Factorial Treatment Structure: Part II

Lukas Meier, Seminar für Statistik

Individual Analyses: Example (Quinn & Keough, 2000)

- Example: Fecundity of limpets.
- Response: Average number of eggs per snail.
- Factors:
  - A: Season (2 levels: Spring / Summer)
  - B: Population density (3 levels: 6 / 12 / 24 limpets under wire mesh guard)
- Per treatment combination: 3 observations

	6	***12***	***24***
Spring	1.17 0.50 1.67	1.50 0.83 1.00	0.67 0.67 0.75
Summer	4.00 3.83 3.83	3.33 2.58 2.75	2.54 1.83 1.63

Individual Analyses: Example (Quinn & Keough, 2000)

- Output of **full model**

```
> fit <- aov(y ~ season * density, data = snails)
> summary(fit)
               Df Sum Sq Mean Sq F value   Pr(>F)
season          1 17.131  17.131 119.373 1.36e-07 ***
density         2  4.001   2.001  13.940 0.000742 ***
season:density  2  1.689   0.845   5.885 0.016552 *
Residuals      12  1.722   0.144
```

Need **individual models** per season

- Output of **individual models**

```
> fit.spring <- aov(y ~ density, data = subset(snails, season == "Spring"))
> fit.summer <- aov(y ~ density, data = subset(snails, season == "Summer"))
>
> summary(fit.spring)
            Df Sum Sq Mean Sq F value Pr(>F)
density      2 0.3445  0.1722   1.104  0.391
Residuals    6 0.9361  0.1560
> summary(fit.summer)
            Df Sum Sq Mean Sq F value  Pr(>F)
density      2  5.346   2.673   20.41 0.00211 **
Residuals    6  0.786   0.131
```

Individual Analyses: Example (Quinn & Keough, 2000)

- Can improve tests by "re-using" MS_E with corresponding df's of **full model** (a better, more precise estimate of σ).
- The more df's we have for MS_E, the **more powerful** the test will be because quantiles of the F-distribution are (much) **smaller**, as seen in previous lecture.
- E.g., for the spring data-set, the "better" F-test is:

$$\frac{MS_{\text{density}}}{MS_E} = \frac{0.1722}{0.144} = 1.2$$

- We have to use an $F_{2,12}$-distribution, hence the p-value is

```
> pf(1.2, 2, 12, lower.tail = FALSE)
[1] 0.334898
```

(which is only a slight improvement here).

Individual Analyses

- Similarly, for testing contrasts etc. we can make use of the global σ estimate given by MS_E of the full model.
- This means: whenever we have an MS_E in our formulas for the **individual** models, we can “plug in” the **global** estimate with the corresponding degrees of freedom.
- This is especially useful if the error df’s of the individual model are small (say below 10).

Single Replicates

Single Replicates

- If we have a factorial experiment with only **one** observation per factor-level combination we **cannot fit a full model** anymore.
- Reason: **Perfect fit**, all residuals are **zero** (or # parameters = # observations).
- Think of two-way ANOVA situation with no replicates. If we have no replicates we have no index k:

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

Cannot distinguish between these two terms in $n = 1$ situation.

- Also: Remember factor $n - 1$ for error in ANOVA table last time (i.e., we implicitly assumed $n > 1$).

Single Replicates

- We can of course still fit a model without interaction term, i.e. a main-effects model only (= additive effects).
- If there is an underlying interaction term, we get an error estimate that is biased upwards (because it contains the error **and** the interaction term).
- Tests will be **conservative** (p-values will be too large).
- See also R-File.

Single Replicates

- Parameter estimates for main-effects model are as for previous model.
- ANOVA table now looks as follows:

Source	*df*	*Sum of squares (SS)*
A	$a-1$	$\sum_{i=1}^{a} b \cdot \hat{\alpha}_i^2$
B	$b-1$	$\sum_{j=1}^{b} a \cdot \hat{\beta}_j^2$
Error	$(a-1)\cdot(b-1)$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\overbrace{(\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j)}\right)^2$
Total	$ab-1$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\bar{y}_{..}\right)^2$

fitted value

$\#observations - 1$

df of total−sum(df *above*)

Single Replicates

- If we have no replicates and **more than two factors** we would typically remove some of the higher-order interaction terms.
- This means: We put them into the error term (the df's of the error term will therefore increase).
- Often: Transformations of the response help getting rid of interactions, see blackboard.
- Alternative: Tukey one-degree-of freedom model for interaction (see next slide).

Tukey One-Degree of Freedom Interaction

- Tukey's idea was to use only **one** additional parameter for the interaction term.
- For the two-factor model it is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \underbrace{\lambda \alpha_i \beta_j} + \epsilon_{ij}$$

Here, interaction actually **is** the product of the main-effects!

- This is a **very special form** of **interaction**.
- Some add-on packages can test $H_0: \lambda = 0$ (we will not discuss this any further).

Contrasts for Factorial Data

Contrasts for Factorial Data

- As we have already seen in the one-way ANOVA situation, **contrasts** can be used to get a more precise picture or to answer more specific question.
- Contrasts can of course also be used for factorial data structure.
- E.g., perform a **pairwise comparison between all possible treatment combinations** (think of having one “huge” cell means model with all the treatment combinations as levels).
- We can also construct **“main-effects” contrasts**.
- See R-File for examples.

Unbalanced Data

Unbalanced Data

- Up to now we assumed that our factorial data is **balanced**, i.e. all the “cells” contained the **same number of replicates** ($= n$).
- This assumption is **crucial** for the **decomposition** of the total variability into the different sources.
- With balanced data we can estimate the effects of a factor by **ignoring** the remaining factors.
- Unbalanced data **destroy** these properties.

- Calculations are more involved. **No** independent estimates anymore and **no** unique decomposition either.

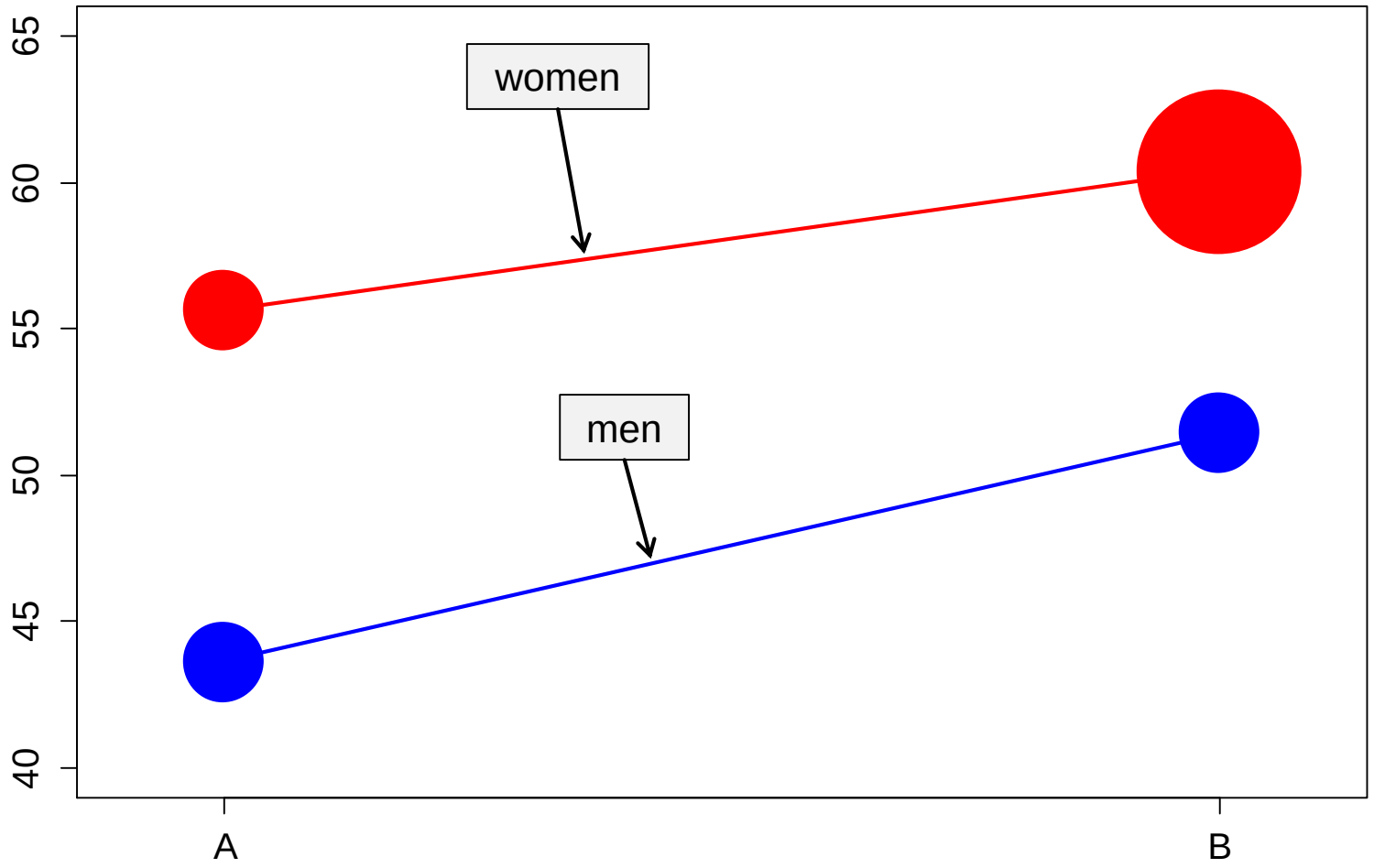
Unbalanced Data: Toy Example

Consider the following data-set for a sports experiment (running time in seconds for a specific track, unpaired data). In red: number of observations.

	Energy Drink A	*Energy Drink B*
Men	40.6, 49.7, 42.1, 42.2, 39.0, 44.2, 44.1, 43.1, 44.7, 46.3 **10**	49.7, 48.1, 49.7, 52.0, 51.5, 49.9, 55.6, 53.0, 53.5, 51.1 **10**
Women	55.7, 61.0, 58.0, 54.1, 51.9, 54.2, 54.4, 55.4, 55.4, 56.1 **10**	62.0, 60.3, 59.9, 61.2, 66.2, 56.5, 59.7, 63.0, 58.4, 61.7, 61.4, 62.6, 56.8, 55.2, 66.1, 60.6, 58.9, 59.1, 56.8, 62.5, 58.5, 61.3, 62.2, 62.5, 60.8, 57.1, 61.6, 65.9, 58.6, 60.6, 56.1, 53.6, 62.4, 62.2, 59.2, 62.9, 57.0, 58.5, 60.9, 63.4, **40**

Unbalanced Data: Toy Example

Interaction plot. Circle area proportional to sample size



Unbalanced Data: Toy Example

- If we simply ignore the “gender” structure and estimate the drink effect we are estimating the wrong effect.
- Why? We have **more women** with **Energy Drink B**.
- Gender and Energy Drink are **correlated** in this example.
- When moving from Drink A to Drink B we are also moving to a different gender structure.
- In the **balanced** case this would **not** happen.

Unbalanced Data: Toy Example

- Hence, we can't estimate the parameters one at a time anymore.
- Parameters have to be estimated **simultaneously** using the principle of least squares (no problem for the computer).
- Parameter estimates estimate the "right thing".
- Similarly: Sum of squares **cannot** be partitioned into different sources anymore.
- Note: There are unbalanced situations that are still "nice".

Unbalanced Data

- "Solution" to decomposition of sum of squares: use **model comparison** approach.
- Remember: SS_B (and the other sum of squares) can be thought of as the **reduction of residual sum of squares** when adding the factor B to our model.
- In the balanced case, it does **not** matter whether we have A (or AB) in our model or not: SS_B is always the same.
- For unbalanced data, SS_B depends on the remaining terms in the model.
- Notation: $SS(B \mid 1, A)$ is **reduction of error sum of squares** when comparing the models $(1, A, B)$ with $(1, A)$.

Unbalanced Data

- **Model terms** in two-way ANOVA situation: $(1, A, B, AB)$

- Could have a look at
  - $SS(A \mid 1)$
  - $SS(B \mid 1, A)$
  - $SS(A \mid 1, B)$
  - $SS(AB \mid 1, A, B)$
  - $SS(\ A \mid 1, B, AB\ )$
  - …

- SS_E or MS_E are typically taken from the **full** model (including **all** terms).

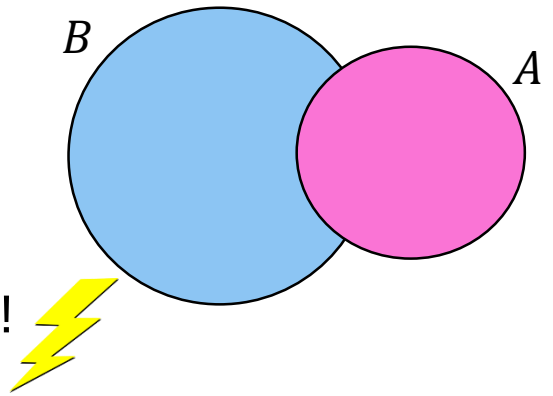
Different Types of Sum of Squares

Terminology from SAS

- **Type I**: **Sequential** sum of squares

  Sequentially build up model

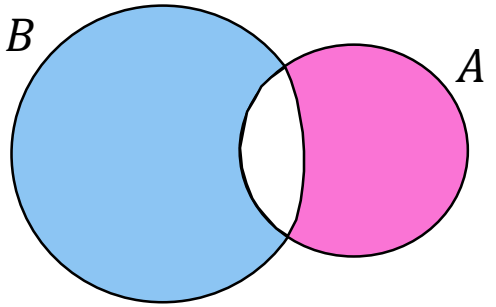
  - $SS(A \mid 1)$
  - $SS(B \mid 1, A)$
  - $SS(AB \mid 1, A, B)$
  - Hence: Depends on **ordering** of factors!
  - R: `aov`



- **Type II**: **Hierarchical / partially sequential** approach

  Control for the influence of the largest hierarchical model not including the term of interest.

  - $SS(A \mid 1, B)$
  - $SS(B \mid 1, A)$
  - $SS(AB \mid 1, A, B)$
  - R: Function `Anova` in package `car`.



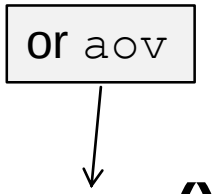
Different Types of Sum of Squares

- **Type III**: **Fully adjusted / marginal** approach

  Control for **all** other terms

  - $SS(\,A \mid 1, B, AB\,)$ (meaningful?)
  - $SS(\,B \mid 1, A, AB\,)$ (meaningful?)
  - $SS(\,AB \mid 1, A, B\,)$
  - R: `drop1`
  - Typically the preferred type.

In R…



R FAQ 7.18: Why does the output from anova() depend on the order of factors in the model?

- In a model such as ~A+B+A:B, R will report the difference in sums of squares between the models ~1, ~A, ~A+B and ~A+B+A:B. If the model were ~B+A+A:B, R would report differences between ~1, ~B, ~A+B, and ~A+B+A:B . In the first case the sum of squares for A is comparing ~1 and ~A, in the second case it is comparing ~B and ~B+A. In a non-orthogonal design (i.e., most unbalanced designs) these comparisons are (conceptually and numerically) different.
- Some packages report instead the sums of squares based on comparing the full model to the models with each factor removed one at a time (the famous 'Type III sums of squares' from SAS, for example). These do not depend on the order of factors in the model. **The question of which set of sums of squares is the Right Thing provokes low-level holy wars on R-help from time to time**.
- **There is no need to be agitated about the particular sums of squares that R reports. You can compute your favorite sums of squares quite easily. Any two models can be compared with anova(model1, model2), and drop1(model1) will show the sums of squares resulting from dropping single terms.**

Recommendations / Comments

- With **balanced data** we always get the **same result**, no matter what type we use.
- For main-effects only models, type II and type III coincide.
- Type I is useful if there is some intrinsic ordering of the terms.
- If there is a significant interaction, tests of the corresponding main-effects are typically difficult to interpret.