DATA2002

Two-way ANOVA

Garth Tarr

THE UNIVERSITY OF
SYDNEY

Two-way ANOVA: adjusting for "blocks"

Multiple comparisons (revision)

Rank-based approaches

Two-way ANOVA: adjusting for "blocks"

Electrode testing

Berry (1987) presents data on skin resistance:

- 5 types of **electrode** were attached to each of 16 **subjects** and the resistance measured;
- **Aim:** do all 5 types perform similarly?
- There may be differences **between the electrode types** and/or **between the subjects**.
    - If there is, this will "add" to the overall variation. **Can we adjust for this?**

```
library(tidyverse)
resist = read_tsv("https://raw.githubuserconter
# glimpse(resist)
# convert from integer to factor
resist$Subject = factor(resist$Subject)
```

```
resist
```

```
## # A tibble: 16 × 6
##    Subject    E1    E2    E3    E4    E5
##    <fct>   <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 1         500   400    98   200   250
##  2 2         660   600   600    75   310
##  3 3         250   370   220   250   220
##  4 4          72   140   240    33    54
##  5 5         135   300   450   430    70
##  6 6          27    84   135   190   180
##  7 7         100    50    82    73    78
##  8 8         105   180    32    58    32
##  9 9          90   180   220    34    64
## 10 10        200   290   320   280   135
## 11 11         15    45    75    88    80
## 12 12        160   200   300   300   220
## 13 13        250   400    50    50    92
## 14 14        170   310   230    20   150
## 15 15         66  1000  1050   280   220
## 16 16        107    48    26    45    51
```

Full path to data: https://raw.githubusercontent.com/DATA2002/data/master/resist.txt

Outliers?

Berry (1987) notes that there may have been interference by a hairy arm:

> "After obtaining the results the experimenters decided that the reason for the two large readings on subject 15 was the excessive amount of hair on those parts of the subject's arm. They concluded that this subject's data should be deleted. Whether these readings are contaminants is not clear; the amount of hair present for the other 78 readings was not assessed relative to these two and no such assessment was made independent of the results."

Table 3
Resistance (in k.ohms)

Subject number	Electrode type				
	1	2	3	4	5
1	500	400	98	200	250
2	660	600	600	75	310
3	250	370	220	250	220
4	72	140	240	33	54
5	135	300	450	430	70
6	27	84	135	190	180
7	100	50	82	73	78
8	105	180	32	58	32
9	90	180	220	34	64
10	200	290	320	280	135
11	15	45	75	88	80
12	160	200	300	300	220
13	250	400	50	50	92
14	170	310	230	20	150
15	66	1000[a]	1050[a]	280	220
16	107	48	26	45	51
Mean	181.7	287.3	258.0	150.4	137.9

[a] Hairy part of arm?

Image source: Berry (1987)

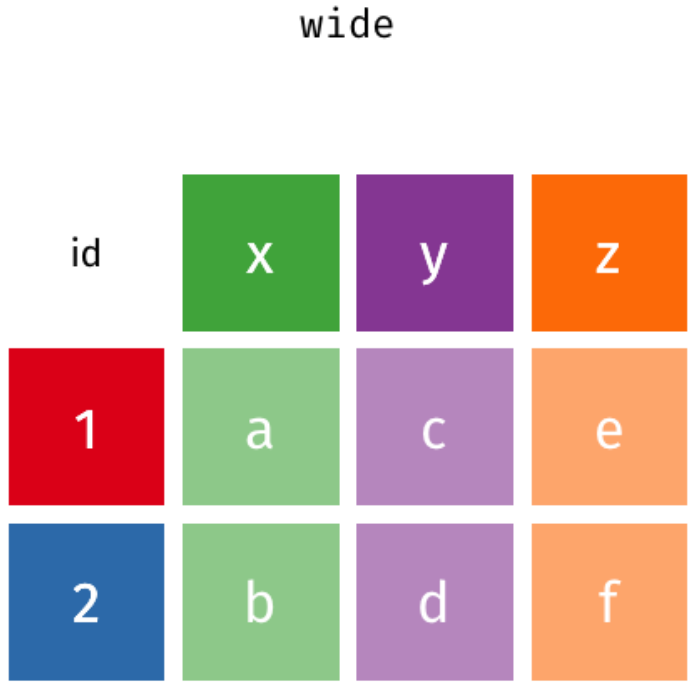
To analyse the data in R we need it in "long" format. I.e. we want a data frame with 3 columns:

- one with the **response**;
- a factor indicating **"treatment"** (i.e. electrode type);
- *another* factor indicating the **Subject**.

```
resist_long = resist %>%
  gather(key = "electrode", # key column name
         value = "resistance", # value column name
         -Subject) # don't gather the Subject column
glimpse(resist_long)
```

```
## Rows: 80
## Columns: 3
## $ Subject    <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, …
## $ electrode  <chr> "E1", "E1", "E1", "E1", "E1", "E1", "E1…
## $ resistance <dbl> 500, 660, 250, 72, 135, 27, 100, 105, 9…
```

```
# alternatively use the newer pivot_longer function
resist_long = resist %>%
  pivot_longer(cols = E1:E5, names_to = "electrode",
               values_to = "resistance")
```



wide

id	x	y	z
1	a	c	e
2	b	d	f

long

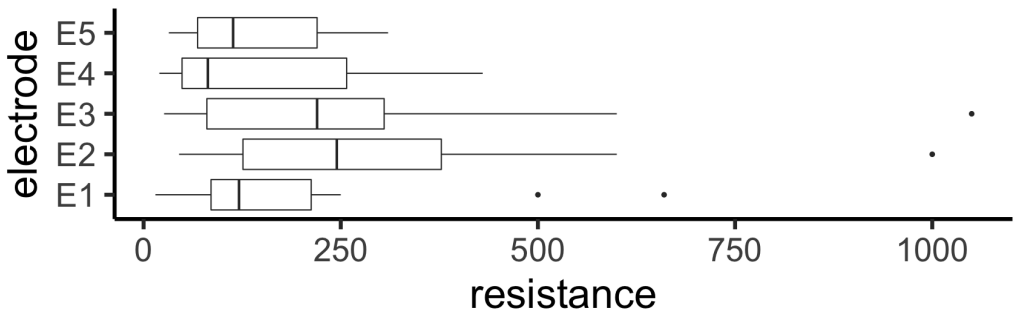
id	key	val
1	x	a
2	x	b
1	y	c
2	y	d
1	z	e
2	z	f

Log transformation

- Let's look at box plots of each **electrode type** *ignoring* **subject**.

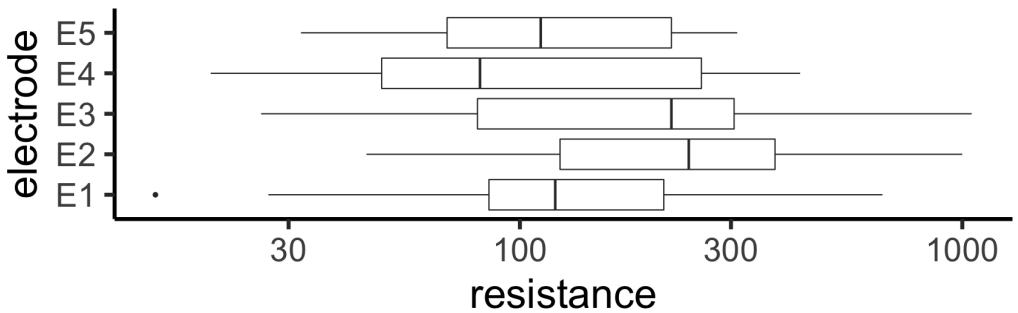
Linear scale

```
ggplot(resist_long, aes(x = electrode,
                        y = resistance)) +
  geom_boxplot() + coord_flip() +
  theme_classic(base_size = 34)
```



Log scale

```
ggplot(resist_long, aes(x = electrode,
                        y = resistance)) +
  geom_boxplot() + scale_y_log10() +
  coord_flip() + theme_classic(base_size = 34)
```



- The log-transformed data looks a little better: let's use that (for the moment).
- Note: `scale_y_log10()` doesn't change the data... it only transforms the plot axis.

Formulas in R

The basic structure of a formula in R is:

```
y ~ x # it's a tilde (squiggly line) between y and x
```

This can be read as "y is a function of x" or "y against x" or "y by x" or "y twiddles x".

Examples we've seen so far:

```
t.test(y ~ group)
ggplot(df) + facet_wrap( ~ group) # one sided formula
ggplot(df) + facet_grid(group1 ~ group2)
aov(y ~ group)
```

Sometimes we might want to consider multiple "explanatory" variables:

```
y ~ x1 + x2 + x3
```

- Define a new variable `y` in our data frame `resist_long` that is the log of the resistance measurement:

```
resist_long$y = log(resist_long$resistance)
```

- Let's start with an ordinary (one-way) ANOVA ignoring **Subject**:

```
fit1 = aov(y ~ electrode, data = resist_long)
summary(fit1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## electrode    4   5.09  1.2719   1.503   0.21
## Residuals   75  63.48  0.8464
```

- This is clearly "not significant".

Adjusting for Subject

We can add `Subject` as an extra factor variable in our **formula** to indicate that it should be used to help "explain" `y`. The formula that we use in the `aov()` is now: `y ~ Subject + electrode`

```
fit2 = aov(y ~ Subject + electrode, data = resist_long)
summary(fit2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Subject     15  33.27  2.2180   4.405 1.77e-05 ***
## electrode    4   5.09  1.2719   2.526     0.05 *
## Residuals   60  30.21  0.5036
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **What is going on here?**

Compare the two tables carefully:

```
summary(fit1) # y ~ electrode
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## electrode    4   5.09  1.2719   1.503   0.21
## Residuals   75  63.48  0.8464
```

```
summary(fit2) # y ~ Subject + electrode
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Subject     15  33.27  2.2180   4.405 1.77e-05
## electrode    4   5.09  1.2719   2.526     0.05
## Residuals   60  30.21  0.5036
```

Decomposition of the residual sum of squares

- The **Residual** sum of squares from `fit1` (63.48, on 75 df) is being *decomposed* into two pieces:
  - the `fit2 Subject` sum of squares (33.27, on **15** df) and
  - the `fit2` **Residual** sum of squares (30.21, on 60 df)
- For `fit2`, the **Residual** sum of squares is much smaller that for `fit1`, but the degrees of freedom is only a little less:
  - this gives a much smaller **Residual Mean Square** (0.5036, compared to 0.8464 for `fit1`);
  - this in turn gives a bigger (*treatment-to-residual*) F-ratio (2.526, compared to 1.503 for `fit1`);
  - **crucially** the p-value has been reduced from 0.21 to 0.05:
    - the effect is now (at least mildly) significant!
- We are now ready to study the mathematics of what we have done here;
  - but first, we return briefly to ordinary ANOVA to learn how to change parameters.

Changing parameters

- For **ordinary one-way ANOVA** we have have written the model as: for $i = 1, \ldots, g$, $j = 1, \ldots, n_i$,

$$Y_{ij} \sim N(\mu_i, \sigma^2)\,.$$

  - There are g unknown mean-value parameters, and 1 unknown variance parameter.

- Another way to write this is, for $i = 1, \ldots, g$, $j = 1, \ldots, n_i$,

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

  where the ε_{ij}'s are iid $N(0, \sigma^2)$.

  - **Again**, there are g unknown mean-value parameters, and 1 unknown variance parameter.

Yet another way...

- A *third* way to write the model is based on expressing each μ_i as
  - an *overall mean* μ (with no subscript) plus
  - an *adjustment* α_i for i-th level of the treatment:

$$\mu_i = \mu + \alpha_i$$

- This leads to the model: for $i = 1, \ldots, g$, $j = 1, \ldots, n_i$,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

- Note there are now $g + 1$ "mean" parameters (sort of): $\mu, \alpha_1, \ldots, \alpha_g$
  - we have "created" another parameter

An extra constraint

- In fact, depending on how μ is defined, the α_i's necessarily obey a certain constraint.
- The overall mean is defined as some kind of (weighted) average of the μ_i's:

$$\mu = \sum_{i=1}^{g} w_i \mu_i \, .$$

  - Then each $\alpha_i = \mu_i - \mu$.
- Necessarily, the same weighted average of the α_i's is:

$$\sum_{i=1}^{g} w_i \alpha_i = \sum_{i=1}^{g} w_i (\mu_i - \mu) = \left(\sum_{i=1}^{g} w_i \mu_i \right) - \mu \sum_{i=1}^{g} w_i = \mu - \mu = 0 \, .$$

- So in fact, knowing $g - 1$ of the α_i's means you also know the final one.

Estimating these new "parameters"

- A common choice for the "weighted average" is

$$\mu = \frac{1}{N}\sum_{i=1}^{g} n_i \mu_i = \frac{\sum_{i=1}^{g} n_i \mu_i}{\sum_{i=1}^{g} n_i},$$

  which is the the *expectation of the grand mean* $\bar{Y}_{\bullet\bullet}$.
- This can be estimated using the *observed* grand mean $\bar{y}_{\bullet\bullet}$.
- Each α_i represents the difference between each group mean and the overall mean,
  - it's thus naturally estimated using the difference

$$\hat{\alpha}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}.$$

The two-way ANOVA model

- The model we shall fit to the electrode data is the following:

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

where

$$\begin{aligned} \mu &= \text{overall mean} \\ \alpha_i &= \text{ adjustment for electrode type } i \text{ for } i = 1, 2, \ldots, g \\ \beta_j &= \text{ adjustment for subject } j \text{ for } j = 1, 2, \ldots, n \end{aligned}$$

and n is the common sample (block) size and the ε_{ij}'s are iid $N(0, \sigma^2)$.

- So each Y_{ij} has a possibly different expectation $\mu_{ij} = \mu + \alpha_i + \beta_j$, but these have an **additive structure**:
  - the ng different means are explained by $1 + (g-1) + (n-1) = g + n - 1$ (free) parameters.

Estimating parameters

- As all "sample sizes" are the same, the overall mean can be thought of as just the mean of the μ_i's.
  - it is naturally estimated using the overall mean $\bar{y}_{\bullet\bullet}$.
- Also, each α_i, the "adjustment" for **electrode type** i, is naturally estimated using the difference

$$\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} .$$

- Similarly, each β_j, the adjustment for **subject** j, is naturally estimated using the difference

$$\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet} .$$

The two-way decomposition

- Each observation therefore, may be notionally split up into 4 pieces:

$$y_{ij} = \underbrace{\bar{y}_{\bullet\bullet}}_{\hat{\mu}} + \underbrace{(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})}_{\hat{\alpha}_i} + \underbrace{(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})}_{\hat{\beta}_j} + \underbrace{(y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})}_{\hat{\varepsilon}_{ij}}$$

- The final part $\hat{\varepsilon}_{ij}$ is the (i, j)-th **residual** or estimated error.
- We can "analyse the variance" here in the same way as the ordinary "one-way" ANOVA model.

Decomposing the total sum of squares

$$
\begin{aligned}
\sum_{i=1}^{g}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2 &= \sum_{i=1}^{g}\sum_{j=1}^{n}\left\{(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})+(\bar{y}_{\bullet j}-\bar{y}_{\bullet\bullet})+(y_{ij}-\bar{y}_{i\bullet}-\bar{y}_{\bullet j}+\bar{y}_{\bullet\bullet})\right\}^2 \\
&= \sum_{i=1}^{g} n(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2 + \sum_{j=1}^{n} g(\bar{y}_{\bullet j}-\bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^{g}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet}-\bar{y}_{\bullet j}+\bar{y}_{\bullet\bullet})^2 \\
&\qquad + \text{cross-product terms which are all zero} \\
&= \text{Treatment sum of squares} + \text{Block sum of squares} + \text{Residual sum of squares}
\end{aligned}
$$

Adjusting for "blocks"

- We have identified some **systematic variation** which can be attributed to the differences between **Subjects** (assuming these contribute *additively*).
- The term **"Block"** again comes from Fisher's agricultural trials, where he adjusted for variation between different blocks of land, in order to compare the (fertiliser) *Treatments* more accurately.
- The net result is that we have a *smaller (more precise) estimate of the error variance* as we have explained an extra part of variation and removed it from the **residual sum of squares**.

About 40,000 grain research plots at Narrabri.

The two-way ANOVA table

The two-way ANOVA table

- With the above sum of squares definitions, the "two-way ANOVA" (not the best name) table is given as follows:

Source of Variation	Sum of squares	df	Mean square	F-ratio
Blocks	Block Sum Sq.	$n-1$		
Treatments	Trt Sum Sq.	$g-1$	$\text{Trt MS} = \dfrac{\text{Trt Sum Sq.}}{g-1}$	$\dfrac{\text{Trt MS}}{\text{Res MS}}$
Residual	Res Sum Sq.	$(n-1)(g-1)$	$\text{Res MS} = \dfrac{\text{Res Sum Sq.}}{(n-1)(g-1)}$	
Total	Total Sum Sq.	$ng-1$		

- The total sum of squares is $\sum_{i=1}^{g}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2$, and the total sample size is $N=ng$.
- Once appropriately coded using R, this can be obtained using `summary(aov(...))`, `anova(aov(...))` or even `anova(lm(...))` or using `broom::tidy(aov(...))`.

The purpose of blocking

- A two-way ANOVA with blocking can be thought of as a generalisation of the paired t-test where each **pair** is a **block**.
- In the paired t-test, the idea is to *remove* the variation "between pairs", to more accurately compare the two treatment levels *within each pair*.
    - the "within pair" difference is then averaged over all pairs to get the "treatment effect".
- We are *not* interested in "testing for a Block effect", we are **only** interested in comparing **Treatments**.
- We are nonetheless *adjusting* for **Blocks**, to more accurately compare Treatments.
- Although the **treatment sum of squares** and **block sum of squares** are *mathematically* identical, they are playing very different *scientific* roles.

Behaviour of sums of squares under the two-way ANOVA model

Two-way ANOVA sums of squares behaviour

- Recall our model: y_{ij} (the observation in block j receiving treatment level i) is modelled as the value taken by

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

for $i = 1, 2, \ldots, g$ and $j = 1, 2, \ldots, n$ where

$$\begin{aligned} \mu &= \text{ overall mean}, \\ \alpha_i &= \text{ adjustment for treatment level } i, \\ \beta_j &= \text{ adjustment for block } j, \\ \varepsilon_{ij} &\sim N(0, \sigma^2)\,, \end{aligned}$$

all random variables are independent and the following constraints are satisfied:

$$\sum_{i=1}^{g} \alpha_i = 0 \text{ and } \sum_{j=1}^{n} \beta_j = 0\,.$$

Averages

- The overall, treatment level and block averages are therefore (due to the constraints):

$$\bar{Y}_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n}\left(\mu + \alpha_i + \beta_j + \varepsilon_{ij}\right) = \mu + \alpha_i + \bar{\varepsilon}_{i\bullet} \quad (\text{free of the } \beta_j\text{'s!})$$

$$\bar{Y}_{\bullet j} = \frac{1}{g}\sum_{i=1}^{g}\left(\mu + \alpha_i + \beta_j + \varepsilon_{ij}\right) = \mu + \beta_j + \bar{\varepsilon}_{\bullet j} \quad (\text{free of the } \alpha_i\text{'s!})$$

$$\bar{Y}_{\bullet\bullet} = \frac{1}{ng}\sum_{i=1}^{g}\sum_{j=1}^{n}\left(\mu + \alpha_i + \beta_j + \varepsilon_{ij}\right) = \mu + \bar{\varepsilon}_{\bullet\bullet}$$

Not examinable

Treatment sum of squares

- The **treatment sum of squares** is

$$\sum_{i=1}^{g} n(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = n\sum_{i=1}^{g}(\alpha_i + \bar{\varepsilon}_{i\bullet} - \bar{\varepsilon}_{\bullet\bullet})^2 \,.$$

- Under the null hypothesis H_0: $\alpha_1 = \ldots = \alpha_g = 0$, this is

$$\underbrace{n\sum_{i=1}^{g}(\bar{\varepsilon}_{i\bullet} - \bar{\varepsilon}_{\bullet\bullet})^2}_{\sim \frac{\sigma^2}{n}\chi^2_{g-1}} \sim n\left(\frac{\sigma^2}{n}\chi^2_{g-1}\right) \sim \sigma^2\chi^2_{g-1}$$

  since under the model the $\bar{\varepsilon}_{i\bullet}$'s are iid normal with variance σ^2/n.

  - This is the same as for one-way ANOVA.

Residual sum of squares

The (i, j)-th residual is

$$Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet} = \ldots = \varepsilon_{ij} - \bar{\varepsilon}_{i\bullet} - \bar{\varepsilon}_{\bullet j} + \bar{\varepsilon}_{\bullet\bullet}.$$

How is the residual sum of squares distributed?

- We have the identity (with $N = ng$),

$$\underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon}_{i\bullet})^2}_{\sim\sigma^2\chi^2_{N-g}} = \underbrace{\sum_{j=1}^{n} g(\bar{\varepsilon}_{\bullet j} - \bar{\varepsilon}_{\bullet\bullet})^2}_{\sim\sigma^2\chi^2_{n-1}} + \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon}_{i\bullet} - \bar{\varepsilon}_{\bullet j} + \bar{\varepsilon}_{\bullet\bullet})^2}_{\sim???}$$

  Roughly speaking this is

$$\text{One-way Res Sum Sq.} = \text{Block Sum Sq. of Errors} \; + \; \text{Two-way Res Sum Sq.}$$

- It can be shown that the two terms on the RHS are independent, so the last double sum **must be** $\sigma^2\chi^2_{N-g-(n-1)} \sim \sigma^2\chi^2_{(n-1)(g-1)}$.

Not examinable

Two-way ANOVA F-ratio

- In summary:
  - the residual sum of squares always follows a $\sigma^2\chi^2_{(n-1)(g-1)}$ distribution (regardless of whether the null hypothesis is true or not);
  - *if the null hypothesis of "no treatment effect" is true*, the treatment sum of squares follows a $\sigma^2\chi^2_{g-1}$ distribution.
- Therefore, *if the null hypothesis is true*, the F-ratio

$$\frac{\text{Treatment mean square}}{\text{Residual mean square}} \sim \frac{\chi^2_{g-1}/(g-1)}{\chi^2_{(n-1)(g-1)}/(n-1)(g-1)} \sim F_{g-1,\,(n-1)(g-1)}\,,$$

- Otherwise, it tends to take *larger* values (as it does for one-way ANOVA).

Further reading

Chapter 13 in Larsen and Marx (2012) is on Randomised Block Designs. The most relevant sections are 13.1 and 13.2. The notation is slightly different to what we have used, but the concepts are discussed in some detail. They also say how to test for block effects, but I still don't think it's something you would sensibly do.

Post hoc tests and multiple comparisons methods

Bonferroni method

- Under our "new" parametrisation, each "treatment difference" e.g. $\alpha_1 - \alpha_2$ is still naturally estimated using the corresponding treatment level mean difference.
- For example, we would estimate
  - α_1 using $\bar{y}_{1\bullet} - \bar{y}_{\bullet\bullet}$
  - α_2 using $\bar{y}_{1\bullet} - \bar{y}_{\bullet\bullet}$
  - $\alpha_1 - \alpha_2$ using $\bar{y}_{1\bullet} - \bar{y}_{2\bullet}$
- Under this two-way model, the corresponding random mean difference is distributed as

$$\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet} \sim N\left(\alpha_1 - \alpha_2,\ \frac{2\sigma^2}{n}\right),$$

since all treatment groups have a common sample size n.

An *individual* t-test and confidence interval

Assume we're testing with a significance level, α.

- We estimate the standard error $\sigma\sqrt{\frac{2}{n}}$ by plugging in $\hat{\sigma}^2$ (the **Residual Mean Square**) as the estimate of σ^2.
- Suppose $c(\alpha)$ satisfies $P(-c(\alpha) \leq t_{(n-1)(g-1)} \leq c(\alpha)) = 1 - \alpha$.
- An **individual** level α t-test for comparing groups 1 and 2 would therefore reject for

$$\frac{|\bar{y}_{1\bullet} - \bar{y}_{2\bullet}|}{\hat{\sigma}\sqrt{\frac{2}{n}}} > c(\alpha)\,.$$

- An **individual** $100(1-\alpha)\%$ confidence interval for $\alpha_1 - \alpha_2$ would be given by

$$\bar{y}_{1\bullet} - \bar{y}_{2\bullet} \pm c(\alpha)\,\hat{\sigma}\sqrt{\frac{2}{n}}\,.$$

Adjusting for multiplicity

- If we are performing k simultaneous comparisons, we replace α with α/k.
- This means that we simply replace $c(\alpha)$ with $c(\alpha/k)$ and
  - perform each t-test at "level α/k"
  - construct each confidence interval at the $100(1-\alpha/k)\%$ confidence level.
- Then the "overall performance" of the k procedures taken "simultaneously" is "at level α": under the model,
  - the probability of incorrectly rejecting *any* of the t-tests is no more than α (family wise error rate)
  - the probability that all true values are included in the corresponding confidence interval is $1-\alpha$

Multiplicity-adjusted p-values

- If we are doing k simultaneous t-tests, we reject each one at "overall level α" if and only if each *individual unadjusted* p-value is less than α/k.
- This is equivalent to rejecting if k **times the unadjusted p-value** exceeds α.
- We thus *define* each "adjusted" p-value as k times the corresponding unadjusted p-value.

The Bonferroni test

- If we are doing *all pairwise comparisons* across g groups then there are $\binom{g}{2}$ of these.
- If **none** of the tests end up rejecting (after adjusting for multiplicity) then this means the *smallest individual p-value* (corresponding to the *most significant pairwise difference*) was greater than $\alpha / \binom{g}{2}$.
- Therefore the p-value for the Bonferroni test is simply $\binom{g}{2}$ times the smallest *unadjusted* p-value.
- If this test rejects, we can "post hoc" identify which comparisons are significant by identifying which *adjusted* p-values are less than α.

Electrode data

- Recall the two-way ANOVA table for the electrode data:

```
fit2 = aov(y ~ Subject + electrode, data = resist_long)
anova(fit2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Subject   15 33.269 2.21797  4.4047 1.768e-05 ***
## electrode  4  5.087 1.27185  2.5258   0.04996 *
## Residuals 60 30.213 0.50355
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

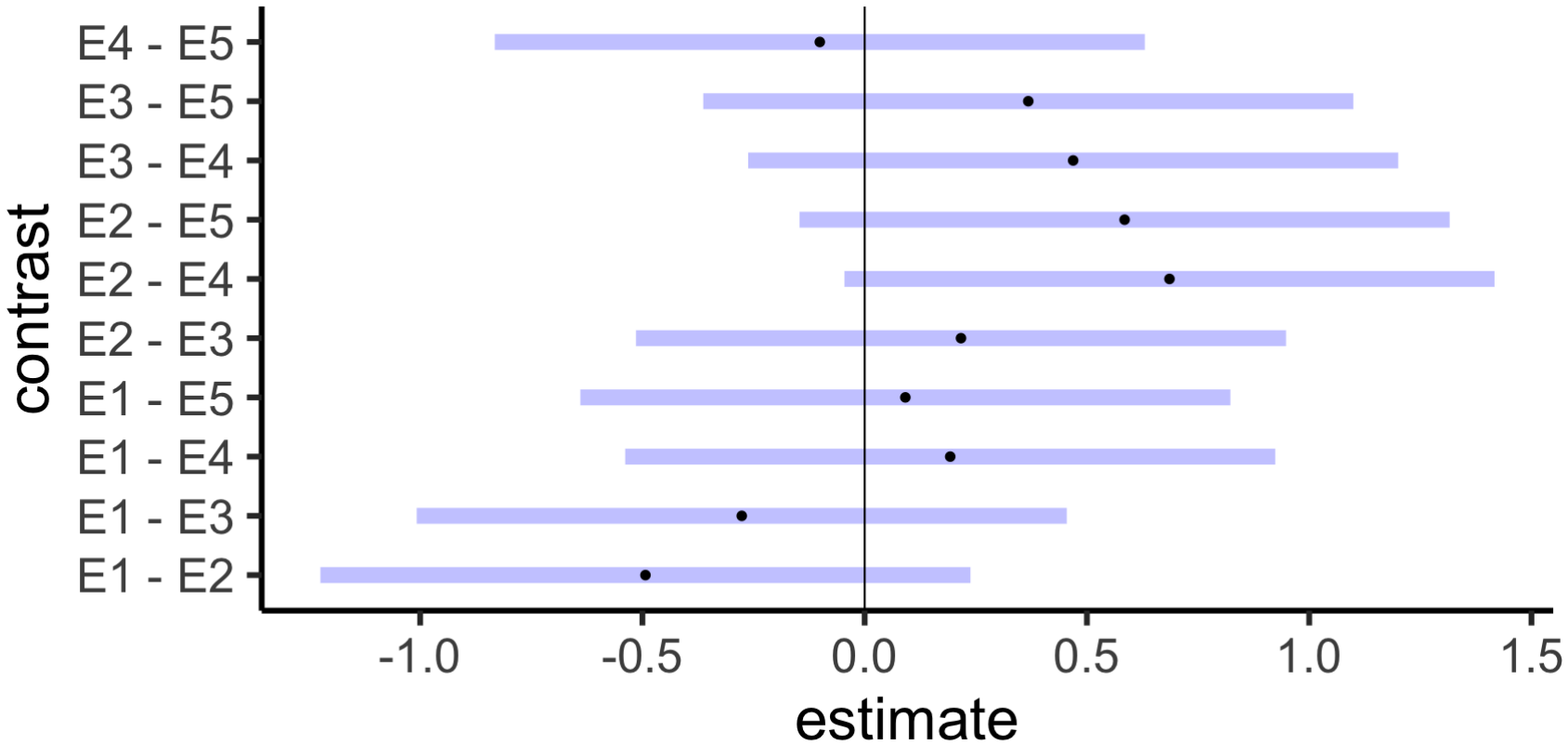
The emmeans package

```
library(emmeans)
fit2_emmeans = emmeans(fit2, ~ electrode)
contrast(fit2_emmeans, method = "pairwise", adjust = "bonferroni")
```

```
##  contrast estimate    SE df t.ratio p.value
##  E1 - E2   -0.4932 0.251 60  -1.966  0.5394
##  E1 - E3   -0.2765 0.251 60  -1.102  1.0000
##  E1 - E4    0.1925 0.251 60   0.767  1.0000
##  E1 - E5    0.0915 0.251 60   0.365  1.0000
##  E2 - E3    0.2167 0.251 60   0.864  1.0000
##  E2 - E4    0.6857 0.251 60   2.733  0.0823
##  E2 - E5    0.5847 0.251 60   2.331  0.2315
##  E3 - E4    0.4690 0.251 60   1.869  0.6645
##  E3 - E5    0.3681 0.251 60   1.467  1.0000
##  E4 - E5   -0.1010 0.251 60  -0.402  1.0000
##
## Results are averaged over the levels of: Subject
## P value adjustment: bonferroni method for 10 tests
```

The emmeans package

```
contrast(fit2_emmeans, method = "pairwise", adjust = "bonferroni") %>%
  plot() + geom_vline(xintercept = 0)
```



Tukey's method (p-values)

- In this case Tukey's method can be applied, and it is exact.

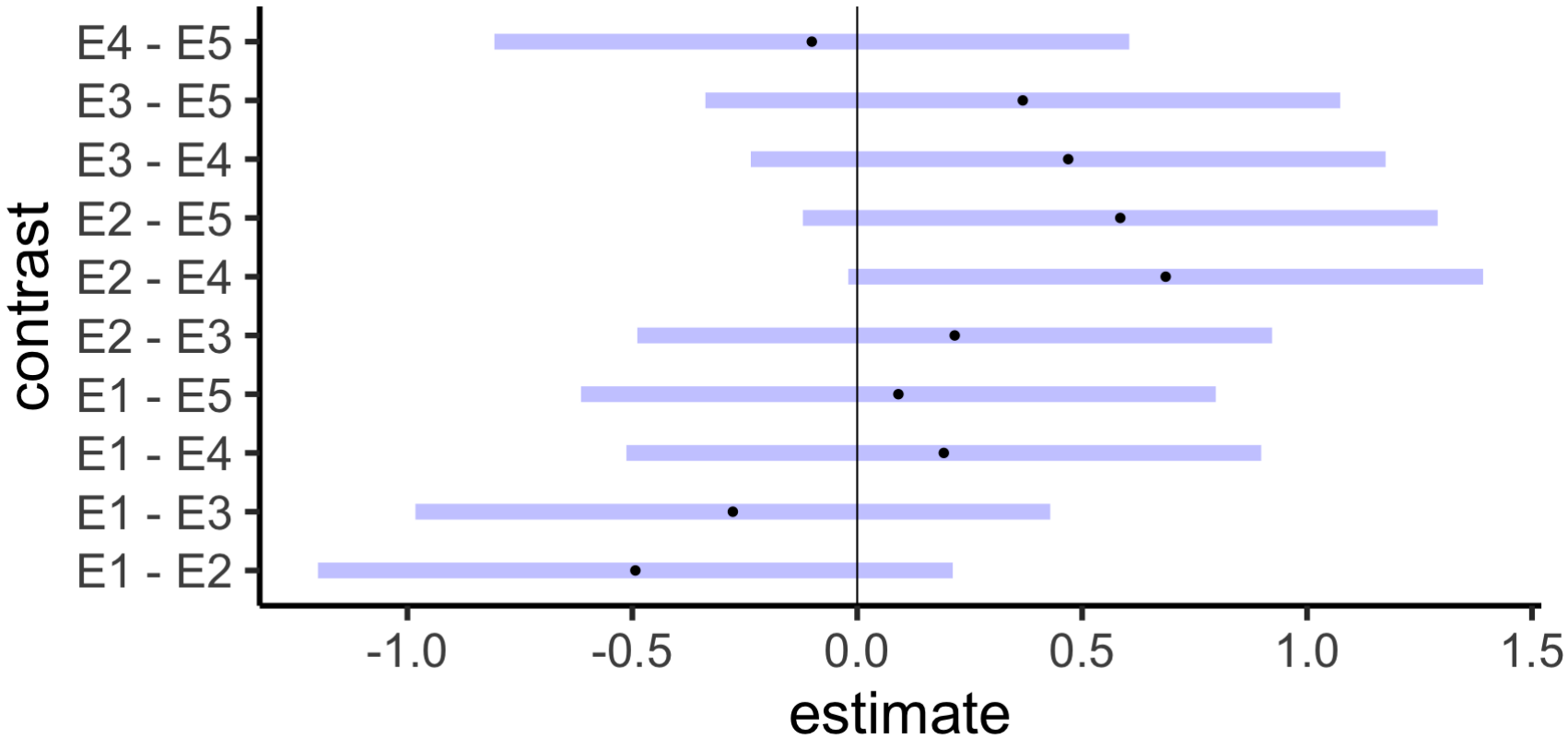
```
# TukeyHSD(fit2, which = "electrode") # an alternative not using emmeans
contrast(fit2_emmeans, method = "pairwise", adjust = "tukey")
```

```
##  contrast estimate    SE df t.ratio p.value
##  E1 - E2   -0.4932 0.251 60  -1.966  0.2950
##  E1 - E3   -0.2765 0.251 60  -1.102  0.8047
##  E1 - E4    0.1925 0.251 60   0.767  0.9390
##  E1 - E5    0.0915 0.251 60   0.365  0.9961
##  E2 - E3    0.2167 0.251 60   0.864  0.9090
##  E2 - E4    0.6857 0.251 60   2.733  0.0607
##  E2 - E5    0.5847 0.251 60   2.331  0.1495
##  E3 - E4    0.4690 0.251 60   1.869  0.3448
##  E3 - E5    0.3681 0.251 60   1.467  0.5875
##  E4 - E5   -0.1010 0.251 60  -0.402  0.9943
##
## Results are averaged over the levels of: Subject
## P value adjustment: tukey method for comparing a family of 5 estimates
```

- The p-value for the "Tukey test" is 0.061.

Tukey's method (visualising CIs)

```
contrast(fit2_emmeans, method = "pairwise", adjust = "tukey") %>%
  plot() + geom_vline(xintercept = 0)
```



Tukey's method (reporting CIs)

```
contrast(fit2_emmeans, method = "pairwise", adjust = "tukey") %>% confint()
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  E1 - E2   -0.4932 0.251 60  -1.1988    0.212
##  E1 - E3   -0.2765 0.251 60  -0.9822    0.429
##  E1 - E4    0.1925 0.251 60  -0.5131    0.898
##  E1 - E5    0.0915 0.251 60  -0.6141    0.797
##  E2 - E3    0.2167 0.251 60  -0.4889    0.922
##  E2 - E4    0.6857 0.251 60  -0.0199    1.391
##  E2 - E5    0.5847 0.251 60  -0.1209    1.290
##  E3 - E4    0.4690 0.251 60  -0.2366    1.175
##  E3 - E5    0.3681 0.251 60  -0.3376    1.074
##  E4 - E5   -0.1010 0.251 60  -0.8066    0.605
##
## Results are averaged over the levels of: Subject
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 5 estimates
```

Scheffe's method

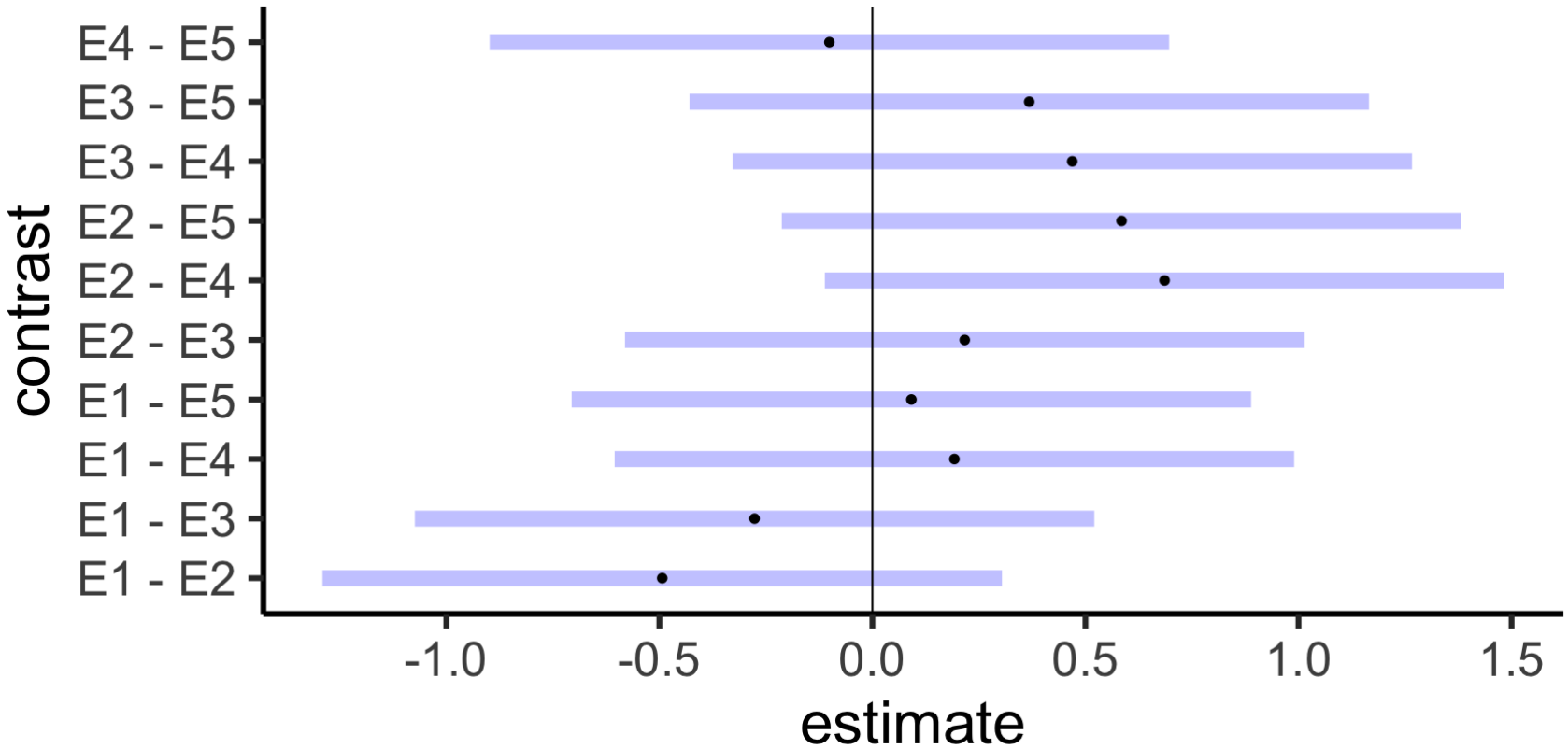
Allows for data snooping, can be used with any number and type of contrasts.

```
contrast(fit2_emmeans, method = "pairwise", adjust = "scheffe")
```

```
##  contrast estimate    SE df t.ratio p.value
##  E1 - E2   -0.4932 0.251 60  -1.966  0.4327
##  E1 - E3   -0.2765 0.251 60  -1.102  0.8743
##  E1 - E4    0.1925 0.251 60   0.767  0.9636
##  E1 - E5    0.0915 0.251 60   0.365  0.9978
##  E2 - E3    0.2167 0.251 60   0.864  0.9446
##  E2 - E4    0.6857 0.251 60   2.733  0.1279
##  E2 - E5    0.5847 0.251 60   2.331  0.2593
##  E3 - E4    0.4690 0.251 60   1.869  0.4851
##  E3 - E5    0.3681 0.251 60   1.467  0.7083
##  E4 - E5   -0.1010 0.251 60  -0.402  0.9968
##
## Results are averaged over the levels of: Subject
## P value adjustment: scheffe method with rank 4
```

Scheffe's method

```
contrast(fit2_emmeans, method = "pairwise", adjust = "scheffe") %>%
  plot() + geom_vline(xintercept = 0)
```



Summary: two-way ANOVA normal model

- Apart from the fact that we have a different $\hat{\sigma}$ (and corresponding degrees of freedom), everything is much the same as in the "one-way ANOVA" normal model.
- Once we have "adjusted for blocks" (in order to get a smaller estimate of the error variance), we adjust the degrees of freedom and then proceed as in the one-way case.

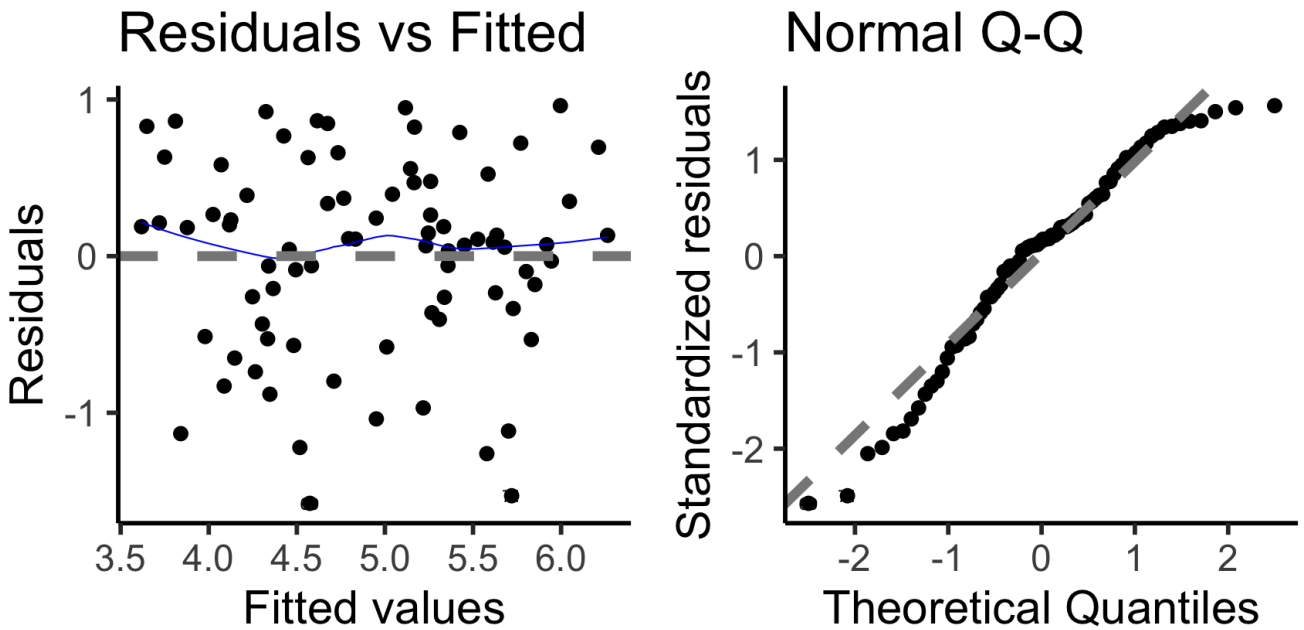
Model checking

- It is customary to check the assumptions underlying the "additive normal model".
- The main things to check are the **normality** and **constant variance** assumption.
- This is usually done by
  - checking that a boxplot or normal QQ plot of **residuals** "looks normal" (i.e. boxplot: symmetric, not too many outliers; QQ plot: the points are "close" to the diagonal line);
  - plotting **residuals** against **fitted values** to check the common variance assumption.
- **Fitted values:** $\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$
- **Residuals:** $r_{ij} = y_{ij} - \hat{y}_{ij}$
- The `fitted.values` and `residuals` can be extracted from the **aov** object:

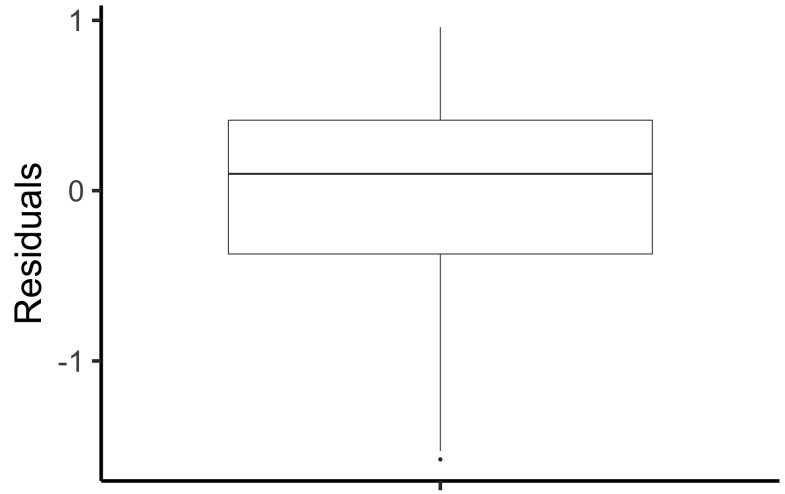
```
resist_long = resist_long %>%
  mutate(
    fitted = fit2$fitted.values,
    resid = fit2$residuals,
  )
```

Residual plots

```
library(ggfortify)
autoplot(fit2, which = 1:2, size = 4,
         ad.size = 3, colour = "black") +
  theme_classic(base_size = 30)
```




```
resist_long %>%
  ggplot(aes(x = "", y = resid)) +
  geom_boxplot() +
  theme_classic(base_size = 40) +
  labs(x = "", y = "Residuals")
```



Possible lack of symmetry?

- The QQ plot and the boxplot suggest that *perhaps* the residuals are not symmetrically distributed about zero. [But they're probably fine.]
- This suggests that we might
  - try a different transformation (remember, we transformed the original data!)
  - try using an alternative method that does not require normality assumptions.
- We explore the second option next!

Adjusting for blocks using ranks

Recap: Kruskal-Wallis test

- The **Kruskal-Wallis** is a rank-based test for a **one-way** layout:
  - each observation is replaced by its *global* rank;
  - a one-way ANOVA F-test is performed on the ranks.
- It can be performed using
  - a "permutation test" approach i.e. repeatedly use a command like

```
F_stat = anova(aov(sample(rank(y))~factor))[1,4] # extract the test statistic
```

  - a χ^2-approximation can be used on the *equivalent* statistic

$$\frac{\text{Treatment sum of squares of the ranks}}{\text{Sample variance of ranks}} = \frac{\text{Treatment sum of squares of the ranks}}{\text{Total sum of squares of the ranks}/(N-1)}.$$

```
kruskal.test(y ~ factor)
```

Friedman test

- The **Friedman test** is a ranks-based test for a **two-way** layout:
  - each observation is replaced by its *within-block* rank;
  - a one-way ANOVA F-test is performed on the ranks.
- This test can *also* be performed using either
  - a permutation-type approach or
  - a χ^2-approximation on an *equivalent* statistic.
- In this case the equivalent statistic is

$$\frac{\text{Treatment sum of squares of the ranks}}{\text{Total sum of squares of the ranks}/n(g-1)}$$

  and has an approximate χ^2_{g-1} distribution under the null hypothesis that all treatments are equivalent.

Larsen and Marx (2012; section 14.5)

Friedman test

- The χ^2-approximation method:

```
friedman.test(y ~ electrode | Subject, data = resist_long)
```

```
##
##     Friedman rank sum test
##
## data:  y and electrode and Subject
## Friedman chi-squared = 5.4522, df = 4, p-value = 0.244
```

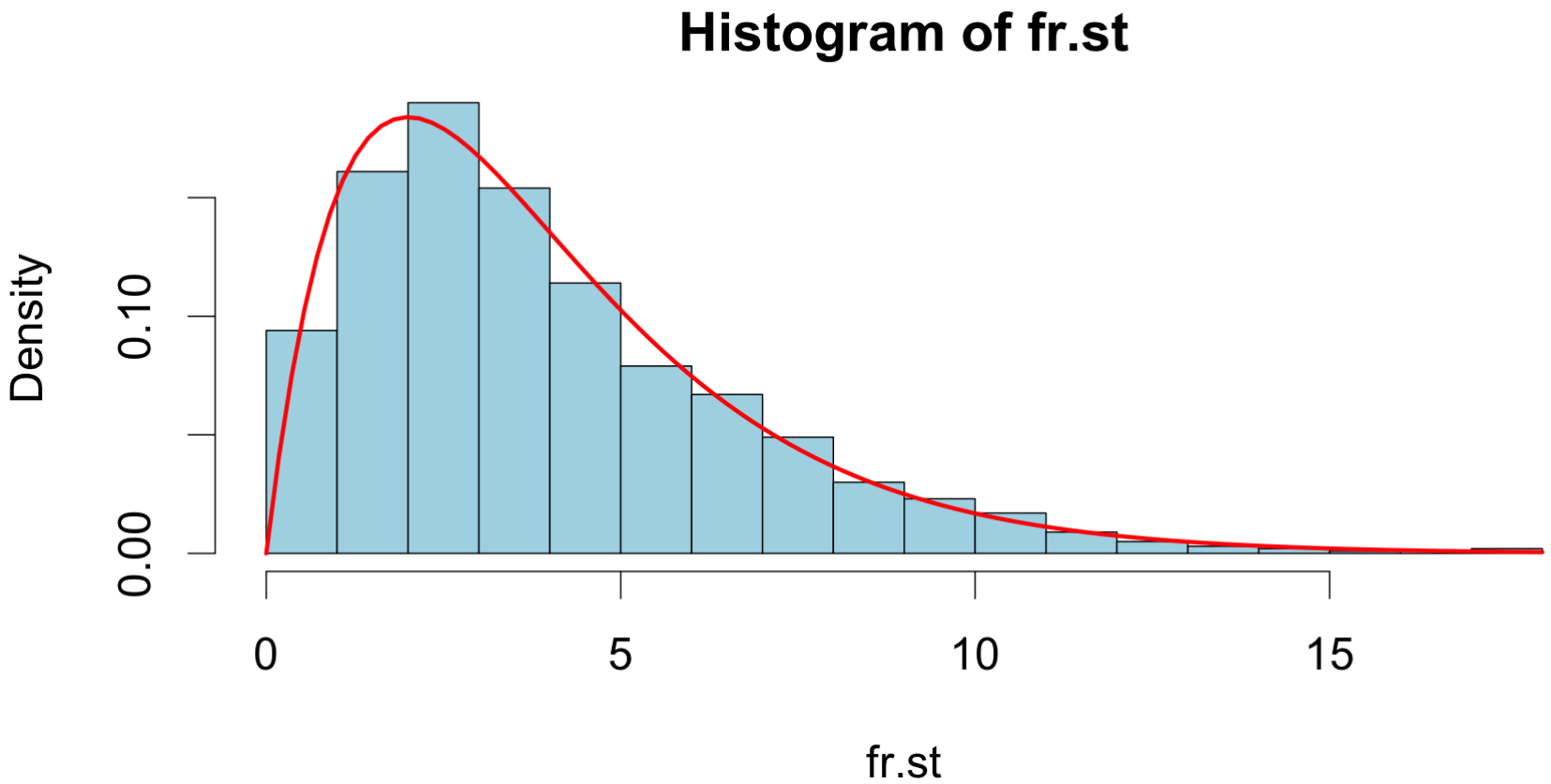
- We can also use a simulation/permutation approach to obtain a p-value:

```
fried.stat = friedman.test(y ~ electrode | Subject, data = resist_long)$statistic
B = 1000
fr.st = vector("numeric", length = B)
for(i in 1:B) {
  fr.st[i] = friedman.test(sample(y) ~ electrode | Subject, data = resist_long)$statistic
}
mean(fr.st>=fried.stat)
```

```
## [1] 0.244
```

Permutation comparison

```
hist(fr.st, breaks = 25, probability = TRUE, col = "lightblue")
curve(dchisq(x, 4), col = "red", add = TRUE, lwd = 3)
```



References

Berry, D. A. (1987). "Logarithmic Transformations in ANOVA". In: *Biometrics* 43.2, pp. 439-456. ISSN: 0006341X, 15410420. DOI: 10.2307/2531826.

Larsen, R. J. and M. L. Marx (2012). *An Introduction to Mathematical Statistics and its Applications*. 5th ed. Boston, MA: Prentice Hall. ISBN: 978-0-321-69394-5.

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.2.3. URL: https://CRAN.R-project.org/package=emmeans.