



NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (8)
Two-way ANOVA [2.7-2.8]

Mette Langaas

To be lectured: January 27, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

ANalysis Of VAriance (ANOVA)

Bingham and Fry (2010): Chapter 2

- The Chi-square distribution and the F-distribution [2.1, 2.3]
- Orthogonality and multivariate transformation formula [2.2, 2.4]
- Normal sample mean and variance [2.5]
- One-way ANOVA [2.6]
- Two-way ANOVA [2.7-2.8]

Today

Primary aim chapter 2: ANalysis Of VAriance – comparing means by studying variability.

Two-way ANOVA : more sums and F, machine example and memory example.

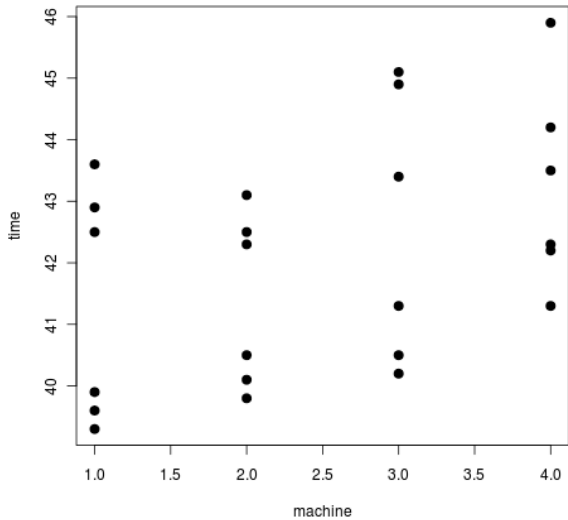
Machine example

WMMY Example 13.6.

- Response: time (s) spent to assemble a product.
- Factor: this is done by four different machines; M_1, M_2, M_3, M_4 .
- Question: Do the machines perform at the same mean rate of speed?

TABLE 13.12 Time, in Seconds, to Assemble Product

Machine	Operator:	1	2	3	4	5	6	Total
1		42.5	39.3	39.6	39.9	42.9	43.6	247.8
2		39.8	40.1	40.5	42.3	42.5	43.1	248.3
3		40.2	40.5	41.3	43.4	44.9	45.1	255.4
4		<u>41.3</u>	<u>42.2</u>	<u>43.5</u>	<u>44.2</u>	<u>45.9</u>	<u>42.3</u>	<u>259.4</u>
Total		163.8	162.1	164.9	169.8	176.2	174.1	1010.9



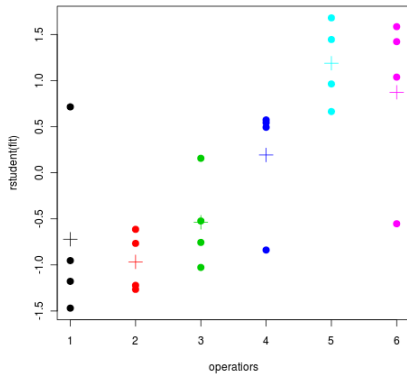
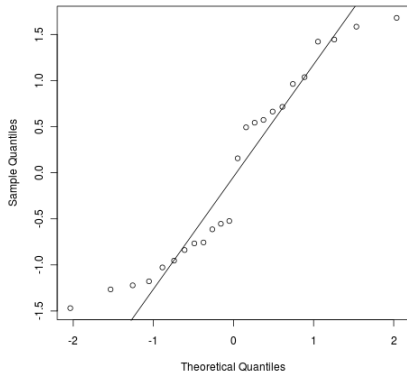
One factor ANOVA

```
> fit <- aov(time~as.factor(machine))  
> summary(fit)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(machine)	3	15.92	5.308	1.61	0.219
Residuals	20	65.94	3.297		

Residuals

Normal Q-Q Plot



Normal plot

- Empirical cumulative distribution (ecdf) is
$$\hat{F}(x) = \frac{\text{number of observations smaller or equal } x}{n}.$$
- If data are normally distributed we would expect that $\hat{F}(x)$ resembles the cdf of the normal distribution.
- x -axis: observations
- y -axis: ecdf (number between 0 and 1).

Normal plot (cont.)

- Then we adjust the scale on the y -axis so that there is a linear relationship when the data are normally distributed (using the invers standard normal cdf).
- To complicate things:
 - R: swaps the x and y axis so the theoretical quantiles are on the x -axis and the data quantiles on the y -axis.
- Anyhow – if the observations lie on a straight line we believe that they are taken from the normal distribution!

The Anderson Darling test

- Should you also test if data are normally distributed?
- Hypotheses
 - H_0 : “Data comes from a normal distribution” vs.
 - H_1 : “Data does not come from a normal distribution”.
- One of the most popular tests is the Anderson-Darling test (package `nortest` for R).
- If the p -value from the AD test is *below* a chosen significance level, e.g. 0.05, we reject H_0 , else we assume data to be normal.
- Not all statisticians agree that normality should be tested, due the lack of power for small sample sizes and robustness towards the normality assumption for various statistical models and methods. But, we may give the AD p -value a quick glance...

Machine example: operators

- The 6 repeated measurements for each machine was in fact made by 6 different operators.
- The operation of the machines requires physical dexterity and differences among the operators in the speed with which they operate the machines is anticipated.
- All of the 6 operators have operated all the 4 machines, and the machines were assigned in random order to the operators= *randomized complete block design*.
- By including a blocking factor called Operator, we will reduce the variation in the experiment that is due to random error. Thus, we reduce variation due to *anticipated factors*.
- By randomizing the order the machines were assigned to the operators we aim to reduce the variation due to *unanticipated factors*.

Two-way ANOVA [2.7]

One-way ANOVA, model assumptions

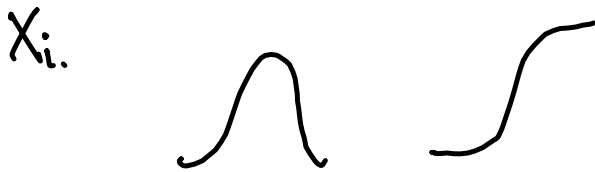
$$X_{ij} = \mu_i + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \text{ i.i.d. } N(0, \sigma^2)$$

\uparrow
 $\mu + \alpha_i$, where $\sum \alpha_i = 0$

$i = 1, \dots, r$
 $j = 1, \dots, n_i$

$$\hat{\varepsilon}_{ij} = X_{ij} - \hat{\mu}_i = e_{ij} \text{ residual}$$

\uparrow
 $X_{i.}$



New model:

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \text{ i.i.d. } N(0, \sigma^2)$$

where

$$\sum_{i=1}^r \alpha_i = 0, \quad \sum_{j=1}^n \beta_j = 0$$

$$i = 1, \dots, r$$

$$j = 1, \dots, n$$

\uparrow balanced

$$\text{Now } X_{.j} = \frac{1}{r} \sum_{i=1}^r X_{ij}$$

$$X_{ij} = X_{..} + (X_{i.} - X_{..}) + (X_{.j} - X_{..}) + (X_{ij} - X_{i.} - X_{.j} + X_{..})$$

square & sum $\Sigma\Sigma$

$$\sum_i \sum_j (X_{ij} - X_{..})^2 = n \sum_i (X_{i.} - X_{..})^2 + r \sum_j (X_{.j} - X_{..})^2$$

SS
SSA
SSB

(r \cdot n - 1)
(r - 1)
(n - 1)

$$+ \sum_i \sum_j (X_{ij} - X_{i.} - X_{.j} + X_{..})^2$$

SSE

(r - 1)(n - 1)

cross terms = 0
(n balanced
set-up)

Proceed in similar manner as for one-way ANOVA.

Effect of factor A:

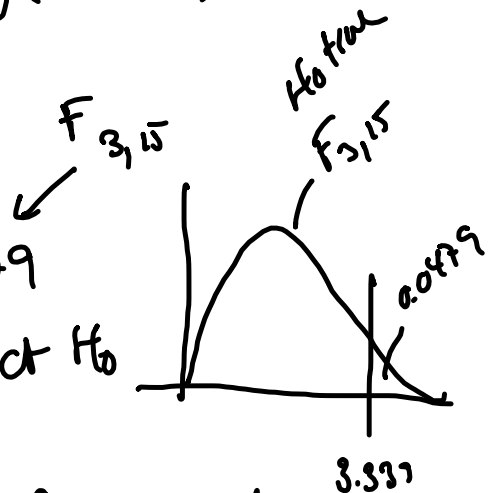
$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0 \text{ vs } H_1: \text{at least one } \alpha_i \neq 0$$

$$F = \frac{SSA / (r-1)}{SSE / ((r-1)(n-1))} \sim F_{(r-1), (r-1)(n-1)}$$

ANOVA output: $F = 3.339$

$$p\text{-value} = 0.0479$$

Conclusion: reject H_0



This model assumes additive effects of treatment and blocks.

Model and Sums of squares

Model

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \text{ for } i = 1, 2, \dots, r \text{ and } j = 1, 2, \dots, n$$

Sums of Squares Identity

$$X_{ij} = X_{..} + (X_{i.} - X_{..}) + (X_{.j} - X_{..}) + (X_{ij} - X_{i.} - X_{.j} + X_{..})$$

$$\sum_{i=1}^r \sum_{j=1}^n (X_{ij} - X_{..})^2 = n \sum_{i=1}^r (X_{i.} - X_{..})^2 + r \sum_{j=1}^n (X_{.j} - X_{..})^2$$

$$+ \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - X_{i.} - X_{.j} + X_{..})^2$$

$$SS = SSA + SSB + SSE$$

$$r \cdot n - 1 = (r - 1) + (n - 1) + (r - 1)(n - 1)$$

Effect of factor A:

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ vs. $H_1 : \text{At least one } \alpha_i \text{ different from } 0$
is then tested based on

$$F_1 = \frac{\frac{SSA}{r-1}}{\frac{SSE}{(r-1)(n-1)}}$$

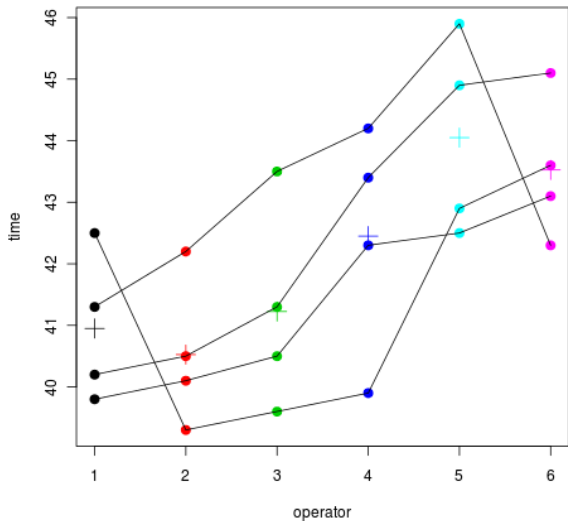
Where H_0 is rejected if $f_1 > f_{\alpha, (r-1), (r-1)(n-1)}$.

Block effect present?

$H_0 : \beta_1 = \beta_2 = \dots = \beta_n = 0$ vs. $H_1 : \text{At least one } \beta_i \text{ different from } 0$
is then tested based on

$$F_2 = \frac{\frac{SSB}{n-1}}{\frac{SSE}{(r-1)(n-1)}}$$

Where H_0 is rejected if $f_2 > f_{\alpha, (n-1), (r-1)(n-1)}$.



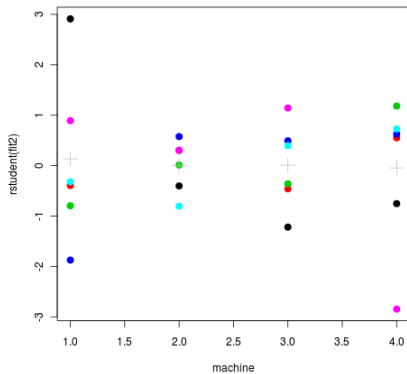
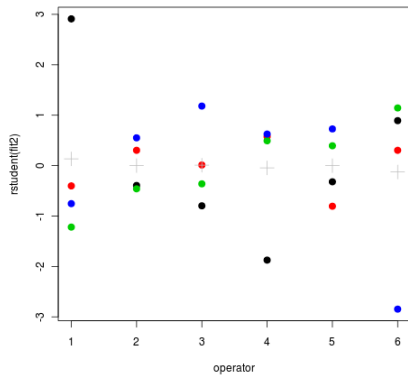
RCBD ANOVA

```
> fit2 <- aov(time~as.factor(machine)+as.factor(operator))
> summary(fit2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(machine)	3	15.92	5.308	3.339	0.04790 *
as.factor(operator)	5	42.09	8.417	5.294	0.00533 **
Residuals	15	23.85	1.590		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals



A second look at the RCBD: additive effects

Previously, randomized complete block design (RCBD) with the machine example:

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where $\sum_{i=1}^r \alpha_i = 0$ and $\sum_{j=0}^n \beta_j = 0$.

This is called *additive effects of treatment and blocks*.

- This means that if we compare two operators there is a constant difference in time to assemble the product,
- or, if we compare machines, these are ranked in the same order of (wrt time) for each operator.

Estimates

$$\hat{\mu} = 42.1208$$

$$\hat{\alpha}_1 = -0.8208$$

$$\hat{\alpha}_2 = -0.7375$$

$$\hat{\alpha}_3 = 0.4458$$

$$\hat{\alpha}_4 = 1.1125$$

$$\hat{\beta}_1 = -1.1708$$

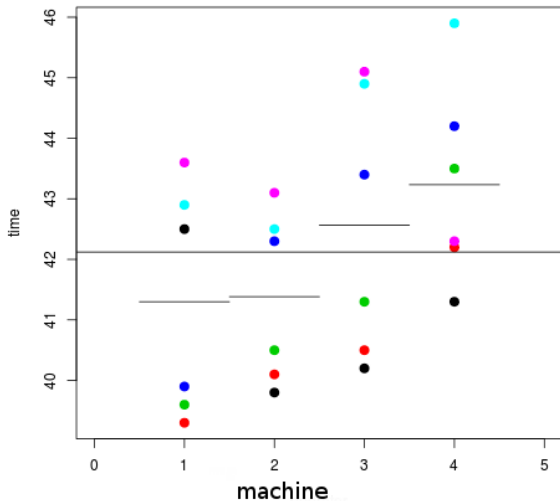
$$\hat{\beta}_2 = -1.5958$$

$$\hat{\beta}_3 = -0.8958$$

$$\hat{\beta}_4 = 0.3292$$

$$\hat{\beta}_5 = 1.9292$$

$$\hat{\beta}_6 = 1.404167$$



Estimates

$$\hat{\mu} = 42.1208$$

$$\hat{\alpha}_1 = -0.8208$$

$$\hat{\alpha}_2 = -0.7375$$

$$\hat{\alpha}_3 = 0.4458$$

$$\hat{\alpha}_4 = 1.1125$$

$$\hat{\beta}_1 = -1.1708$$

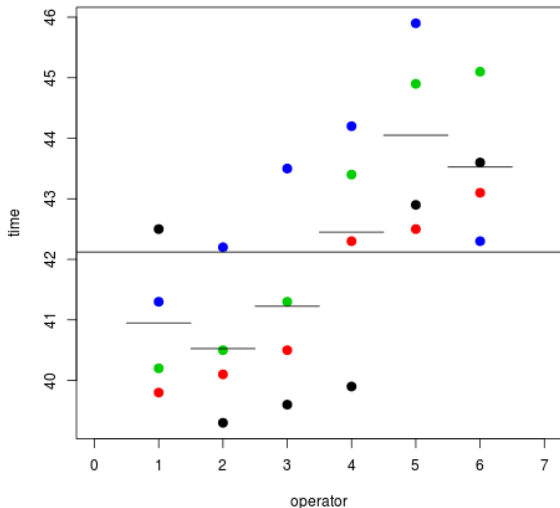
$$\hat{\beta}_2 = -1.5958$$

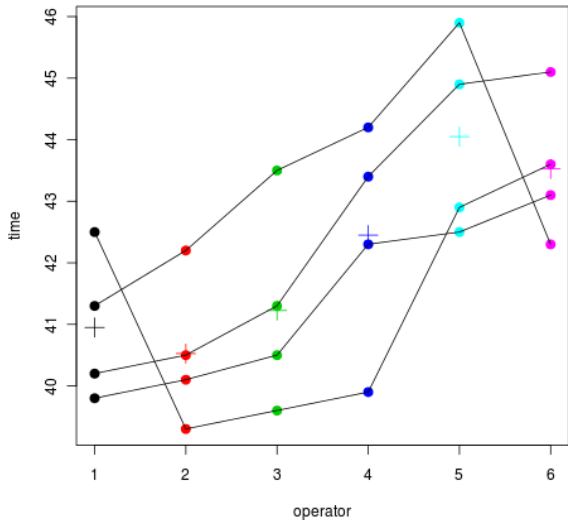
$$\hat{\beta}_3 = -0.8958$$

$$\hat{\beta}_4 = 0.3292$$

$$\hat{\beta}_5 = 1.9292$$

$$\hat{\beta}_6 = 1.404167$$





Interaction effect?

But, it could be interactions present. What if one of the operators really could not manage one of the machines?

Model with interaction between treatment and block:

$$X_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$$

where $\sum_{i=1}^r \gamma_{ij} = \sum_{j=1}^n \gamma_{ij} = 0$ in addition to $\sum_{i=1}^r \alpha_i = 0$ and $\sum_{j=1}^n \beta_j = 0$.

But, since we only have one observation for each combination of i and j , we can not separate γ_{ij} and ε_{ij} .

Interaction effect?

$$SSE = \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - X_{.i} - X_{.j} + X_{..})^2$$

$$E\left(\frac{SSE}{(r-1)(n-1)}\right) = \sigma^2 + \frac{\sum_{i=1}^r \sum_{j=1}^n \gamma_{ij}^2}{(n-1)(r-1)}$$

A large value of SSE will either mean that we have an interaction term present, or that σ^2 is large. We can not assess interaction in a RCBD. We need more than one observation for each observation to distinguish between γ_{ij} and ε_{ij} .

Age and memory

- Why do older people often seem not to remember things as well as younger people? Do they not pay attention? Do they just not process the material as thoroughly?
- One theory regarding memory is that verbal material is remembered as a function of the degree to which it was processed when it was initially presented.
- Eysenck (1974) randomly assigned 50 younger subjects and 50 older (between 55 and 65 years old) to one of five learning groups.
- After the subjects had gone through a list of 27 items three times they were asked to write down all the words they could remember.

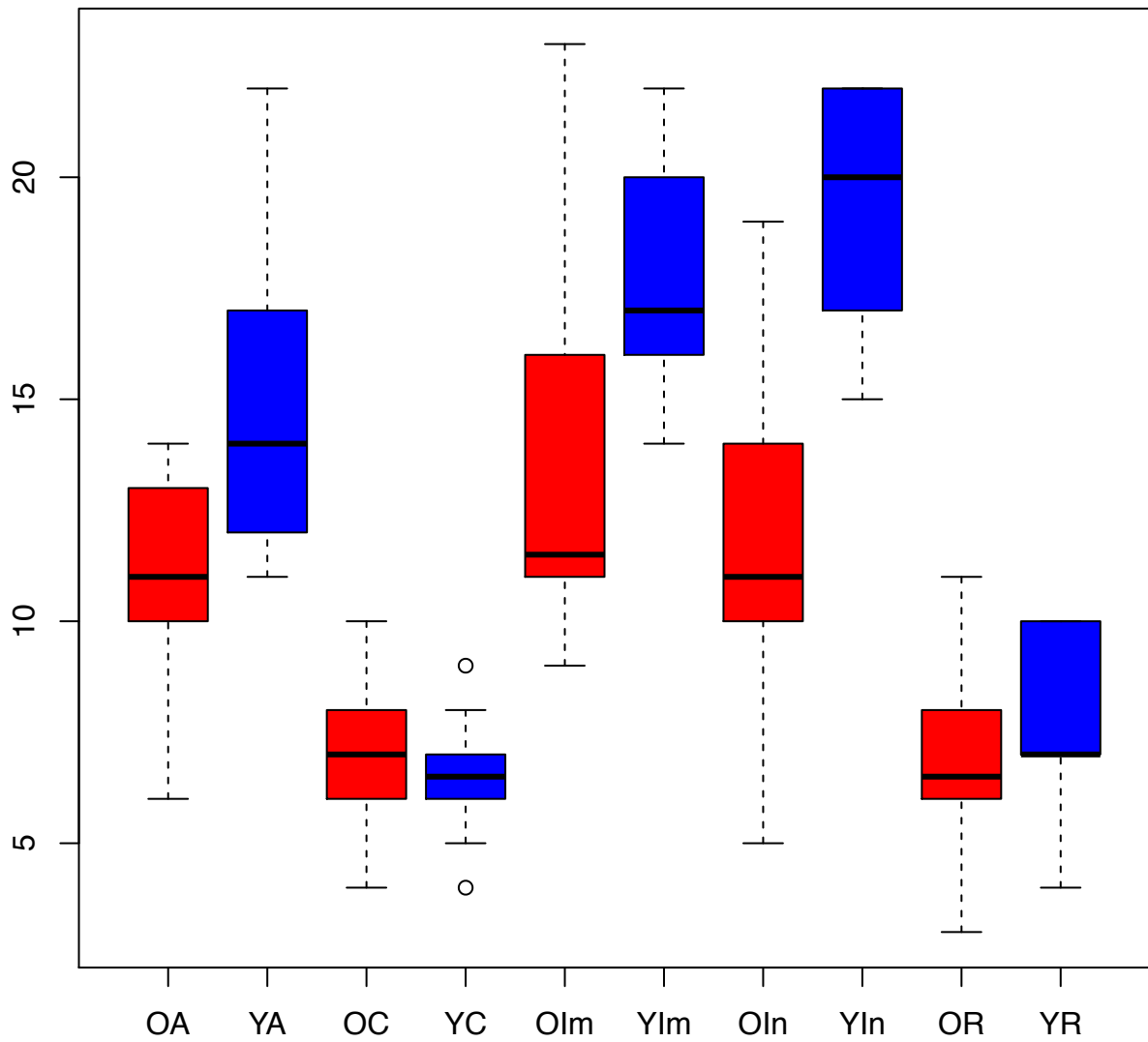
Eysenck study of recall of older and younger subjects under conditions of differential processing, Eysenck (1974) and presented in Howell (1999).

The Age and Memory data set

- Number of words recalled: After the subjects had gone through the list of 27 items three times they were asked to write down all the words they could remember.
- Age: Younger (18-30) and Older (55-65).

The Age and Memory data set: Process

- The Counting group was asked to read through a list of words and count the number of letters in each word. This involved the lowest level of processing.
 - The Rhyming group was asked to read each word and think of a word that rhymed with it.
 - The Adjective group was asked to give an adjective that could reasonably be used to modify each word in the list.
 - The Imagery group was instructed to form vivid images of each word, and this was assumed to require the deepest level of processing.
- None of these four groups was told they would later be asked to recall the items.
- Finally, the Intentional group was asked to memorize the words for later recall.



Model and Sums of Squares

Model:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

for $i = 1, 2, \dots, r$ and $j = 1, 2, \dots, n$ and $k = 1, \dots, m$

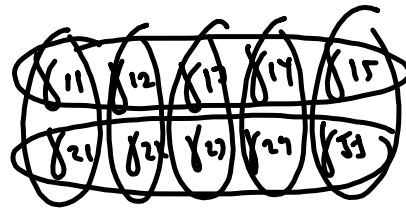
$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

Two-way ANOVA w/interactions [2.8]

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \begin{array}{l} i=1, \dots, r \\ j=1, \dots, n \\ k=1, \dots, m \end{array}$$

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0, \quad \sum_i \gamma_{ij} = 0 \quad \forall j$$

$$\varepsilon_{ijk} \text{ i.i.d. } N(0, \sigma^2) \quad \sum_j \gamma_{ij} = 0 \quad \forall i$$



$$X_{i..} = \frac{1}{n \cdot m} \sum_j \sum_k X_{ijk}$$

$$X_{.j.} = \frac{1}{r \cdot m} \sum_i \sum_k X_{ijk}$$

$$X_{ij.} = \frac{1}{m} \sum_k X_{ijk}$$

$$X_{...} = \frac{1}{n \cdot m \cdot r} \sum_i \sum_j \sum_k X_{ijk}$$

$$SS = \sum_i \sum_j \sum_k (X_{ijk} - X_{...})^2$$

$$= n \cdot m \sum_i (X_{i..} - X_{...})^2 \leftarrow SSA$$

$$+ r \cdot m \sum_j (X_{.j.} - X_{...})^2 \leftarrow SSB$$

$$+ m \sum_i \sum_j (X_{ij.} - X_{i..} - X_{.j.} + X_{...})^2 \leftarrow SS^{(AB)}$$

$$+ \sum_i \sum_j \sum_k (X_{ijk} - X_{ij.})^2 \leftarrow SSE$$

$$SS = SSA + SSB + SS^{(AB)} + SSE$$

$$rnm - 1 \quad r - 1 \quad r - 1 \quad (r - 1)(n - 1) \quad rn(m - 1)$$

→ Place into ANOVA table.

Two-way ANOVA

Source	df	SS	Mean Square	F
Treatments	$r - 1$	SST	$MST = \frac{SST}{r-1}$	MST/MSE
Blocks	$n - 1$	SSB	$MSB = \frac{SSB}{n-1}$	MSB/MSE
Interaction	$(r - 1)(n - 1)$	SSI	$MSI = \frac{SSI}{(r-1)(n-1)}$	MSI/MSE
Residual	$rn(m - 1)$	SSE	$MSE = \frac{SSE}{rn(m-1)}$	
Total	$rmn - 1$	SS		

Table 2.7 Two-way ANOVA table with interactions

Two-way ANOVA questions

There are three main questions that we might ask in two-way ANOVA:

- Does the response variable depend on Factor A?
- Does the response variable depend on Factor B?
- Does the response variable depend on Factor A differently for different values of Factor B, and vice versa?

All of these questions can be answered using hypothesis tests, first we test the interaction.

Effect of interaction AB

$$H_0^A : \gamma_{11} = \gamma_{12} = \dots = \gamma_{rn} = 0 \text{ vs.}$$

H_1 : At least one γ_{ij} different from 0

is then tested based on

$$F_3 = \frac{\frac{SS(AB)}{(r-1)(n-1)}}{\frac{SSE}{rn(m-1)}}$$

Where H_0 is rejected if $f_3 > f_\alpha, (r-1)(n-1), rn(m-1)$.

What do we do after testing for interaction?

- If the interaction is significant (we reject H_0^{AB}).
 - Then it is not recommended to test for main effects (that is, the marginal contributions of the two factors A and B separately). This is since the interpretation of the marginal “main effect” is unclear in the presence of interaction. How can we “separate out” the effect of A from the interaction?
 - Instead, it is usually preferable to examine contrasts in the treatment combinations.
- If the interaction is not found to be significant (do not reject H_0^{AB}).
 - We are then interested in the main effects. These can now be tested within the complete model.

Effect of factor A:

$H_0^A : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ vs. $H_1 : \text{At least one } \alpha_i \text{ different from } 0$

is then tested based on

$$F_1 = \frac{\frac{SSA}{r-1}}{\frac{SSE}{rn(m-1)}}$$

Where H_0^A is rejected if $f_1 > f_{\alpha, (r-1), rn(m-1)}$.

Effect of factor B:

$H_0^B : \beta_1 = \beta_2 = \dots = \beta_n = 0$ vs. $H_1 : \text{At least one } \beta_i \text{ different from } 0$

is then tested based on

$$F_2 = \frac{\frac{SSB}{n-1}}{\frac{SSE}{rn(m-1)}}$$

Where H_0^B is rejected if $f_2 > f_{\alpha, (n-1), rn(m-1)}$.

R - syntax

Additive model: $\text{Words} \sim \text{Age} + \text{Process}$

w/interaction: $\text{Words} \sim \underbrace{\text{Age} + \text{Process} + \text{Age}:\text{Process}}_{\text{Age} * \text{Process}}$

the
same

Age * Process

Eysenck ANOVA

```
> res <- aov(Words~as.factor(Age)*as.factor(Process))
> summary(res)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(Age)	1	240.2	240.2	29.936	3.98e-07	***
as.factor(Process)	4	1514.9	378.7	47.191	< 2e-16	***
as.factor(Age):as.factor(Process)	4	190.3	47.6	5.928	0.000279	***
Residuals	90	722.3	8.0			

```
---
```

Next: maybe want to compare different combinations of age and process?