Applied Biostatistics

https://moodle.epfl.ch/course/view.php?id=15590

- Research process
- Basic experimental design ideas
- Analysis of variance
- ANOVA modeling with R

Research process

- *Question* of scientific interest
- Decide : *what data* to collect (and how)
- Collect and *analyze* the data
- Conclusions, generalizations : *inference* about the population of interest
- *Communication* and diffusion of results

Generic Question : Does a'treatment' cause an 'effect' ?

Exemples :

- Does smoking cause cancer,cardiac illness, *etc* ?
- Does consuming oat bran decrease cholesterol ?
- Does echinacea prevent illness ?
- Does exercise slow the aging process ?

Approach the question

- One simple method for resolving this type of question is to
  Une méthode simple pour résoudre ce type de question consiste à *compare two groups* of study subjects :
  - *Control group* : gives a base level for comparison
  - *Treatment group* : group receiving the 'treatment'

Types of studies

- *Experimental studies* : subjects assigned to groups (treatment, control) by the investigator
  - *randomization* : protects against bias in group assignments
  - *'blind', 'double-blind'* : protects against bias in the evaluation of results
  - *placebo* : artificial treatment
- *Observational study* : subjets 'assign' themselves to groups
  - *Confounding factor* : a factor associated with both the treatment *and* the result

Comments

- With a well-planned and executed controled experiment, it is possible to infer *causation*
- This is *NOT possible* with observational studies due to the presence of confounding factors
- When there are confounding factors, it is not possible to say whether the observed difference between the groupsis due to the *treatment* or to the *confounding factor*
- However, it is not always possible to carry out an experimental study, for *pratical* and/or *ethical* reasons

Example : Hibernation

- General question : *How do changes in an animal's environment induce hibernation ?*
- What changes should be studied *? ?*
  - temperature
  - photoperiod (daylight duration)
- What measures to take ?
  - nerve enzymatic activity (Na+K+ATP-ase)
- What animal to study ?
  - golden hamster, 2 organs

Specific question

- General question : *How do changes in an animal's environment induce hibernation ?*
- *Specific question :* What is effect of changing daylight duration on the enzyme concentration of the sodium pump in two golden hamster organs ?

Sources of variability

- Variability due to the conditions of interest (wanted)
  - Duration (long or short)
  - Organ (heart or brain)
- Variability of the response (NOT wanted) : measurement error
  - Preparation of the enzyme suspension
  - Instrument calibration/standardization
- Variability in experimental units (NOT wanted)
  - biological differences between hamsters
  - environmental differences

Types of variability

- Systematic, expected (wanted)
- Random variation (can manage this)
- Systematic, unexpected (NOT wanted)
  - biased results
  - *e.g.*, what time the measurements are made

Questions for the hibernation study

- *Long or short :* Is there an effect of daylight duration on enzyme concentration ?
- *Heart vs. Brain :* Are the concentrations differents ?
- *Interaction :* Is the *difference* in enzyme concentration (long/short) *different* for heart and brain ?
- *Hamsters :* Variability between hamsters ?
- *Measurement error :* What is the error due to the measurement process for enzyme concentration ?

Completely randomized experiment

- Concentrated on 1 organ (heart, for example)
- Randomization : use a *random mechanism* to assign hamsters to long or short days
- 'Random' $\neq$ 'haphazard' or 'arbitrary'
- *Balanced :* assign the *same* number of hamsters to long and short days
- Example (8 hamsters) :
  Long : 4, 1, 7, 2
  SHort : 3, 8, 5, 6

(Complete) Randomized block design

- Assume that the hamsters have come from 4 *different litters*, 2 hamsters per litter
- We expect that hamsters born in the same litter are *more similar* to each other than hamsters from a different litter
- For each pair of hamsters {colitrandomly assign short or long to one member of each pair
- Example (toss a fair coin, for example) :
  S, L // L, S // S, L // S, L

Factorial experiment

- Compare two (or more) sets of conditions in the *same experiment* : long/ short AND heart/brain
- In this example, there are 4 combinations of conditions :
  - Long/Heart, Long/Brain, Short/Heart, Short/Brain
- Example (2 coin tosses, for example) :
  L/H : 7, 2 L/B : 4, 1
  S/H : 3, 5 S/B : 8, 6

Replication, Randomization, Blocking

- These are the 'big three' of experimental design
- **Replication** –to reduce random variation of the test statistic, increases generalizability
- **Randomisation** – to reduce/remove bias
- **Blocking** – to reduce unwanted variation
- Idea here is that units within a block are similar to each other, but *different between blocks*
- 'Block what you can, randomize what you cannot'

Trees

- A study is carried out to examine the growth of a certain type of tree at an altitude of 675 meters
- The variable of interest is the measure of the base (in cm) during a period of 10 years
- According to a theory, the mean should be at least 1.75
- For a random sample of 10 measures (trees), we have $\overline{x} = 2$ cm, $s_x = 0.5$ cm

Steps in hypothesis testing (I)

1. Identify the population parameter
   Here, the parameter of interest is the population mean of the base measure during 10 years
2. Formulate the NULL and ALT hypotheses
   $H : \mu = 1.75$
   $A : \mu > 1.75$
3. Calculate the TS
   $t_{obs} = (2 - 1.75)/(.5/\sqrt{10}) = 1.58$

Steps in hypothesis testing (II)

4. Calculate the p-value
   $p_{obs} = P(T_9 > 1.58) = .07$
   $(t_{9,0.90} = 1.383 < t_{obs} < t_{9,0.95} = 1.833)$
5. (Optional) *Decision rule* : REJECT the NULL hypothesis H if $p_{obs} \leq \alpha$
   $\alpha = 0.05$, thus DO NOT REJECT H (barely !)

More trees

- Now, suppose that we are interested in knowing whether the mean base measure is the same for trees at 675 meters and 825 meters
- We have another random sample of 10 trees at 825 meters, for which $\overline{y} = 2.65$ cm, $s_y = 1.15$ cm
- How could we test the null hypothesis that the means are equal, against the alternative that they are different ? **...**

Review : Test for comparing 2 (independent) means

Supposing equal variances :

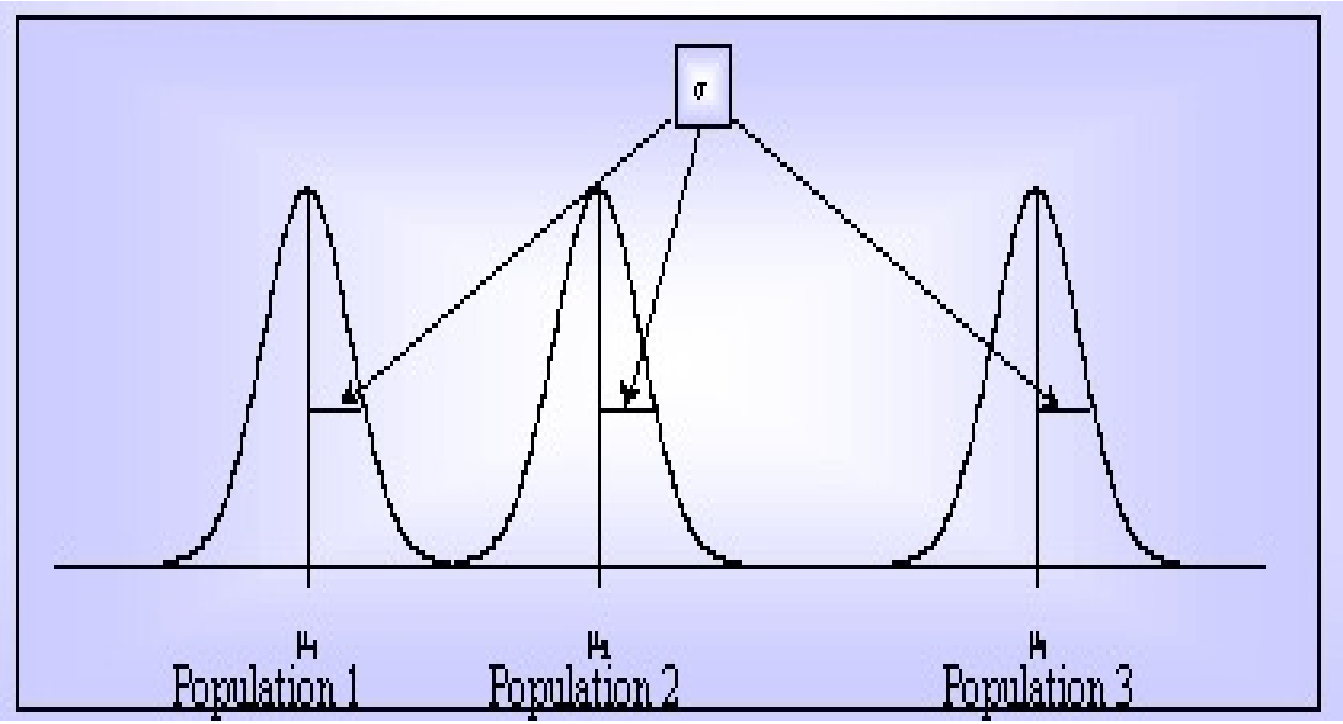
- T = obs. diff / SE(obs. diff) $= \Delta / \sqrt{\widehat{Var}(\hat{\Delta})}$;
  $\hat{\Delta} = \bar{y} - \bar{x}$; $Var(\hat{\Delta}) = \sigma^2/n + \sigma^2/m = \frac{n+m}{nm}\sigma^2$
- We can estimate the variances *separately* :
  $s_x^2 = ((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2)/(n-1)$
  $s_y^2 = ((y_1 - \bar{y})^2 + \cdots + (y_m - \bar{y})^2)/(m-1)$
- When the variances are *equal*, we can combine the two estimators : $s_p^2 = ((n-1)s_x^2 + (m-1)s_y^2)/(n+m-2)$
- $$\Longrightarrow \; t_{obs} = \frac{\bar{y} - \bar{x}}{\sqrt{s_p^2(n+m)/(nm)}} \sim t_{n+m-2} \text{ under } H$$

Trees one more time !

- You guessed it ! ! Now we are also interested in trees at 975 meters
- We want to make a *three-way* comparison
- We have another random sample ($n = 10$), with $\overline{z} = 2.5$ cm, $s_z = 1$ cm
- How could we test the null hypothesis that the *three* means are equal, against the alternative that they are not equal ? (Suppose that we have not done the two-way test) **...**

ANOVA

- Abbreviation for *AN*alysis *O*f *VA*riance
- BUT : it is a test of difference between *means*
- The idea :



Principle

- The total variation (sum of squares of differences from the mean) has 2 components :
  - *individual* fluctuations : *within-groupe* variability (due to error)
  - Fluctuations *between the groups* : *inter-groupe* variability (due to the treatment)
- Within-group variability $>$ Between-group variability
  $\Longrightarrow$ at least 2 means are different
- General principle :
  - Decompose the total sum of squared deviations into 2 parts
  - Test whether the mean square *between* is (significantly) bigger than the mean square *within*

Hypothesis test

- Notation :
  - k groups
  - n_i is the sample size for group i
  - observations x_{ij} (observation j of group i)
- $H : \mu_1 = \mu_2 = \cdots = \mu_k$
  $A : \exists\, \mu_i \neq \mu_j$ (at least one mean is different from the others)
- ANOVA is a rather *robust* test (results not greatly affected by mild deviation from assumptions)

The models

- $\epsilon_{ij} \sim$ iid $N(0, \sigma^2)$
- Under H, the model is :

$$x_{ij} = \mu + \epsilon_{ij}$$

- Under A, the model is :

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where α_i est *the effect of level i of factor A on the variable X*

  - Model is *overparameterized* :

Uses more parameters than necessary

- Need to introduce a constraint so that the model is *identifiable*, such as $\sum_i \alpha_i = 0$ (sum to zero contrasts) or $\alpha_1 = 0$ (treatment contrasts)

Sum of Squares

- Goal : test difference between means of two (or more) groups
  - Between SS measures the difference
- The difference must be measured relative to the variance within the groups
  - Within SS
- *F-test* : considers the ratio of B/W
- The larger F is, the more significant the difference

The ANOVA procedure

- Subdivide observed total sum of squares into several components
- Pick appropriate significance point for a chosen Type I error α from an F table
- Compare the observed components to test the NULL hypothesis

PAUSE

Parameter estimation

- Under $H : x_{ij} = \mu + \epsilon_{ij}$:

$$\hat{\mu} = \overline{x} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}, \quad n = \sum_{i=1}^{k} n_i$$

- Under $A : x_{ij} = \mu + \alpha_i + \epsilon_{ij}$:

$$\hat{\mu} + \hat{\alpha}_i = \overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

which gives us $\hat{\alpha} = \overline{x}_i - \overline{x}$
$\hat{\epsilon}_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \hat{\mu} - \hat{\alpha}_i = x_{ij} - \overline{x} - (\overline{x}_i - \overline{x}) = x_{ij} - \overline{x}_i$

Decomposition of the total variation

- The model under A : $x_{ij} = \mu + \alpha_i + \epsilon_{ij}$
- with estimators : $x_{ij} = \overline{x} + (\overline{x}_i - \overline{x}) + (x_{ij} - \overline{x}_i)$
- $\Longrightarrow \ (x_{ij} - \overline{x}) = (\overline{x}_i - \overline{x}) + (x_{ij} - \overline{x}_i)$
- with sum of squares :
  $(x_{ij} - \overline{x})^2 = (\overline{x}_i - \overline{x})^2 + (x_{ij} - \overline{x}_i)^2 + 2(\overline{x}_i - \overline{x})(x_{ij} - \overline{x}_i)$
- and sums for individuals (j) :
  $\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2 =$
  $n_i(\overline{x}_i - \overline{x})^2 + \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2 + 2(\overline{x}_i - \overline{x})\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)$

Décomposition, cont.

- Thus, $2(\overline{x}_i - \overline{x}) \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)$, since $\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i) = 0$ $(E(\epsilon_{ij} = 0))$
- Therefore,

$$\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2 = n_i(\overline{x}_i - \overline{x})^2 + \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2$$

- with the sums for the factor levels :

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2 = \sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2$$

- $\Longrightarrow \quad SSE_{total} \quad = \quad SSE_{treatments} \quad + \quad SSE_{error}$

Principle of the test

- 1-way (1 factor) analysis of variance tests the effect of the factor A having k levels on the means of a quantitative (continuous) variable X
- The tested hypotheses :
  $H : \mu_1 = \mu_2 = \cdots \ \mu_k = \mu$ vs. $A : \exists\, \mu_i \neq \mu_j$
- Test whether the ratio of the 2 estimators for the variance is close to 1
- The associated variance estimators (or *mean squares*) sont :
  - Total variance : $SS_{total}/(n-1)$
  - Variance due to factor A (MS_{trts}) : $SS_{trts}/(k-1)$
    $\Longrightarrow$ estimator of σ^2 if H is true
  - Residual variance (MS_{error}) : $SS_{errorr}/(n-k)$
    $\Longrightarrow$ estimator of σ^2 for either model

Test statistic

- Under H, $SS_{trts}/(k-1)$ and $SS_{error}/(n-k)$
  $\Longrightarrow$ *estimators of the same parameter* σ^2
- Thus, (under H), the ratio $\frac{SS_{trts}/(k-1)}{SS_{error}/(n-k)} \approx 1$
- Under A, at least one $\alpha_i \neq 0$ and $SS_{error}/(n-k)$ is a unique estimator of σ^2 ; $SS_{trts}/(k-1) >> SS_{error}/(n-k)$
- Thus, (under A), the ratio $\frac{SS_{trts}/(k-1)}{SS_{error}/(n-k)}$ *much bigger than 1*
- $\Longrightarrow$ *1-sided test*
- $F_{obs} = \frac{SS_{trts}/(k-1)}{SS_{error}/(n-k)} = MS_{trts}/MS_{error}$
- Null distribution of this test statistic is the Fisher distribution with $k-1$ (num) and $n-k$ (denom) degrees of freedom (df)

ANOVA table

Tableau d'ANOVA

source	df	SS	MS $(=SS/\text{df})$	F	p-value
treatments	$k-1$	SS_{trts}	$SS_{trts}/(k-1)$	MS_{trts}/MS_{error}	$P(F_{obs} > F_{k-1,n-k})$
error	$n-k$	SS_{error}	$SS_{error}/(n-k)(=\hat{\sigma}^2)$		
total (corr.)	$n-1$	SS_{total}			

What does it mean when we reject H ?

- The null hypothesis H is a joint one : that *all* population means are equal
- When we reject the null hypothesis, it does NOT mean that all means are different !
- It means that *at least one* mean is different
- In order to find out which is/are different, we can carry out 'post-hoc'/*a posteriori* tests (pairs of t-tests, Tukey's Honest Significant Difference, *etc.*)

Model formulas in R

- A simple *model formula* in R looks something like: **`yvar ~ xvar1 + xvar2 + xvar3`**
- We could write this model (algebraically) as

$$Y = a + b_1 * x_1 + b_2 * x_2 + b_3 * x_3$$

- By default, an intercept is included in the model - you don't have to include a term in the model formula
- If you want to leave the intercept out:

```
yvar ~ -1 + xvar1 + xvar2 + xvar3
```

More on model formulas

- We can also include *interaction terms* in a model formula:

```
yvar ~ xvar1 + xvar2 + xvar3
```

Examples

- `yvar ~ xvar1 + xvar2 + xvar3 + xvar1:xvar2`
- `yvar ~ (xvar1 + xvar2 + xvar3)^2`
- `yvar ~ (xvar1 * xvar2 * xvar3)`

More on model formulas

- The generic form is **`response ~ predictors`**
- The predictors can be **`numeric`** or **`factor`**
- Other symbols to create formulas with *combinations of variables* (e.g. *interaction*s)
  - **`+`** to *add* more variables
  - **`-`** to *leave out* variables
  - **`:`** to introduce *interactions* between two terms
  - **`*`** to include *both interactions and the terms*

    (**`a*b`** is the same as **`a+b+a:b`**)
  - **`^n`** *adds all terms* including interactions up to order n
  - **`I()`** treats what's in () as a *mathematical expression*

Interpreting R output

```
> chicks.aov <- aov(Weight ~ House + Protein*LP*LS)
> summary(chicks.aov)
              Df  Sum Sq Mean Sq F value    Pr(>F)
House          1  708297  708297 15.8153 0.0021705 **
Protein        1  373751  373751  8.3454 0.0147366 *
LP             2  636283  318141  7.1037 0.0104535 *
LS             1 1421553 1421553 31.7414 0.0001524 ***
Protein:LP     2  858158  429079  9.5808 0.0038964 **
Protein:LS     1    7176    7176  0.1602 0.6966078
LP:LS          2  308888  154444  3.4485 0.0687641 .
Protein:LP:LS  2   50128   25064  0.5596 0.5868633
Residuals     11  492640   44785
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

Numerical and graphical analysis

- Tables of group means:

		Groundnut	Soybean	Mean
Level of	0	6676	7452	7064
protein	1	6893	6961	7927
	2	6719	6624	6671
Mean		6763	7012	6887

		G-nut	Soy	Level of protein			Mean
				0	1	2	
Level of	0	6537	6752	6750	6595	6588	6644
fish	1	6989	7273	7379	7259	6755	7131
Mean		6763	7012	7064	6927	6671	6887

Pairs of tests : why not ?

Why do ANOVA instead of performing tests (z or t) for each pair of samples ?

- For m (independent) comparisons, the probability of rejecting at least one H is given by : $\alpha_m = 1 - (1 - \alpha)^m$
- For $\alpha = 0.05$, we have :
  - 3 tests $\Longrightarrow$ P(Type I error) $= 0.14$
  - 5 tests $\Longrightarrow$ P(Type I error) $= 0.23$
  - 10 tests $\Longrightarrow$ P(Type I error) $= 0.4$
  - 21 tests $\Longrightarrow$ P(Type I error) $= 0.66$

Assumptions

- *Independance* : The k samples are independent ; the set of n is allocated *randomly* entre les k levels of the factor A, with n_i individuals receiving treatment i.
- *Homoscedasticity* : The k populations have the same variance ; the factor A acts only on the *mean* of the variable X and doesn't change its variance
- *Normality* : The quantitative variable X is normally distributed in the k populations (or CLT applies for n_i 'sufficiently large')

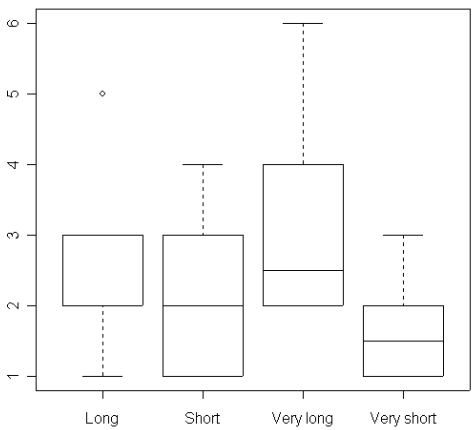
Model assessment : Normality

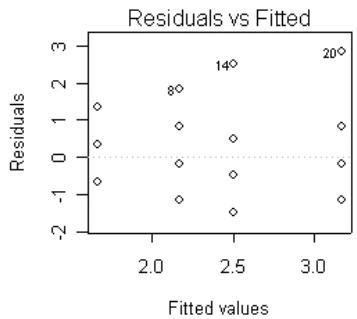
- Boxplots of the observations (or residuals) should be symmetric
- Graph of sample means vs. sample variances should not show any pattern
- QQ-plot (normal) of the observations (or residuals) should fall on a straight line
- Look to see whether there are any outliers or unusual observations

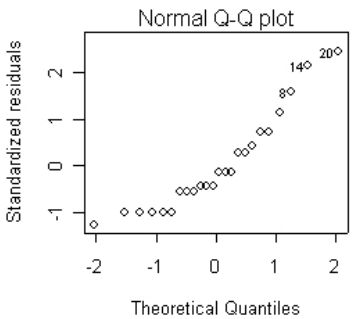
Model assessment : Homogeneity of variance

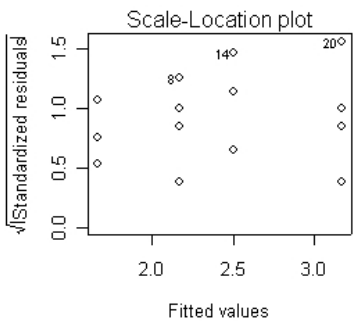
- Boxplots of the observations should have a similar variability
- Variability of the residuals should be similar in the graph of residuals vs. group means
- There are also formal hypothesis tests (*e.g.*, Bartlett, Levene), but they are not very useful as diagnostics

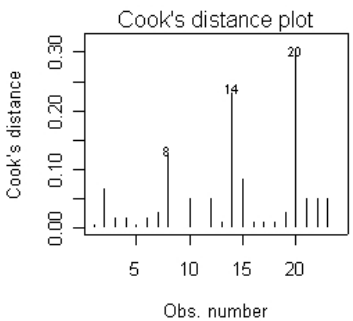
Some diagnostic graphs











Model assessment : Independence

- Graph : residuals vs. group means, might be able to indicate autocorrelation (for example)
- Typically this issue is treated during experiment planning, by randomization and/or other methods

Summary : numerical and graphical analysis

- Design plot
- Boxplots of outcome for each factor
- Interaction plots
- Write out model, assumptions, define all parameters
- anova table
- Plots for assumption checking/model assessment

R output – ANOVA

```
> redcell.aov<-aov(Folate~Group)
> summary(redcell.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
Group        2  15516    7758  3.7113 0.04359 *
Residuals   19  39716    2090
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

ANOVA : after the test

- Once the ANOVA conditions have been verified and the analysis carried out, two conclusions are possible :
  - REJECT H
  - there is not enough evidence to reject H (DO NOT REJECT)
- If we do not reject H, we conclude that there are no significant differences in the means of the groups
- If we do reject H, we want to *identify the levels of the factor* that lead to the significant results

Multiple comparisons

- Comparison of means for pairs of treatments
- These are made *after* a significant ANOVA
- Types of comparisons
  - planned (*a priori*) : independent of the ANOVA results ; the theory predicts which treatments should be different
  - unplanned (*a posteriori*) : comparisons are determined *based on the ANOVA results*
- $H : \mu_i = \mu_j$ vs. $A : \mu_i \neq \mu_j$
- Test statistic

$$t = \frac{\overline{y}_i - \overline{y}_j}{\sqrt{\hat{\sigma}^2 \left(1/n_i + 1/n_j\right)}}$$

- $(\hat{\sigma}^2 = MS_{error})$; df $= df_{error}$

Method of Bonferroni – global control

- For k comparisons, the probability of not rejecting a true null $H = (1 - \alpha)^k$
- $\Longrightarrow$ we must control the (total) error α *while adjusting for the number comparisons*
- To maintain the global α_e at level α, we must *adjust α for each comparison by the number of comparisons*
- In this way, α_e becomes independent of the number of comparisons
- The most simple : Bonferroni method

$$\alpha' = \alpha / k,$$

where k = number of comparisons (tests)

- $p_{adjusted} = \min(kp, 1)$
- The method of Bonferronni assures that *the global level is at most the desired level*

Multiple comparisons : Tukey Honest Significant Differences

- Interested in simultaneous confidence intervals or tests for differences in the mean outcome X for pairs of levels of a factor
- To test all pairwise comparisons among means using the Tukey HSD, calculate HSD for each pair of means :

$$q_s = \frac{M_i - M_j}{\sqrt{MSW / n_{group}}},$$

where M_k is the mean of group k, $M_i > M_j$

- For hypothesis testing, the value q_s is compared to a q value from the *studentized range distribution* (difference between largest and smallest sample means divided by pooled sample SD$\sqrt{2/n}$)
- Reject the null at level α if $q_s > q_\alpha$
- CI : $(\overline{y}_i - \overline{y}_j) \pm \frac{q_{\alpha;k;N-k}}{\sqrt{2}} \hat{\sigma}_e \sqrt{\frac{2}{n}}$; $i, j = 1, \ldots, k$, $i \neq j$
- k = number of populations ; N = total sample size

Factorial crossing

- Compare 2 (or more) sets of conditions in the *same experiment*
- Designs with factorial treatment structure allow you to measure *interaction* between two (or more) sets of conditions that influence the response
- Factorial designs may be either observational or experimental

Interaction

- Interaction is very common (and very important) in science
- Interaction is a *difference of differences*
- Interaction is present if the effect of one factor *is different* for different levels of the other factor
- *Main effects can be difficult to interpret in the presence of interaction*, because the effect of one factor *depends* on the level of the other factor

Two-way ANOVA

- Simultaneously study factor A with I levels and factor B with J levels
- For each pair of levels (A, B) :
  - there is a sample
  - all samples are of the *same size* n (balanced design)
- Assumptions :
  - measures in each population are normally distributed
  - variances in each population are equal (homoscedasticity)
  - samples are obtained independently at random from the populations

Full model

- *Full model* : with interactions
- $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$
- $E[\epsilon_{ijk}] = 0$, $Var(\epsilon_{ijk}) = \sigma^2$, $Cov(\epsilon_{ijk}, \epsilon_{i'j'k'}) = 0$ si $(ijk) \neq (i'j'k')$

ANOVA table

source	df	SS	MS	F
A	$I-1$	$nJ\sum_{i=1}^{I}(\overline{y}_{i\cdot\cdot} - \overline{y}_{\cdots})^2$	SS_A/df_A	MS_A/MS_{err}
B	$J-1$	$nI\sum_{j=1}^{J}(\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdots})^2$	SS_B/df_B	MS_B/MS_{err}
AB	$(I-1)(J-1)$	$n\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ij\cdot} - \overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot j\cdot} + \overline{y}_{\cdots})^2$	SS_{AB}/df_{AB}	MS_{AB}/MS_{err}
error	$IJ(n-1)$	$\sum_{k=1}^{n}\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ijk} - \overline{y}_{ij\cdot})^2$	SS_{err}/df_{err}	
total (corr.)	$nIJ-1$	$\sum_{k=1}^{n}\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ijk} - \overline{y}_{\cdots})^2$		

Two-way ANOVA

- Associated with each sum of squares (SS)
  - Corresponding degrees of freedom (df)
  - Corresponding mean square (MS) = SS/df
- The mean squares are compared using F ratios to test various effects
  - First – test for a significant *interaction* between the factors
  - If there is an interaction, it may not be reasonable to test for significant risk or age differences

Hypothesis tests

- Test for interaction
  $H : \gamma_{ij} = 0,\ i = 1, \ldots, I,\ j = 1, \ldots, J$
- Test statistic :
  $F_{AB} = MS_{AB}/MS_{error} \sim F_{(I-1)(J-1),IJ(n-1)}$ under H
- Test for effect of factor A
  $H : \alpha_i = 0,\ i = 1, \ldots, I$
- Test statistic :
  $F_A = MS_A/MS_{error} \sim F_{I-1,IJ(n-1)}$ under H
- Test for effect of factor B
  $H : \beta_j = 0,\ j = 1, \ldots, J$
- Test statistic :
  $F_B = MS_B/MS_{error} \sim F_{J-1,IJ(n-1)}$ under H

Additive model

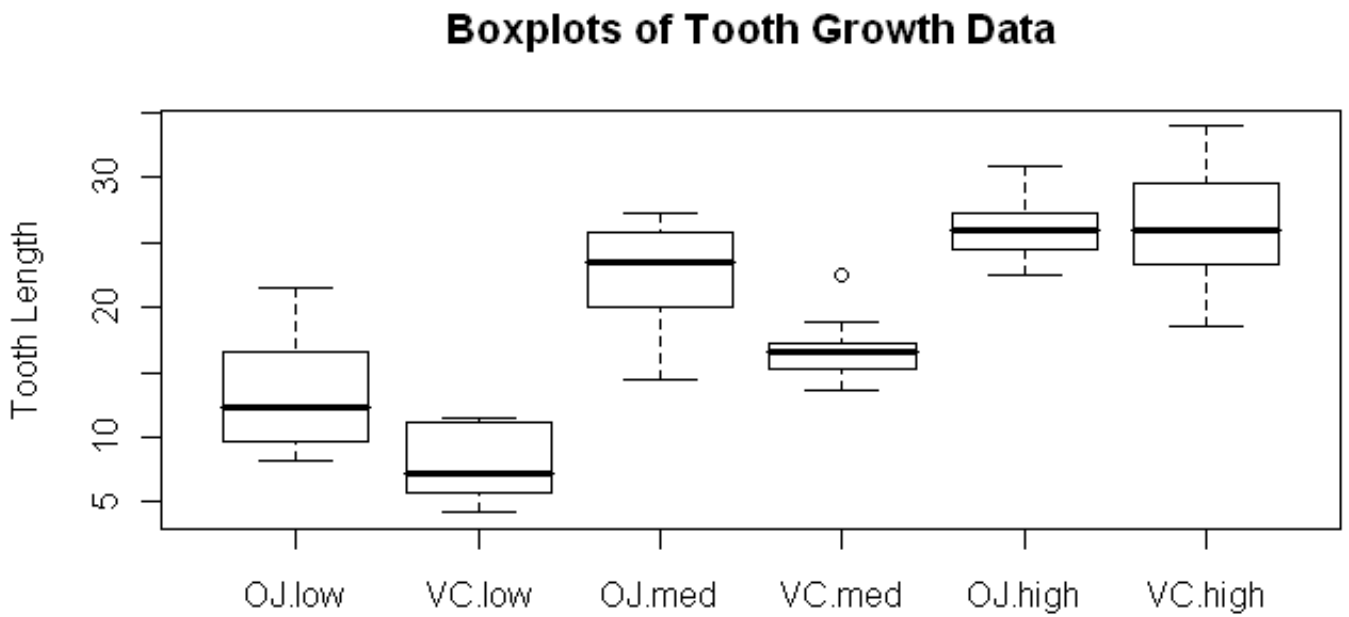
- *Additive model* : without interactions
- $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$
- $E[\epsilon_{ijk}] = 0$, $Var(\epsilon_{ijk}) = \sigma^2$, $Cov(\epsilon_{ijk}, \epsilon_{i'j'k'}) = 0$ if $(ijk) \neq (i'j'k')$

ANOVA table

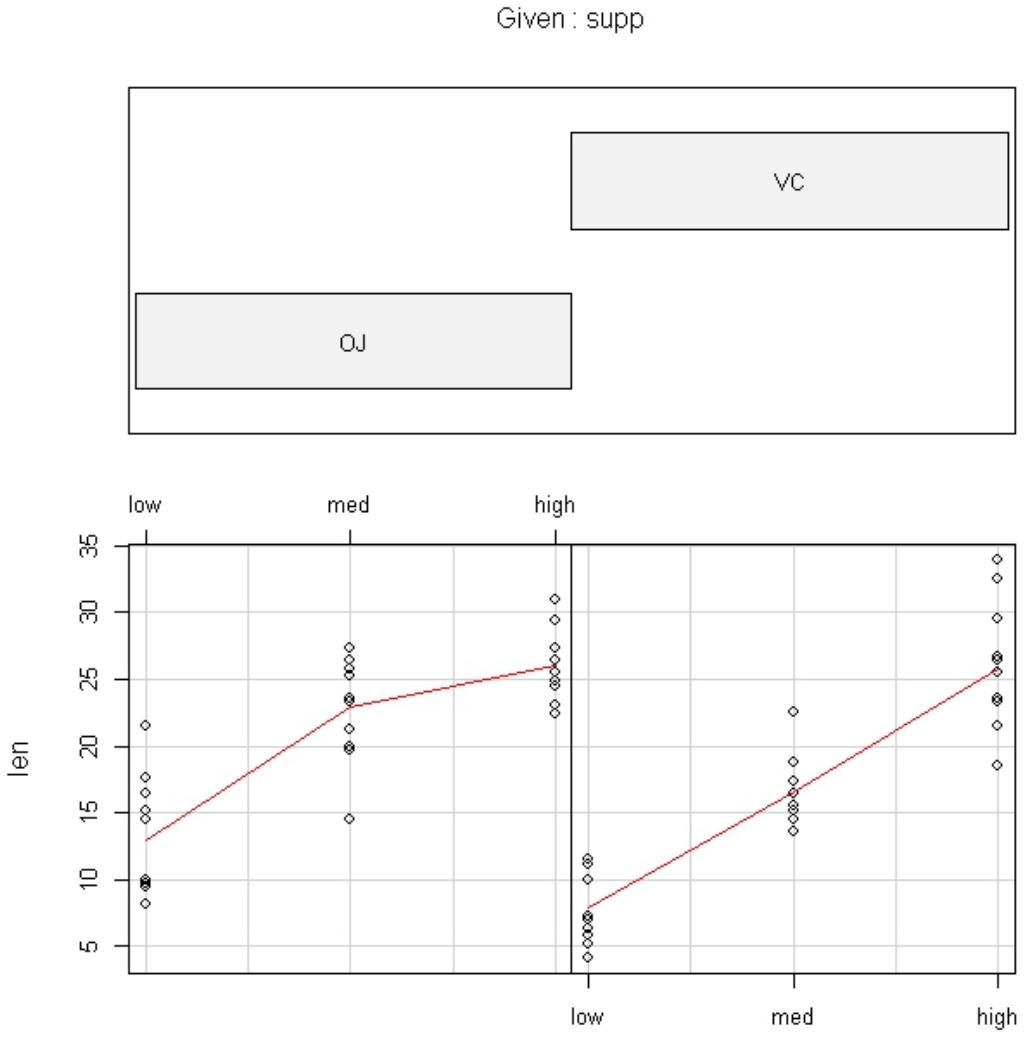
source	df	*SS*	*MS*	*F*
A	$I-1$	$nJ\sum_{i=1}^{I}(\overline{y}_{i\cdot\cdot} - \overline{y}_{\cdots})^2$	SS_A/df_A	MS_A/MS_{err}
B	$J-1$	$nI\sum_{j=1}^{J}(\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdots})^2$	SS_B/df_B	MS_B/MS_{err}
error	$nIJ-I-J+1$	$\sum_{k=1}^{n}\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ijk} - \overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot j\cdot} + \overline{y}_{\cdots})^2$	SS_{err}/df_{err}	
total (corr.)	$nIJ-1$	$\sum_{k=1}^{n}\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ijk} - \overline{y}_{\cdots})^2$		

Example : ToothGrowth

- "The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)."

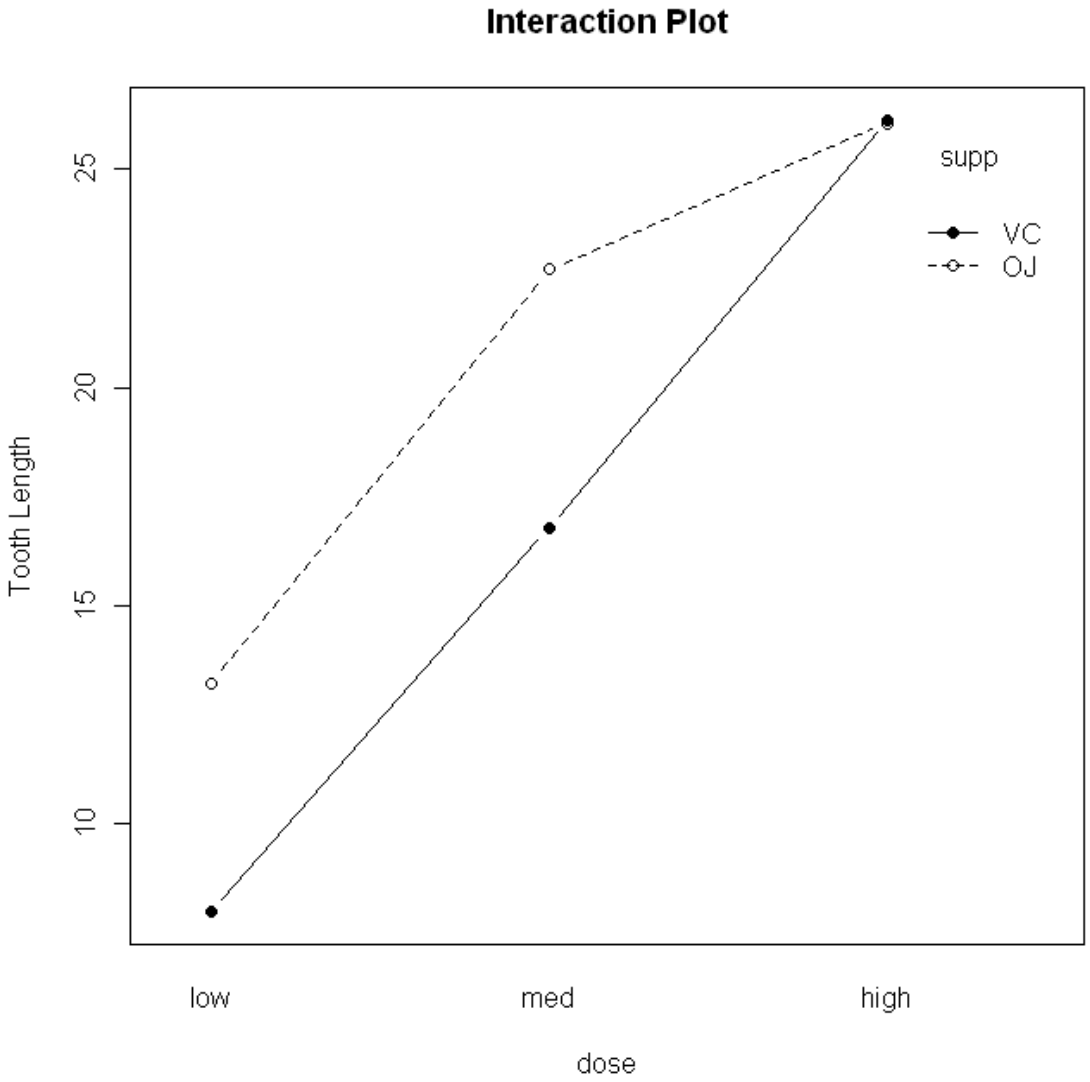


Example : ToothGrowth, cont.



ToothGrowth data: length vs dose, given type of supplement

Example : ToothGrowth, cont.



Example : ToothGrowth, cont.

```
> aov.out = aov(len ~ supp * dose, data=ToothGrowth)
> summary(aov.out)

             Df Sum Sq Mean Sq F value   Pr(>F)
supp          1  205.3   205.3  15.572 0.000231 ***
dose          2 2426.4  1213.2  92.000  < 2e-16 ***
supp:dose     2  108.3    54.2   4.107 0.021860 *
Residuals    54  712.1    13.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
’ 1
```

Unbalanced designs

- When all sample sizes are equal, the main effects and interactions can be estimated *independently*
- That's because of the orthogonality of the sub-spaces that correspond to the different model effects
- This is no longer the case when the sample sizes are different (unbalance case)
- For an unbalanced design, effect estimation must be *adjusted* (for the other effects in the model) : the estimated values depend on the other terms in the model and their order of entry

Example : ToothGrowth (unbalanced)

	L	M	H
VC	4.2 11.5 7.3	16.5 16.5 15.2 17.3	23.6 18.5
OJ	15.2 21.5 17.6 9.7	19.7 23.3	25.5 26.4 22.4 24.5

Example : supp

```
> # full interaction model with
> # supp entering first
>
> fit1 <-
    lm(len ~ supp + doselev + supp:doselev,
      data=toothun)
> anova(fit1)
Analysis of Variance Table

Response: len
              Df Sum Sq Mean Sq F value    Pr(>F)
supp           1 174.46  174.46 17.3664 0.0011049
doselev        2 375.75  187.87 18.7012 0.0001495
supp:doselev   2  17.70    8.85  0.8808 0.4377931
Residuals     13 130.60   10.05
```

Example : doselev

```
> # full interaction model with doselev
> # entering first
>
> fit2 <-
    lm(len ~ doselev + supp + supp:doselev,
      data=toothun)
> anova(fit2)
Analysis of Variance Table

Response: len
              Df Sum Sq Mean Sq F value    Pr(>F)
doselev        2 396.08  198.04 19.7131 0.0001158
supp           1 154.13  154.13 15.3428 0.0017685
doselev:supp   2  17.70    8.85  0.8808 0.4377931
Residuals     13 130.60   10.05
```