0.60
rows	0.88
interaction	10^{-13}

Time-course experiments

- Obesity study in mice
- Two groups:
  - untreated
  - treated with a drug
- Feed them a lot
- Observe body mass over time

- Is there a difference between the two groups?



Time-course experiments

- You can do ANOVA
- $p = 5\times10^{-5}$

- But

- Data are correlated
- ANOVA doesn't recognize numerical variables (time)
- You don't know where the change is

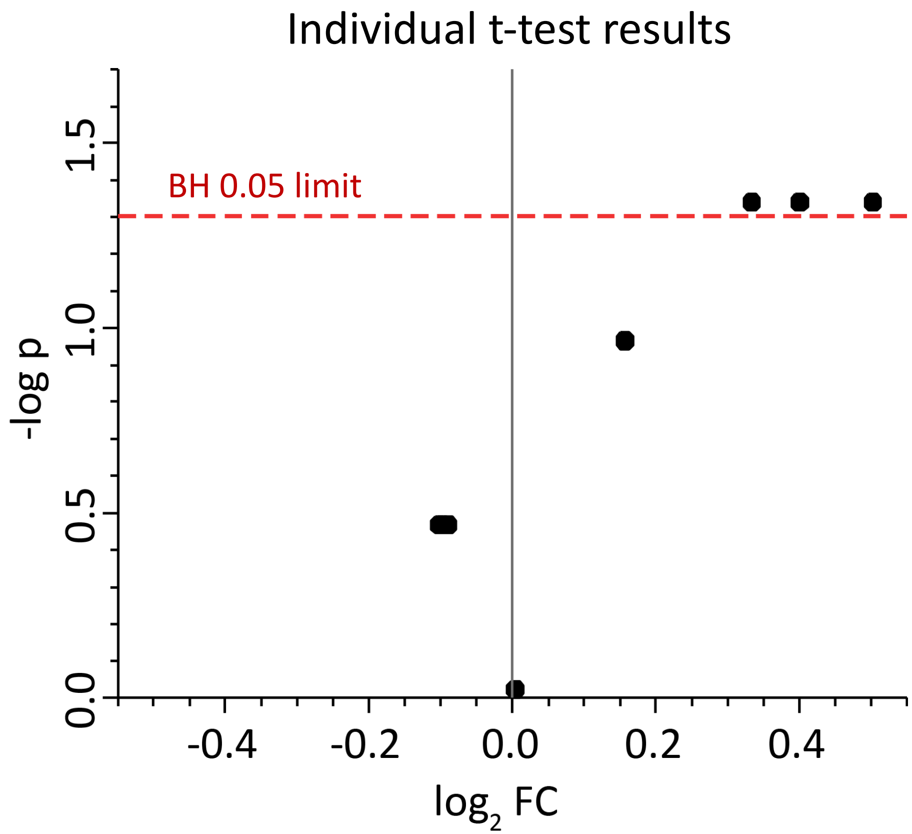
```
> dat = read.table('http://tiny.cc/time_course', header=T)
> dat.lm = lm(Mass ~ Treatment + Time + Treatment*Time, dat)
> anova(dat.lm)

               Df  Sum Sq Mean Sq F value    Pr(>F)
Treatment       1  85.538  85.538 20.1508 4.481e-05
Time            7 272.465  38.924  9.1694 3.825e-07
Treatment:Time  7 230.738  32.963  7.7652 2.907e-06
```

Time-course experiments

- What about t-test at each time point?
- Works well!
- Three time points are significantly different

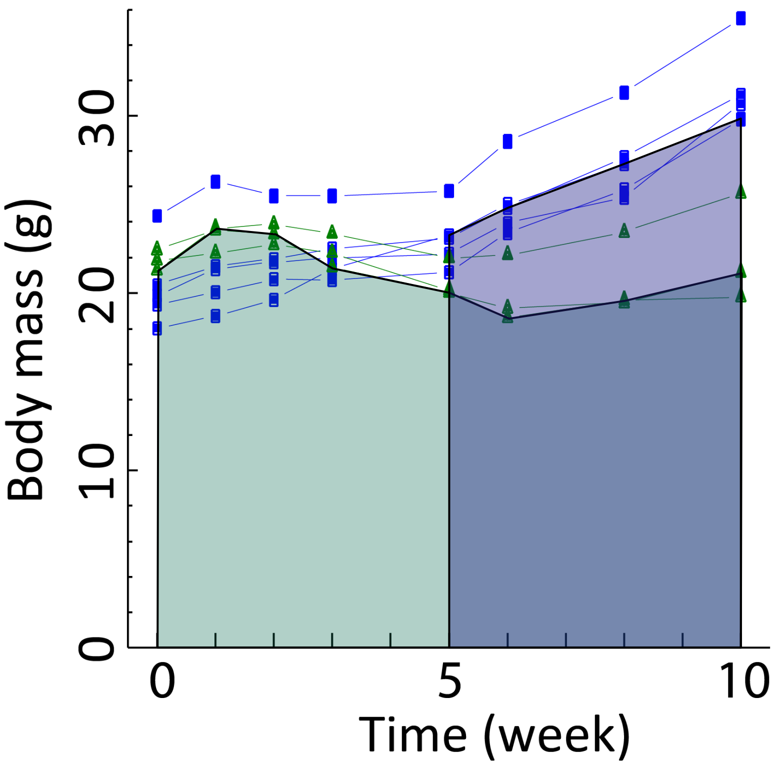
- But: misses point-to-point correlation



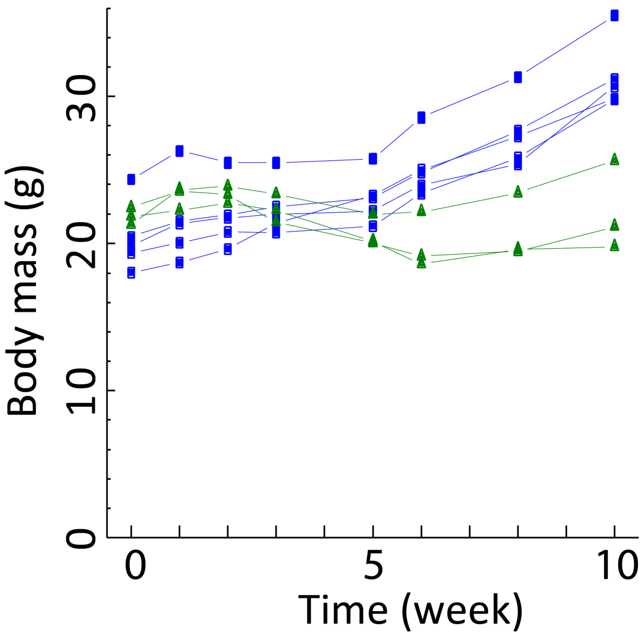
Better approach: build a model

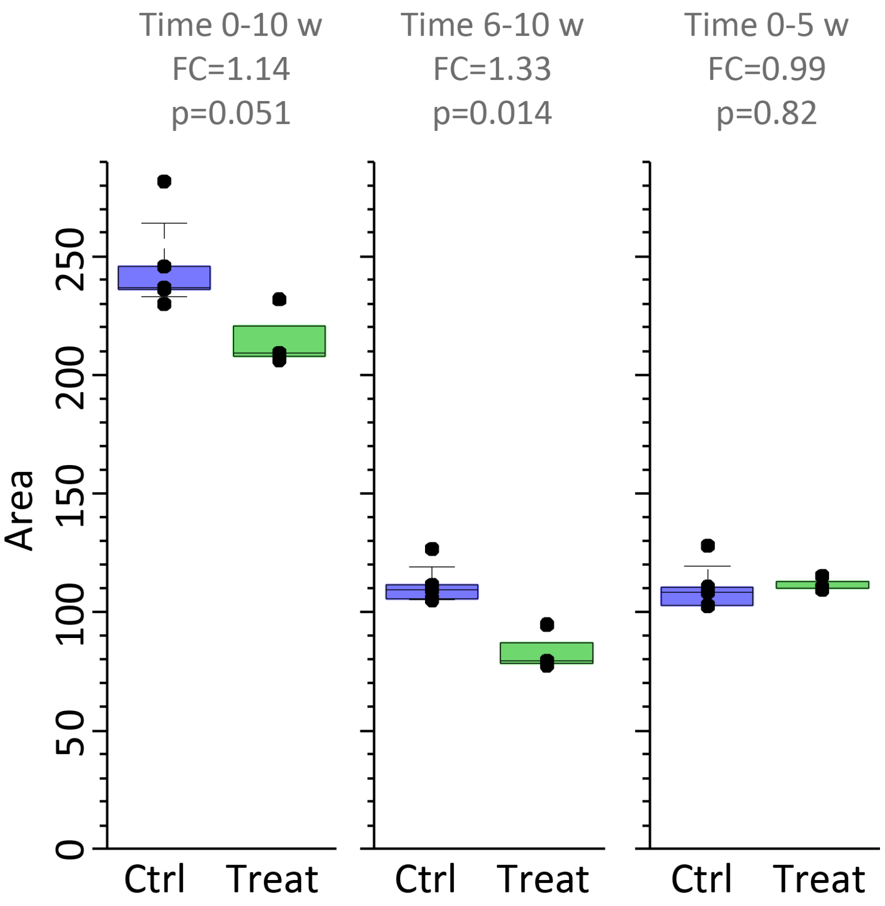
- First: understand your data
- Build a model and reduce time-course curves to just one number
- Do a t-test or similar test on these numbers

- Very simple: area under each curve
- This gives us 4 vs. 3 areas



Compare area under the curve





Chi-square or G-test vs. ANOVA

	WT	KO1	KO2	KO3
G1	50, 54, 48	61, 75, 69	78, 77, 80	43, 34, 49
S	172, 180, 172	175, 168, 166	162, 167, 180	178, 173, 168
G2	55, 50, 63	45, 41, 38	47, 49, 43	59, 50, 45

Fisher's test / Chi-square test / G-test
Experiment outcome: category
Table contains counts

	English	Scottish	Welsh	N. Irish
White	19.1, 20, 21	22.3, 21.2, 25.6	18.1, 19.2, 22.7	15.6, 16.7, 15
Black	21.1, 20, 20.5	21.1, 27.5, 23	22.5, 18.5, 19	19.1, 17.7, 13.5
Grey	20, 21, 17	18.6, 20.1, 19.7	15, 18, 22	12, 18.1, 20.3

ANOVA
Experiment outcome: measurement (could be counts)
Table contains measurements

G-test or ANOVA?

Bacterial antibiotic resistance

- Four strains
- Grown in normal medium and two antibiotic concentrations
- Dilution plating, count colonies

	WT	KO1	KO2	KO3
No antibiotic	77, 51, 92	50, 83, 16	70, 111, 78	121, 147, 110
Conc. 1	83, 51, 40	66, 18, 49	95, 109, 52	75, 116, 109
Conc. 2	11, 7, 31	69, 41, 21	85, 51, 60	95, 128, 116

Outcome is measurement, not category
This is not a contingency table!

Perhaps ANOVA
Need to check normality

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