

## Comparison of means of multiple groups

When we have many (more than 2) groups, we may think to perform  $t$ -tests for  $\mu_1 = \mu_2$ , then  $\mu_1 = \mu_3$ , then  $\mu_2 = \mu_3 \dots$

## Comparison of means of multiple groups

When we have many (more than 2) groups, we may think to perform  $t$ -tests for  $\mu_1 = \mu_2$ , then  $\mu_1 = \mu_3$ , then  $\mu_2 = \mu_3 \dots$   
Why is this not appropriate? not optimal?

## Comparison of means of multiple groups

When we have many (more than 2) groups, we may think to perform  $t$ -tests for  $\mu_1 = \mu_2$ , then  $\mu_1 = \mu_3$ , then  $\mu_2 = \mu_3 \dots$   
Why is this not appropriate? not optimal?

If we perform many tests, we need to correct probability levels. If we perform 20 tests, each with 5% probability of being positive, we suspect some may become positive just for chance. . .

## Comparison of means of multiple groups

When we have many (more than 2) groups, we may think to perform  $t$ -tests for  $\mu_1 = \mu_2$ , then  $\mu_1 = \mu_3$ , then  $\mu_2 = \mu_3 \dots$

Why is this not appropriate? not optimal?

If we perform many tests, we need to correct probability levels. If we perform 20 tests, each with 5% probability of being positive, we suspect some may become positive just for chance. . .

Tests are not independent. . .

## Comparison of means of multiple groups

When we have many (more than 2) groups, we may think to perform  $t$ -tests for  $\mu_1 = \mu_2$ , then  $\mu_1 = \mu_3$ , then  $\mu_2 = \mu_3 \dots$

Why is this not appropriate? not optimal?

If we perform many tests, we need to correct probability levels. If we perform 20 tests, each with 5% probability of being positive, we suspect some may become positive just for chance. . .

Tests are not independent. . .

Proper way to correcting for this: analysis of variance.

# Analysis of variance

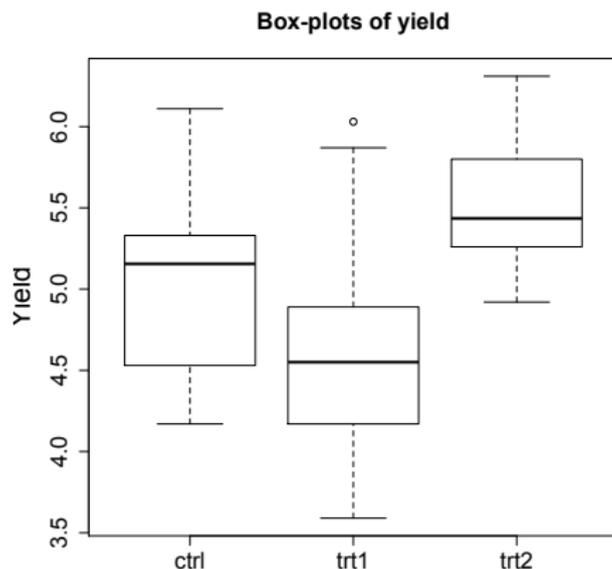
From R.R. Sokal and F.J. Rohlf, *Biometry*, 2nd Edition (1981):

*A knowledge of analysis of variance is indispensable to any modern biologist and, after you have mastered it, you will undoubtedly use it numerous times to test scientific hypotheses. However, the analysis of variance is more than a technique for statistical analysis. Once understood, analysis of variance provides an insight into the nature of variation of natural events, into Nature in short, which is possibly of even greater value than the knowledge of the method as such. However, like other models in science, analysis of variance may create constructions of nature in the mind of the scientist that give rise to misleading or unproductive conclusions.*

# One-way analysis of variance

We have measured a response variable  $Y$  in  $g$  groups.

**Example:** Yield under a control and two treatments<sup>1</sup>



**Question:** Is yield different under control and treatments?

<sup>1</sup>data from Dobson, A. J. (1983) *An Introduction to Statistical Modelling* in the standard R library

## Formulae in analysis of variance

$$Y_{ik} \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, g \text{ (groups)}$$

$$H_0: \mu_i \equiv \mu, \sigma^2 > 0.$$

$$H_1: \text{differences among } \mu_i, \sigma^2 > 0.$$

**Note;** if  $g = 2$  test on equality of means ( $t$ -test).

Method: decomposition of variance

$$\sum_{i,k} (Y_{ik} - \bar{Y})^2 = \sum_i n_i (\bar{Y}_i - \bar{Y})^2 + \sum_i \sum_k (Y_{ik} - \bar{Y}_i)^2.$$

total variation	variation among groups	variation within groups
-----------------	---------------------------	----------------------------

$\bar{Y}_i$ : mean in group  $i$ ,       $\bar{Y}$  general mean.

# Hypothesis testing in analysis of variance

Under  $H_0$  (all  $\mu_i$  equal)

$$\frac{1}{\sigma^2} SS_{\text{among}} = \frac{1}{\sigma^2} \sum_i n_i (\bar{Y}_i - \bar{Y})^2 \sim \chi^2(g-1)$$

$$\frac{1}{\sigma^2} SS_{\text{within}} = \frac{1}{\sigma^2} \sum_i \sum_k (Y_{ik} - \bar{Y}_i)^2 \sim \chi^2((n-1)g)$$

and they are independent.

*[Formulae true with equal size  $n$  of all groups]*

Hence the distribution of

$$\frac{SS_{\text{among}}/(g-1)}{SS_{\text{within}}/((n-1)g)}$$

is known and is called  $F_{g-1, (n-1)g}$ .

# Analysis of variance: computations

**Recipe for hypothesis testing:** Compute

$$\frac{SS_{\text{among}}/(g-1)}{SS_{\text{within}}/((n-1)g)};$$

if it is below the  $(1 - \alpha)$ -quantile of  $F_{g-1, (n-1)g}$ , accept  $H_0$  at level  $\alpha$ ; otherwise, reject it.

**Example:** Analysis of Variance Table

Response: weight

	Df	Sum Sq	Mean Sq	F value	Pr(> F)
group	2	3.7663	1.8832	4.8461	0.01591 *
Residuals	27	10.4921	0.3886		

## Analysis of variance: $R^2$

A related quantity

$$R^2 = \frac{SS_{\text{among}}}{SS_{\text{total}}} = \frac{\text{'variance' 'explained' by model}}{\text{total 'variance'}}.$$

Necessarily  $0 \leq R^2 \leq 1$ .

High  $R^2$  and rejection of  $H_0$  are different (though related) facts.

R-squared in the example: 0.2641

## Paired comparisons

Rejecting  $H_0$ , we find that **not all** group means are equal. But which are different from which ones?

We can compare the means of any given pair of groups by a  $t$ -test for equality of means. By default R compares each group to one taken as reference<sup>2</sup>. Results in the example:

Coefficients:

	Estimate	Std. Error	$t$ value	$\Pr(>  t )$	
(Intercept)	5.0320	0.1971	25.527	$< 2e - 16$	***
grouptrt1	-0.3710	0.2788	-1.331	0.1944	
grouptrt2	0.4940	0.2788	1.772	0.0877	.

grouptrt1 is the difference between the mean of trt1 and that of the reference (ctrl).

No treatment appears different from the control at the 5% level.

---

<sup>2</sup>The first one in alphabetical order

## Paired comparisons

But which is the correct level of significance?

We may be comparing only groups whose means appear particularly different. If there are 6 groups, there would be 15 possible comparisons. The probability that at least one of them would be beyond the nominal 5% level is much higher. . .

One solution is called Tukey's HSD (Honest Significant Difference) that looks at the distribution of maximum difference:

	diff	lwr	upr	p adj
trt1-ctrl	-0.371	-1.0622161	0.3202161	0.3908711
trt2-ctrl	0.494	-0.1972161	1.1852161	0.1979960
trt2-trt1	0.865	0.1737839	1.5562161	0.0120064

We can conclude that the yield with treatment 2 is significantly different (higher) than with treatment 1.

## 2-way analysis of variance

We may be considering the effect of two factors ( $A$  and  $B$ ) on our response variable, for instance fertilizer and variety on maize yield; or therapy and sex on cholesterol level.

The most general model would be

$$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2), k = 1, \dots, n_{ij}$$

$\mu$  is the general mean,  $\alpha_i$  the effect of level  $i$  of factor  $A$ ,  $\beta_j$  the effect of level  $j$  of factor  $B$ ;  $(\alpha\beta)_{ij}$  their 'interaction'.

$n_{ij}$  number of observations at combination  $(i, j)$  of the factors  
*[much easier if  $n_{ij}$  are all the same].*

What does the 'interaction' mean? Without that term, the effect of the two factors would be additive; on the other hand, it may well be that variety 1 grows better with fertilizer 1 than with 2, while variety 2 has the opposite trend. Such features are captured by the interaction term.

## Variance decomposition

$$SS_{\text{total}} = SS_A + SS_B + SS_{A \times B} + SS_{\text{within}}.$$

First, test of interaction:

$$\frac{SS_{A \times B} / (a - 1)(b - 1)}{SS_{\text{within}} / (ab(n - 1))}$$

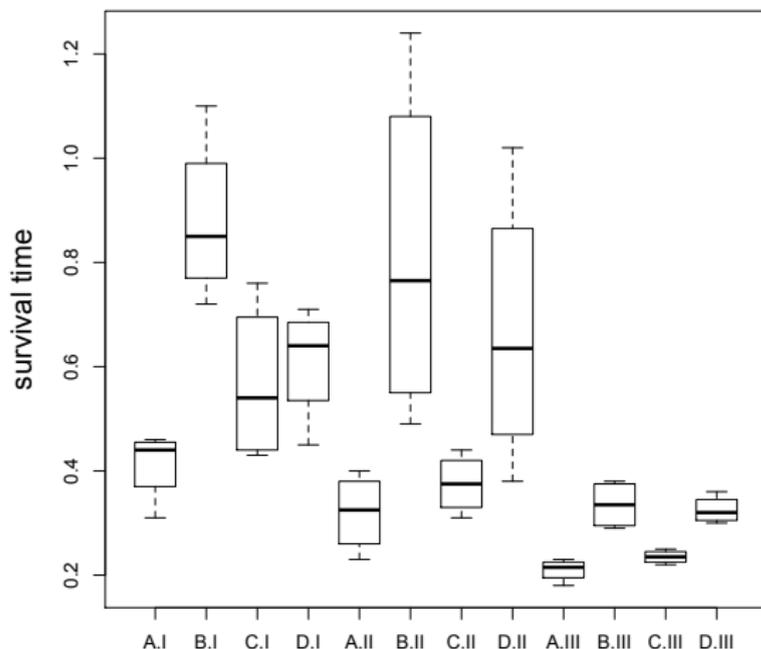
$a$  = number of levels of factor  $A$ ;  $b$  = number of levels of factor  $B$ ;  
 $n$  = number of replicates of each combination of the two factors.

If the interaction is significant, it is difficult to talk of the 'effect' of factor  $A$  or factor  $B$ ...

If the interaction is not significant, then one proceeds with an additive model with factors  $A$  and  $B$ .

## An example

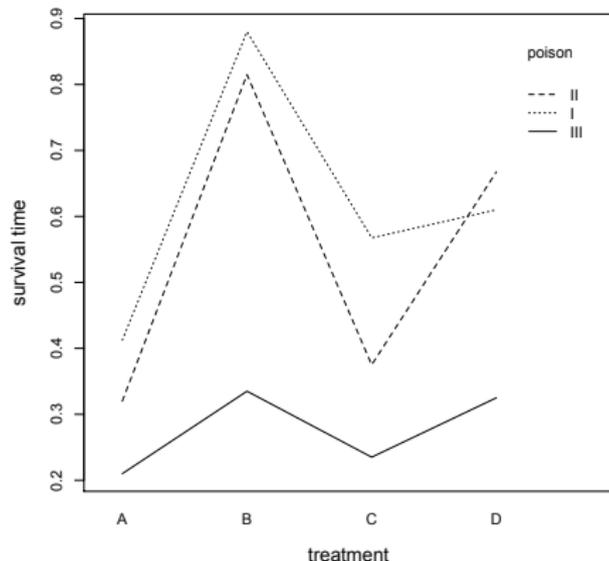
I consider a dataset (found somewhere on the Web) that presents the survival time of 48 rats after having received a poison (first factor) of 3 different types (I, II or III), and one of 4 different treatments (second factor) named *A* to *D*. See the data in the figure:



## Interaction plot

An “interaction plot” is a visual guidance to see interaction:  
y-axis: mean response (survival time); x-axis: first factor (treatment);  
lines: second factor (poison).

If additive model, parallel lines. Lines in opposite directions:  
relevant interaction.



In the example, lines not far from parallel: weak interaction.

## Test for interaction

**Note:** if  $n = 1$  (no replicates) impossible testing for interactions (# of parameters = # of interactions); analysis possible only using additive model, after an interaction plot.

Results of complete 2-way anova:

Response: time

	Df	Sum Sq	Mean Sq	F value	Pr(> F)	
poison	2	1.03301	0.51651	23.2217	3.331e-07	***
treat	3	0.92121	0.30707	13.8056	3.777e-06	***
poison:treat	6	0.25014	0.04169	1.8743	0.1123	
Residuals	36	0.80072	0.02224			

poison:treat are the interaction terms. Relevant  $F$  test is ratio between Mean Sq of interaction and of Residuals (within).

The result shows that interaction is not significant.

## Testing for factor effects

Repeat analysis of variance without interaction. Result:

Response: time

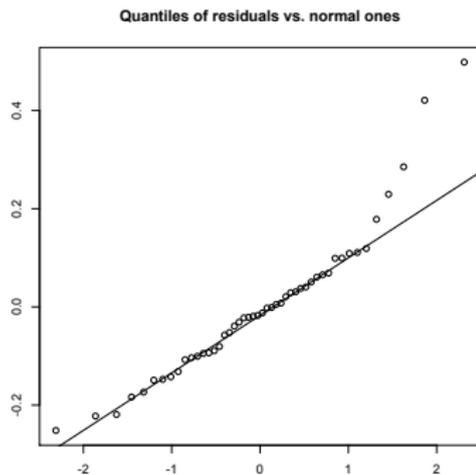
	Df	Sum Sq	Mean Sq	F value	Pr(> F)	
poison	2	1.03301	0.51651	20.643	5.704e-07	***
treat	3	0.92121	0.30707	12.273	6.697e-06	***
Residuals	42	1.05086	0.02502			

The variation relative to the two factors is the same as before. The residual variation ( $SS_{\text{within}}$ ) has been increased of what was previously attributed to interaction (not present in the tested additive model).

$F$  values are slightly different from before. The two factors still highly significant: mean effect of treatments, and of poisons is **not** the same!

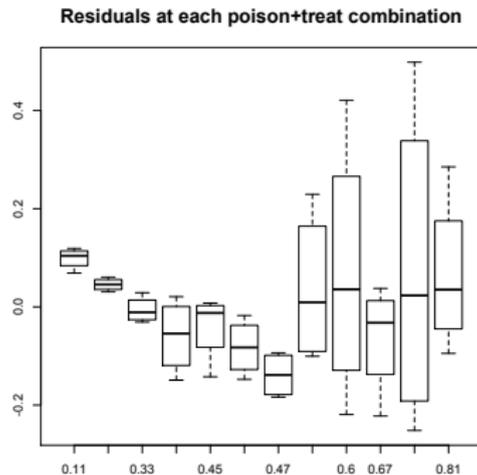
# How to check assumptions (visually)

Check for normality:  
QQ-plot of residuals



If from a normal distribution,  
points should lie on a line.

Check for equal variances:  
Box-plots of residuals

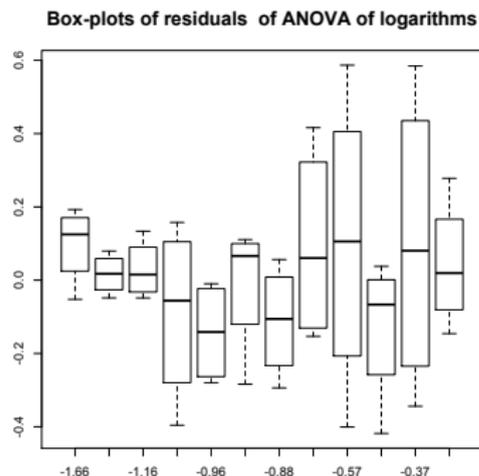
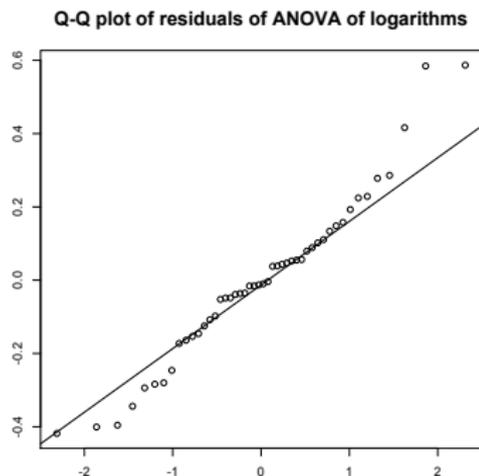


If equal variances, boxes should  
have approximately the same width.

# Transformations of data to correct the problems

A classical transformation is the **logarithm**.

Repeat the **analysis of variance on the logarithm** of survival, and plot the residuals of the regression:

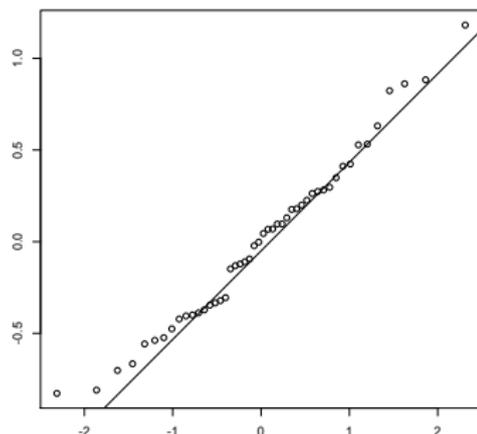


Slightly better...

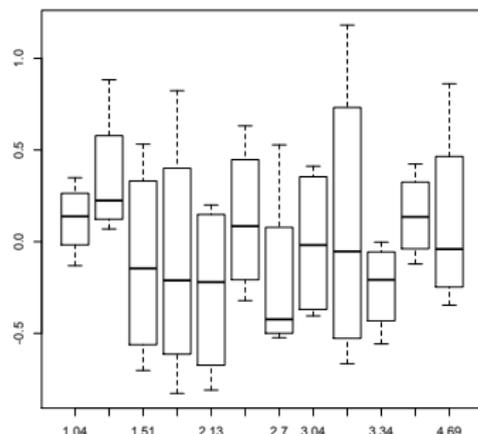
# Transformations. II

Use the inverse:  $1/\text{survival} \approx \text{mortality}$

Q-Q plot of residuals of ANOVA of inverses



Box-plots of residuals of ANOVA of logarithms



This seems the best one so far. . . . present the analysis on this.

# ANOVA of 1/survival on factors

Analysis of variance table. Response:1/ time

	Df	Sum Sq	Mean Sq	F value	Pr(> F)	
poison	2	34.877	17.4386	71.708	2.865e-14	***
treat	3	20.414	6.8048	27.982	4.192e-10	***
Residuals	42	10.214	0.2432			

Estimate of effects of each level of the variable.

	Estimate	Std. Error	t value	Pr(>  t )	
(Intercept)	2.6977	0.1744	15.473	<2e-16	***
poisonII	0.4686	0.1744	2.688	0.01026	*
poisonIII	1.9964	0.1744	11.451	1.69e-14	***
treatB	-1.6574	0.2013	-8.233	2.66e-10	***
treatC	-0.5721	0.2013	-2.842	0.00689	**
treatD	-1.3583	0.2013	-6.747	3.35e-08	***

*Intercept = poison I + treat A. Other rows are relative differences.*

## Tukey's multiple comparisons of means

### Poisons

	diff	lwr	upr	p adj
II-I	0.4686413	0.04505584	0.8922267	0.0271587
III-I	1.9964249	1.57283950	2.4200103	0.0000000
III-II	1.5277837	1.10419824	1.9513691	0.0000000

### Treatments

	diff	lwr	upr	p adj
B-A	-1.6574024	-2.1959343	-1.11887050	0.0000000
C-A	-0.5721354	-1.1106673	-0.03360355	0.0335202
D-A	-1.3583383	-1.8968702	-0.81980640	0.0000002
C-B	1.0852669	0.5467351	1.62379883	0.0000172
D-B	0.2990641	-0.2394678	0.83759598	0.4550931
D-C	-0.7862029	-1.3247347	-0.24767096	0.0018399

*All poisons are different. All treatments are different, except B with D and (perhaps) A with C.*

## Do we need a 2-way analysis ?

Analyse the effect of treatment without considering poisons:

	Df	Sum Sq	Mean Sq	F value	Pr(> F)	
treat	3	20.414	6.8048	6.6401	0.0008496	***
Residuals	44	45.091	1.0248			

*Still an effect of treatment, but much lower!*

Tukey's multiple comparisons of means

	diff	lwr	upr	p adj
B-A	-1.6574024	-2.76085841	-0.5539464	0.0012809
C-A	-0.5721354	-1.67559147	0.5313206	0.5156472
D-A	-1.3583383	-2.46179432	-0.2548823	0.0103775
C-B	1.0852669	-0.01818908	2.1887230	0.0554193
D-B	0.2990641	-0.80439193	1.4025201	0.8871316
D-C	-0.7862029	-1.88965888	0.3172532	0.2418871

*Only differences between A and B, and A and C.*

Performing the correct analysis (2-way), results are different.