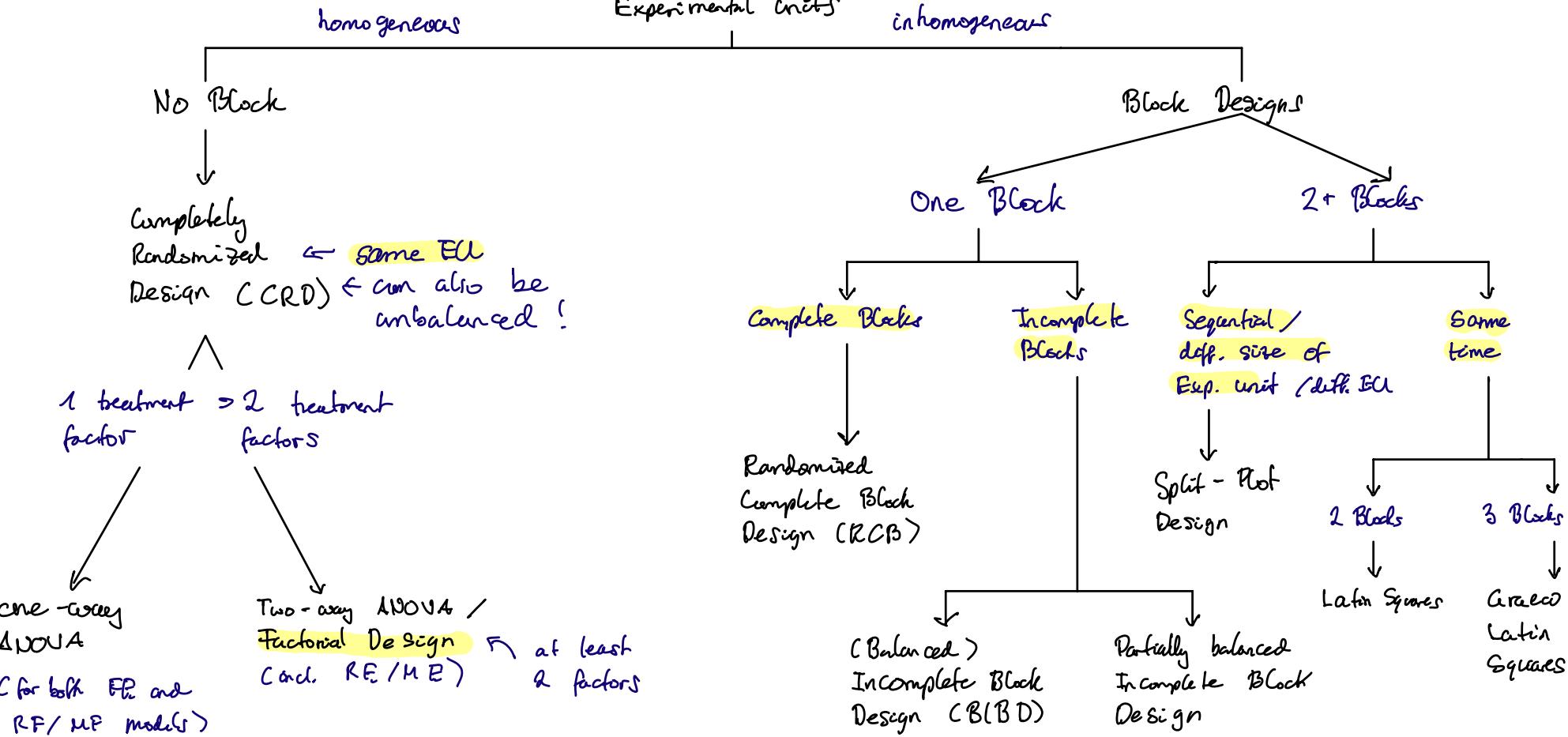
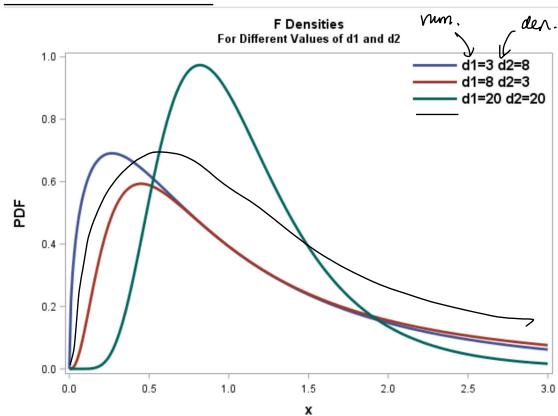


ANOVA

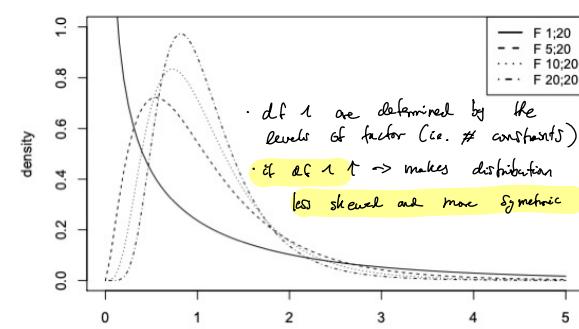
Experimental units



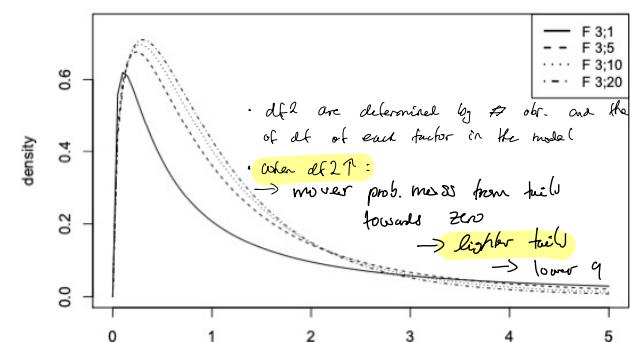
F-Distribution



- Effect of increasing df_1



- Effect of increasing df_2



Basics of Testing

- T-test: one-sample t-test (\approx test mean)

$$t = \frac{\bar{d} - \mu_0}{s/\sqrt{n}},$$

where \bar{d} is the mean of the data (here the differences d_1, d_2, \dots, d_n), n is the sample size, and s is the sample standard deviation (of the differences)

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2}.$$

- Two-sample T-test: for comparing two means

$$t = \frac{\bar{y}_{2\bullet} - \bar{y}_{1\bullet}}{s_p \sqrt{1/n_1 + 1/n_2}},$$

where $\bar{y}_{1\bullet}$ and $\bar{y}_{2\bullet}$ are the means of the first and second samples, n_1 and n_2 are the sample sizes, and s_p^2 is the pooled estimate of variance defined by

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_{1\bullet})^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_{2\bullet})^2}{n_1 + n_2 - 2}.$$

- For a factor with g levels $\rightarrow \frac{g(g-1)}{2}$ pairwise comparisons exists

- For the balanced case $n_1 = n_2$

- Under H_0 , $t \sim T_{n_1+n_2-2}$

- For the case where factor only has two levels, i.e. $g=2$

$$\rightarrow F_{1, n-1} = T_{n-1}^2$$

- F-Ratio \rightarrow always one-sided test

- General form: (for regression) \downarrow predictors

$$F = \frac{n-p}{r} \frac{\Delta SSE}{SST} \sim F_{r, n-p}$$

\uparrow constraints

- For Anova compares the variation (sum-of-squares test)

$$F\text{-ratio} = \frac{MS_{\text{treat}}}{MS_{\text{error}}} \sim F_{g-1, N-g}$$

- For $g=2 \rightarrow F = \text{squared } t\text{-test} : F_{1, n} = t^2$

- Test $H_0: \mu_1 = \mu_2 \dots = \mu_g$ vs. $H_A: \mu_k \neq \mu_j$ for one pair

\hookrightarrow do F-test for every factor in a factorial design

- ANOVA seen as model comparison

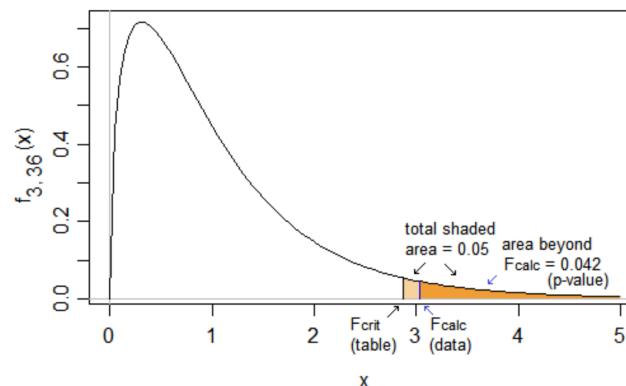
- Can be seen as in regression \rightarrow compare full to reduced model

Mit jeder Erweiterung des Modells verringert sich die Summe der quadrierten Residuen:

Modell	Residuen-SS	df	Reduktion-SS	df
$y_{ij} = e_{ij}$	SSR(0) = 100256.1	24		
$y_{ij} = \mu + e_{ij}$	SSR(μ) = 190.6	23	$R(\mu) = 100065.5$	1
$y_{ij} = \mu_i + e_{ij}$	SSR(μ_i) = 101.1	18	$R(\mu_i \mu) = 100155.0$	6
			$R(\mu_i \mu) = 89.5$	5

Relation of F to p-value (given α)

Change in df 1	df 1 \uparrow	No change	df 1 \downarrow
Change df 2 \	df 2 \uparrow	means $n \uparrow$ $\rightarrow F \uparrow \rightarrow p \downarrow$	not clear
df 2 \uparrow	not clear		
No change	depends on expl. power	No change	depends on expl. power
df 2 \downarrow	not clear	means $n \downarrow$ $\rightarrow F \downarrow \rightarrow p \uparrow$	not clear



Interpretation:

The **F-value** says us how far away we are from the hypothesis "we can not distinguish between error and treatment", i.e. "Treatment is not relevant according to our data"!

A big F-value implies that the effect of the treatment is relevant!

Remark 3 A small F-value does NOT imply that the hypothesis $A_i = 0 \forall i$ is true. (We just can not conclude that it is false!)

1. Study Setup

- **Experimental study**: can control some predictors, can control treatment assignment
 \rightarrow randomization \rightarrow allows for causal interpretation
- **Observational study**: uncontrolled (analyze after fact)
 - cross-sectional study (snapshot of population)
 - cohort study
 - case control study

- Experimental vs. Measurement units

- a) **Experimental units (EU)**: where treatment is applied
 \hookrightarrow should be able to receive treatment independently
- b) **Measurement units (MU)**: where response is measured
 \rightarrow pof. EU \neq MU

- **Randomization**: random allocation of EU to the diff. treatments \rightarrow pof. confounders are "averaged out"
 \rightarrow can get causal effect

- **Blocking**: restricted randomization scheme
 - Block: subset of EU that is more homogeneous than the entire set

- \rightarrow to block we randomize within these groups
- \rightarrow Var(EU) \downarrow but also df \downarrow
- \rightarrow Block what you can / randomize what you cannot

- **Experimental Error**: diff. EU give diff. responses to the same treatment \rightarrow want to estimate this error to make statistical inference \rightarrow need multiple replicates

- **Blinding**: evaluators don't know which treatment is given to which EU
 \hookrightarrow double-blinding: neither evaluators nor subjects know

2. Control Treatments

- std. treatment as a baseline
- Null treatment: no treatment as baseline

3. General Guidelines for Experiment Design

1. Statement of Problem / Hypothesis
2. Select response
3. Determine Source of Variation in response
 - factors of interest (treatment factors)
 - nuisance factors (blocking, randomization)
 - factors that can be held constant
4. Choose proper design + randomization scheme

2. Completely Randomized Designs (CRD)

- assume homogeneous RVE
- defined by: # replicates + # treatments \rightarrow very flexible
- Balanced: all treatment groups have same n_i
- Cell means model: each group has its own expected value:

$$Y_{ij} \sim N(\mu_i, \sigma^2) \rightarrow \text{one mean per treatment}$$

\hookrightarrow constant var across groups

$$\rightarrow Y_{ij} = \mu_i + \varepsilon_{ij} \quad (\text{where } \mu_i \text{ is the mean for treatment group } i)$$

\rightarrow Can rewrite as linear regression:

$$\cdot \text{Use: } \mu_i = \mu + \alpha_i \quad \uparrow \text{the } i\text{th treatment effect}$$

$$\rightarrow \text{results in: } Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

\hookrightarrow we predict every observation as deviation from the global mean μ .

\rightarrow But introduced an additional parameter

\rightarrow Need add. constraint for identifiability

- Side constraint (on α_i)

a) Sum of treatment effects = 0 (R: contr.sum)

$$\alpha_0 = -(\alpha_1 + \dots + \alpha_g), \sum_{i=1}^g \alpha_i = 0$$

b) Sum of weighted treatment effects = 0 (R: need to do manually)

$$\sum_{i=1}^g \alpha_i n_i = 0$$

c) Reference level: Set $\mu = \mu_0 \rightarrow \alpha_0 = 0$,
(in R: contr.treatment / default or R)
 $\alpha_1 = \mu_1 - \mu_0$
 $\alpha_2 = \mu_2 - \mu_0$

- Parameter Estimation

* LS estimation: $\hat{\mu}, \hat{\alpha}_i = \underset{\mu, \alpha_i}{\operatorname{argmin}} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \hat{\mu}_i$

\rightarrow pred./fitted values per group: $\hat{\mu}_i = \mu + \hat{\alpha}_i$
value depends on side constraint

- Variance Estimation

* $\hat{SS}_E = \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2$ just like RSS

* $MSE = \sigma^2 = \hat{MS}_E = \frac{1}{N-g} \hat{SS}_E = \frac{1}{N-g} \sum_{i=1}^g (n_i - 1) S_i^2$

Estimation Accuracy

Standard errors for the parameter estimates (using the sum of weighted treatment effects constraint).

Parameter	Estimator	Standard Error
μ	$\hat{\mu}_i$	σ/\sqrt{N}
μ_i	$\hat{\mu}_i$	$\sigma/\sqrt{n_i}$
α_i	$\hat{\alpha}_i = \bar{y}_i - \bar{y}_..$	$\sigma \sqrt{\frac{1}{n_i} \frac{1}{N}}$
$\mu_i - \mu_j = \alpha_i - \alpha_j$	$\hat{\alpha}_i - \hat{\alpha}_j$	$\sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

Therefore, a 95% confidence interval for α_i is given by

$$\hat{\alpha}_i \pm \frac{0.975}{N-g} \cdot \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{N}} \quad \rightarrow \text{just for weighted TE}$$

97.5% quantile of t_{N-g} distribution
 $N-g$ degrees of freedom because of the degrees of freedom of MS_E

SS Decomposition

$$\sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \underbrace{\sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y}_{..})^2}_{SS_{Trt}} + \underbrace{\sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_{..} - \hat{y}_{ij})^2}_{SS_E}$$

grand mean
var between groups
var between groups

$= n_i \cdot \sum_{i=1}^g (\bar{y}_i - \bar{y}_{..})^2$

ANOVA Table (One-way ANOVA)

Source	df	Sum of squares (SS)	Mean Squares (MS)	F-ratio
Treatments	$g-1$	SS_{Trt}	$MS_{Trt} = \frac{SS_{Trt}}{g-1}$	$\frac{MS_{Trt}}{MS_E}$
Error	$N-g$	SS_E	$MS_E = \frac{SS_E}{N-g}$	

```
## Call:
## aov(formula = y ~ treatment, data = meat) one-way ANOVA
##
## Residuals:
##   Min   1Q Median   3Q   Max
## -0.500 -0.225  0.110  0.210  0.320
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.4800    0.1965 38.064 2.49e-10 ***
## treatmentCO2 -4.1200    0.2779 -14.825 4.22e-07 ***
## treatmentMixed -0.2200    0.2779 -0.792  0.451
## treatmentVacuum -1.9800    0.2779 -7.125 9.95e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
## 
## Residual standard error: 0.3404 on 8 degrees of freedom
## Multiple R-squared:  0.9726, Adjusted R-squared:  0.9623
## F-statistic: 94.58 on 3 and 8 DF, p-value: 1.376e-06
```

① $RSE = \sqrt{MS_E} = \sqrt{\frac{1}{N-g} SS_E}$

② F-stat: on the factor treatment

- Expected MS:

$$E(MS_{Trt}) = EMS_{Trt} = \sigma^2 + \sum_{i=1}^g n_i \alpha_i^2 / (g-1)$$

The important things to get from this expression are

- When all of the α_i 's are zero, the mean square for treatments also estimates σ^2 .
- When some of the α_i 's are nonzero, the mean square for treatments tends to be bigger than σ^2 .

$$\hookrightarrow E[MS_{Trt}] = \sigma^2$$

Checking Model Assumptions

- Homoscedasticity: compare residuals vs. fitted values \rightarrow we want equal Var between treatments of the same factor!
- Normality of errors: check residuals \rightarrow QQ Plot
- Error independence \rightarrow difficult to check other than for serial correlation (time)
- Error mean zero \rightarrow TA plot

But works with multicollinearity (for one-way case)

2. Specific Differences (Contrasts)

- To understand exactly how the group means differ \rightarrow should define contrast a priori!

- Contrast: encodes the H₀ on the form that

$$H_0: \sum_{i=1}^g c_i \mu_i = 0 \quad (\text{sum of constraint})$$

\hookrightarrow constraint ensures that contrast is orthogonal to the overall mean

- Estimation: estimate true value of contrast

$$\sum_{i=1}^g c_i \mu_i \quad \text{with} \quad \sum_{i=1}^g c_i \bar{y}_i \quad (\text{in R: ght})$$

\rightarrow every contrast has an associated SS:

$$SS_C = \sum_{i=1}^g (c_i \bar{y}_i)^2 \quad \text{since only 1 contrast}$$

$\rightarrow 1 \text{ df}$
 $\rightarrow MS_C = SS_C / nc$

\hookrightarrow this is just squared t-stat for the null:

$$H_0: \sum_{i=1}^g c_i \mu_i = 0 \quad (\text{without } MS_E)$$

\rightarrow So under H₀: $\frac{MS_C}{MS_E} \sim F_{1, N-g}$

- Interpretation

Treatments were:

- Commercial plastic wrap (ambient air)
 - Vacuum package
 - 1% CO, 40% O₂, 59% N
 - 100% CO₂
- Current techniques
- New techniques

Possible questions and their corresponding contra-

Comparison	Corresponding contrast $c \in \mathbb{R}^4$
④ New vs. Old	$\left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$
⑤ New vs. Vacuum	$(0, -1, \frac{1}{2}, \frac{1}{2})$
CO ₂ vs. Mixed	$(0, 0, -1, 1)$
Mixed vs. Commercial	$(-1, 0, 1, 0)$

④ Average of 1) + 2) vs. Average 1) + 2)

⑤ 2) vs. Average of 3) + 4)

In R:

```
fit.gb <- glht(fit, linfct = mcp(TeachingStrategy = c(1, -1/2, -1/2)))
confint(fit.gb)
and obtained the following output:
Simultaneous Confidence Intervals
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = score ~ TeachingStrategy, data = df)
Quantile = 1.9935
95% family-wise confidence level
Linear Hypotheses:
Estimate lwr upr
1 == 0 -0.1223 -0.6173 0.3727
```

↑ 1 contrast
NOT 3!

⇒ use $F_{1, N-g}$

- **Orthogonal Contrasts**: two contrasts are called orthogonal if $\sum_{i=1}^g c_i c_i^* = 0 \rightarrow$ cf C' C is diagonal!

for balanced case $n_i = n \rightarrow \sum_{i=1}^g c_i c_i^* = 0$

- means we have independent info on two contrast → should not reuse same contrast twice
- if there are g groups → $g-1$ orthogonal contrasts → have completely exhausted the data
- for orthogonal contrasts get nice partitioning of SS:

$$SS_{C^{(1)}} + SS_{C^{(2)}} + \dots + SS_{C^{(g-1)}} = SS_{\text{Trt}}$$

3. Multiple Testing

• **P-value**: $= P_{H_0}(|T| \geq |t_{\text{null}}|) \sim U(0,1)$ (under H_0)

• Errors:

H_0	Rejects	H_1
Type I = α	correct	Type II = β

Decision

H_0	Type I = α	correct
H_1	correct	Type II = β

→ as $n \uparrow$ or eff. size $\uparrow \rightarrow \beta \downarrow \rightarrow (1-\beta) \uparrow$ but α same

• discoveries = rejecting H_0
 • false discoveries = type I error
 • power = $1 - \text{Type I error}$ ($1-\beta$)
 $(m = \text{total # of hypotheses})$

• **Comparisonwise error rate**: Type I error rate of 1 test

• **FWER**: $P(V \geq 1) = 1 - (1 - \alpha)^m$

- for independent test = α
- for pos. dependent tests
 $\rightarrow \text{FWER} < \alpha$ (as); for neg. dependence $\rightarrow \text{FWER} > \alpha$ (ind.)

• **FDR**: weaker error control than FWER
 C_{FWER} also controls FDR)

$$\text{FDR} = E[C_{\text{FDP}}] = E[V/R]$$

expectation of
false discoveries

- **Procedures to control FWER**

Recommendations

- Planned contrasts: Bonferroni
- All pairwise comparisons: Tukey HSD
- Comparison with a control: Dunnett
- Unplanned contrasts: Scheffé

Bonferroni (Conservati v)	alpha* = alpha/m Or PW* = PW·m	→ correction in two ways	Controls FWER No assumptions needed Also gives Simultaneous CI
Bonferroni- Holm	Sort p-values from small to large: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$ where $p_{(j)} \leq p_{(j+1)} \dots \leq p_{(m)}$ For $j = 1, 2, \dots$: Reject null hypothesis if $p_{(j)} \leq \frac{\alpha}{(m-j+1)}$. Stop when you reach the first non-significant p-value.	Less conservative More power (uniformly) "step down procedure" Cannot be used to generate CLs R: p.adjust	
Scheffe	Theory: <ul style="list-style-type: none"> • $SS_c \leq (g-1)MS_{\text{Trt}}$ for any contrast c (because $SS_{\text{Trt}} = SS_c + \dots$) • Hence, $\frac{SS_c}{MS_E} \leq (g-1)\frac{MS_{\text{Trt}}}{MS_E}$ for any contrast c. • Therefore, $\max_c \frac{SS_c / (g-1)}{MS_E} \leq \frac{MS_{\text{Trt}}}{MS_E} \sim F_{g-1, N-g}$ under $H_0: \mu_1 = \dots = \mu_g$. <p>R: <ul style="list-style-type: none"> • Calculate F-ratio (MS_c/MS_E) as if "ordinary" contrast. • Use $(g-1) \cdot F_{g-1, N-g, 1-\alpha}$ as critical value (instead of $F_{1, N-g, 1-\alpha}$). </p>	Controls search over any possible contrast → data fishing allowed Honest p-values Low power	
HSD: Tukey Honest Sig. differences	<ul style="list-style-type: none"> • Start with statistics of t-test (here for the balanced case where HSD gives exact p-values) • Use the distribution of $\frac{ \bar{y}_i - \bar{y}_j }{\sqrt{MS_E} \sqrt{\frac{1}{n} + \frac{1}{n}}}$ • Means: "How does the maximal difference between groups behave?" • If all the means are equal (H_0), this follows the so called studentized range distribution (R: ptukey). 	Simultaneous CI for differences between all pairs p-values with controlled FWER	<p>R: TukeyHSD or multcomp</p> <p>HSD > Bonferroni if all pairwise comparisons are of interest If only subset > may want to use BF</p> <p>F test can be significant although HSD yields only insign. Pairings (and vice versa)</p>
MCC: Multi Comparisons with Control	<ul style="list-style-type: none"> • Often: Compare all treatments with a (specific) control treatment. • Hence, do $g-1$ (pairwise) comparisons with the control group. • Dunnett procedure constructs simultaneous confidence intervals for the differences $\mu_i - \mu_g, i = 1, \dots, g-1$ (assuming group g is control group). • R: Use package multcomp. 	Get smaller p-values than with HSD as we have fewer test to correct for	<p>Should not cond. On sign. F-test</p>

Exemplars:

1) Bonferroni / Bonferroni-Holm

```
library(multcomp)
mat.contr <- rbind(c(1, -2, 1), c(1, 0, -1))
circ.mc <- glht(circ.fit, linfct = mcp(Type = mat.contr))
summary(circ.mc, test = adjusted("bonferroni"))
...
## Simultaneous Tests for General Linear Hypotheses
## Multiple Comparisons of Means: User-defined Contrasts
##
## Fit: aov(formula = Y ~ Type, data = circ)
##
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## 1 == 0 -25.200 4.503 -5.596 0.000234 ***
## 2 == 0 2.400 2.600 0.923 0.748310
## ...
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```

We see that the contrast $c = (1, -2, 1)$ is highly significant. Hence, we reject the null hypothesis $H_0: \mu_2 = \frac{1}{2}(\mu_1 + \mu_3)$. There is a significant difference between the mean of type 2 and the means of the other types (using a significance level of 0.05).

→ Could do do it by hand:

We could also do the tests by hand, by computing the value of the F-ratio for each contrast (which follows an $F_{1, N-g}$ -distribution) and test whether the corresponding null hypothesis has to be rejected.

Test	Contrast	c_1	c_2	c_3	$\sum_i c_i \bar{y}_i$	$\sum_i (c_i^2/n_i)$	$SS_c = (\sum_i c_i \bar{y}_i)^2 / \sum_i (c_i^2/n_i)$
T2 vs. other	c	1	-2	1	-25.2	1.2	529.2
T1 vs. T3	c^*	1	0	-1	2.4	0.4	14.4

Using the SS_c we calculate the MS_c for the contrasts. By dividing MS_c by MS_E , we obtain the value of the F-ratio for a contrast:

$$MS_c = \frac{SS_c}{1} = SS_c$$

$$\text{F-ratio} = \frac{MS_c}{MS_E} \sim F_{1, N-g} = F_{1, 12}$$

3) Scheffé

```
plant.glht.scheffe <- glht(fit.plant,
linfct = mcp(group = c(1/2, -1, 1/2)))
## p-value according to Scheffé (g = 3, N - g = 27)
pf((summary(plant.glht.scheffe)$test$stat)^2 / 2, 2, 27,
lower.tail = FALSE)
```

```
## 1
## 0.05323
```

If we use a significance level of 5% we do not get a significant result (with the more extreme contrast $c = (0, -1, 1)$ we would be successful).

4) Tukey HSD

a)

```
TukeyHSD(fit.plant)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = weight ~ group, data = PlantGrowth)
##
## $group
## diff lwr upr p adj
## trt1-ctrl -0.371 -1.0622 0.3202 0.3909
## trt2-ctrl 0.494 -0.1972 1.1852 0.1980
## trt2-trt1 0.865 0.1738 1.5562 0.0120 → abs this sign.
```

Each line in the above output contains information about a specific pairwise comparison. For example, the line trt1-ctrl says that the comparison of level trt1 with ctrl is not significant (the p-value is 0.39). The confidence interval for the difference $\mu_2 - \mu_1$ is given by [-1.06, 0.32]. Confidence intervals can be visualized by simply calling plot.

b)

With the function TukeyHSD() we can compare pairs of treatment means.

```
TukeyHSD(circ.fit, which = "Type", conf.level = 0.95)
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Y ~ Type, data = circ)
##
## $Type
## diff lwr upr p adj
## 2-1 11.4 4.463555 18.336445 0.0023656
## 3-1 -2.4 -9.336445 4.536445 0.6367043 → abs 0 → nf sign. diff
## 3-2 -13.8 -20.736445 -8.863555 0.0005042
```

The result can be interpreted as follows:

Type 2 is significantly different from the other two types at level $\alpha = 0.05$. The difference between Type 1 and Type 3 is not significantly different from 0.

5) MCC → comparison with control group

```
levels(PlantGrowth[, "group"])
## [1] "ctrl" "trt1" "trt2"

With multcomp we simply set group = "Dunnett".
plant.glht.dunnett <- glht(fit.plant, linfct = mcp(group = "Dunnett"))
summary(plant.glht.dunnett)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
## ...
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## trt1 - ctrl == 0 -0.371 0.279 -1.33 0.32
## trt2 - ctrl == 0 0.494 0.279 1.77 0.15
## (Adjusted p values reported -- single-step method)
```

We get smaller p-values than with the Tukey HSD procedure because we have to correct for less tests (there are more comparisons between pairs than there are comparisons to the control treatment).

4. Factorial Treatment Structure

- exists if the g treatments are a combo of >1

Setup:

- Factor A with a levels
- Factor B with b levels
- $n > 1$ replicates for every combination - balanced
- Total of $N = a \cdot b \cdot n$ observations

- Crossed Factors: Factors are crossed if all combos are seen

- Two-way ANOVA model with interaction

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

Typically, **sum-to-zero constraints** are being used, i.e.

- $\sum_{i=1}^a \alpha_i = 0, \sum_{j=1}^b \beta_j = 0$. $\rightarrow a-1$ and $b-1$ degrees of freedom
- $\sum_{i=1}^a (\alpha\beta)_{ij} = 0, \sum_{j=1}^b (\alpha\beta)_{ij} = 0$. $\rightarrow (a-1) \cdot (b-1)$ degrees of freedom

→ can look at interaction plot to empirically see the effect

- Main effects: avg. effects after changing the level of a factor

- Interaction effect: contains leftovers after having considered main effects

- Additive model: does NOT include an interaction
↳ only parallel lines for diff. levels

⇒ Transformation can often help to transform multiplicative interaction effects into additive effects → use log

⇒ interactions often a question of scale of response

Parameter Estimation

Estimates for the balanced case (with sum-to-zero constraints) are:

Parameter	Estimator
μ	$\hat{\mu} = \bar{y}_{..}$
α_i	$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{..}$
β_j	$\hat{\beta}_j = \bar{y}_{..j} - \bar{y}_{..}$
$(\alpha\beta)_{ij}$	$(\hat{\alpha}\beta)_{ij} = \bar{y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$

• balanced design = orthogonal design

↳ main effects can be estimated as in one-way ANOVA

- SS Decomposition: can again decompose SS

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

- ANOVA Table (two-way ANOVA)

Source	df	SS	MS	F
A (cells)	$a-1$	SS_A	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
B	$b-1$	SS_B	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
AB	$(a-1) \cdot (b-1)$	SS_{AB}	$\frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
Error	$ab \cdot (n-1)$ (replicates)	SS_E	$\frac{SS_E}{(n-1)ab}$	

$$= (Total obs - 1) - \text{all other df used up!}$$

• can apply F-test → move from bottom to top (i.e. here start with interaction)

→ usually interested in interaction effect only

F-Tests: Overview

Interaction AB

- $H_0: (\alpha\beta)_{ij} = 0$ for all i, j
- $H_A: \text{At least one } (\alpha\beta)_{ij} \neq 0$
- Under $H_0: \frac{MS_{AB}}{MS_E} \sim F_{(a-1)(b-1), ab(n-1)}$

Main effect A

- $H_0: \alpha_i = 0$ for all i
- $H_A: \text{At least one } \alpha_i \neq 0$
- Under $H_0: \frac{MS_A}{MS_E} \sim F_{(a-1), ab(n-1)}$

Main effect B

- $H_0: \beta_j = 0$ for all j
- $H_A: \text{At least one } \beta_j \neq 0$
- Under $H_0: \frac{MS_B}{MS_E} \sim F_{(b-1), ab(n-1)}$

usually only interested in interaction effects

} for main effects estimate separate models and reuse MS_E from full model
↳ separate at every level of factor

• generally if $df_E = 0 \rightarrow$ need to remove one of the other factors!

• Always include all factors, even if not significant

→ analyses must follow the randomization used in experiment

ANOVA with > 2 factors

Source	df	F-ratio
A	$a-1$	$\frac{MS_A}{MS_E}$
B	$b-1$	$\frac{MS_B}{MS_E}$
C	$c-1$	$\frac{MS_C}{MS_E}$
AB	$(a-1)(b-1)$	$\frac{MS_{AB}}{MS_E}$
AC	$(a-1)(c-1)$	$\frac{MS_{AC}}{MS_E}$
BC	$(b-1)(c-1)$	$\frac{MS_{BC}}{MS_E}$
ABC	$(a-1)(b-1)(c-1)$	$\frac{MS_{ABC}}{MS_E}$
Error	$abc(n-1)$	

- No. of possible interactions with multiple factors

Table: How many interactions are possible?

k-way	factor n	1	2	3	4	5	6	7	8	9	10	11
1-way		1	2	3	4	5	6	7	8	9	10	11
2-way			1	3	6	10	15	21	28	36	45	55
3-way				1	4	10	20	35	56	84	120	165
4-way					1	5	10	35	70	126	210	330
5-way						1	6	21	56	126	252	462
6-way							1	7	28	86	210	462
7-way								1	8	36	120	330
8-way									1	9	45	165
9-way										1	10	55
10-way											1	11

$$\text{Formula: } \frac{n!}{(n-k)!k!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} = \binom{n}{n-k} = \binom{n}{k}$$

n-way interaction: 1

(n-1)-way interaction: n

- Single Replicates: only one obs. per cell → zero df error

• cannot fit full model → no interaction!

→ can fit model with main effects only but error estimate will be biased ↑ as act. interaction effect $\in SS_{AB}$ in → test will be more conservative → OK in SSE

But if we expect interaction → should not do experiments with 1 obs. per cell. Alternatives:

a) Transformation: can get rid of interaction

b) Tukey One-Degree of Freedom Interaction:

→ introduce multiplicative interaction with new para

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{i(j)} \beta_j + \epsilon_{ijk}$$

5 Unbalanced Data

• Unbalanced data → no unique decomposition of SS anymore

→ add. observation will bias the estimate automatically in direction of larger group → will take all the SS → variance not the same anymore for diff. groups!

• Summation estimation needed (still using Principles of LS)

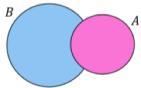
→ use model comparison approach with 3 diff. ways:

- Types of SS = like diff. Model comparisons

- Type I: Sequential sum of squares

Sequentially build up model

- $SS(A|1)$
- $SS(B|1, A)$
- $SS(AB|1, A, B)$
- Hence: Depends on ordering of factors!
- R: aov
- ↳ not good!



- Type II: Hierarchical / partially sequential approach

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

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



- Type III: Fully adjusted / marginal approach

(gen. recommended approach)

Control for all other terms

- $SS(A|1, B, AB)$ (interpretation?)
- $SS(B|1, A, AB)$ (interpretation?)
- $SS(AB|1, A, B)$
- R: drop1
- (make sure that you'll use options (contrasts = c("contr.sum", "contr.sum")))
- The default type for many other software packages.

need to specify contrasts for comparison

• For balanced data \rightarrow all the same

• For main effects only \rightarrow II & III the same

• Since SS for interaction of highest order is estimated last in all approaches \rightarrow always get same SS

• I & II both lead to tests that depend on sample size of the individual cell

• Type III test test unweighted mean \rightarrow does not depend on n's

→ Example: Type III SS

2) Recommended Approach:

use MSE of the full model for testing

```
## Type III sum of squares
drop1fit, scope = ~., test = "F", data = running)
## or: Anova(fit, type = "III", data = running)
```

```
## Single term deletions
##
## Model:
## y ~ method * drink
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 507 146
## method 1 1352 1859 236 176.21 < 2e-16
## drink 1 484 991 192 63.09 3.3e-11
## method:drink 1 29 536 148 3.79 0.056
```

Now the row method tests the null hypothesis

$$H_0: \frac{1}{2} \cdot \mu_{11} + \frac{1}{2} \cdot \mu_{12} = \frac{1}{2} \cdot \mu_{21} + \frac{1}{2} \cdot \mu_{22},$$

which is nothing else than saying that the "row-averages" of the μ_{ij} 's is constant.
Here, the actual sample sizes don't play a role anymore.

6. RCB: Complete Block Designs

• Restricted Randomization: assign treatments to ECs, independently of the other blocks

• Complete as every block sees every treatment

• Std case: $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$
→ also work > factor

Example:

Two-way +
Block

Source	df
Block	r - 1
A	a - 1
B	b - 1
AB	(a - 1) · (b - 1)
Error	(ab - 1) · (r - 1) ← "Leftovers"
Total	rab - 1 ← # observations - 1

→ use blocks to decrease var by making the blocks more homogeneous

- Relative Efficiency: Can compare Var of CRD with RCB

$$RE = \frac{\hat{\sigma}_{\text{RCB}}^2}{\hat{\sigma}_{\text{CRD}}^2} \stackrel{=}{\leftarrow} \text{using weighted avg: } w \cdot MS_{\text{RCB}} + (1-w) \cdot MS_{\text{CRD}}$$

↳ just MSE from the RCB

↳ can use this to identify N/r, i.e. how many more replicates would be needed on CRD setup to get same precision

• Quick check: See if Block helps explain

$$\frac{MS_{\text{Block}}}{MS_{\text{E}}} > 1 \Leftrightarrow \text{rel. efficiency} > 1$$

- If deleting a block / factor \rightarrow take df/SS to Replicates
→ need to calculate MS_E / F-Ratios

7. Multiple Block Factors

- Latin Squares: 2 block factors each with g levels
→ g^2 ECs
→ get $g \times g$ matrix / "square"
 - Each treatment appears once in each row and column
→ each treatment appears once for each level of the two factors
 - Analysis: usual main effect model

Tire position	Front Left	Front Right	Rear Left	Rear Right
Car 1	A	B	C	D
Car 2	B	C	D	A
Car 3	C	D	A	B
Car 4	D	A	B	C

→ Orthogonal Design

• ideally should use a random Latin Square

2) Greek Latin Squares: 3 block factors

- On top of Lat. Square setup, each latin letter occurs exactly once with each greek letter
- Analysis: analogous with 3 block factors

8. (Balanced) Incomplete Block Designs

• When we cannot see all treatments in each block

• Disconnected Design: occurs when we can build two groups of treatments / split them up

→ not possible to estimate all treatment differences any more

• BIBD: balanced when all treatment pairs occur together in the same block equally often ($>$ times)

↳ Var of treatment diff. will be constant

• Notation:

- g : # treatments
- b : # blocks
- k : # units per block ($k < g$)
- r : # replicates per treatment
- N : total ECs: $N = b \cdot k = g \cdot r$

→ due to balancedness get same precision/var for each of the treatment differences

- Unreduced BIBD: for every $k \in g$ get a BIBD

$$\binom{g!}{k!} = \frac{g!}{k!(g-k)!} \rightarrow \text{will give } b$$

- Nec. but not sufficient condition for BIBD

$$\frac{(k-1)}{r-1} \Rightarrow \# \text{ times a pair of treatments can appear}$$

↳ if $\frac{k-1}{r-1} \neq \text{integer} \rightarrow$ no BIBD exists

- Analysis of BIBD: model like std. block design

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

→ since not observe all treatment & block combinations normal estimates do NOT work

→ use Type III SS to test treatment effects adjusted for block effects

- Intrablock Analysis: TE while controlling BE

- Interblock Analysis: Compare different blocks

- RCB vs BCBD: BIBD can only beat RCB if

Var reduction (smaller var in the incomplete blocks) more than compensates for loss in efficiency
→ if var of the blocks is equal \rightarrow RCB is better

- Partially BIBD: when no BCBD is available, i.e. some treatment pair occur more often than others

→ can run experiment with fewer blocks

• same analysis as for BCBD

BUT not the same var per treatment difference

→ less efficient than a BCBD

9. Random Effects → want to learn about population

• Treatment effects are not fixed but random sampled from a population of pot. treatments

• Question: we are interested in Var / Expectation of the treatment population

- Random Effects Model (One-way ANOVA with RE)

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

Parameter

with

- α_i i.i.d. $\sim N(0, \sigma_{\alpha}^2)$
- ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$

effect of machine

In R: lmer

No need for constraints on α_i or σ^2 as already assume mean 0 in the distribution

- Properties of RE model:
 - $E[Y_{ij}] = \mu$
 - $\text{Var}[Y_{ij}] = \sigma_\alpha^2 + \sigma^2$ can split the Var components
 - $\text{Cor}(Y_{ik}, Y_{il}) = \begin{cases} 0 & i \neq k \\ \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2} & i = k, j \neq l \\ 1 & i = k, j = l \end{cases}$

Intraclass Correlation (ICC) \rightarrow corr. since from same population

\rightarrow if j is fixed then the diff. i are correlated as they come from the same population (e.g. same raters)

- ICC is large if $\hat{\sigma}_\alpha^2 \gg \hat{\sigma}^2$, i.e. obs. from the same group (e.g. machine) are very similar to each other but groups are quite different
- \rightarrow the more RE two obs. share \rightarrow bigger ICC

Comparison FE vs. RE

Term	Fixed effects model	Random effects model
α_i	fixed, unknown constant	α_i i.i.d. $\sim N(0, \sigma_\alpha^2)$
Side constraint on α_i	needed	not needed
$E[Y_{ij}]$	$\mu + \alpha_i$ <small>for one</small>	(μ) but $E[Y_{ij} \alpha_i] = \mu + \alpha_i$
$\text{Var}(Y_{ij})$	σ^2	$\sigma_\alpha^2 + \sigma^2$
$\text{Corr}(Y_{ij}, Y_{kl})$	$= 0 (j \neq l)$	$\begin{cases} 0 & i \neq k \\ \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2} & i = k, j \neq l \\ 1 & i = k, j = l \end{cases}$

Generic ANOVA table: two RE

Source	DF	EMS
A	$a - 1$	$\sigma^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$
B	$b - 1$	$\sigma^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$
AB	$(a - 1)(b - 1)$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
Error	$N - ab = ab(n - 1)$	σ^2

\rightarrow when testing RE need to be careful!

One-way RE ANOVA: can use same statistic as with FE model

\Rightarrow 2 (random) factors: need to compare neighboring EMS

\rightarrow when there are 3 or more RE \rightarrow could be that we can find NO exact F-test

\rightarrow test hypothesis for RE:

If $\sigma_\mu^2 = 0$, i.e. the variance of all possible population means is zero = all levels have same mean for response

\rightarrow Estimation: RE models are estimated via REML

- Example: One-way ANOVA w/ RE

Model: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, α_i i.i.d. $\sim N(0, \sigma_\alpha^2)$, ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$.

In R using the function `lmer` in the package `lmerTest` (or `lme4`).

> `fit.lme <- lmer(weight ~ 1 | sire, data = animals)`, group structure

Linear mixed model fit by REML [`lmerMod`]

Formula: weight ~ 1 | sire

Data: animals

REML criterion at convergence: 358.2

Scaled residuals:

Min 1Q Median 3Q Max

-1.9593 -0.7459 -0.1581 0.8143 1.9421

Random effects:

Groups Name Variance Std.Dev.

sire (Intercept) 116.7 10.81

Residual 463.8 21.54

Number of obs: 40, groups: sire, 5

Fixed effects:

Estimate Std. Error t value

(Intercept) 32.350 5.911 5.48

Check if model was interpreted correctly

$\hat{\mu}$ Estimate of population mean!

lmer test also shows p-value in R

σ^2 Estimate of population mean!

Meaning: a random effect across sire

convert to σ^2 / n with this

convert to σ^2 / 5 with this

convert to σ^2 / 1 with this

convert to $\sigma^2</$

11. Mixed Effects Model (FE + RE mixed)

$$Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk(i)} + \epsilon_{l(ijk)}$$

where

- α_i are the **fixed effects** of background
- $\beta_{j(i)}$ are the **random effects** of rater (within background)
- γ_k are the **fixed effects** of cheese type
- $(\alpha\gamma)_{ik}$ is the (**fixed**) interaction effect between **background** and **cheese type**
- $(\beta\gamma)_{jk(i)}$ is the (**random**) interaction between **rater** and **cheese type**.

Interaction between RE & FE \rightarrow RE

Size of p-values: when we use pure FE instead of RE
 \rightarrow p-values should be larger since less uncertainty

- **Check:**
 - a) Normality of all errors \rightarrow QQ plots
 - b) Homoscedasticity \rightarrow TA plot
 - c) correct interpretation of grouping structure

- **Analysis:** (in R: using lmer)

```
library(lmerTest)
fit <- lmer(y ~ background * choc + (1 | rater:background) +
             (1 | rater:background:choc), data = chocolate)
```

- Anova (fit) \rightarrow get table with p-value for FE
- Summary (fit) \rightarrow get estimates for FE
- confint (fit, oldnames = False) \rightarrow get approx CI for FE wef. + Var components
 \hookrightarrow need to know encoding scheme for interpretation
- usually not interested in RE estimates but more how they change meaning of FE
- FE become population (e.g. machine effect across all possible workers)

```
anova(fit)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## background    263   263     1    18 27.61 5.4e-05
## choc          4219  1406     3    54 147.74 < 2e-16
## background:choc 64    21     3    54  2.24  0.094
```

We see that the interaction is not significant but there is a significant effect of background and chocolate type. This is what we basically already observed in the interaction plot. There, the profiles were quite parallel, but raters with urban background rated higher on average than those with rural background. In addition, there was a clear difference between different chocolate types.

Can we get an intuitive idea about the denominator degrees of freedom that are being used in these tests?

- The main effect of background can be thought of as a two-sample t-test with two groups having 10 observations each. Think of taking one (average) value per rater. Hence, we get $2 \cdot 10 - 2 = 18$ degrees of freedom.
- As we allow for a rater specific chocolate type preference, we have to check whether the effect of chocolate is substantially larger than this rater specific variation. The rater specific variation can be thought of as the interaction between rater and chocolate type. As rater is nested in background, rater has 9 degrees of freedom in each background group. Hence, the interaction has $((9+9) \cdot 3 = 54)$ degrees of freedom.
- The same argument as above holds true for the interaction between background and chocolate type.

\Rightarrow if we use a FE model instead of RE, in general, p-values will be smaller \rightarrow underestimate the uncertainty!

- **Example:** 1 RE (per worker) + FE (machine) + random interaction

```
options(contrasts = c("contr.treatment", "contr.poly"))
library(lmerTest)
fit <- lmer(score ~ Machine + (1 | Worker) + (1 | Worker:Machine),
            data = Machines)
anova(fit)

## Type III Analysis of Variance Table with Satterthwaite's method
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## Machine    38.1    19     2    10 20.6 0.00029
## Worker:Machine 13.909  3.730
## Worker      (Intercept) 22.858  4.781
## Residual       0.925  0.962
## Number of obs: 54, groups: Worker:Machine, 18; Worker, 6
##
## Fixed effects:
##             Estimate Std. Error df t value Pr(>|t|)
## (Intercept)  52.36     2.49  8.52 21.06 1.2e-08
## MachineB     7.97     2.18 10.00   3.66  0.0044
## MachineC    13.92     2.18 10.00   6.39 7.9e-05
## ...
## Random effects:
## Groups      Name        Variance Std.Dev.
## Worker:Machine (Intercept) 13.909  3.730
## Worker      (Intercept) 22.858  4.781
## Residual           0.925  0.962
## Number of obs: 54, groups: Worker:Machine, 18; Worker, 6
## ...
## Correlation of Fixed Effects:
## (Intercept) MachineB MachineC
## MachineB   -0.000
## MachineC   -0.000
```

- **Need to check:**

- assumptions on the error
- assumptions on the RE
- assumptions on the random interaction

- **Special case:** when we have certainly NO interaction estimation will be same with usual ANOVA

- We could also interpret subject as a **fixed block factor** ANOVA.

```
> fit2 <- aov(effort ~ Type + Subject, data = ergostool)
> summary(fit2)
  Df Sum Sq Mean Sq F value Pr(>F)
Type  3  81.19  27.065 22.356 3.93e-07 ***
Subject 8  66.50  8.313  6.866 0.000106 ***
Residuals 24 29.06  1.211
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> coef(fit2)
(Intercept) TypeT2 TypeT3 TypeT4 SubjectS SubjectT4 SubjectT9
6.555556  3.888888  2.222222  0.666667 0.2500000 1.0000000 1.7500000
Subject6 Subject3 Subject7 Subject1 Subject2 2.0000000 2.5000000 4.0000000 4.0000000
```

\rightarrow get same F-value / p-value

\rightarrow we just model subject as a fixed block factor

12. Split Plot Designs

- for when some factors are more costly to vary than others
- 2 (or more) experiments on 1 with different ECs or different size
- **Setups:** \rightarrow two independent randomizations

- **Whole/Plot Plots:** the large EC \rightarrow get the whole-plot factor treatment
 \rightarrow like randomized blocks for split-plots

- **Split/Sub Plots:** The small EC \rightarrow get the split-plot factor treatment

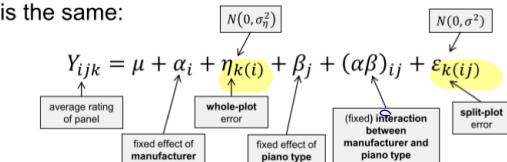
\rightarrow also a restricted randomization

\rightarrow split plots confound whole plots with incomplete blocks

- To identify: follow the path of randomization

- **Model:** Mixed model formulation with two diff. errors

The model is the same:



We get the following output:

```
> fit.lme <- lmer(Y ~ B + V * N + (1 | B:V), data = oats)
> anova(fit.lme)
```

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
B	4675.0	935.0	5	10	5.280	0.01244 *
V	526.1	263.0	2	10	1.485	0.27239
N	20020.5	6673.5	3	45	37.686	2.458e-12 ***
V:N	321.8	53.6	6	45	0.303	0.93220

\hookrightarrow obs. from **some** whole plot share **some** whole-plot error \rightarrow Not independent

\hookrightarrow not a lot of obs. on whole-plot level with this type of design

\hookrightarrow not a lot of df (Den df \rightarrow just MSE on whole plot level) left on the whole plot level

\hookrightarrow less powerful tests (compared to split-plot level)

- **Checks:**

- a) visual
- check normality of both error terms \rightarrow QQ Plot

More complicated designs:

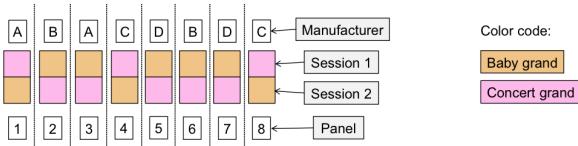
- a) Split-block + add block: RCB for whole plot + Split-Plot
- b) Split-Split Block \rightarrow just with 3 levels
- c) Can also have more than 1 factor, e.g. two-factorial on whole-plot level

→ if we estimate without taking into account special structure → overly optimistic estimates
→ too small p-values

- Example: Pianos (Time as split plot)

Example II: Pianos

- The whole plots are the 8 panels.
- The whole-plot factor is the manufacturer.
- The split plots are the 2 session time-slots.
time can be split plot
- The split-plot factor is the piano type (baby vs. concert grand).



The model is the same:

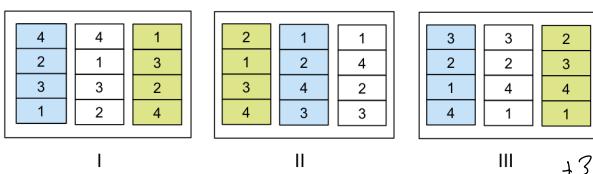
$$Y_{ijk} = \mu + \alpha_i + \eta_{k(j)} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{k(ij)}$$

average rating of panel
fixed effect of manufacturer
whole-plot error
fixed effect of piano type
fixed interaction between manufacturer and piano type
split-plot error

- Example: Oats (Split-Plot + RCB)

Overview of data:

- 6 different blocks (B) → have one more layer
- 3 different varieties (V)
- 4 different nitrogen treatments (N)
- Response (Y): Yields (in 1/4 lbs per sub-plot, each of area $\frac{1}{80}$ acre).



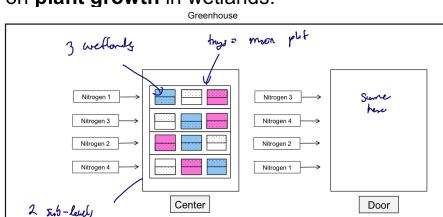
We get the following output:

```
> fit.lme <- lmer(Y ~ B + V * N + (1 | B:V), data = oats)
> anova(fit.lme)
Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom
Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
B 4675.0 935.0 5 10 5.280 0.01244 ** low power
V 526.1 263.0 2 10 1.485 0.27239
N 20020.5 6673.5 3 45 37.686 2.458e-12 *** → only evidence here
V:N 321.8 53.6 6 45 0.303 0.93200
```

- Example: Wetlands (Split-Split Plot)

The experiment studied the effect of

- nitrogen (4 levels of nitrogen)
 - weed (3 levels)
 - clipping treatments (2 levels: clipping / no clipping)
- on plant growth in wetlands.



Example IV: Weed Biomass in Wetlands

We use the following model:

```
> fit <- lmer(pct.nonweed.biomass ~ table + (1 | tray) + weed * nitrogen * clipping + (1 | wetland),
+             data = wetland)
> anova(fit) → for global F-test
```

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
table	0.00	0.00	1	3	0.538	0.7213
weed	1186.82	593.41	3	8	555.4531	2.61e-09 ***
nitrogen	36.73	12.24	3	11	46.10	0.03765 *
clipping	125.45	125.45	1	12	117.4290	1.49e-07 ***
weed:nitrogen	157.57	26.26	6	8	24.5814	9.66e-05 ***
weed:clipping	0.25	0.12	2	12	0.1149	0.89246
nitrogen:clipping	0.74	0.25	3	12	0.2293	0.87419
weed:nitrogen:clipping	4.82	0.80	6	12	0.7514	0.62033

split-plot error (tray is always unique)
↓ for subplot
↓ for wetland
↳ no need for interaction bc also uniquely identified (can eval.)

no evidence
↓ due to test - from the error!

ANOVA Table for Split-Plot Design

Example setup:

- 90 subjects
 - each takes one pizza
 - whole-plot factor in 6 packages (split-plot-factor)
- $\Rightarrow 90 \times 6 \text{ observations} = 540$

The model considered is

$$Y_{ijk} = \mu + \beta_j + \eta_{k(j)} + \alpha_i + (\alpha\beta)_{ij} + \varepsilon_{k(ij)},$$

where $\eta_{k(j)}$ is the whole-plot error (per person). The ANOVA skeleton is given by

Plot level	Source	df	MS	F
Whole plots	B	2	MS _B	MS _B / MS _η
	Residual	87	MS _η	
Split plots	A	5	MS _A	MS _A /MS _E
	AB	10	MS _{AB}	MS _{AB} /MS _E
	Residual	435	MS _E	
	Total	539 (540 - 1)		

Testing

- Whole-Plot Factor (B): use MS_η (based on the whole-plot error)
- Split-Plots (and interaction): use MS_E (as usual)

→ if we do not analyze as split-plot design cannot distinguish MS_η from MS_E

↪ Comparison on split-plot factor will be more precise than on whole-plot factor

→ effect on Test: Wrong analysis

Plot level	Source	df	Effect on F
N/A	B	b-1	too large (big a lot)
N/A	A	a-1	too small
N/A	AB	(a-1)(b-1)	too small

13. Add-on Testing: Power

Confidence Interval:

$$P(T_e \leq \theta \leq T_u) = \gamma \quad \forall \theta \in \mathbb{R}$$

Tests: quantify evidence against H_0 (using p-value)

P-value: $P(T \text{ obtaining equal or more extreme result } T) \text{ than observed } (H_0 \text{ is true})$

Type I: $P(T \geq t_{\text{crit}} | H_0) = \alpha$ (sign. level)

Type II: $P(T \leq t_{\text{crit}} | H_1) = \beta$

Power: $P(\text{reject } H_0 \text{ given that a certain setting in } H_1 \text{ holds}) = 1 - \beta$

Power curve: shows $P_{1-\beta}$ (power) for diff. levels of θ

Power level: ideally want power $\geq 80\%$

↪ "low power": chances are high that we will NOT get a sign. result, even though H_1 holds

Influencing Factors:

a) Design of the experiment (balanced, unbalanced...)

b) Sign. level α (taken as given)

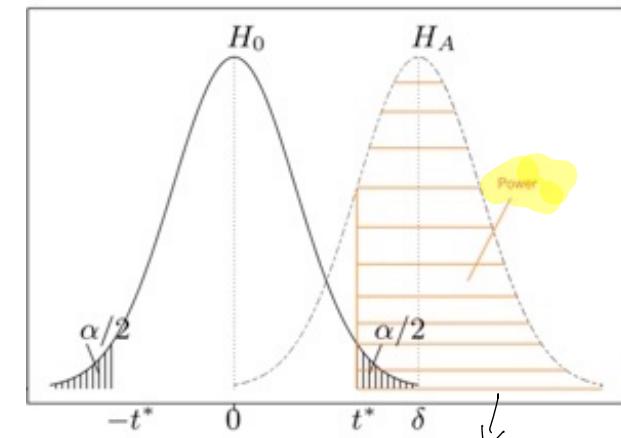
↪ trade-off with β

c) Parameter setting under H_1

d) Sample size $n \rightarrow \text{or not} \rightarrow (1-\beta)T^*$

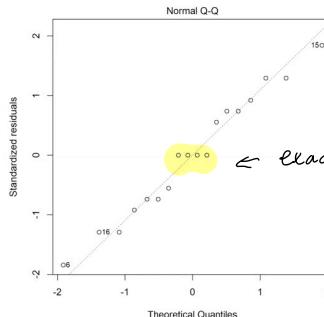
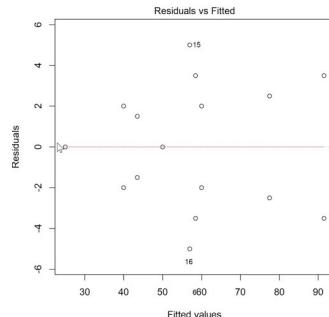
↪ choose n st the desired power level!

→ can simulate power of an experiment easily before doing it



i.e. the prob. of getting a sign. result, given that H_1 is true → just a "thought experiment"

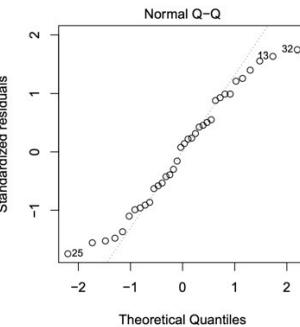
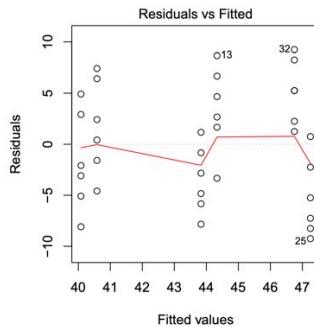
- Add. Plots: Checking Model Assumptions



exactly same residuals
→ not normal

CK

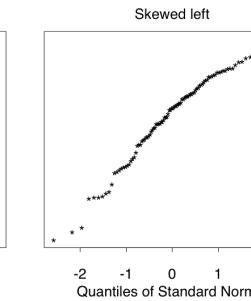
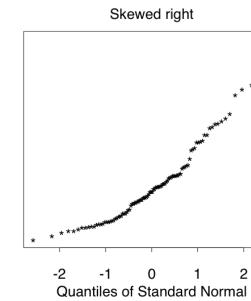
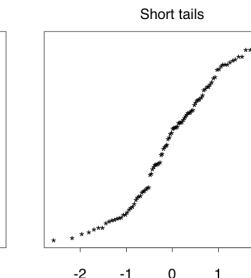
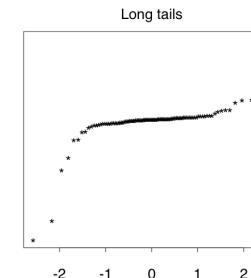
CK



Not good
→ mean ≠ 0

not sure / Ok
→ light tails

QQ - Plots - Violations of Normality



- TA Plots: Violation of Homoscedasticity

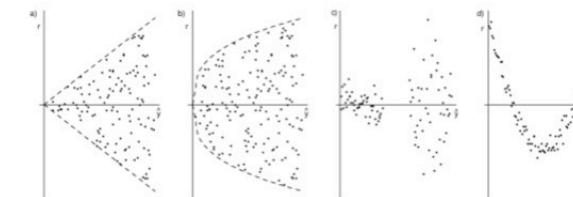


Figure 1.18: a) Linear increase in standard deviation, b) non-linear increase in standard deviation, c) 2 groups with differing variance, d) missing quadratic term.