

GROUP COMPARISONS

Treatments

Also called the *levels* of a *factor*, they identify the groups to be compared. We measure a property in each group, the dependent variable, and ask whether the group means are "all the same" or "different". If they "differ", then the treatments had an effect.

Knockout



Wildtype



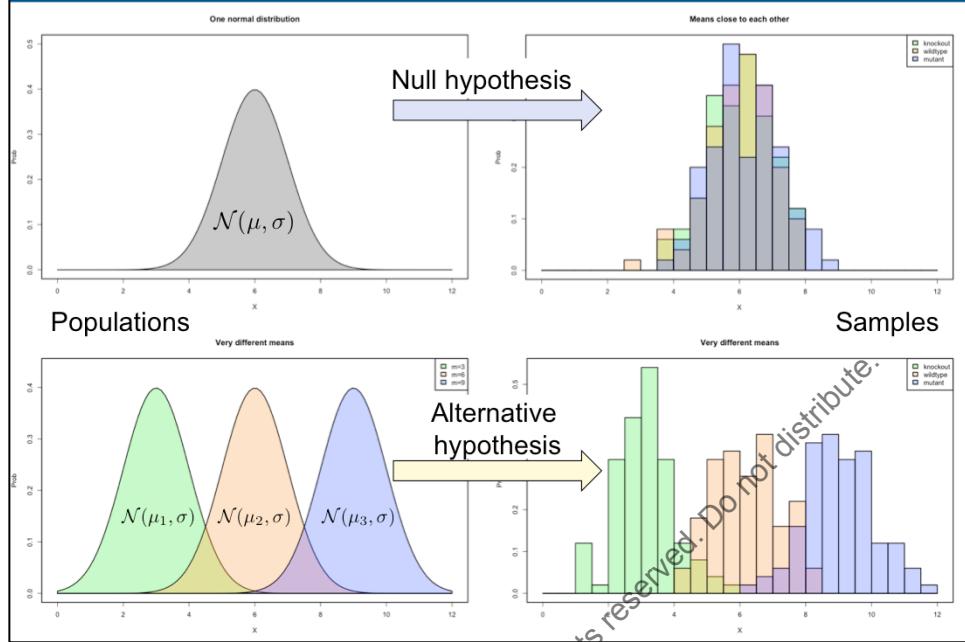
Mutant



Dai et al, *Nature Biotech.*
20, 251 - 255 (2002)

We use ANOVA to test whether the means of groups of observations are different. Somewhat counterintuitively, the technique relies on comparing the variances rather than the means.

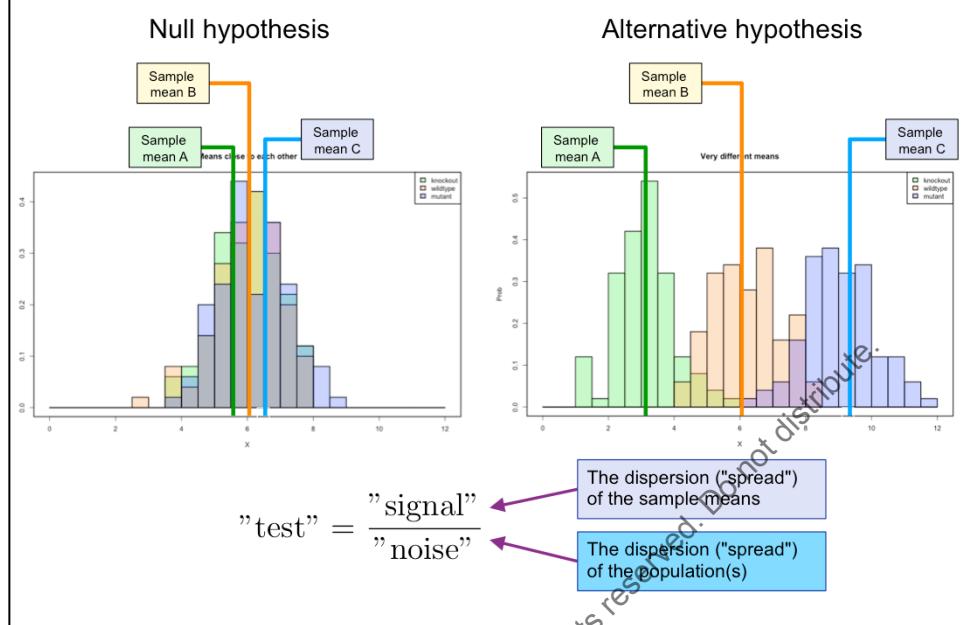
ASSUMPTIONS



The ANOVA technique relies on the following assumptions:

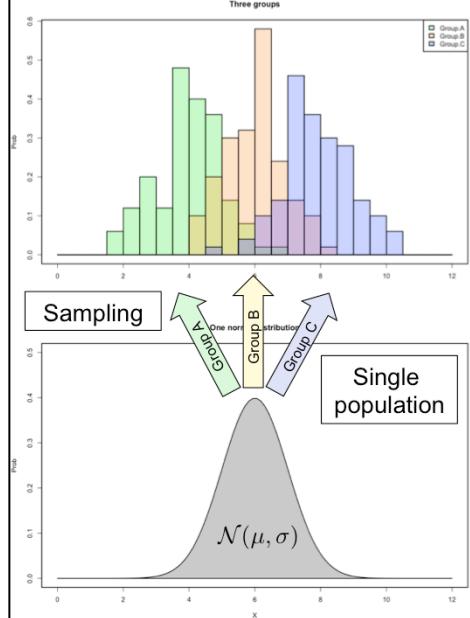
- 1) The groups are sampled from Normal distributions
 - 2) The variance of the group populations are equal ("homoscedastic")
- The null hypothesis is that the group samples come from the same underlying Normal population, i.e. all group means are "the same". The null hypothesis is rejected if at least one of the group means is different from the others.

WHAT SHALL WE TEST?



The test statistic is a function that, loosely speaking, provides an estimate of how big the "effect" is, in other words, quantifies the "signal" over a "noise". A good test statistic for comparing several group means shall measure "how far" the group means are from each other, relative to the "noise" inherent in the underlying population(s). We will make this intuitive notion exact in the following.

THE NULL HYPOTHESIS



$$\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$\begin{aligned}\bar{x}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} \\ s_j^2 &= \frac{1}{(n_j - 1)} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2\end{aligned}$$

Estimates from k groups

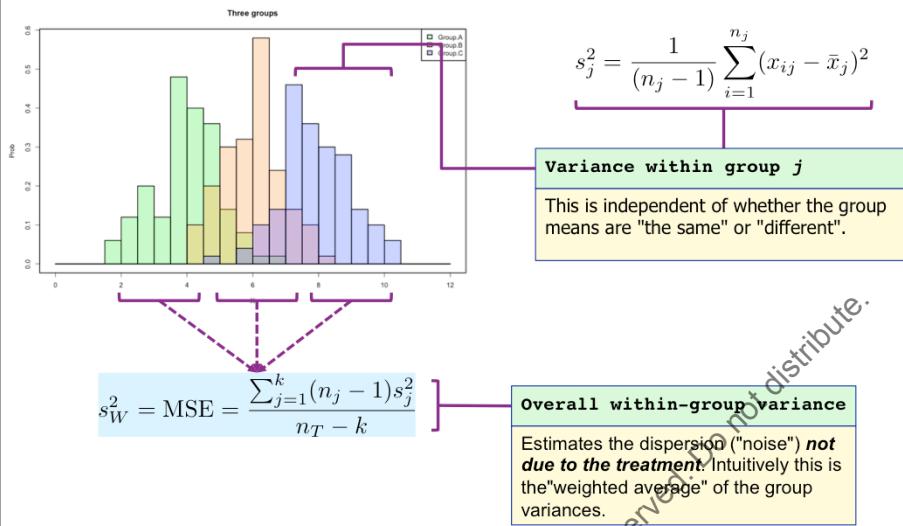
Mean and variance of the j -th sample.
There are two ways to estimate the population variance σ^2 from them:

"Within-groups"	"Between-groups"
Average the group variances	Variance of group means (SEM squared)

If the two estimates are "close", we accept H_0 .

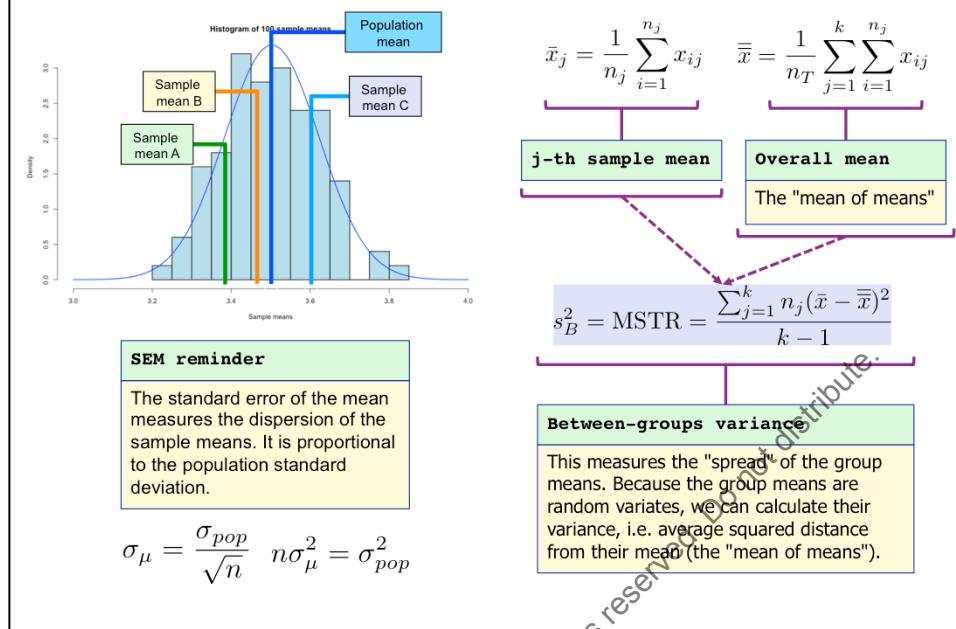
If the null hypothesis is "true", then the differences between the samples are due to the random nature of the sampling alone.

WITHIN-GROUPS VARIANCE



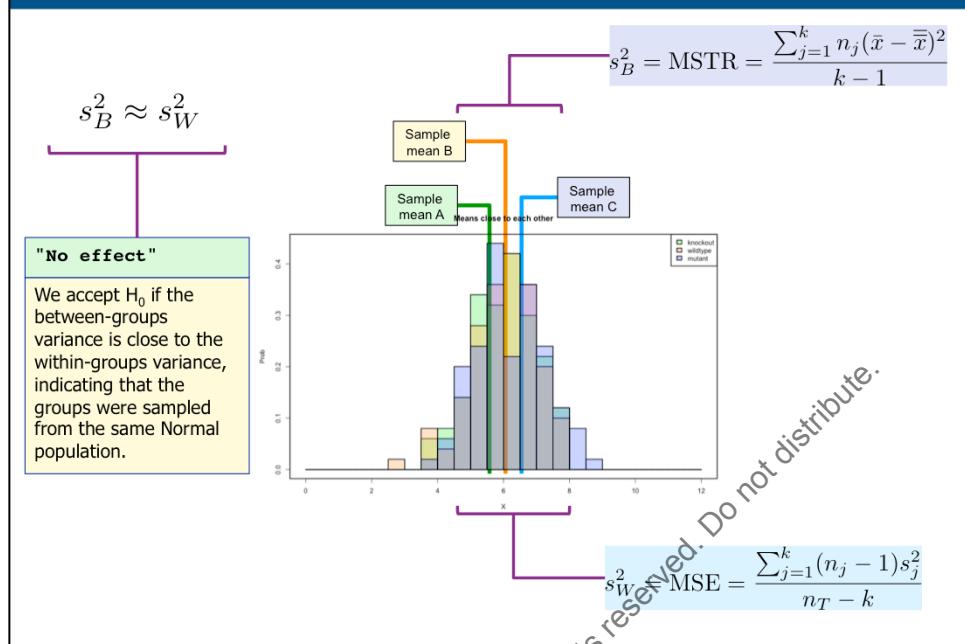
The within-group variance (or mean squared error, MSE) measures part of the overall variance which would be present in the data anyway, independently of whether the group means are the same or not. This provides a kind of "baseline": if the between-group variance (see next slide) is comparable to it, then there is no big difference between the group means.

BETWEEN-GROUPS VARIANCE



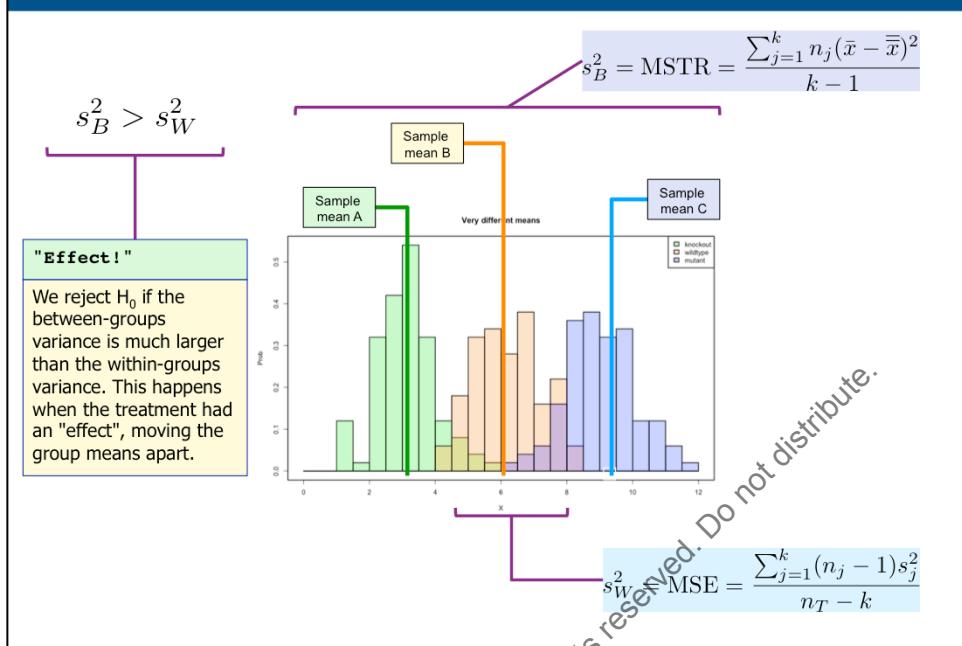
The MSTR, or "mean square treatment", measures the "spread" (dispersion) of the group means. Under the null hypothesis, all groups are sampled from the same (Normal) population and the distribution of the group means will be Normal with a variance that is the population variance divided by the sample size. (The square root of the variance of the group means is called the "standard error of the mean" or SEM for short.)

TESTING THE NULL HYPOTHESIS



If there are no differences between the group means, then the between- and within-groups variances are comparable.

THE ALTERNATIVE HYPOTHESIS



If the group means are different, the between-groups variance will be much larger than the within-groups variance. In other words, the "variance due to treatment" is large, "there is an effect".

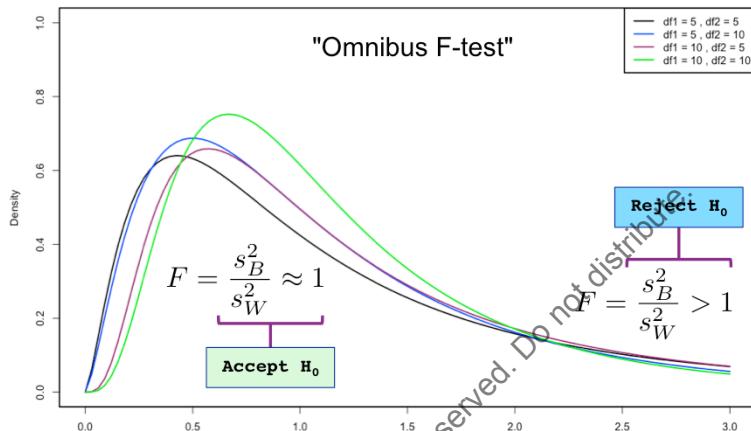
COMPARING VARIANCES



R. A. Fisher

$$F = \frac{s_1^2}{s_2^2}$$

PDF of the F distribution



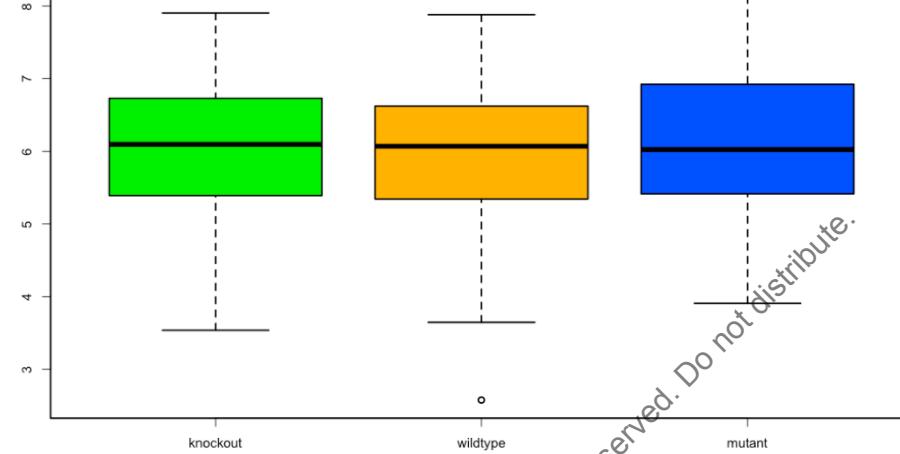
R. A. Fisher (who invented ANOVA) designed the F-test to compare two variances. The F-test can be used in many situations when you want to compare variances; it plays a central role in ANOVA to decide whether to keep the null hypothesis that there are no differences between the group means, i.e. the between-group variance is roughly the same as the within-group variance. This test in ANOVA is affectionately called the "omnibus F-test" because, in a way, it tests whether all group means being the same "at once".

NOT MUCH EFFECT HERE...

Try me!

Load and plot the data

```
dc <- read.csv("close.csv")
cols <- c("green","orange","blue")
boxplot(dc, col=cols)
```



Our first example dataset. We can already see that not much is happening. However, this must be tested.

"WIDE" AND "TALL" TABLES *Try me!*

knockout	wildtype	mutant
6.373	5.843	6.349
7.356	5.832	4.758
5.644	6.080	6.912
...

Data frame rearrangement

```
sc <- stack(dc)
head(dc)
head(sc)
```

Making "tall" tables

`stack()` puts the data frame columns on top of each other, the factor levels are *indicated* in the `ind` column. This is needed by R's ANOVA function. You can change the column names if you wish.

values	ind
6.373	knockout
7.356	knockout
5.644	knockout
...	...
5.843	wildtype
5.832	wildtype
6.080	wildtype
...	...
6.349	mutant
4.758	mutant
6.912	mutant
...	...

R's ANOVA function expects a 2-column data frame where one of the columns contains the measurements and the other column the group assignments (which are "levels" of a "factor", i.e. a categorical explanatory variable). Very often we keep our data in a "wide" format data frame, with one column for each of the groups. The `stack()` function converts such data frames into the "tall" format expected by the ANOVA function. By default the column names will be "values" and "ind" (for "indicator variable"), but this you can override.

MY FIRST ANOVA

Try me!

Homoscedasticity tests

```
bartlett.test(values ~ ind, sc)
bartlett.test(dc)
fligner.test(values ~ ind, sc)
fligner.test(dc)
```

Are the variances equal?

This is an important ANOVA assumption which can be checked using Bartlett's test or the Fligner-Killeen test. Both run on "wide" or "tall" tables as well.

ANOVA... at last!

```
ac <- aov(values ~ ind, sc)
summary(ac)
```

Interpreting the results

The omnibus F-test retains H_0 , the group means are not significantly different. The between-groups variance (**0.9023**) is even a bit smaller than the within-groups variance (**1.0065**).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	2	1.8	0.9023	0.896	0.409
Residuals	297	298.9	1.0065		

Variances and sums-of-squares

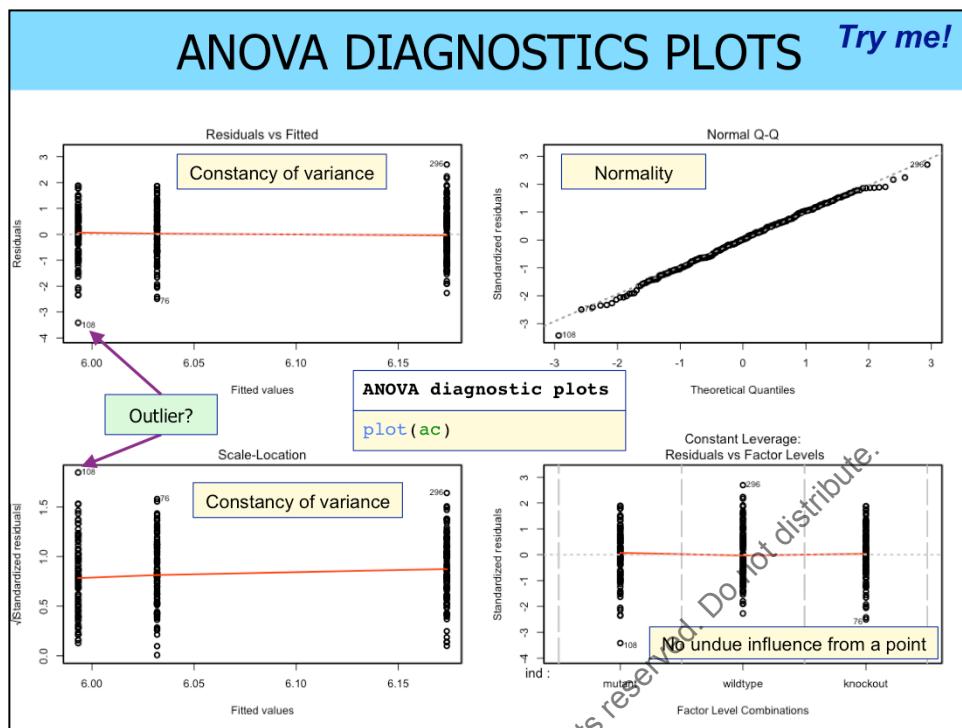
```
source("sumsq.R")
calc.vars(dc)
```

Check!

I provide this utility function to calculate the variances.

ANOVA assumes that the group variances are equal ("homoscedasticity"). There are at least two tests for this, Bartlett's test which we discussed in the "Statistics in R" course, or the Fligner-Killeen test. Practically it is not a great drama if the group variances are slightly different.

Then we run the ANOVA function `aov()` and save the results in `ac`. You can print a summary of the results. For comparison, I have written a function `calc.vars` that takes the (wide) input data frame and calculates the sums-of-squares and the between- and within-variances directly from the dataset. The ANOVA summary calls the variances "Mean Sq" which is a little bit confusing.

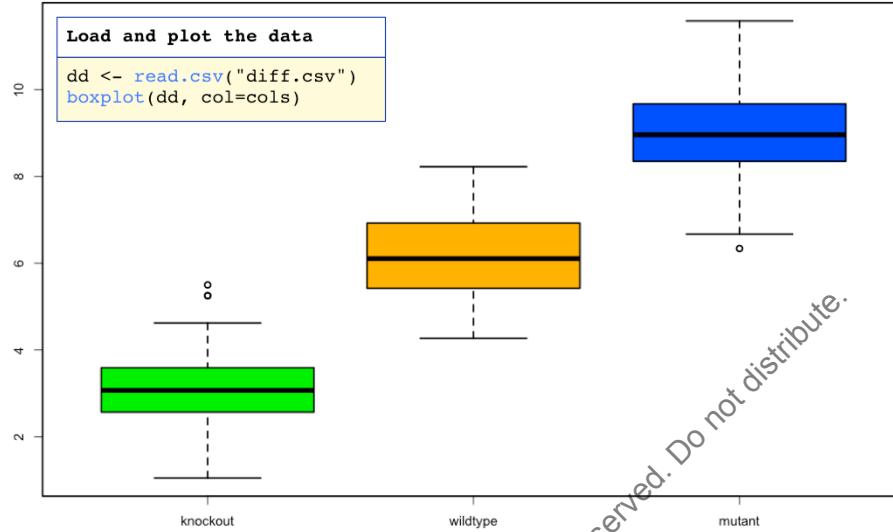


If you invoke the `plot()` function on the output of `aov()` then R gives you a nice selection of standard diagnostic plots. Refer to the R help on how to configure which plots should be generated.

DIFFERENT GROUP MEANS

Try me!

Very different means



Let us see what happens if the group means are obviously different.

ANOVA AGAIN

Try me!

Preparation

```
sd <- stack(dd)  
bartlett.test(dd)  
fligner.test(dd)
```

ANOVA

```
ad <- aov(values ~ ind, sd)  
summary(ad)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	2	1758.4	879.2	908	<2e-16 ***
Residuals	297	287.6	1.0		

Interpreting the results

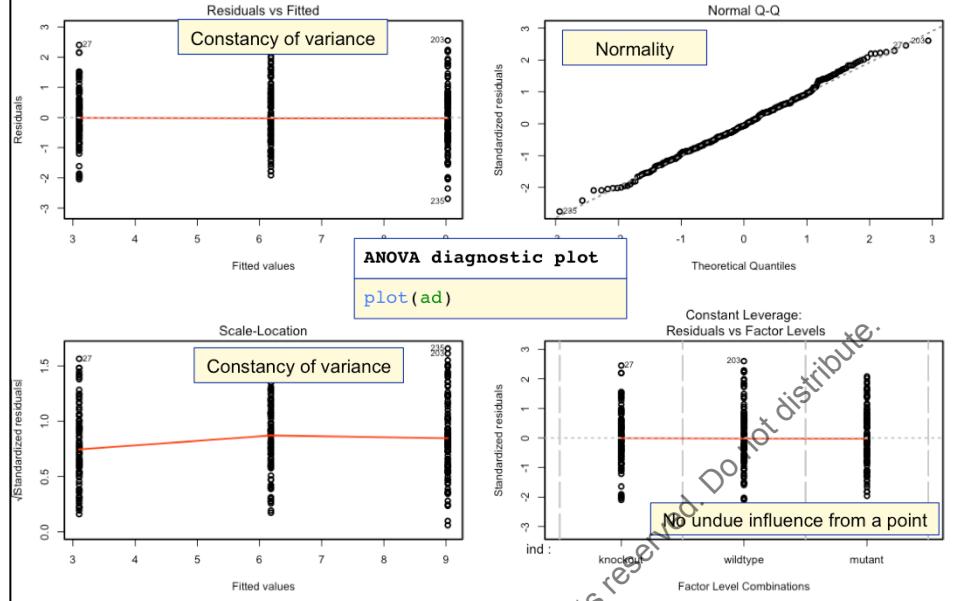
The between-groups variance (879.2) is much, much larger than the within-groups variance (1.0). The omnibus F-test rejects H_0 , the group means are obviously significantly different.

Variances and sums-of-squares

```
calc.vars(dd)
```

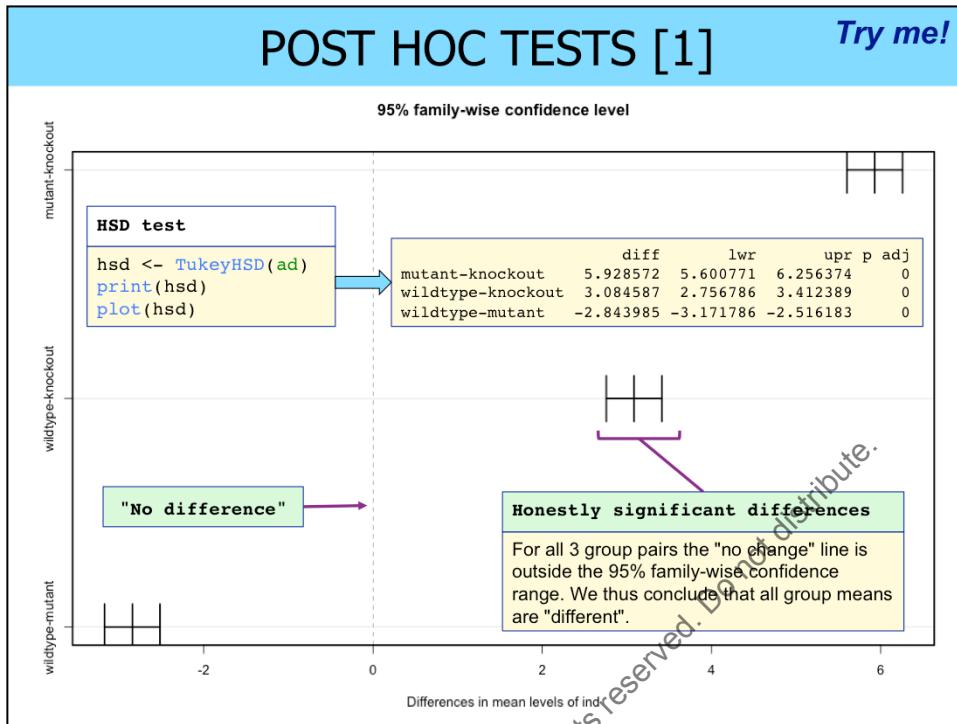
It is very instructive to compare the ANOVA summary to that of the previous data set. You can also calculate the sums-of-squares and within-/between- variances with calc.vars just to make sure...

ANOVA DIAGNOSTICS PLOT *Try me!*



POST HOC TESTS [1]

Try me!



Once the omnibus F-test told us that not all group means are equal, we must check which are indeed different from each other. Such tests are called "post hoc" (i.e. "after the fact") because we run them only if the null hypothesis of the omnibus F-test is rejected. There are several such tests; in this course we use "Tukey's Honestly Significant Differences" (HSD). The test function returns an object which can be `print()` -ed to get a summary or `plot()`-ted.

POST HOC TESTS [2]

Try me!



Adjusted t-test

```
pairwise.t.test(sd$values, sd$ind, p.adj="BH")
```

knockout mutant

mutant <2e-16 -
wildtype <2e-16 <2e-16

$$FWER = Pr(FP \geq 1)$$

$$FDR = \frac{FP}{FP + TP}$$

Multiple test corrections

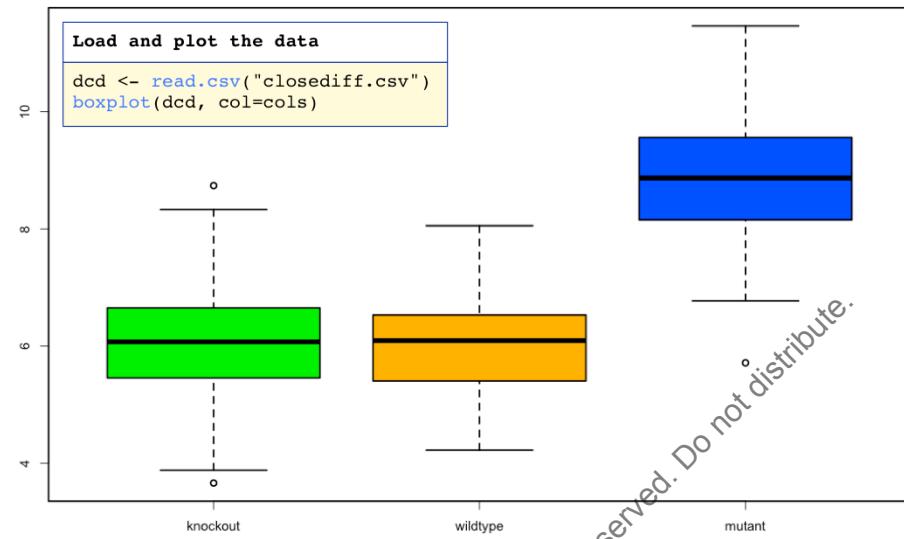
"bonferroni", "holm", "hochberg",
"hommel" control the *familywise error rate*.
Bonferroni is too strict and its usage is not
recommended.

"BH" (Benjamini-Hochberg) and "BY" control
the *false discovery rate*.

Biostatisticians preach at every street corner that comparing the means of multiple groups of observations with the t-test is a deadly sin. Luckily, redemption is nigh: R offers the `pairwise.t.test()` function which corrects for the multiple testing problem. Several methods are available, we usually recommend the Benjamini-Hochberg procedure.

SOME GROUPS ARE EQUAL... *Try me!*

Some means are different



Our third example dataset consists of 3 groups where two seem to have the same mean and the third group appears to be different.

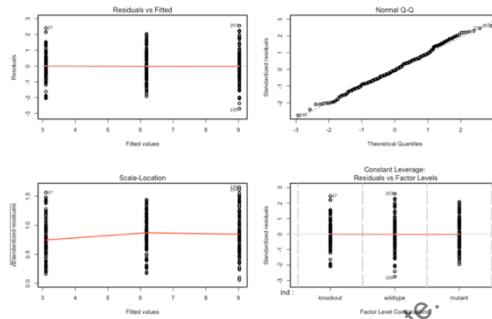
©VBCF/BioComp, András Aszódi 2018 All rights reserved. Do not distribute.

ANOVA OMNIBUS-TEST

Try me!

Preparation

```
scd <- stack(dcd)
bartlett.test(dcd)
fligner.test(dcd)
```



ANOVA

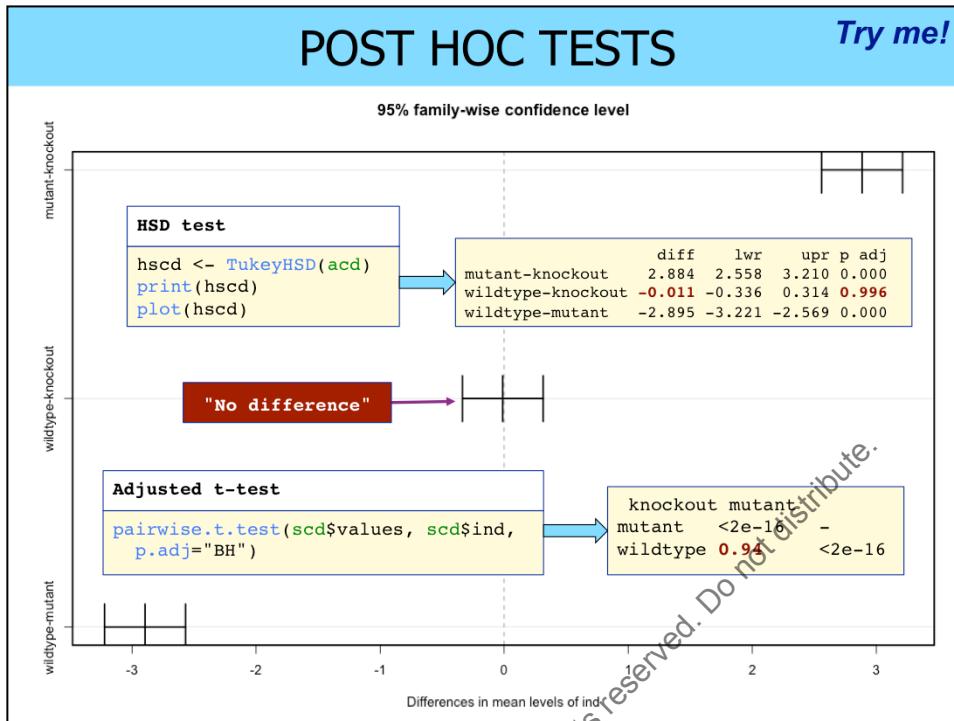
```
acd <- aov(values ~ ind, scd)
summary(acd)
```

```
Df Sum Sq Mean Sq F value Pr(>F)
ind      2   556.8  278.42  291.1 <2e-16 ***
Residuals 297  284.0     0.96
```

By now this procedure should be second nature for you ☺

POST HOC TESTS

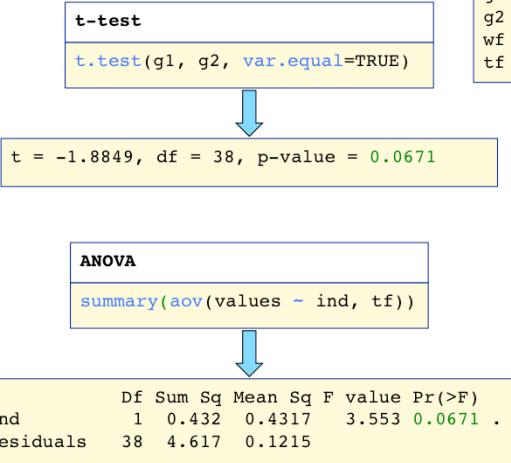
Try me!



We try both the HSD and the adjusted t-tests. Unsurprisingly they deliver the same result.

SPECIAL CASE: THE t-TEST

Try me!



Create synthetic data

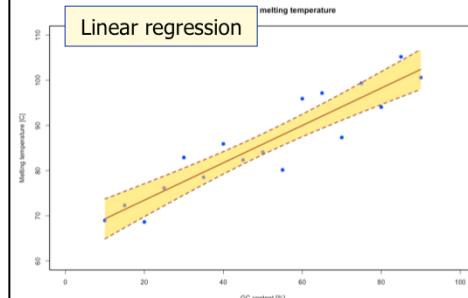
```

set.seed(137)
g1 <- rnorm(20, mean=1.7, sd=0.4)
g2 <- rnorm(20, mean=2.1, sd=0.4)
wf <- data.frame(grp1=g1, grp2=g2)
tf <- stack(wf)
  
```

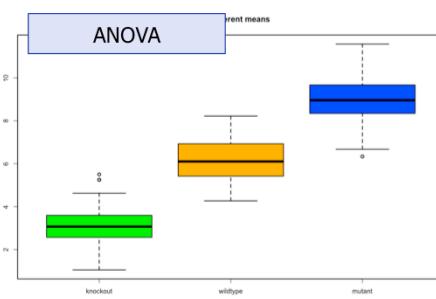
values	ind
1.85	grp1
2.25	grp1
...	...
1.53	grp2
1.98	grp2
...	...

This is not a mathematically rigorous proof, just an indication that the t-test is a special case of the one-way ANOVA. Note that the synthetic data come from Normal distributions with equal variances ("homoscedasticity") which is an ANOVA requirement. The `t.test()` function in R assumes unequal variances by default and runs "Welch's t-test" in that case, leading to slightly different p-values and fractional degrees of freedom. Hence we must specify `var.equal=TRUE` to make sure the equal-variance "classic" Student's t-test computation is used.

LINEAR REGRESSION AND ANOVA



Explanatory variable(s)	Metric
"Signal"	Linear equation
"Noise"	Residual error



Explanatory variable(s)	Categorical
"Signal"	Between-group variance
"Noise"	Within-group variance

Linear regression and ANOVA are conceptually related. Both approaches construct a model to "explain" the "spread" (i.e. the "change") of a response variable. In linear regression the explanatory variable is metric (i.e. the values are real numbers), and the model is a linear equation that describes the "systematic" change in the response due to changes in the explanatory variable. What the linear equation cannot explain is treated as "noise". In ANOVA the explanatory variable (the "treatment") is categorical (the values are the group membership labels), and the model describes the change in the group means due to group membership. The "spread" of values in the groups is the "within-group variance" which models the "noise" that remains unexplained.

ANOVA AS LINEAR REGRESSION

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i \quad \boxed{\text{Bivariate linear regression}}$$

$$y_i = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \epsilon_i \quad \boxed{\text{Three-group ANOVA}}$$

Indicator variables

These help to convert the non-numeric ANOVA factor levels (which are not even ordered!) to numbers.

$$I_1 = \begin{cases} 1 & \text{if group } = 2, \\ 0 & \text{otherwise} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if group } = 3, \\ 0 & \text{otherwise} \end{cases}$$

Group	I_1	I_2	Group mean
1	0	0	β_0
2	1	0	$\beta_0 + \beta_1$
3	0	1	$\beta_0 + \beta_2$

By using "indicator variables" we can re-formulate ANOVA as a multivariate linear regression problem. In general if we have k groups we will need $k-1$ indicator variables. The slide shows how to set these up in the case of three groups. The ANOVA model equation shows that the first group mean will be estimated by the intercept β_0 , the other group means will be the intercept plus the coefficient of the corresponding indicator variable.

ANOVA AS LINEAR REGRESSION [1] *Try me!*

Indicator variables by hand

```
ld <- within(sd, {  
  m <- as.integer(ind=="mutant")  
  w <- as.integer(ind=="wildtype")  
})
```

Group	"m"=I ₁	"w"=I ₂
"knockout"	0	0
"mutant"	1	0
"wildtype"	0	1

values	ind	m	w
3.461	knockout	0	0
...
7.012	wildtype	0	1
...
10.936	mutant	1	0
...

We add the "m" and "w" columns to the "sd" stacked data frame and return the result in a new dataframe "ld". The "m" and "w" columns are the indicator variables I₁ and I₂, but I named them like this to signify that they are associated with the "mutant" and "wildtype" columns.

ANOVA AS LINEAR REGRESSION [2] Try me!

ANOVA as linear regression

```
lr <- lm(values ~ m + w, ld)
summary(lr)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.1008	0.0984	31.51	<2e-16 ***
m	5.9286	0.1392	42.60	<2e-16 ***
w	3.0846	0.1392	22.16	<2e-16 ***

F-statistic: 908 on 2 and 297 DF, p-value: < 2.2e-16

Remember?

The ANOVA omnibus F statistic value was also 908.

Group means

```
apply(dd, MARGIN=2, mean)
```

	knockout	wildtype	mutant
	3.100759	6.185347	9.029331

Check the coefficients!

3.1008 + 3.0846
3.1008 + 5.9286

We perform multivariate linear regression on the dataframe with the indicator variables, and then check if the coefficients estimated the group means correctly. Also note that the F-statistic value from the linear regression is the same as what we got from the ANOVA (many slides before).

ANOVA AS LINEAR REGRESSION [3] Try me!

ANOVA as ANOVA ☺

```
ad <- aov(values ~ ind, sd)
summary(ad)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	2	1758.4	879.2	908	<2e-16 ***
Residuals	297	287.6	1.0		

Reminder

Same F statistic value as in the linear regression: **908**.

Sums-of-squares of the regression

```
summary.aov(lr)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
m	1	1282.6	1282.6	1324.6	<2e-16 ***
w	1	475.7	475.7	491.3	<2e-16 ***
Residuals	297	287.6	1.0		

Check the sums of squares!

```
calc.vars(dd)
```

ss.between var.between ss.within var.within

1758.3631450 879.1815725 287.5895181 0.9683149

1282.6 + 475.7 = 1758.4

We can see that the sums of squares match perfectly. For the "between-groups" sum you need to add the individual sums-of-squares for the two indicator variables. The "within-group" sum of square (the "noise") is obviously the same, whether you run ANOVA or the linear regression with indicator variables.

ANOVA AS LINEAR REGRESSION [4] Try me!

Reminder

ANOVA as linear regression

```
lr <- lm(values ~ m + w, ld)
summary(lr)
```



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.1008	0.0984	31.51	<2e-16 ***
m	5.9286	0.1392	42.60	<2e-16 ***
w	3.0846	0.1392	22.16	<2e-16 ***

F-statistic: **908** on 2 and 297 DF, p-value: < 2.2e-16

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.1008	0.0984	31.51	<2e-16 ***
indmutant	5.9286	0.1392	42.60	<2e-16 ***
indwildtype	3.0846	0.1392	22.16	<2e-16 ***

F-statistic: **908** on 2 and 297 DF, p-value: < 2.2e-16

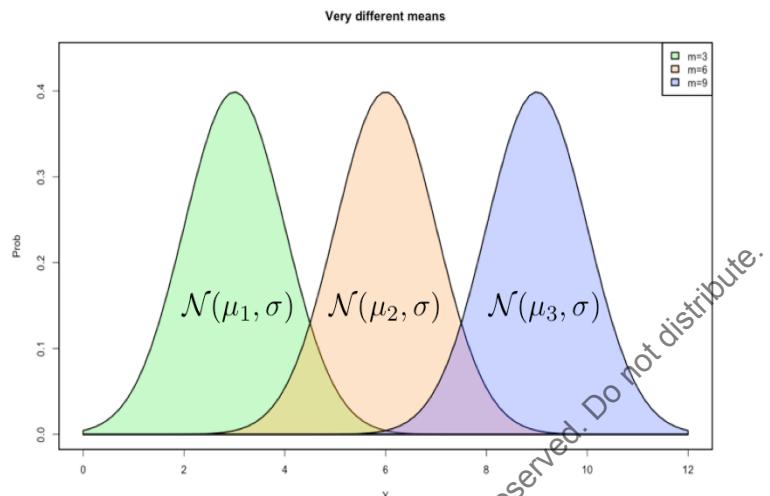
Automatic "regression"

```
ar <- lm(values ~ ind, sd)
summary(ar)
```

The `lm` function in R is smart enough to recognise a "tall" ANOVA table, and performs the conversion to the indicator variables we have seen two slides before automatically. They will just be named "indXXX", "indYYY", etc. where "XXX", "YYY" are the factor levels.

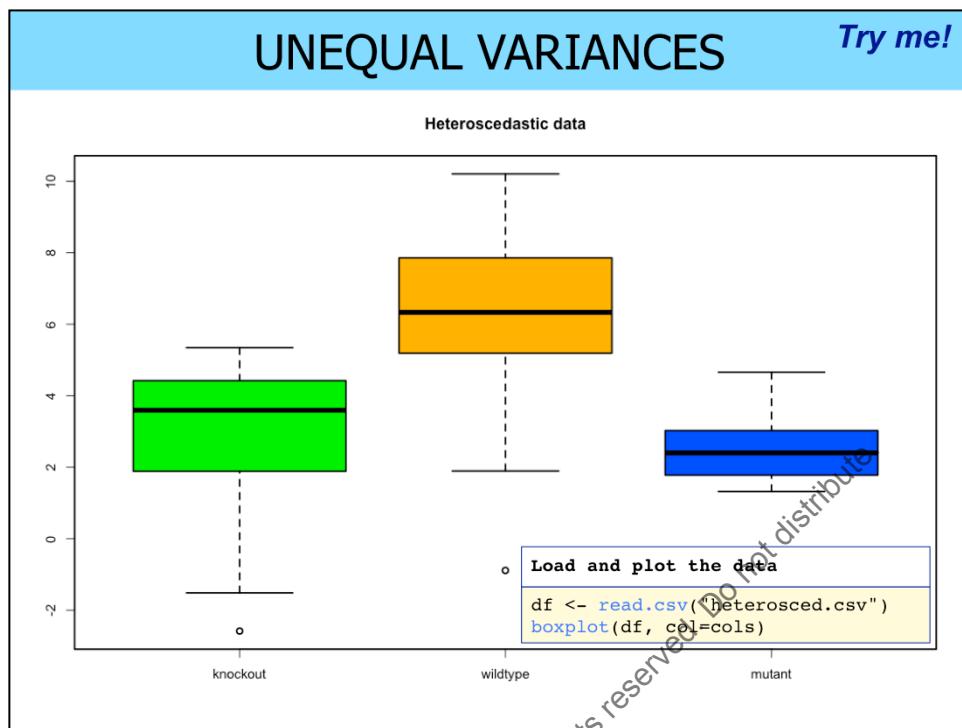
RELAXING THE ASSUMPTIONS

- 1) The population variances are equal
- 2) The underlying populations are Normal



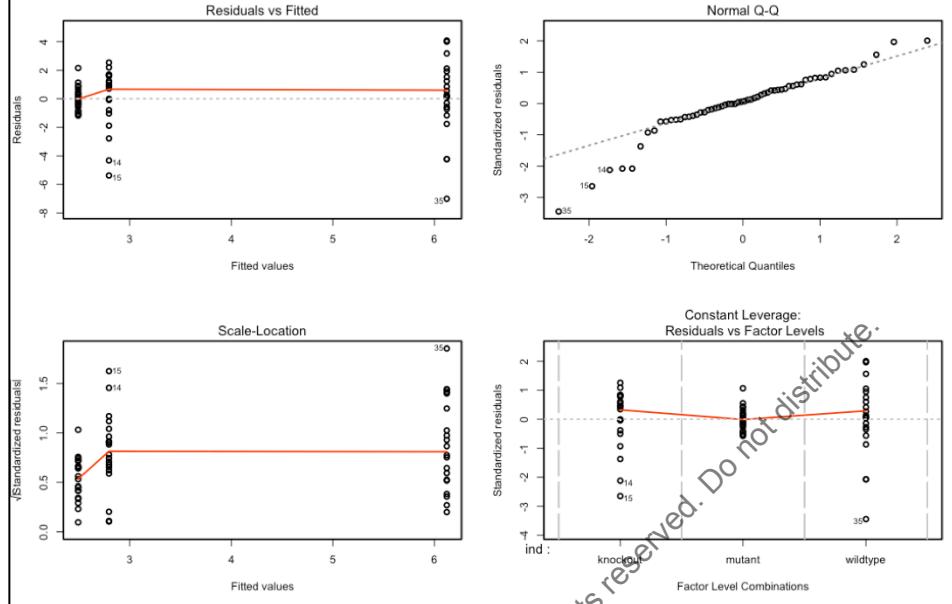
The standard assumption of ANOVA, namely that

- 1) The groups are sampled from Normal distributions
 - 2) The variance of the group populations are equal ("homoscedastic")
- can be relaxed.



Real data sets often do not care about the homoscedasticity requirement of ANOVA and the group variances turn out to be markedly different. Not all is lost though.

ANOVA DIAGNOSTICS PLOT *Try me!*



©VBCF/BioComp, András Aszódi 2018. All rights reserved. Do not distribute.

WELCH'S ONE-WAY TEST

Try me!

Homoscedasticity?

```
bartlett.test(df)  
fligner.test(df)
```

p = 2.519e-05
p = 0.02722

Heteroscedasticity!

This violates the ANOVA assumption of equal variances.

Welch's test

```
sf <- stack(df)  
print(oneway.test(values ~ ind, sf))
```

```
data: values and ind  
F = 15.305, num df = 2.000, denom df = 30.602, p-value = 2.472e-05
```

...because Welch's one-way test can be used to deal with heteroscedastic datasets.

HETEROSKEDASTIC POST-HOC TEST *Try me!*

Games-Howell test

```
source("gameshowell.R")
games.howell(sf$ind,sf$values)
```

Groups
The column indicating the group membership.

Observations
The column containing the measured values (the observations).

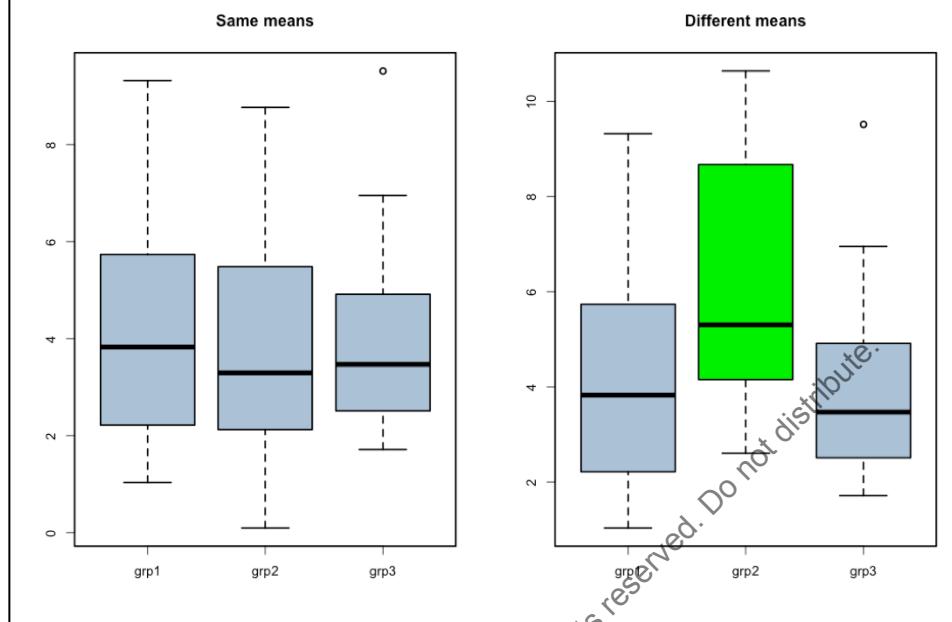
groups	Mean	Difference	Standard Error	t	df	p	upper limit	lower limit
1 knockout:wildtype	3.327	0.555	4.240	35.789	0.000	5.244	1.409	
2 knockout:mutant	-0.303	0.366	0.585	24.930	0.829	0.987	-1.594	
3 wildtype:mutant	-3.630	0.459	5.592	22.626	0.000	2.003	-5.258	

<https://rpubs.com/aaronsc32/games-howell-test>

The Games-Howell post-hoc test is applicable whenever the group variances are not equal to each other (heteroskedastic). What's more, it works also if the distributions are not Normal and you use Kruskal's non-parametric test of equality of group means instead of a one-way ANOVA.

The Games-Howell test is not part of the standard R distribution at the time of this writing. You can find it in the "userfriendlyscience" package or source my "gameshowell.R" script which contains the implementation by Aaron Schlegel (see also the URL on the slide).

RELAXING THE NORMALITY ASSUMPTION



"Classical" ANOVA relies on the assumption that the samples come from Normal populations because the F-test at the heart of the procedure makes this assumption. Luckily, there is a non-parametric test developed by Kruskal that can be used to compare the means of non-Normally distributed groups.

TESTING SEVERAL MEANS

Try me!

grp1	grp2	grp3
4.6	3.7	2.3
5.6	1.5	6.1
...

gd
5.2
3.0
...

Data preparation (done by AA for you)

```
set.seed(137)
g1 <- rgamma(20, shape=2, scale=2.0)
g2 <- rgamma(20, shape=2, scale=2.0)
g3 <- rgamma(20, shape=2, scale=2.0)
gd <- rgamma(20, shape=2,
             scale=2.0) + 1.5
```

"Different" means

This sample is "shifted" by 1.5

The second group now has a different mean

"Same" means

3 samples from the same population

p-values

$p = 0.8128$

$p = 0.008564$

Try Kruskal's test

```
d.same<-read.csv("kruskal_same.csv")
d.diff<-read.csv("kruskal_diff.csv")
kruskal.test(d.same)
kruskal.test(d.diff)
```

Here we make up two datasets of 3 groups of Gamma-distributed samples each. In the `d.same` dataset all 3 columns come from the same Gamma distribution. In the `d.diff` dataset we exchange the second group with Gamma samples that were "shifted" by 1.5 so that they have the same shape but a different mean. You can load the datasets from the files "kruskal_same.csv" and "kruskal_diff.csv", respectively.

POST-HOC TEST REFERENCE

Test name	R function name	Remarks
Tukey's "honest significant differences"	<code>TukeyHSD</code>	<i>Comes with nice plots</i>
Pairwise t-tests	<code>pairwise.t.test</code>	<i>Offers various corrections: Bonferroni, Benjamini-Hochberg, ...</i>
Other post-hoc tests	<code>PostHocTest</code> in the DescTools package	<i>Corrections as above, plus Scheffé, Newman-Keuls, Duncan, ...</i>
Games-Howell test	Several implementations	<i>Heteroskedastic data</i>

There are a large variety of post-hoc tests that differ in their way of compensation for the multiple comparison problem. This table helps you get started but it does not constitute an exhaustive reference.

MORE THAN ONE "TREATMENT"



So far we have looked at "one-way" ANOVA, which means there was one "treatment", i.e. one categorical explanatory variable. Often we conduct experiments where two or even more factors influence the outcome. In this example, Farmer Joe is measuring the growth of his pigs depending on their genotype ("wildtype", "mutant", "knockout" like in our previous example) AND on the type of food they get ("corn" and "swill").

THE DATASET

Try me!

Load the data

```
d2 <- read.csv("a2.csv")
```

Genotype	Feed	
	corn	swill
The Plan		
knockout		
mutant		
wildtype		



genotype	feed	growth
"knockout"	"corn"	4.384
...
"knockout"	"swill"	1.307
"mutant"	"corn"	4.116
...
"mutant"	"swill"	1.866
"wildtype"	"corn"	5.614
...
"wildtype"	"swill"	3.738

Farmer Joe's observations on the growth of his pigs as a function of their genotype AND the type of feed they got are collected in a single table. He has several observations for each and every combination of genotype and feed.

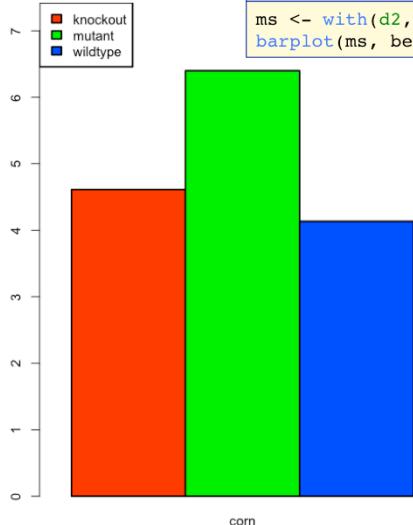
©VBCF/BioComp, András Aszódi, 2018 All rights reserved. Do not distribute.

THE GROUP MEANS

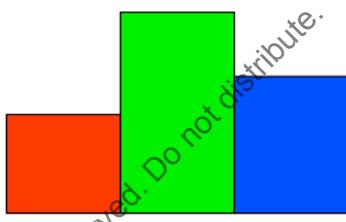
Try me!

Plot the means

```
ms <- with(d2,tapply(growth,list(genotype,feed),mean))  
barplot(ms, beside=T, col=rainbow(3))
```

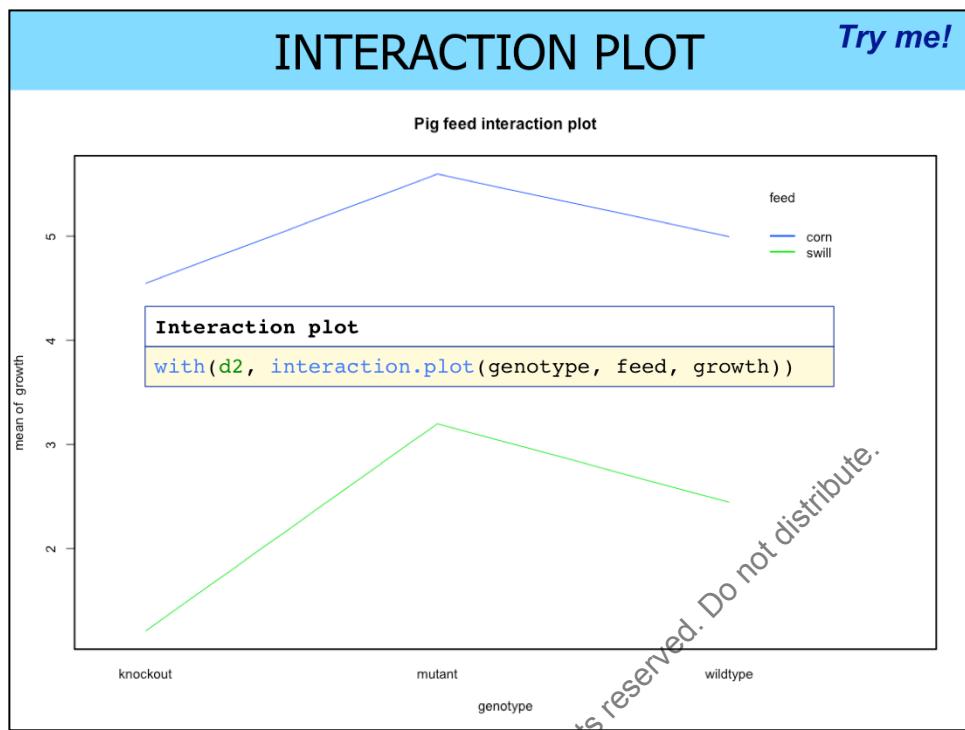


Means	corn	swill
knockout	4.55	1.21
mutant	5.59	3.20
wildtype	4.99	2.44



We calculate the means of the observations for each genotype and feed combination using the `tapply()` function. A matrix 'ms' is returned that contains the means, and the columns and rows are nicely named. The 'barplot' function can automatically make a plot from this matrix.

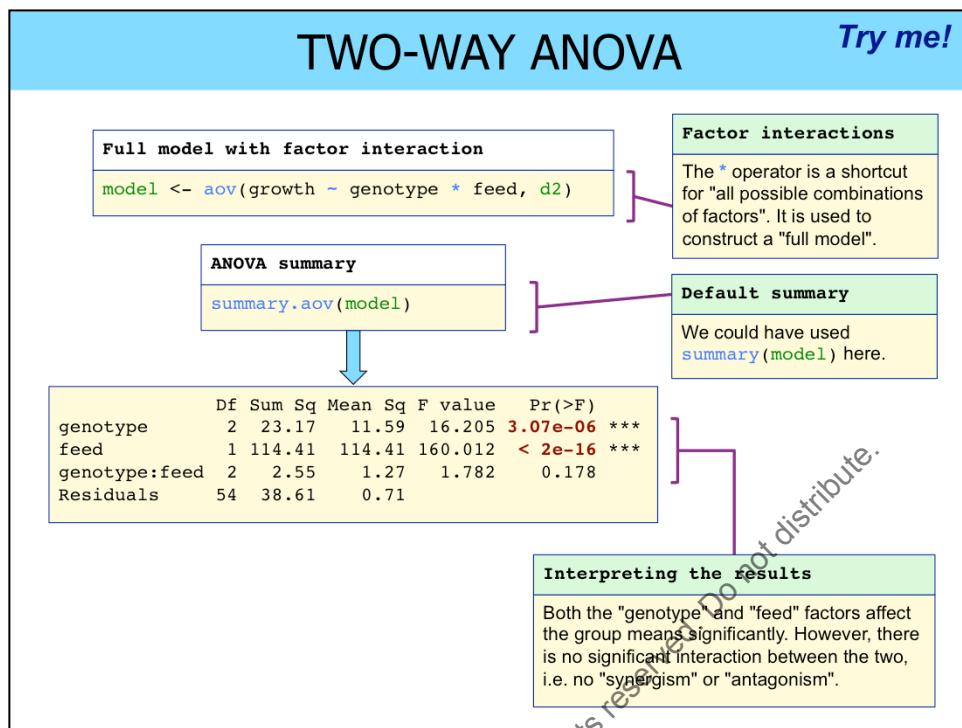
©VBCF/BioComp, András Zolnai-Beregi All rights reserved. Do not distribute.



This is the standard R function to plot two-way ANOVA datasets. The plot on the slide was made with the additional options:

`lty="solid", col=c("blue","green"), main="Pig feed interaction plot"`

©VBCF/BioComp, András Aszodi 2018 All rights reserved. Do not distribute.



Because two factors (genotype and feed) can affect the growth, we run a "two-way" ANOVA. A "full model" is constructed that checks the main effects of the factors on their own and also a possible interaction (synergism or antagonism) between them. Note that the input data are already in a "tall" format so we do not need to rearrange them.

AN ALTERNATIVE VIEW

Try me!

The "intercept"

The mean at the alphabetically first factor levels ("knockout" and "corn"). All other effects are relative to this mean.

Linear model summary

`summary.lm(model)`

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.5468	0.2674	17.004	< 2e-16 ***
genotypemutant	1.0485	0.3782	2.773	0.00762 **
genotypewildtype	0.4477	0.3782	1.184	0.24165
feedswill	-3.3376	0.3782	-8.826	4.69e-12 ***
genotypemutant:feedswill	0.9419	0.5348	1.761	0.08385 .
genotypewildtype:feedswill	0.7856	0.5348	1.469	0.14763

Interpreting the main effects

For the "genotype" factor, the "mutant" mean is significantly different from "knockout" (the first), "wildtype" is not. For the "feed" factor, "swill" is significantly different from "corn".

Interactions

None of the factor levels show significant interactions.

Because ANOVA is just another kind of linear model, we can ask R explicitly to summarise the result as if it had been generated by the `lm()` function. The advantage is that the effects of the factors and their interactions are presented in more detail. Because the changes are given relative to the first level of each factor (which is determined alphabetically, so the first level of "genotype" is "knockout" and the first level of "feed" is "corn" in our example), sometimes the `lm`-style summary is a bit difficult to interpret at first sight.

(SIMPLE) MODEL REDUCTION *Try me!*

Model without factor interaction

```
noint.model <- aov(growth ~ genotype + feed, d2)
```

ANOVA summary

```
summary.aov(noint.model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
genotype	2	23.17	11.59	15.77	3.71e-06 ***
feed	1	114.41	114.41	155.66	< 2e-16 ***
Residuals	56	41.16	0.73		

No interactions

The `+` operator indicates that we are interested in the independent main effects of the factors only.

Interpreting the results

Obviously both the "genotype" and "feed" factors remained significant. The residual sum of squares increased only slightly from 38.61 to 41.16, further indicating that dropping the interaction term was desirable.

Especially in more complex situations with lots of factor levels or many interactions (e.g. three-way ANOVA) it is advisable NOT to leave all possible interactions in the model. Model reduction is a complex topic and in this course only a simple example can be given. We have seen before that the full model was not necessary because there was no significant interaction between the factors. Leaving the interaction out gives rise to a simpler model which is almost always preferable.

THE SIMPLIFIED MODEL

Try me!

Linear model summary

summary.lm(noint.model)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.2588	0.2214	19.240	< 2e-16	***
genotypemutant	1.5194	0.2711	5.605	6.62e-07	***
genotypewildtype	0.8405	0.2711	3.100	0.00302	**
feedswill	-2.7617	0.2214	-12.477	< 2e-16	***

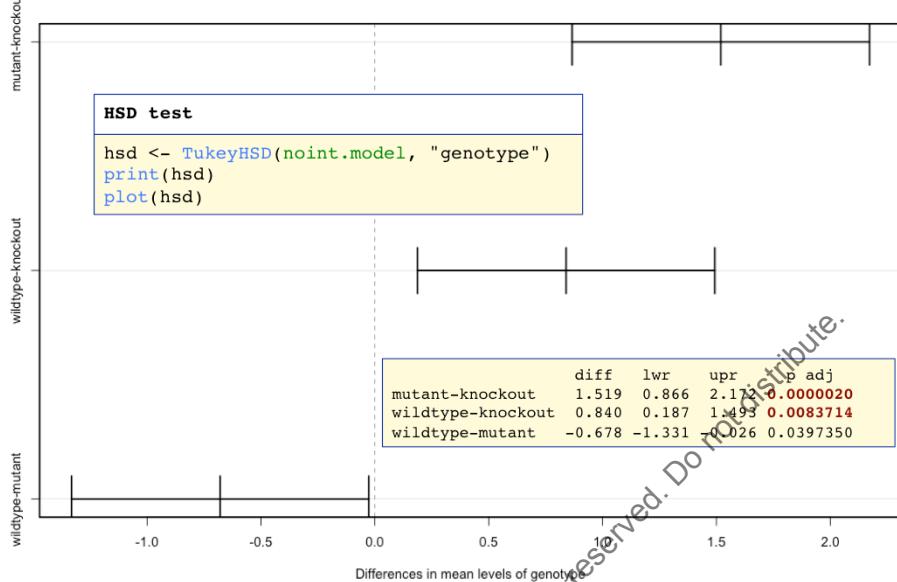
Main effects only

In the simplified model now all effects are significant. Their pairwise differences are also quite different from each other (relative to the std.error). This is an adequate minimal model.

The "lm summary" shows nicely that all effects are significant, giving rise to an adequate minimal model.

TWO-WAY POST HOC TESTS *Try me!*

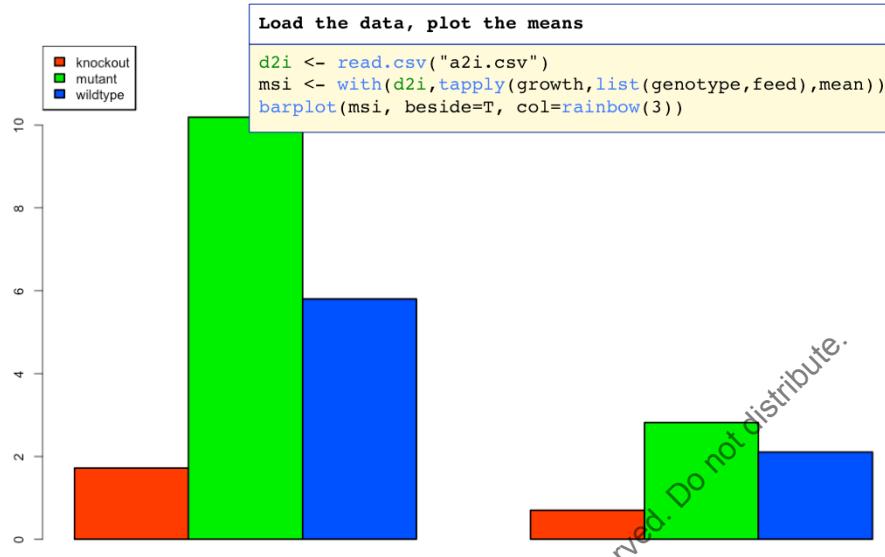
95% family-wise confidence level



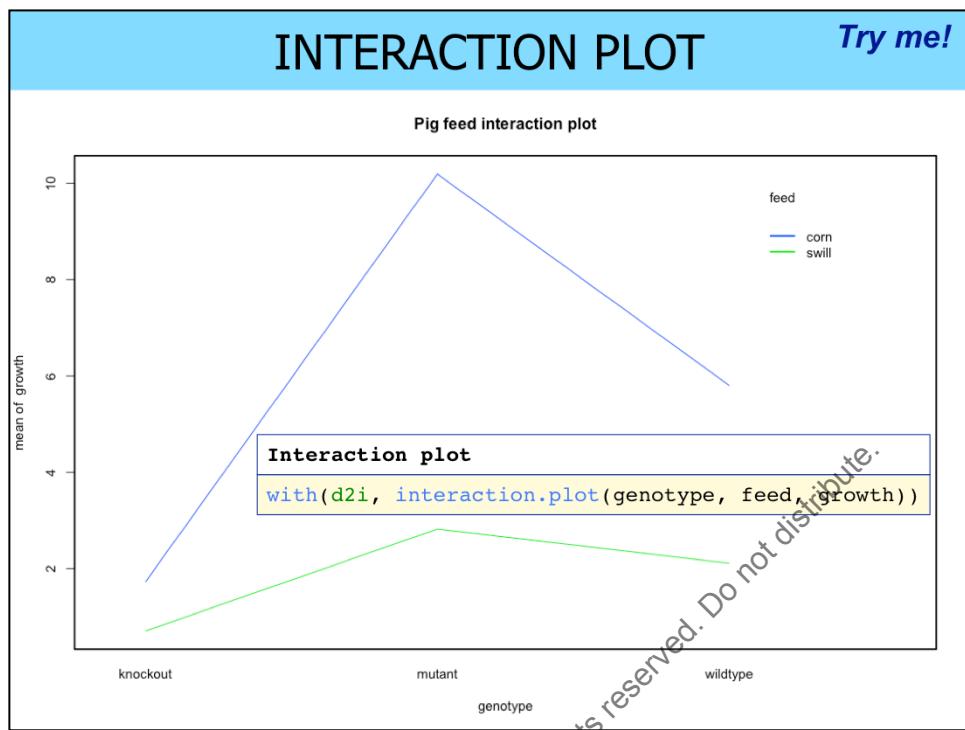
The post-hoc tests work the same way as with the one-way ANOVA.
Only the HSD test results are shown.

A DIFFERENT DATASET

Try me!



These measurements were made after Farmer Joe's publication was rejected by the Journal of Irreproducible Results. Referee #2 wanted to see a synergistic effect between genotype and feed.



This is the standard R function to plot two-way ANOVA datasets. The plot on the slide was made with the additional options:

`lty="solid", col=c("blue","green"), main="Pig feed interaction plot"`

TWO-WAY ANOVA

Try me!

Full model with factor interaction

```
model <- aov(growth ~ genotype * feed, d2i)
```

ANOVA summary

```
summary.aov(model)
```

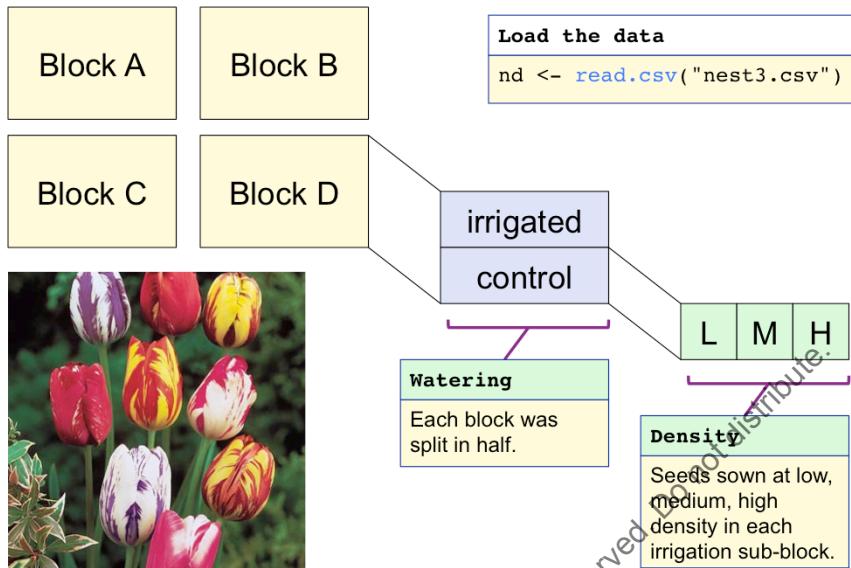
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
genotype	2	280.22	140.11	121.10	< 2e-16 ***
feed	1	243.28	243.28	210.27	< 2e-16 ***
genotype:feed	2	101.67	50.84	43.94	4.71e-12 ***
Residuals	54	62.48	1.16		

Interpreting the results

This time there is significant interaction between the two factors.

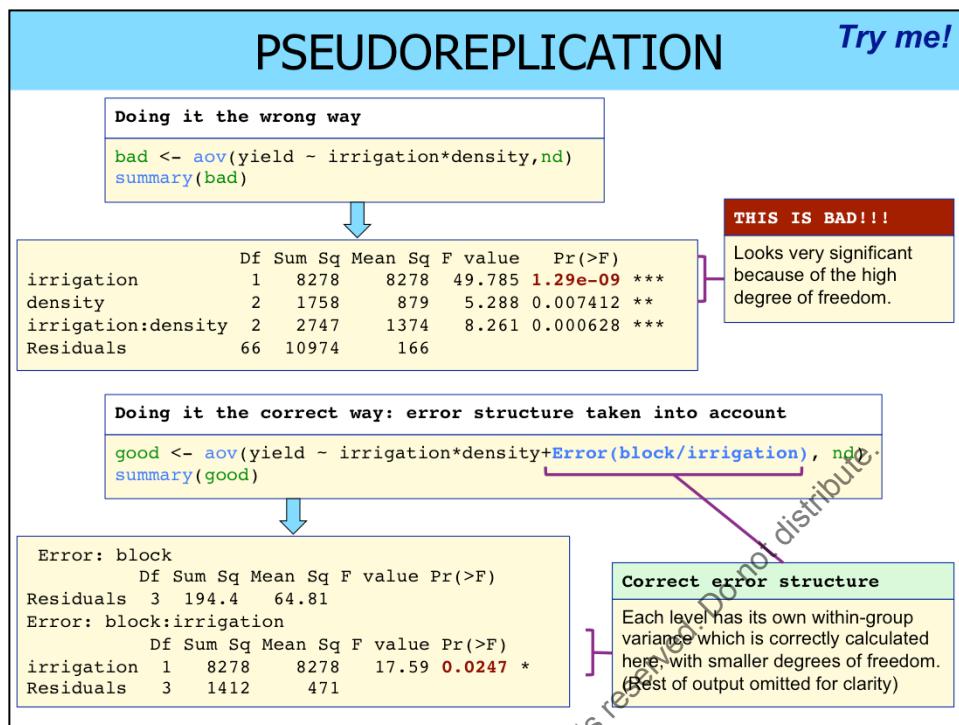
The full model was needed because the genotype:feed interaction is indeed significant, there is a "synergy".

NESTED DESIGN ("SPLIT-PLOT")



Source: Ripley, The R book, p. 519ff.

We stay with agriculture but switch to plants. These experiments often involved "split-plot" designs whereby a big plot is subdivided in smaller blocks, and then within the blocks more subdivisions can be made, each level of subdivision corresponds to one treatment. In this example (which is a simplified version of Prof. Ripley's "splityield.txt" data), tulips are sown in 4 blocks A,B,C,D. Each block is subdivided into two, one half is watered ("irrigated"), the other stays dry as the "control". Finally, each irrigation sub-block is further divided into 3, where the seeds are sown at low, medium and high density. The response variable is the yield, which is the number of tulips that survived. (In the original dataset there was another level of subdivision which I omitted for simplicity.)



Calculating a standard ANOVA that simply assumes that each group has the same variance can go wrong in a nested design because the algorithm "sees" more degrees of freedom than what is actually there. This is called "pseudoreplication". In our example, if we look at the irrigation treatment, there are only 8 data points (because the lower subdivisions should be "merged"), and the correct degree of freedom should be 3, not 66. Pseudoreplication leads to unrealistically low p-values (remember, if you have enough data points, even the smallest changes appear to be significant). We correct for pseudoreplication by adding the Error term, starting with the uppermost division level (in our example the "block" factor), then the smaller and smaller ones, separated by / -es. The lowest level does not need to be specified.