Analysis of Variance

ANOVA — **AN**alysis **O**f **VA**riance

- "generalized t-test"—compares means
- two versions
  - single ANOVA — compare groups along 1 dim., e.g. school classes
  - multiple ANOVA — compare groups along > 1 dim., e.g. school classes and sex

Typical application of multiple ANOVA (MANOVA):

compare processing times for **two** syntactic structures under **two** phonological conditions

Multiple ANOVA

Like single ANOVA

- compares means of different groups
- based on F distribution

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

  - always positive
  - two kinds of df: df_{s_1}, df_{s_2}
  - value 1 indicates same variance, values near 0 or $+\infty$ indicate diff.
- uses F distribution: compare variances among means vs. overall variances
- ANOVA, MANOVA $\neq$ F-test!
- expects near-normal distributions in all groups,
- sd's in all groups roughly equal ($sd_i/sd_j \leq 2$)

Multiple ANOVA

Why Multiple ANOVA? —why not two 1-way ANOVA's?

- efficient
  - in number of experiments, subjects needed
- combining two experiments into one improves accuracy (increases n, decreases SE)
- opportunity to study **interaction**
  - age and subtype of cancer have independent effects (on mortality)
    — but these are **reversed** in some combinations, e.g. breast cancer among young women

Interaction requires care!

MANOVA Example

Withaar & Stowe investigated effects of syntax and phonology on processing time

Task: read sentences word-by-word on computer screen, press button to see following word. Times between button presses are measured.

Syntax: difference between relative clauses where

- rel. pronouns are understood **subjects**

  de bakker die de tuinmannen verjaagt

- rel. pronouns are understood **objects**

  de bakker die de tuinmannen verjagen

Phonology: rhyming vs. non-rhyming words
Longoni, Richardson & Aiello showed that word lists with rhyming elements take longer to process

Syntax, Rhyme, Reaction Times

Design: Four kinds of sentences shown:

	Syntax	
Phonology	**Obj. Rel.**	**Subj. Rel**
nonrhyming	nonrhym. obj.rel	nonrhym. subj. rel.
rhyming	rhym. obj. rel.	rhym. subj. rel.

"Extras": W&S also controlled for subject's attention span, and for which sentences were shown (no similar sentences shown to same persons).

Measurement: time needed for last word in relative clause

Data: Means, SD's of Four Groups

rhyming(y/n)	process time obj-rel.	process time subj-rel.
non-rhyming		
Mean	1581.86	1265.90
StdDev	341.82	316.89
rhyming		
Mean	1494.51	1250.55
StdDev	382.45	198.30
Grand Total		
Mean	1538.19	1258.23
StdDev	360.75	261.03

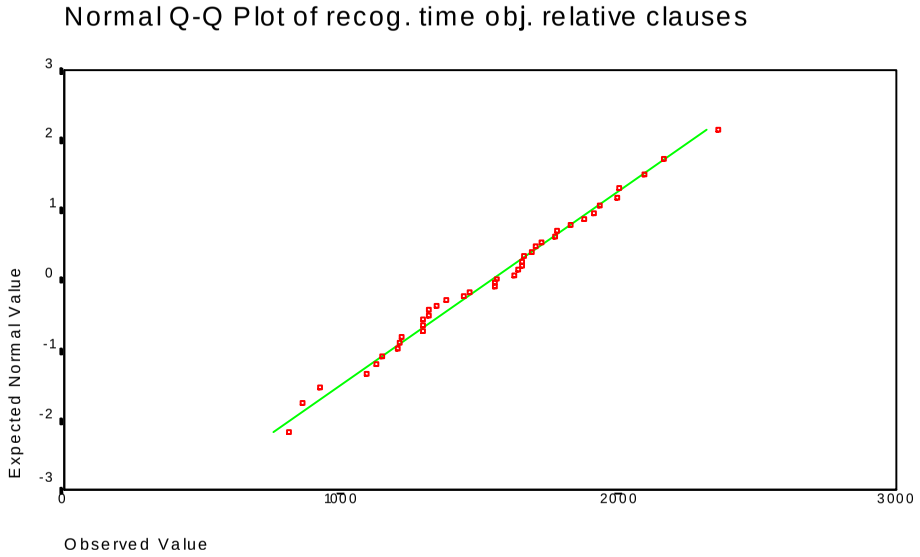
N.B. no sd is twice as large as another (but it is close!)

MANOVA question: are means significantly different?

MANOVA: Normality

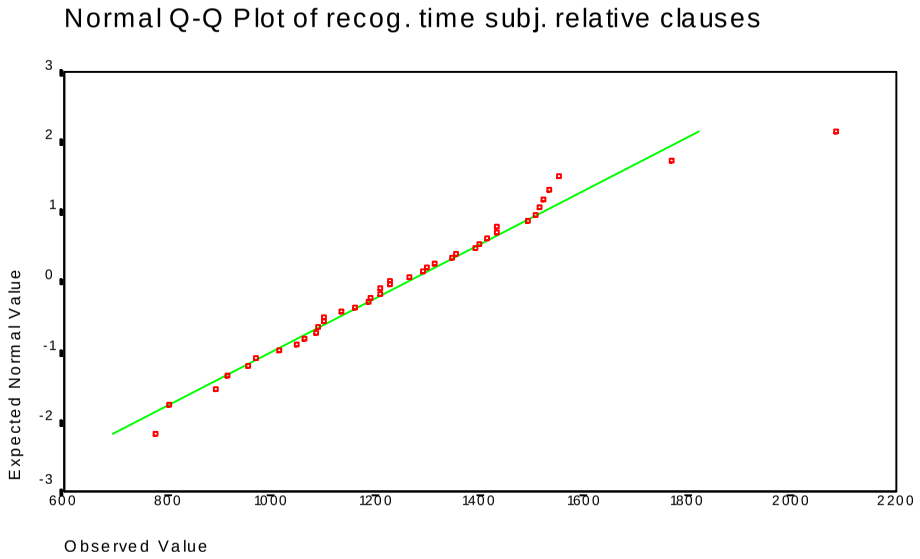
Look at data: are distributions normal?

Rhymed and unrhymed object relatives



MANOVA: Normality

Rhymed and unrhymed subject relatives



Remark: longest reaction time good candidate for elimination (worth checking on)

MANOVA Qestions

Asks **two/three** questions simultaneously:

1. Is rhyme affecting word processing time?
2. Do relative clause types affect processing time? and

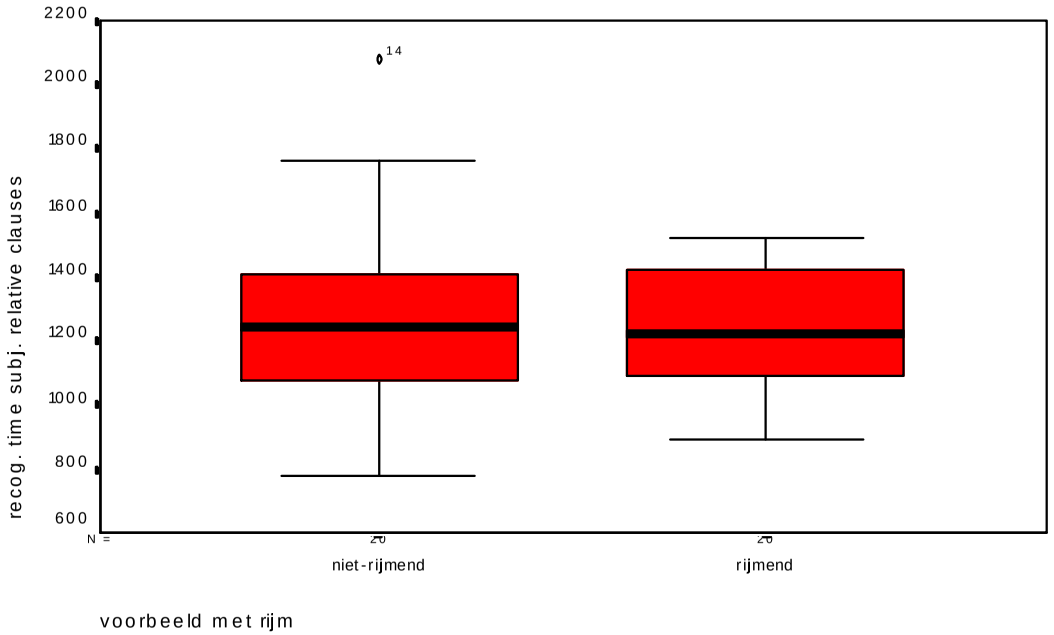
3. Do the effects interact, or are they independent?

Questions 1,2 might have been asked in separate (single) ANOVA designs (but these would have been more costly in number of subjects).

Question 3 is new to MANOVA.

Visualizing MANOVA Qestions

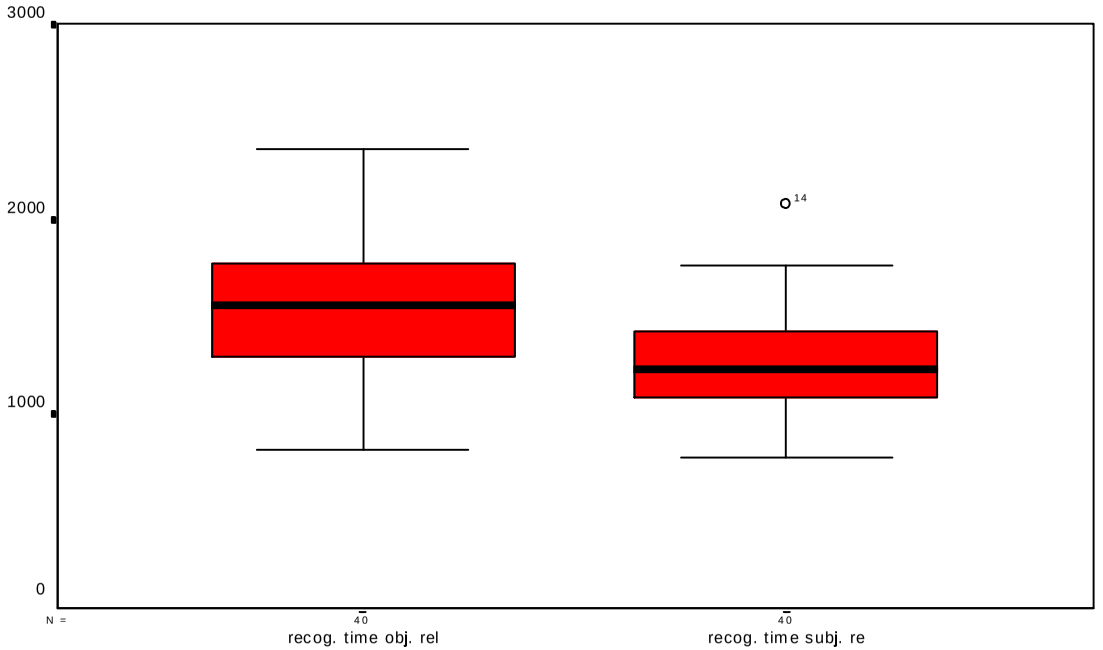
Question 1. Is rhyme affecting processing time?



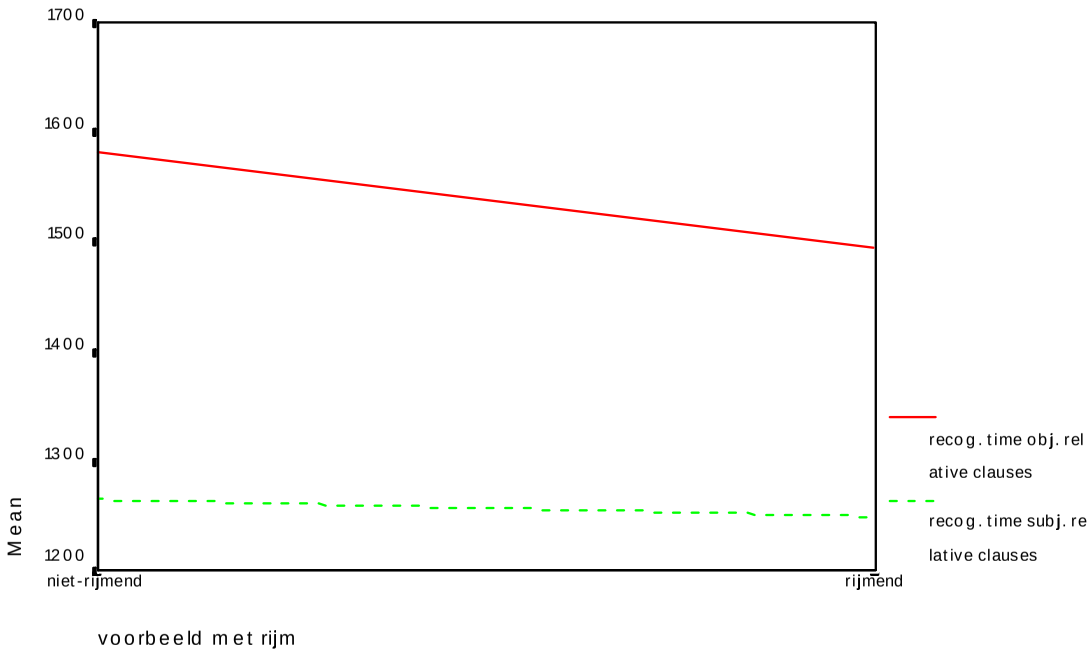
N.b. similar box plots for rhyme in object relatives.

Visualizing MANOVA Qestions

Question 2. Do relative clause types affect processing time?



Visualizing Interaction



If **no** interaction, lines should be parallel. In fact, rhyming speeds processing of object relatives.

MANOVA will measure this exactly

MANOVA Results

Calculations compare mean group (variance) and mean individual variance as ANOVA.

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

SPSS terminology:

within-subj.	between-subj.	
Phon.	**Obj. Rel.**	**Subj. Rel**
no rhyme	¬rhym. o-rel	¬rhym. s-rel.
rhyme	rhym. o-rel.	rhym. s-rel.

Invoke `ANOVA, repeated measures`

MANOVA Results

Invoke ANOVA, repeated measures

"Between-Subjects" (Row) effects (Rhyme)

```
* * * * * * Analysis of Variance -- design 1 * * * * * *

Tests of Between-Subjects Effects.

 Tests of Significance for T1 using UNIQUE sums of squares
 Source of Variation      SS      DF      MS      F  Sig of F

 WITHIN+RESIDUAL       6332920    38   166656
 RIJM                    52734     1    52734    .32   .577
```

—rhyme doesn't significantly affect processing speed

MANOVA Results

"Within-Subjects" (Column) effects (Syntax)

```
Tests involving 'SYNTAX' Within-Subject Effect.

 Tests of Significance for T2 using UNIQUE sums of squares
 Source of Variation      SS      DF      MS      F    Sig of F

 WITHIN+RESIDUAL       1321219    38    34769
 SYNTAX                1567532     1  1567532   45.08    .000
 RIJM BY SYNTAX          25917     1    25917     .75    .393
```

—syntax has profound effect; no interaction (in spite of graph!)

Multiple Analysis of Variance

MANOVA — **M**ultiple **AN**alysis **O**f **VA**riance

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

MANOVA: Another Perspective

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

$$\begin{aligned} x_{i,j} &= \mu + \epsilon_{i,j} \\ x_{i,j} &= \mu + \alpha_i + \epsilon_{i,j} \end{aligned}$$

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

$$\begin{aligned} x_{i,j} &= \mu + \epsilon_{i,j} \\ x_{i,j} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j} \end{aligned}$$

MANOVA Models

$$
\begin{aligned}
x_{i,j} &= \mu + \epsilon_{i,j} \\
x_{i,j} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j}
\end{aligned}
$$

first model

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

second model

- real group effect(s)
- each datapoint 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

Next – Nietparametrische Toets

RuG