Statistiek II

John Nerbonne

Dept of Information Science
j.nerbonne@rug.nl
based also on H.Fitz's work

March 10, 2010

university of
groningen

Last week: one-way ANOVA

- generalized *t*-test to compare means of more than two groups
- example:
  - (a) compare frequencies of stylistic elements in three book reviews
  - (b) compare Dutch proficiency test results of four groups of foreigners
- assumptions of
  - (i) normality
  - (ii) and similar standard deviations in each group
  - (iii) independent samples
- partitioning of total variance (SST) into between-groups variance (SSG) and error variance (SSE): SST = SSG + SSE
- based on F-distribution: $F = \frac{\text{MSG}}{\text{MSE}}$

Today: **factorial** ANOVA

Factorial ANOVA

Like one-way ANOVA, but more than one factor (aka *n*-way ANOVA)

- compares means of different groups
- based on F-distribution:

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

    - always positive
    - two kinds of degrees of freedom: df_{s_1}, df_{s_2}
    - value of 1 indicates same variance, values near 0 or $+\infty$ indicate difference
- uses F-distribution: compare variances among means with random variability inside the groups
- one-way ANOVA, *n*-way ANOVA $\neq$ F-test!
- assumes near-normal distribution in all groups
- standard deviations in all groups roughly equal ($\frac{\mathrm{sd}_i}{\mathrm{sd}_j} \leq 2$)

Factorial ANOVA

Why two-way ANOVA?—why not just two one-way ANOVAs?

- efficient in the number of experiments and subjects needed

Factorial ANOVA

Why two-way ANOVA?—why not just two one-way ANOVAs?

- efficient in the number of experiments and subjects needed

Suppose we want to measure effect of calcium and magnesium intake on blood pressure

Two-way ANOVA:

	Calcium		
Magnesium	L	M	H
L	1	2	3
M	4	5	6
H	7	8	9

- two-way design results in 9 groups
- assign 9 subjects to each group
- hence 81 subjects required

Factorial ANOVA

Why two-way ANOVA?—why not just two one-way ANOVAs?

- efficient in the number of experiments and subjects needed

Suppose we want to measure effect of calcium and magnesium intake on blood pressure

Two one-way ANOVAs:

	Calcium		
Magnesium	L	M	H
M	1	2	3

	Magnesium		
Calcium	L	M	H
M	1	2	3

- two one-way designs result in 6 groups
- assign 15 subjects to each group
- hence <u>90</u> subjects required
- and: only 15 subjects per level compared with 27 in two-way design

Factorial ANOVA

Why two-way ANOVA?—why not just two one-way ANOVAs?

- efficient in the number of experiments and subjects needed
- combining two experiments into one improves accuracy:
  - increases number of data points per level
  - decreases SE (standard error of the mean):

    standard deviation of sample mean: $\frac{\sigma}{\sqrt{n}}$
    in one-way ANOVA: $\frac{\sigma}{\sqrt{15}} = 0.26\sigma$
    in two-way ANOVA: $\frac{\sigma}{\sqrt{27}} = 0.19\sigma$

Hence, sample mean responses are less variable in two-way design

Factorial ANOVA

Why two-way ANOVA?—why not just two one-way ANOVAs?

- efficient in the number of experiments and subjects needed
- combining two experiments into one improves accuracy (increases n, decreases SE)
- opportunity to study **interaction**:

  E.g., age and subtype of cancer have independent effects on mortality:

  - breast cancer more treatable than other forms of cancer
  - in general, cancer more treatable with young age

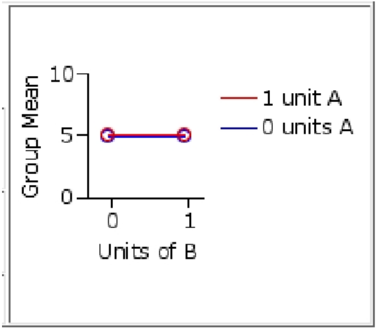
  but these are **reversed** in some combinations, e.g., breast cancer in young women particularly aggressive and dangerous.

Interaction requires care!

Types of interaction

Two drugs A and B administered in doses 0 and 1.

Dependent measure: blood level of some hormone.

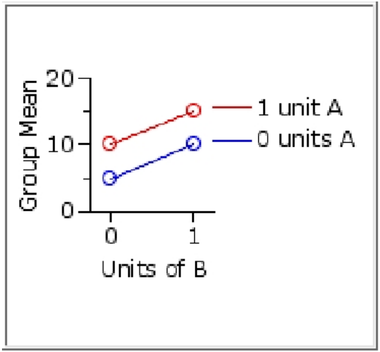


- drugs show no effect
- either separately or in combination
- no interaction
- null hypothesis true

Types of interaction

Two drugs A and B administered in doses 0 and 1.

Dependent measure: blood level of some hormone.

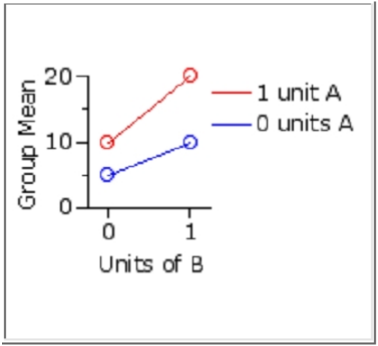


- both drugs have an effect
- combined effect is additive
- no interaction

Types of interaction

Two drugs A and B administered in doses 0 and 1.

Dependent measure: blood level of some hormone.

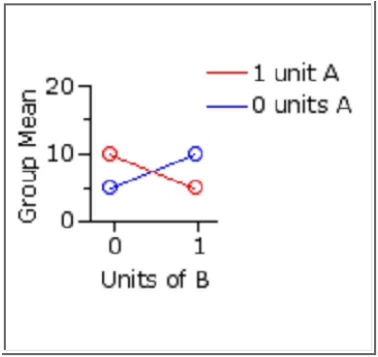


- both drugs have an effect
- combined effect is stronger than the sum of separate effects
- interaction

Types of interaction

Two drugs A and B administered in doses 0 and 1.

Dependent measure: blood level of some hormone.



- both drugs have the same effect as previously!
- when combined, the two drugs cancel each other out
- interaction

Factorial ANOVA: partitioning the variance

As in one-way ANOVA:

SST	=	SSG	+	SSE
Total **S**um of **S**quares		**G**roup **S**um of **S**quares		**E**rror **S**um of **S**quares

SSG: aggregate measure of differences between groups
SSE: aggregate measure of random variability inside groups

But: in *n*-way ANOVA **several** factors contribute to between-groups variance (SSG)

To measure the effect of different factors, we partition the SSG into **components** which correspond to these factors

Factorial ANOVA: partitioning the variance

For example: two factors A & B, then SSG partitions into:

$$\underset{}{\text{SSG}} = \underset{\substack{\text{main effect}\\\text{factor A}}}{\text{SS}_A} + \underset{\substack{\text{main effect}\\\text{factor B}}}{\text{SS}_B} + \underset{\substack{\text{interaction}\\\text{effect}}}{\text{SS}_{A\times B}}$$

In one-way ANOVA: $$\text{SSG} = \sum_{i=1}^{I} N_i(\overline{x}_i - \overline{x})^2$$

In factorial ANOVA: $\text{SS}_A = \sum_{i=1}^{I_A} N_i(\overline{x}_i - \overline{x})^2$ where I_A is the number of levels in factor A.

Note: three factors—A, B, C—induce four interaction sum of squares: $\text{SS}_{A\times B}$, $\text{SS}_{A\times C}$, $\text{SS}_{B\times C}$, $\text{SS}_{A\times B\times C}$

Factorial ANOVA: degrees of freedom

Degrees of freedom are partitioned similarly:

$$\mathrm{DFT} = \underbrace{(\mathrm{DF}_A + \mathrm{DF}_B + \mathrm{DF}_{A\times B})}_{\mathrm{DFG}} + \mathrm{DFE}$$

In one-way ANOVA: $\mathrm{DFG} = I - 1$ (I = number of groups)

In two-way ANOVA:

$$\mathrm{DFG} = \underbrace{(I_A - 1)}_{\mathrm{DF}_A} + \underbrace{(I_B - 1)}_{\mathrm{DF}_B} + \underbrace{(I_A - 1)\cdot(I_B - 1)}_{\mathrm{DF}_{A\times B}} = I_A I_B - 1 = \underline{I - 1}$$

Factorial ANOVA: mean squares

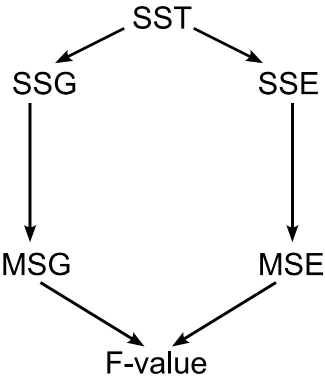
In factorial ANOVA we obtain several mean squares between groups (here 3):

$$\begin{aligned} \mathrm{MS}_A &= \frac{\mathrm{SS}_A}{\mathrm{DF}_A} && \text{(factor A)} \\ \mathrm{MS}_B &= \frac{\mathrm{SS}_B}{\mathrm{DF}_B} && \text{(factor B)} \\ \mathrm{MS}_{A\times B} &= \frac{\mathrm{SS}_{A\times B}}{\mathrm{DF}_{A\times B}} && \text{(interaction)} \end{aligned}$$

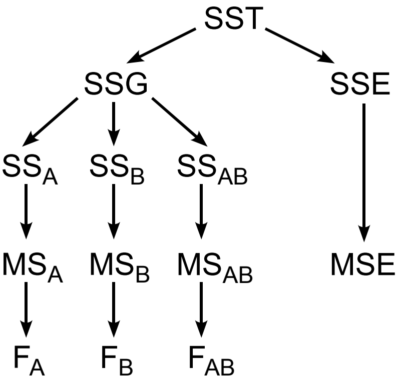
Hence, there are also **three** F-values—F_A, F_B, and $\mathrm{F}_{A\times B}$—for which we test significance!

Factorial ANOVA schematically

One-way ANOVA:

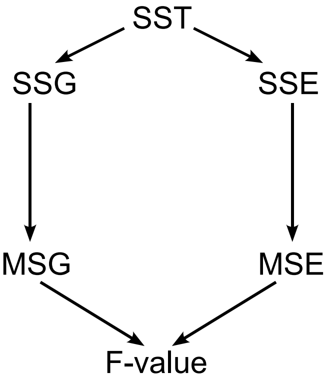


Two-way ANOVA:

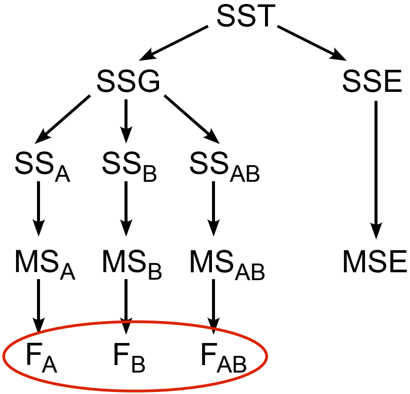


Factorial ANOVA schematically

One-way ANOVA:



Two-way ANOVA:



Factorial ANOVA example

Many studies with children and adults (across languages) show:

Object-relative clauses are more difficult to produce/comprehend than **subject**-relative clauses.

For example:

Obj-Rel: There is the man [that the dog bit __ at the park yesterday].
Subj-Rel: There is the boy [that __ hit the cricket ball over the fence].

Kidd, Brandt, Lieven & Tomasello (*Language and Cognitive Processes*, 22(6), 2007) investigated what makes object-relative clauses **easier** to process.

Factorial ANOVA example

Task: 3–4 year-old children had to repeat sentences with relative clauses from an experimenter ('parrot game')

Two kinds of **lexical** manipulations:

- Pronominal subjects versus full NPs:

  This is the boy that you saw at the shop on Saturday.
  This is the boy that the man saw at the shop on Saturday.

- Animate versus inanimate head nouns:

  This is the football that he kicked in the garden yesterday.
  This is the dog that he kicked in the garden yesterday.

Animacy, pronouns and repetition accuracy

Design: Four kinds of sentences shown:

	Head noun	
RC-subject	Animate	Inanimate
pronominal	pronoun + anim. head	pronoun + inanim. head
full NP	NP + animate head	NP + inanimate head

Extras: KBLT controlled test sentences for length in words and syllables. Each child saw four different items of each type.

Measure: exact repetitions were scored as 1
minor modifications (e.g., tense/aspect) as 0.5
ungrammatical or different syntax as 0

KBLT data and design

Label	Head noun	RC-subject	Score
ANP	animate	NP	0.23
ANP	animate	NP	0.19
ANP	animate	NP	0.14
ANP	animate	NP	0.14
INP	inanimate	NP	0.25
⋮	⋮	⋮	⋮
APro	animate	pronoun	0.63
⋮	⋮	⋮	⋮
IPro	inanimate	pronoun	0.58

There are four sentence types, and four different tokens per type.

Examples:		
	ANP:	*...the dog that the man kicked...*
	INP:	*...the toy that the man kicked...*
	APro:	*...the dog that he kicked...*
	IPro:	*...the toy that he kicked...*

KBLT data and design

Label	Head noun	RC-subject	Score
ANP	animate	NP	0.23
ANP	animate	NP	0.19
ANP	animate	NP	0.14
ANP	animate	NP	0.14
INP	inanimate	NP	0.25
⋮	⋮	⋮	⋮
APro	animate	pronoun	0.63
⋮	⋮	⋮	⋮
IPro	inanimate	pronoun	0.58

Two-way ANOVA “by item” with head noun animacy and RC-subject type as **factors**.

Dependent variable: score, represents average repetition accuracy of 48 kids (3–4 years of age).

Data: means and SDs of four groups

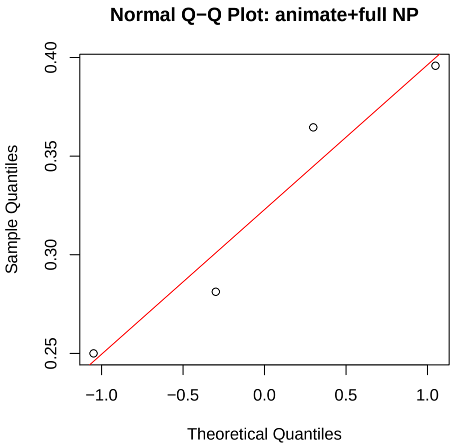
Dependent Variable:score

Animacy	Subject	Mean	Std. Deviation	N
animate	NP	.172	.045	4
	pro	.633	.026	4
	Total	.402	.249	8
inanimate	NP	.323	.069	4
	pro	.625	.091	4
	Total	.474	.178	8
Total	NP	.247	.097	8
	pro	.629	.062	8
	Total	.438	.212	16

Note: SDs not approximately equal (because data is streamlined):
$2\times(\text{animate}+\text{NP}) \leq (\text{inanimate}+\text{pro})$

Factorial ANOVA question: are means significantly different?

Check normality of data



Check normality assumption for all groups!

Multiple questions

Two-way ANOVA asks **two/three** questions simultaneously:

1. Is head noun animacy affecting repetition accuracy?
2. Does lexical type of subject NP affect repetition accuracy?
3. Do the two effects interact, or are they independent?

Questions 1 & 2 might have been asked in separate one-way ANOVA designs (but these would have been more costly in number of subjects)

Question 3 is new to two-way ANOVA

Multiple null hypotheses in *n*-way ANOVA

In our example: each of the two factors has two levels

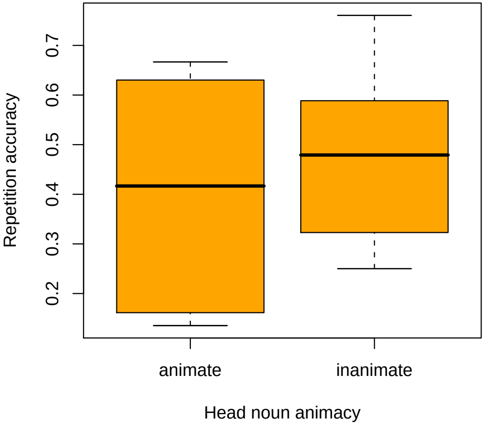
Factor A:	animacy of the head noun
Levels in A:	animate or inanimate
Factor B:	lexical type of relative clause subject
Levels in B:	pronoun or full NP

Three null hypotheses:

1. There is no difference in the means of factor *animacy*
2. There is no difference in the means of factor *subject type*
3. There is no interaction between factors *animacy* and *subject type*

Visualizing factorial ANOVA questions

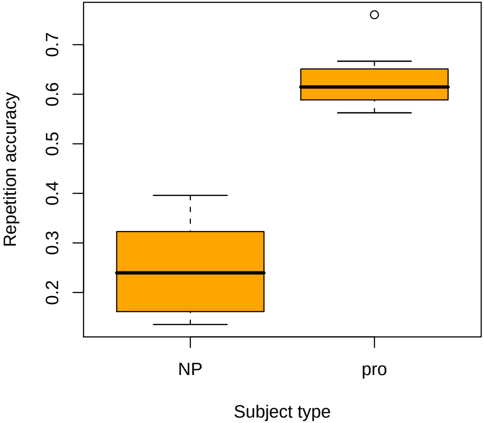
Question 1: Is head noun animacy affecting repetition accuracy?



Little skew, similar medians, large overlap: probably not significant

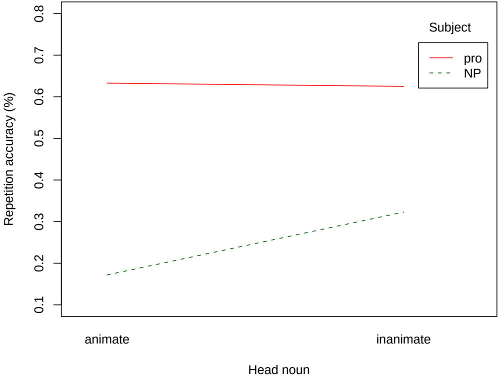
Visualizing factorial ANOVA questions

Question 2: Does lexical type of subject NP affect repetition accuracy?



Little skew, different medians, no overlap: very likely significant

Visualizing interaction



If **no** interaction, lines should be roughly parallel.

It looks like inanimate head nouns facilitate the processing of object-relative clauses with full-NP RC-subjects. Two-way ANOVA will measure this exactly.

Factorial ANOVA results

Calculations compare mean group variance and mean individual variance as ANOVA

$$F = \frac{\text{MSG}}{\text{MSE}}$$

SPSS terminology:

		between-subjects	
between-subjects	RC-subject	Animate head	Inanimate head
	full NP	animate, NP	inanimate, NP
	pronoun	animate, pronoun	inanimate, pronoun

Invoke: `General linear model` → `Univariate` → `fixed factors`

Factorial ANOVA results

```
Response: Perc
                Df   Sum Sq   Mean Sq    F value     Pr(>F)
     Animacy     1  0.02051   0.02051     5.2065    0.04154 *
          NP     1  0.58220   0.58220   147.7608  4.188e-08 ***
  Animacy:NP     1  0.02523   0.02523     6.4045    0.02639 *
   Residuals    12  0.04728   0.00394
           —
Signif. codes: 0 ***  0.001 **     0.01 *       0.05 .
```

1. Animacy of the head noun has a significant effect on repetition accuracy of object-relative clauses (despite the boxplot earlier)
2. The type of RC-subject noun phrase has a profound effect on repetition accuracy
3. Significant interaction: the difference in repetition accuracy between object-relatives with pronominal and full-NP subjects is significantly smaller when head nouns are inanimate.

Measuring effect size

We found significant main effects for both factors, and an interaction effect.

Note: because factors only have two levels here, no need to do post-hoc tests.

Additional question: How **large** are the effects we found, i.e. how meaningful are the results?

Effect size indicates the amount of variability in the dependent variable that can be accounted for by the independent variable.

Note: effect size is **not** the same as the ANOVA *p*-value:
Smaller *p*-value does not mean a bigger effect, because *p* depends on the sample size (as well as the effect size).

Measuring effect size

Effect size for one-way ANOVA: $\eta^2 = \frac{\mathsf{SSG}}{\mathsf{SST}}$ (‘**eta-squared**’)

η^2 indicates proportion of variance in the dependent variable accounted for by differences between the levels of the factor.

η^2 not suitable for *n*-way ANOVA because SST depends on presence of other factors!

Effect size for two-way ANOVA: $\eta_p^2 = \frac{\mathsf{SS}_A}{\mathsf{SS}_A + \mathsf{SSE}}$ (‘**partial eta-squared**’)

In other words: from SSG we take the portion of the variance that can be attributed to factor A, and from SST we take that same portion plus the random within-groups variability.

Measuring effect size

In our example:

$$\eta_p^2 = \frac{SS_{animacy}}{SS_{animacy}+SSE} = \frac{0.021}{0.021+0.04} = \underline{0.3}$$

$$\eta_p^2 = \frac{SS_{subject}}{SS_{subject}+SSE} = \frac{0.582}{0.582+0.04} = \underline{0.925}$$

$$\eta_p^2 = \frac{SS_{interaction}}{SS_{interaction}+SSE} = \frac{0.025}{0.025+0.04} = \underline{0.348}$$

Rule of thumb:

$\eta_p^2 < 0.1$	weak effect
$0.1 \leq \eta_p^2 < 0.6$	medium-sized effect
$\eta_p^2 \geq 0.6$	large effect

Factorial analysis of variance

Factorial analysis of variance:

- "generalized *t*-test"—compares means
- compares groups along > 1 dimensions, e.g., school classes and gender
- assumes normal distributions, similar SDs in each group
- typical application: compare processing times for **two** syntactic structures under **two** phonological conditions (factorial design)
- compares variance among means vs. general variance (F-score)
- efficient in the use of subjects and experiment time
- allows (and forces!) attention to potential **interaction**

Factorial ANOVA: another perspective

Recall that ANOVA seeks evidence for α_i (in comparison of models):

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

Similarly, factorial ANOVA asks **separately** for significance of α_i, β_j, and **interaction** $(\alpha\beta)_{ij}$, comparing models:

$$x_{ij} = \mu + \epsilon_{ij}$$
$$x_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

Factorial ANOVA models

$$x_{ij} = \mu + \epsilon_{ij}$$
$$x_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

First model:

- no group effects
- each data point represents error (ϵ) around a mean (μ)

Second model:

- real group effect(s)
- each data point represents error (ϵ) around an overall mean (μ), combined with one or two group adjustments (α_i and β_j)
- possibly group effects involve interaction ($\alpha\beta_{ij}$)

Next week

Next week we will look at

- repeated measures ANOVA with one factor
- factorial ANOVA with one within-subjects factor