Designs with two factors

Last time: Completely randomized designs

▶ Two equivalent parametrizations:

1. Group means parametrization:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

2. Sum-to-zero parametrization:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad \sum_{i=1}^{a} \tau_i = 0, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

▶ $i \in \{1, \dots a\}$ indexes group membership
▶ $j \in \{1, \dots, n_i\}$ indexes observations within a group

Types of factors: treatments vs blocks

Factors (variables) can be of two types

- **Treatments:** We are interested in how they affect the response. We can control how we assign them to the experimental units.

- **Blocks** We are not interested in how they affect the response, but suspect that they affect the response. We can control them when we're performing the experiments, but often can't control them outside of the experimental design.

Additive vs interaction models

- **Additive model:** The effects of the factors on the response don't depend on levels of other factors

- **Interaction model:** The effects of the factors on the response can depend on levels of other factors

Complete and balanced designs

- ▶ The design is *complete* if all the combinations of the factors are observed

- ▶ The design is *balanced* if all the combinations of the factors are observed the same number of times: r

- ▶ The models we will see in this part of the course assume that the designs are **complete and balanced**

Example: Painting line

▶ An engineer working at a painting line wants to know if different types of paints have different average drying times

▶ They will perform the experiments on four different days

▶ The engineers think that the weather may affect the outcome: that is, the day may affect the outcome

▶ The treatment is the type of paint, the block is the day

Balanced complete randomized block designs with replicates

- ▶ We assign experimental units randomly to blocks and treatments
- ▶ The block has b levels
- ▶ The treatment has a levels
- ▶ We collect r observations for each combination of block and treatment: there are r replicates
- ▶ Total sample size is $N = rba$

Terminology

- The term "balanced complete randomized block designs" usually refers to the case $r = 1$, which is the most common

- Designs with $r > 1$ are often referred to as "balanced complete randomized block designs with replicates"

Additive vs interaction models

- Additive models assume that the treatment effect does not depend on the block (or the other treatment, if there are two treatments)

- Block designs are often additive; this assumption comes from prior knowledge of the experimenters

- If $r = 1$, we don't have enough data to properly test if an interaction is needed

- If $r > 1$ we can test if an interaction is needed

Example: Painting line

- Day (block): $b = 4$ days
- Type of paint (treatment): $a = 3$ types
- $r = 1$ replicate
- $N = rba = 12$ observations

	Day 1	Day 2	Day 3	Day 4
Type A	2.60	2.10	3.5	2.65
Type B	3.25	2.65	3.8	3.50
Type C	2.60	2.05	3.1	2.62

Additive model

$$Y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}, \qquad \sum_{i=1}^{a} \tau_i = \sum_{j=1}^{b} \beta_j = 0, \qquad \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

- $i \in \{1, ..., a\}$ indexes treatments, $j \in \{1, ..., b\}$ indexes blocks, and $k \in \{1, ..., r\}$ indexes replicates
- As before, we assume that the errors are independent and normal, and that the variance σ^2 doesn't depend on the treatment or block
- We also assume an additive relationship between treatment and block

Additivity

- Treatment effect doesn't depend on blocks; block effect doesn't depend on treatment

- **Example:** $a = 2$ treatments and $b = 2$ blocks, table with expected values

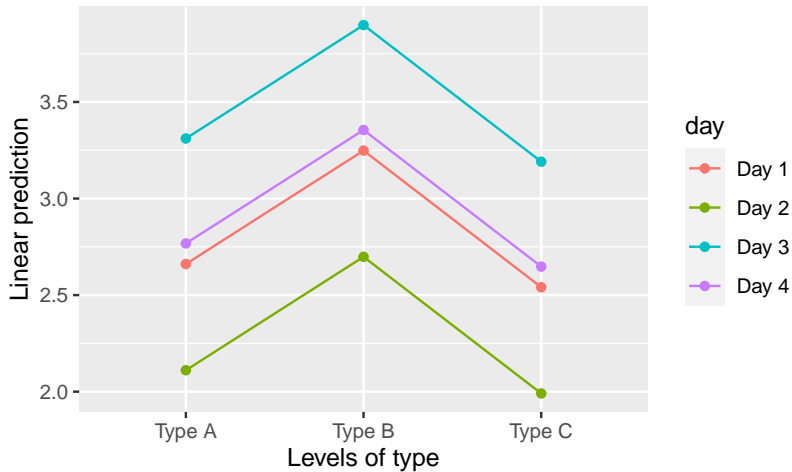|	Block 1	Block 2
Treat 1	$\mu + \tau_1 + \beta_1$	$\mu + \tau_1 + \beta_1$
Treat 2	$\mu + \tau_2 + \beta_2$	$\mu + \tau_2 + \beta_2$

Average treatment effect in block 1: $\mathbb{E}(y_{11k} - y_{21k'}) = \tau_1 - \tau_2$
Average treatment effect in block 2: $\mathbb{E}(y_{12k} - y_{22k'}) = \tau_1 - \tau_2$
Average block effect in treat 1: $\mathbb{E}(y_{11k} - y_{12k'}) = \beta_1 - \beta_2$
Average block effect in treat 2: $\mathbb{E}(y_{21k} - y_{22k'}) = \beta_1 - \beta_2$

Example: estimated effects with additive model

Estimating parameters: Additive design

▶ Let

$$\text{total mean} = \overline{Y} = \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} Y_{ijk}$$

$$\text{treat } i \text{ mean} = \overline{Y}_{i.} = \frac{1}{br} \sum_{j=1}^{b} \sum_{k=1}^{r} Y_{ijk}$$

$$\text{block } j \text{ mean} = \overline{Y}_{.j} = \frac{1}{ar} \sum_{i=1}^{a} \sum_{k=1}^{r} Y_{ijk}$$

▶ Point estimates

$$\widehat{\mu} = \overline{Y}, \qquad \widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}, \qquad \widehat{\beta}_j = \overline{Y}_{.j} - \overline{Y},$$

and

$$\widehat{\sigma^2} = \frac{1}{N-1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j)^2$$

Example: Reading in data

```r
mat = matrix(c(2.60, 2.10, 3.50, 2.65,
               3.25, 2.65, 3.80, 3.5,
               2.60, 2.05, 3.10, 2.62),
       nrow = 3,
       ncol = 4,
       byrow = TRUE)
rownames(mat) = c("Type A", "Type B", "Type C")
colnames(mat) = c("Day 1", "Day 2", "Day 3", "Day 4")
```

Example: Converting into convenient (long) format

```
library(tidyverse)
df = as.data.frame(mat)
df$type = rownames(df)
df = df %>% pivot_longer(!type,
                         names_to = "day",
                         values_to = "outcome")
df$type = factor(df$type); df$day = factor(df$day)
```

Example: model with sum-to-zero constraint

```
# sum-to-zero
options(contrasts = c("contr.sum", "contr.poly"))
# fit model
mod = aov(outcome ~ day + type, data = df)
# point estimates
dummy.coef(mod)

## Full coefficients are
##
## (Intercept):        2.868333
## day:              Day 1       Day 2       Day 3       Day 4
##               -0.05166667 -0.60166667  0.59833333  0.05500000
## type:             Type A      Type B      Type C
##                -0.1558333   0.4316667  -0.2758333
```
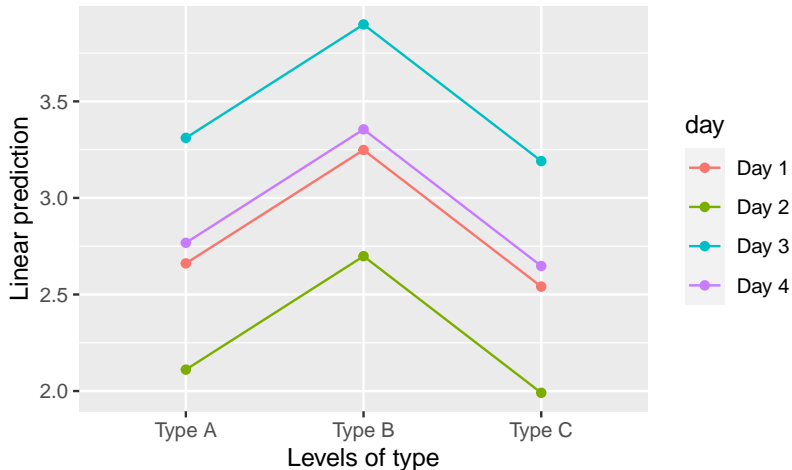
Example: 95% confidence intervals

```
##                  2.5 %      97.5 %
## (Intercept)  2.7760320  2.96063468
## day1        -0.2115373  0.10820396
## day2        -0.7615373 -0.44179604
## day3         0.4384627  0.75820396
## type1       -0.2863672 -0.02529951
## type2        0.3011328  0.56220049
```

No intervals for $\tau_a = -\sum_{i=1}^{a-1} \tau_i$ and $\beta_b = -\sum_{j=1}^{b-1} \beta_j$

Effect plots

```
library(emmeans)
emmip(mod,  day ~ type)
```

Global F tests

Now we have two F tests

▶ One for τ_i

$$H_{0,\tau} : \tau_i = 0, \text{ for all } i \in \{1, ..., a\}$$
$$H_{1,\tau} : \text{ at least one } \tau_i \neq 0$$

▶ Another one for β_j:

$$H_{0,\beta} : \beta_j = 0, \text{ for all } j \in \{1, ..., b\}$$
$$H_{1,\beta} : \text{ at least one } \beta_j \neq 0.$$

Sums of squares

$$\mathrm{SS}_{\mathrm{total}} = \mathrm{SS}_{\mathrm{treat}} + \mathrm{SS}_{\mathrm{block}} + \mathrm{SS}_{\mathrm{error}},$$

where

$$\mathrm{SS}_{\mathrm{total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (\overline{Y}_{ijk} - \overline{Y})^2$$

$$\mathrm{SS}_{\mathrm{treat}} = br \sum_{i=1}^{a} (\overline{Y}_{i.} - \overline{Y})^2$$

$$\mathrm{SS}_{\mathrm{block}} = ar \sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y})^2$$

$$\mathrm{SS}_{\mathrm{error}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j)^2$$

Under $H_{0,\tau}$, $\overline{Y}_{i.} \approx \overline{Y}$ and $\mathrm{SS}_{\mathrm{treat}}$ will be small; same story with $H_{0,\beta}$ and $\mathrm{SS}_{\mathrm{block}}$

ANOVA table

There are two F statistics: one for the treatment and another one for the block

	df	SS	MS	F	p
Treat	$a-1$	SS_{treat}	$\text{MS}_{\text{treat}} = \frac{\text{SS}_{\text{treat}}}{a-1}$	$f_{\tau,\text{obs}} = \frac{\text{MS}_{\text{treat}}}{\text{MS}_{\text{error}}}$	p_τ
Block	$b-1$	SS_{block}	$\text{MS}_{\text{block}} = \frac{\text{SS}_{\text{block}}}{b-1}$	$f_{\beta,\text{obs}} = \frac{\text{MS}_{\text{block}}}{\text{MS}_{\text{error}}}$	p_β
Error	ν	SS_{error}	$\text{MS}_{\text{error}} = \frac{\text{SS}_{\text{error}}}{\nu}$		

where $\nu = N - a - b + 1$,

$$p_\tau = P(F_{a-1,\nu} > f_{\tau,\text{obs}}), \qquad p_\beta = P(F_{b-1,\nu} > f_{\beta,\text{obs}}),$$

and $F_{\alpha,\nu}$ is notation for the F-distribution with α and ν degrees of freedom.

Example: ANOVA table

```
summary(mod)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## day          3 2.1771  0.7257   42.50 0.000195 ***
## type         2 1.1468  0.5734   33.58 0.000552 ***
## Residuals    6 0.1024  0.0171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Block and treatment effect are significant at the $\alpha = 0.05$
significance level

Checking model assumptions with residuals

▶ We can write the model as

$$\varepsilon_{ijk} = Y_{ijk} - (\mu + \tau_i + \beta_j) \overset{\mathrm{ind}}{\sim} N(0, \sigma^2)$$

▶ If the model is correct,

$$\frac{Y_{ijk} - (\mu + \tau_i + \beta_j)}{\sigma} \sim N(0, 1)$$

▶ The errors ε_{ijk} have a variance that doesn't depend on the expected value $\mu + \tau_i + \beta_j$

Residuals

- We don't observe ε_{ijk}.

- However, since we can estimate μ, τ_i, β_j, and σ^2, we can find the residuals

$$e_{ijk} = Y_{ijk} - (\widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j) \approx N(0, \widehat{\sigma^2}),$$

so

$$e_{ijk}/\widehat{\sigma} \approx N(0, 1)$$

and the residuals e_{ijk} should have a variance that doesn't depend much on the fitted values $\widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j$

Checking model assumptions, in practice

▶ We can check the assumption of normality of errors by looking at a qq-plot comparing the quantiles of $N(0,1)$ to the empirical quantiles of $e_{ijk}/\widehat{\sigma}$

▶ We can check the hypothesis of equality of variances by looking at a plot of fitted values vs residuals: if the assumption is satisfied, the variance of the residuals shouldn't depend much on the fitted values

▶ Checking independence is hard unless we have a variable that allows us to check for space- or time-dependence, for example

▶ If $r = 1$, we can't test if we need an interaction. If $r > 1$, we can check whether an interaction term is needed or not through a hypothesis test.
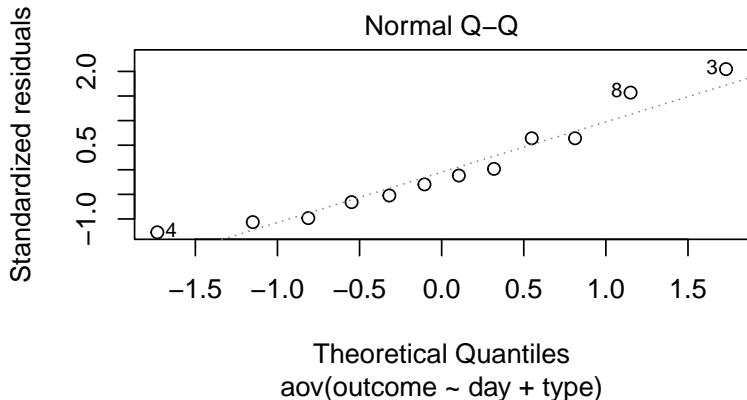
Example: checking equality of variances

```
plot(mod, which = 1)
```


Residuals vs Fitted

aov(outcome ~ day + type)
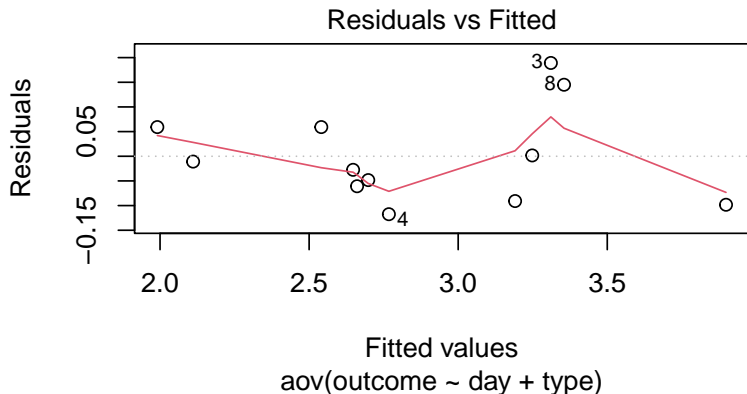
Example: checking normality of residuals

```
plot(mod, which = 2)
```


Normal Q–Q

Standardized residuals vs Theoretical Quantiles

aov(outcome ~ day + type)

Pairwise comparisons

▶ The global *F* test tells us that there are differences between treatments, but doesn't tell us where the differences are

▶ We can do pairwise tests to compare treatments; however, we have to be careful and control the type I error rate appropriately

▶ We can do that with `TukeyHSD`, as we did with completely randomized designs

Example: `TukeyHSD`

```
TukeyHSD(mod, which = "type")

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = outcome ~ day + type, data = df)
##
## $type
##                  diff        lwr        upr      p adj
## Type B-Type A  0.5875  0.3039958  0.8710042 0.0017250
## Type C-Type A -0.1200 -0.4035042  0.1635042 0.4459693
## Type C-Type B -0.7075 -0.9910042 -0.4239958 0.0006345
```

Significant differences between types A and B and types B and C at
the $\alpha = 0.05$ significance level

Example: Batteries (Montgomery)

▶ A company is producing batteries that need to work at different temperatures and for different types of devices

▶ Battery life is suspected to depend on the temperature and the material that the device is made of (type A, B, or C)

▶ The company wants to know how temperature and the device affect battery life; they also want to know if there is an interaction between temperature and type of device

▶ There are 36 observations: **3 replicates** for each combination of type and temperature. Therefore, we can test if we need an interaction.

▶ If there is an interaction, it means that the effect of temperature depends on the type of device

Plotting the data

Does the effect of the temperature depend on the type of device?
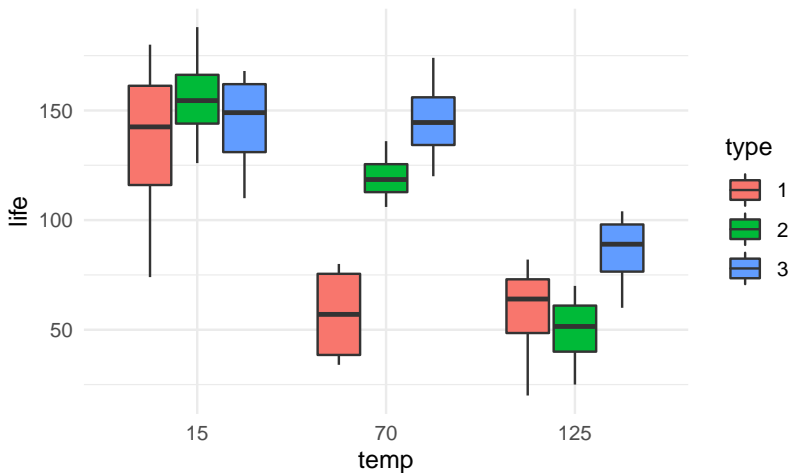
Model with interaction

$$Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

with sum-to-zero constraints

$$\sum_{i=1}^{a} \tau_i = \sum_{j=1}^{b} \gamma_j = 0, \qquad \sum_{i=1}^{a} (\tau\gamma)_{ij} = \sum_{j=1}^{b} (\tau\gamma)_{ij} = 0$$

▶ $i \in \{1, ..., a\}$ indexes the first treatment, $j \in \{1, ..., b\}$ indexes the second treatment, and $k \in \{2, ..., r\}$ indexes replicates (important: $r > 1$)

▶ The errors are independent and normal, the variance σ^2 doesn't depend on the treatment

▶ $(\tau\gamma)_{ij}$ is the interaction term

Point estimation

Let

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} Y_{ijk}, \qquad \overline{Y}_{ij} = \frac{1}{r} \sum_{k=1}^{r} Y_{ijk}$$

$$\overline{Y}_{i.} = \frac{1}{br} \sum_{j=1}^{b} \sum_{k=1}^{r} Y_{ijk}, \qquad \overline{Y}_{.j} = \frac{1}{ar} \sum_{i=1}^{a} \sum_{k=1}^{r} Y_{ijk}$$

Then,

$$\widehat{\mu} = \overline{Y}, \qquad \widehat{(\tau\gamma)}_{ij} = \overline{Y}_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\gamma}_j$$

$$\widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}, \qquad \widehat{\gamma}_j = \overline{Y}_{.j} - \overline{Y}$$

and

$$\widehat{\sigma^2} = \frac{1}{N-1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\gamma}_j - \widehat{(\tau\gamma)}_{ij})^2$$

Example: point estimation

Here's how to find estimates and confidence intervals in R

```
battery = read.csv("https://vicpena.github.io/doe/battery.csv")
battery$type = factor(battery$type)
battery$temp = factor(battery$temp)
options(contrasts = c("contr.sum", "contr.poly"))
mod = aov(life ~ type*temp, data = battery)
dummy.coef(mod)
confint(mod)
```

Not printing the results here for concreteness

# Interpreting coefficients with interactions

▶ Interpreting the main effects τ_i and γ_j is difficult when there are interactions, since the effect of a treatment depends on the levels of the other treatment. . .

▶ The main effects τ_i and γ_j are "average effects", in the sense that

$$\mu + \tau_i = \frac{1}{b} \sum_{j=1}^{b} \mathbb{E}(Y_{ijk}), \qquad \mu + \gamma_j = \frac{1}{a} \sum_{i=1}^{a} \mathbb{E}(Y_{ijk})$$

▶ In practice, most of us just look at interaction plots to avoid confusing ourselves

Example: Interaction plot

```
emmip(mod, type ~ temp)
```
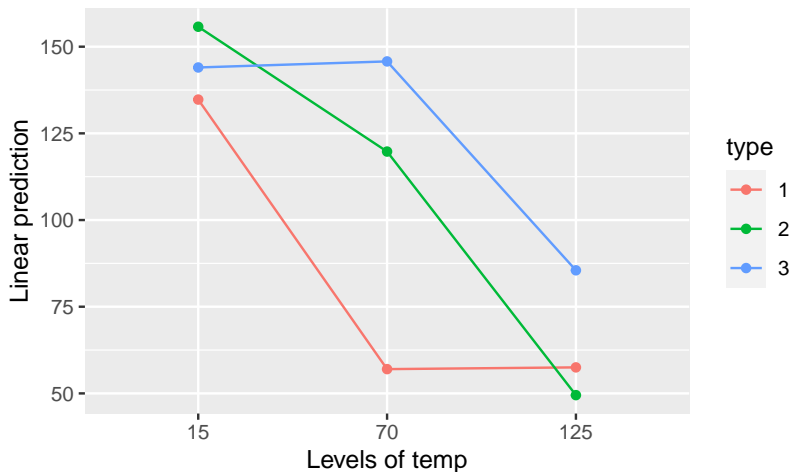
Global F tests

We have three F tests

▶ One for τ_i

$$H_{0,\tau} : \tau_i = 0, \text{ for all } i$$
$$H_{1,\tau} : \text{ at least one } \tau_i \neq 0$$

▶ One for γ_j

$$H_{0,\gamma} : \gamma_j = 0, \text{ for all } j$$
$$H_{1,\gamma} : \text{ at least one } \gamma_j \neq 0$$

▶ And another one for $(\tau\gamma)_{ij}$

$$H_{0,\tau\gamma} : (\tau\gamma)_{ij} = 0, \text{ for all } i, j$$
$$H_{1,\tau\gamma} : \text{ at least one } (\tau\gamma)_{ij} \neq 0$$

Sums of squares

$$\mathrm{SS}_{\mathrm{total}} = \mathrm{SS}_\tau + \mathrm{SS}_\gamma + \mathrm{SS}_{\tau\gamma} + \mathrm{SS}_{\mathrm{error}},$$

where

$$\mathrm{SS}_{\mathrm{total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (\overline{Y}_{.j} - \overline{Y})^2$$

$$\mathrm{SS}_\tau = br \sum_{i=1}^{a} (\overline{Y}_{i.} - \overline{Y})^2$$

$$\mathrm{SS}_\gamma = ar \sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y})^2$$

$$\mathrm{SS}_{\tau\gamma} = r \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{Y}_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y})^2$$

$$\mathrm{SS}_{\mathrm{error}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \overline{Y}_{ij})^2$$

ANOVA table

	df	SS	MS	F	p
τ	$a-1$	SS_τ	$\mathrm{MS}_\tau = \frac{\mathrm{SS}_\tau}{a-1}$	$f_{\tau,\mathrm{obs}} = \frac{\mathrm{MS}_\tau}{\mathrm{MS}_{\mathrm{error}}}$	p_τ
γ	$b-1$	SS_γ	$\mathrm{MS}_\gamma = \frac{\mathrm{SS}_\gamma}{b-1}$	$f_{\gamma,\mathrm{obs}} = \frac{\mathrm{MS}_\gamma}{\mathrm{MS}_{\mathrm{error}}}$	p_γ
$(\tau\gamma)$	$(a-1)(b-1)$	$\mathrm{SS}_{\tau\gamma}$	$\mathrm{MS}_{\tau\gamma} = \frac{\mathrm{SS}_{\tau\gamma}}{(a-1)(b-1)}$	$f_{\tau\gamma,\mathrm{obs}} = \frac{\mathrm{MS}_{\tau\gamma}}{\mathrm{MS}_{\mathrm{error}}}$	$p_{\tau\gamma}$
Error	ν	$\mathrm{SS}_{\mathrm{error}}$	$\mathrm{MS}_{\mathrm{error}} = \frac{\mathrm{SS}_{\mathrm{error}}}{\nu}$		

▶ If the interaction is not significant, we can fit an additive model that excludes it

Example: ANOVA table

```
summary(mod)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## type          2  10728    5364   7.960  0.00192 **
## temp          2  39115   19557  29.020 1.88e-07 ***
## type:temp     4   9670    2417   3.587  0.01803 *
## Residuals    27  18196     674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```
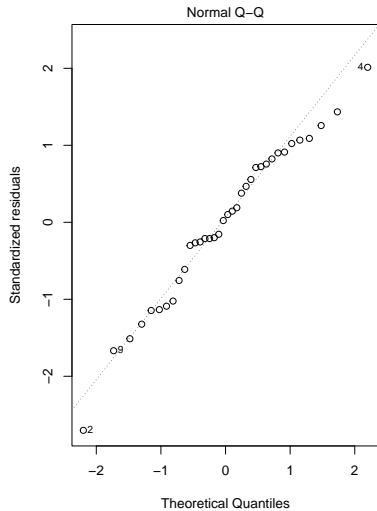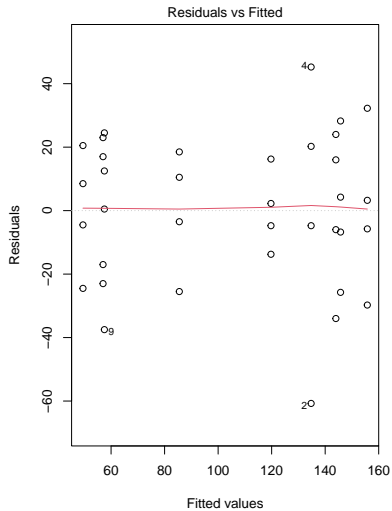
Significant treatment effects and interaction at $\alpha = 0.05$.

Assumptions

▶ We still have assumptions of independence, normality, and equality of variances

▶ We can check them as we did with the blocked design

▶ Our model has an interaction: if it isn't significant, we can drop it from our model and fit an additive model

Example: checking assumptions

Pairwise comparisons

- As before, global tests don't tell us where the differences are
- For that, we can still use `TukeyHSD`

Two factors: summary

- Factors can be treatments or blocks
- Models with two factors can be either additive or have an interaction
- Models with a treatment and a block are often assumed to be additive, but they don't have to be: if you have $r > 1$, you can test whether there is a significant interaction

More than two factors

- ▶ What we have seen can be extended to models that have more than two factors

- ▶ In that case, we will have higher-order interactions: interactions with 3, 4, or more variables

- ▶ In practice, we usually don't consider interactions of order 4 or higher because they're hard to interpret...

Example: 3 factor model

▶ Maxwell and Delaney (2003) present a study whose goal was to quantify the effect of medication, psychological counseling, and dieting on blood pressure

▶ Response: blood pressure

▶ Three types of medication: drug X, drug Y, drug Z

▶ Psychological counseling: yes or no

▶ Special diet: yes or no

▶ Sample size: 72 subjects, $r = 6$ replicates for each treatment combination

▶ Balanced, complete design with replicates

Example: looking at the data

```
df = read.csv("http://vicpena.github.io/doe/threeway.csv")
ggplot(df) +
  aes(x = drug, y = bp) +
  geom_point() +
  facet_grid(feed ~ diet)
```
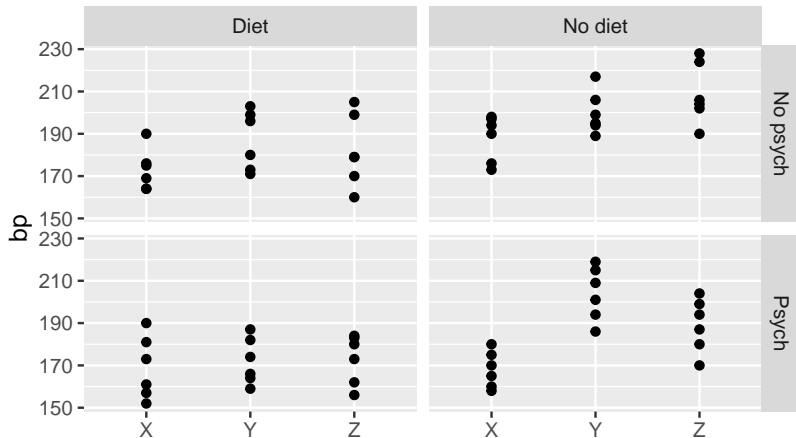
Three way ANOVA model

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl},$$

where $\varepsilon_{ijkl} \overset{\text{ind}}{\sim} N(0, \sigma^2)$.

- ▶ Global mean: μ
- ▶ Main effects: $\alpha_i, \beta_j, \gamma_k$
- ▶ Two-way interactions: $(\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}$
- ▶ Three-way interaction: $(\alpha\beta\gamma)_{ijk}$
- ▶ There are sum-to-zero constraints on the effects
- ▶ The total sum of squares can be decomposed and we can find an ANOVA table

Example: ANOVA table

```
mod = aov(bp ~ drug*feed*diet, data = df)
summary(mod)

##                Df Sum Sq Mean Sq F value   Pr(>F)
## drug            2   3675    1837  11.729 5.02e-05 ***
## feed            1   2048    2048  13.072 0.000615 ***
## diet            1   5202    5202  33.204 3.05e-07 ***
## drug:feed       2    259     129   0.827 0.442457
## drug:diet       2    903     451   2.882 0.063815 .
## feed:diet       1     32      32   0.204 0.652937
## drug:feed:diet  2   1075     537   3.431 0.038834 *
## Residuals      60   9400     157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Significant main effects and three-way interaction at $\alpha = 0.05$

Example: Residuals

```
par(mfrow = c(1,2))
plot(mod, which = c(1,2))
```
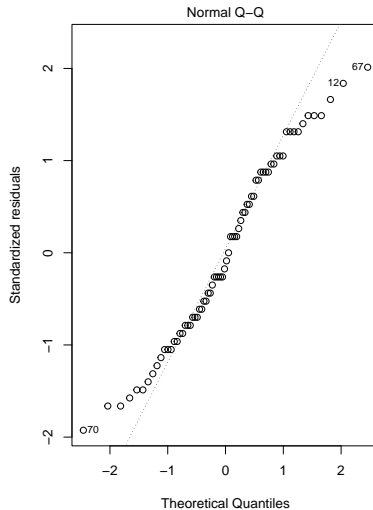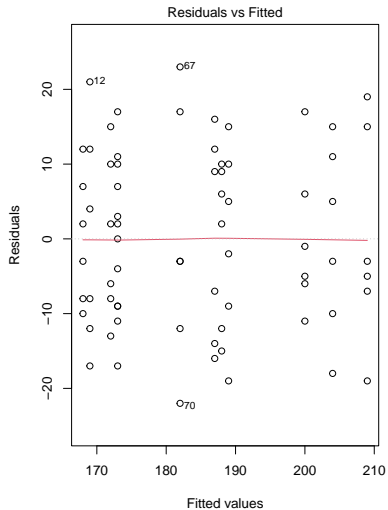
Example: interaction plot

```
library(emmeans)
emmip(mod, ~ drug | feed + diet)
```