

STAT 516 Lec 05

One-way analysis of variance (review-ish)

Karl Gregory

2024-02-22

Rust inhibitors example

Data from Kutner et al. (2005).

Ten experimental units assigned to each of four brands of rust inhibitors.

```
link <- url("https://people.stat.sc.edu/gregorkb/data/KNNLrust.txt")
rust <- read.csv(link,col.names=c("score","brand","rep"),sep = "")
head(rust)
```

Y_{ij}

	score	brand	rep
1	39.0	1	2
2	46.7	1	3
3	43.8	1	4
4	44.2	1	5
5	47.7	1	6
6	43.6	1	7

treatment group

$\text{lm}(\text{score} \sim \text{brand})$

Do the brands differ in effectiveness? Is there a best brand?

Randomized experiments comparing treatments

Start with N experimental units (EUs), e.g. subjects, mice, etc.

Randomly assign each EU to one of a treatment groups.

Measure on each EU after treatment a response Y .

Compute the average of the responses in each treatment group...

Questions we'd like to answer:

- ▶ Is the response mean the same in all treatment groups?
- ▶ If not, then which pairs of means are different?

One-way ANOVA setup

treatment effects model.

Consider the model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad j = 1, \dots, (n_i), \quad i = 1, \dots, a,$$

Annotations: τ_i is circled with an arrow pointing to "treatment effects"; i is circled with an arrow pointing to "treatment group"; τ is labeled as "tau".

where

- ▶ Y_{ij} is the response for EU j in treatment group i .
- ▶ μ represents an overall or baseline mean.
- ▶ τ_i is the treatment effect for treatment i .
- ▶ The ε_{ij} are independent $\text{Normal}(0, \sigma^2)$ error terms.

Of central interest are the hypotheses

$$H_0: \tau_i = 0 \text{ for all } i \quad \text{versus} \quad H_1: \text{At least one } \tau_i \text{ is nonzero.}$$

If we reject H_0 , we may wish to sort/compare the treatments.

Alternative “cell means model” setup

An alternate version of the model is

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, a,$$

where

- ▶ Y_{ij} is the response for EU j in treatment group i .
- ▶ μ_i represents the mean of treatment group i .
- ▶ The ε_{ij} are error terms distributed as $\text{Normal}(0, \sigma^2)$.

In this version of the model the central hypotheses become

$$H_0: \mu_1 = \dots = \mu_a \quad \text{versus} \quad H_1: \mu_i \neq \mu_j \text{ for some } i \neq j.$$

Goals in one-way ANOVA

Under the one-way ANOVA setup

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, a,$$

where $\varepsilon_{ij} \stackrel{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$, we wish to

1. Visualize the data.
2. Estimate the parameters $\mu, \tau_1, \dots, \tau_a$.
3. Estimate the error term variance σ^2 .
4. Decompose the variation in the Y_{ij} as signal plus noise.
5. Test whether there is any difference in treatment group means.
6. Sort/compare the treatment means if there is any difference.
7. Check whether the model assumptions are satisfied.

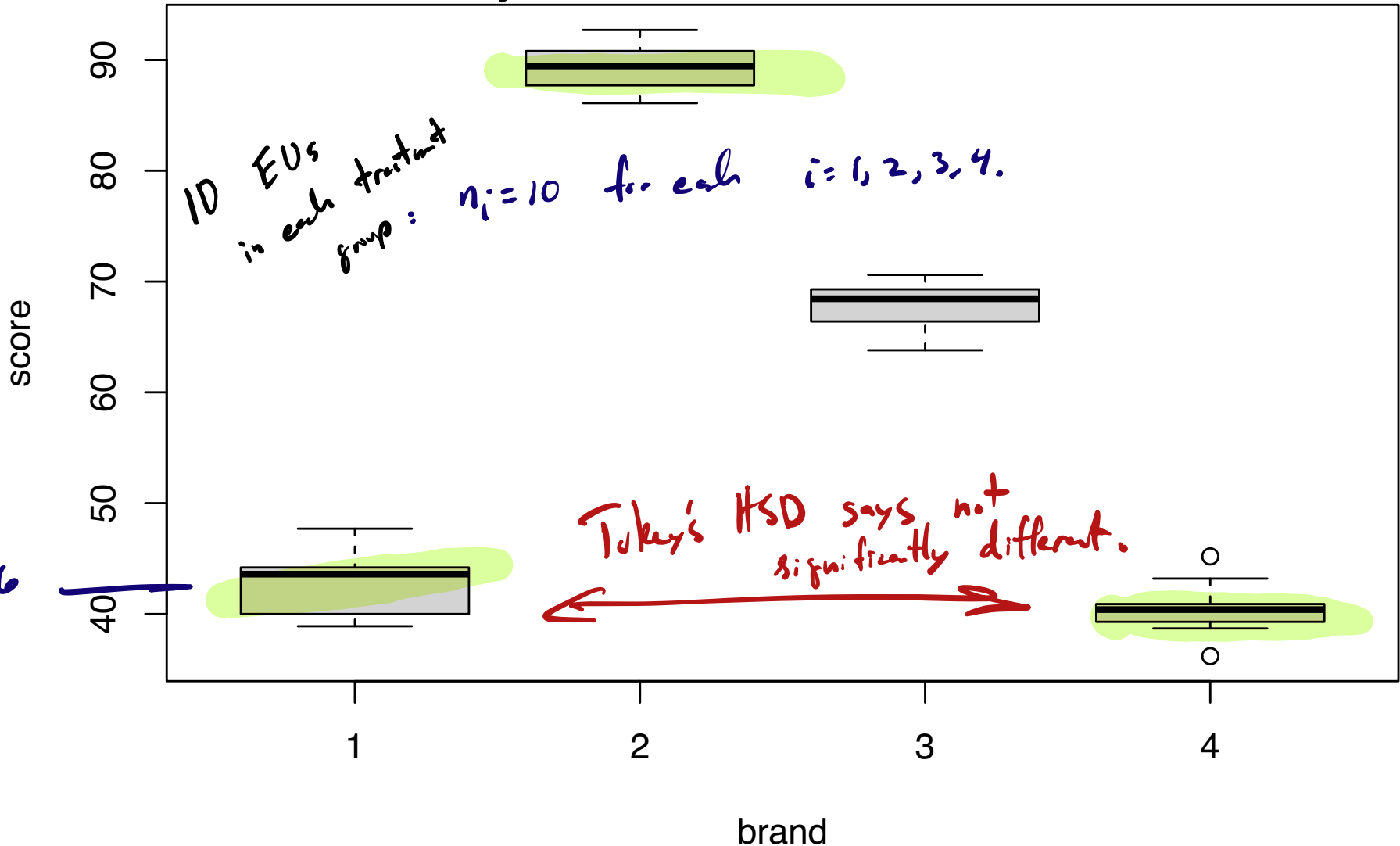
owing to the treatment effect
owing to variation among
EUs within the
groups

Rust inhibitors example (cont)

Visually compare the means of several treatment groups with boxplots.

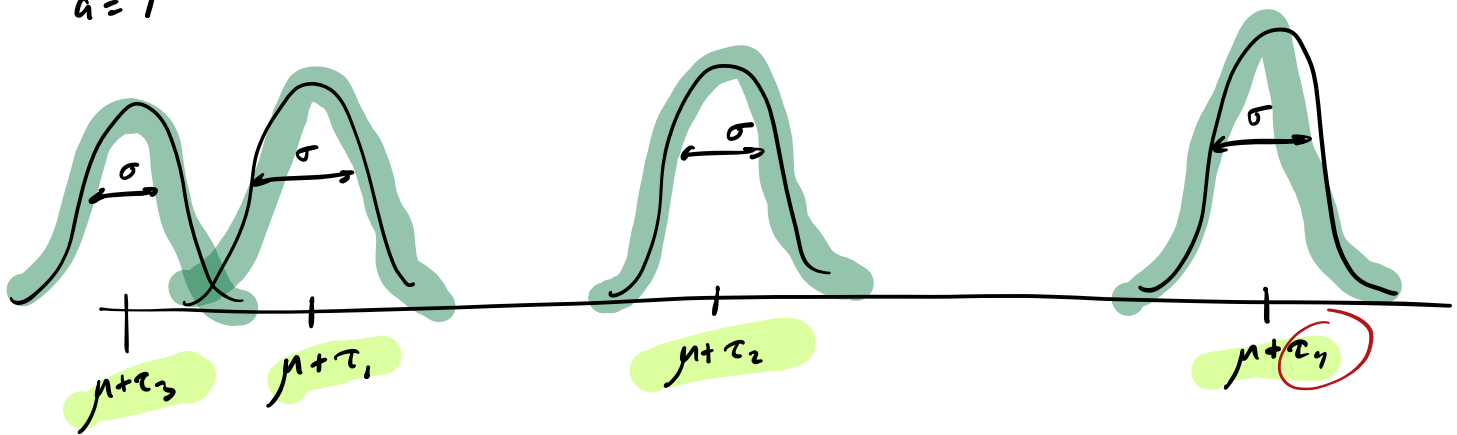
```
boxplot(score ~ brand, data = rust)
```

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$



$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Normal}(0, \sigma^2)$$

$a=4$



Note: I have $a+1$ parameters $\mu, \tau_1, \dots, \tau_a$ which describe a treatment means.

To uniquely identify μ and τ_1, \dots, τ_a , impose a constraint:

$$1) \tau_a = 0$$

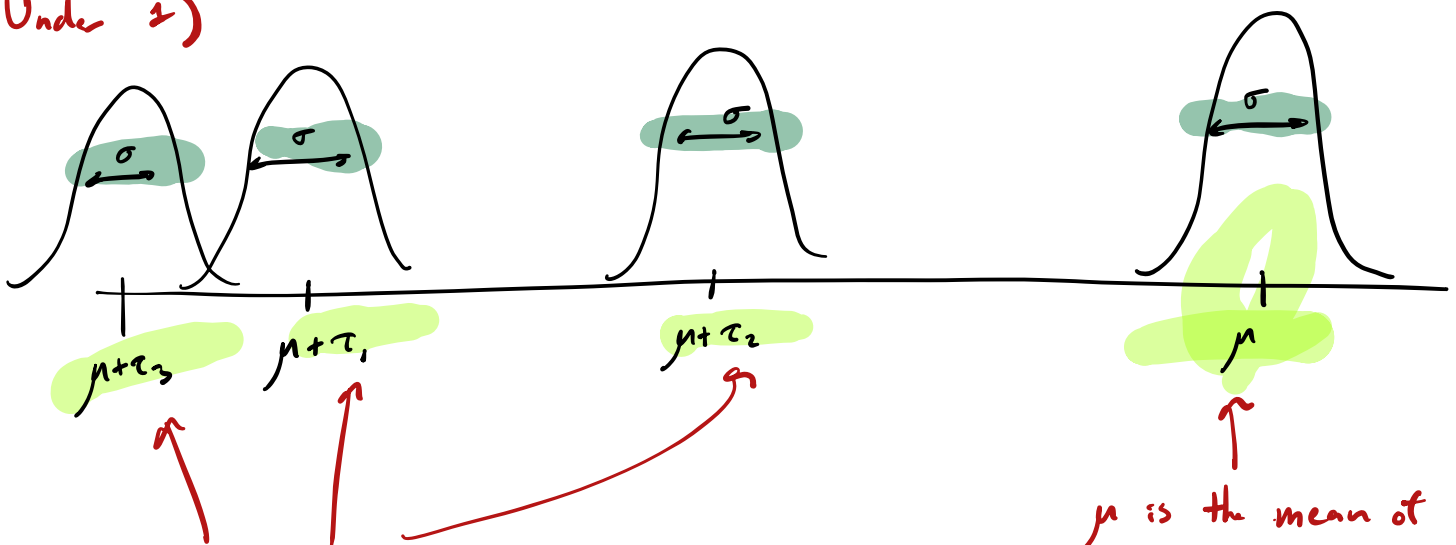
\Downarrow

Gives μ a "baseline" interpretation

$$2) \sum_{i=1}^a n_i \tau_i = 0$$

Don't worry about this one

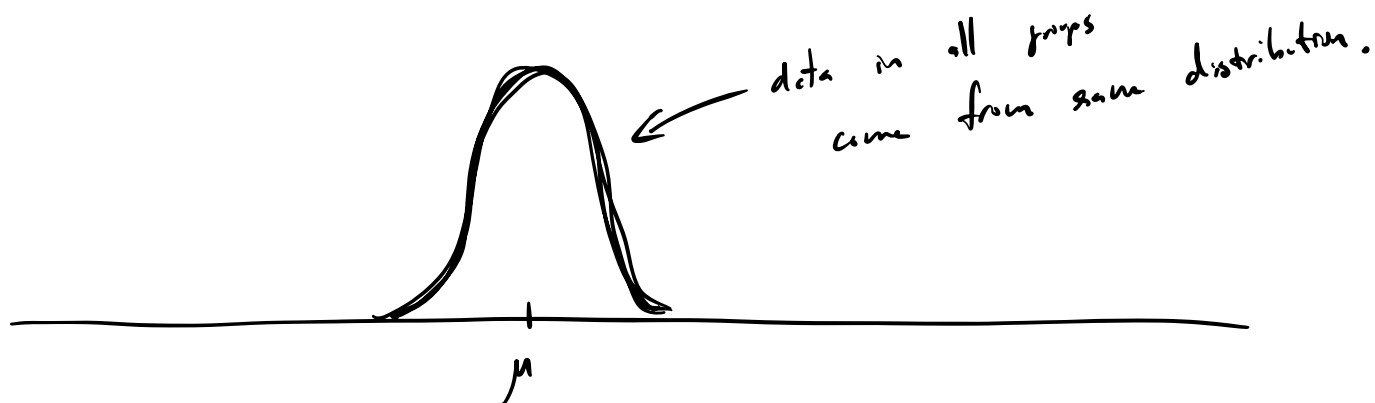
Under 1)



τ_1, τ_2, τ_3 are deviations from treatment group a .

μ is the mean of treatment group a .

Under $H_0: \tau_i = 0$ for all i



Treatment effect estimation in one-way ANOVA

Let $N = n_1 + \dots + n_a$ and define $\bar{Y} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}$ as well as

total #
of EUs

$$\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{for } i = 1, \dots, a.$$

treatment means

overall mean

The two most common ways to estimate $\mu, \tau_1, \dots, \tau_a$:

1. Deviations from baseline parameterization:

Set $\hat{\tau}_a = 0$ and $\hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}_{a\cdot}$ for $i = 1, \dots, a - 1$ and set $\hat{\mu} = \bar{Y}_{a\cdot}$.

2. Deviations from overall mean parameterization:

Set $\hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}_{..}$ for $i = 1, \dots, a$ and $\hat{\mu} = \bar{Y}_{..}$.

Under either parameterization, we have $\hat{\mu} + \hat{\tau}_i = \bar{Y}_{i\cdot}$ for $i = 1, \dots, a$.

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

Example :

$$a = 3$$

$$n_1 = n_2 = n_3 = 2$$

$$\begin{array}{l} \text{group 1} \\ \text{group 2} \\ \text{group 3} \end{array} \left(\begin{array}{l} Y_{11} = \mu + \tau_1 + \varepsilon_{11} \\ Y_{12} = \mu + \tau_1 + \varepsilon_{12} \\ Y_{21} = \mu + \tau_2 + \varepsilon_{21} \\ Y_{22} = \mu + \tau_2 + \varepsilon_{22} \\ Y_{31} = \mu + \tau_3 + \varepsilon_{31} \\ Y_{32} = \mu + \tau_3 + \varepsilon_{32} \end{array} \right)$$

Deviations from baseline:

Estimate $\hat{\mu} = \bar{Y}_3.$

$$\hat{\tau}_1 = \bar{Y}_1. - \bar{Y}_3.$$

$$\hat{\tau}_2 = \bar{Y}_2. - \bar{Y}_3.$$

$$\hat{\tau}_3 = 0$$

$$\begin{array}{l} \hat{Y}_{11} = \hat{\mu} + \hat{\tau}_1 = \bar{Y}_3. + (\bar{Y}_1. - \bar{Y}_3.) = \bar{Y}_1. \\ \hat{Y}_{12} = \hat{\mu} + \hat{\tau}_1 = \bar{Y}_3. + (\bar{Y}_1. - \bar{Y}_3.) = \bar{Y}_1. \\ \hat{Y}_{21} = \hat{\mu} + \hat{\tau}_2 = \bar{Y}_3. + (\bar{Y}_2. - \bar{Y}_3.) = \bar{Y}_2. \\ \hat{Y}_{22} = \hat{\mu} + \hat{\tau}_2 = \bar{Y}_3. + (\bar{Y}_2. - \bar{Y}_3.) = \bar{Y}_2. \\ \hat{Y}_{31} = \hat{\mu} + \hat{\tau}_3 = \bar{Y}_3. \\ \hat{Y}_{32} = \hat{\mu} + \hat{\tau}_3 = \bar{Y}_3. \end{array}$$

Derivatives from overall mean:

$$\text{Set } \hat{\mu} = \bar{y}_{..}, \quad \hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{..}$$

$$\hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{..}$$

$$\hat{\tau}_3 = \bar{y}_{3.} - \bar{y}_{..}$$

$$\hat{y}_{11} = \hat{\mu} + \hat{\tau}_1 = \bar{y}_{..} + (\bar{y}_{1.} - \bar{y}_{..}) = \bar{y}_{1.}$$

$$\hat{y}_{12} = \hat{\mu} + \hat{\tau}_1 = \bar{y}_{..} + (\bar{y}_{1.} - \bar{y}_{..}) = \bar{y}_{1.}$$

$$\hat{y}_{21} = \hat{\mu} + \hat{\tau}_2 = \bar{y}_{..} + (\bar{y}_{2.} - \bar{y}_{..}) = \bar{y}_{2.}$$

$$\hat{y}_{22} = \hat{\mu} + \hat{\tau}_2 = \bar{y}_{..} + (\bar{y}_{2.} - \bar{y}_{..}) = \bar{y}_{2.}$$

$$\hat{y}_{31} = \hat{\mu} + \hat{\tau}_3 = \bar{y}_{..} + (\bar{y}_{3.} - \bar{y}_{..}) = \bar{y}_{3.}$$

$$\hat{y}_{32} = \hat{\mu} + \hat{\tau}_3 = \bar{y}_{..} + (\bar{y}_{3.} - \bar{y}_{..}) = \bar{y}_{3.}$$

Why more than one way to estimate $\mu, \tau_1, \dots, \tau_a$?

same model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i=1, \dots, a, \quad j=1, \dots, n_i.$$
$$y_{ij} = (\mu + c) + (\tau_i - c) + \varepsilon_{ij}$$

Problem is: I cannot uniquely identify the values of μ and τ_1, \dots, τ_a .

Solution: Constrain the values in some way

(i) set $\tau_a = 0$. (ii) $\sum_{i=1}^a \tau_i = 0$

Rust inhibitors example (cont)

R uses by default the deviations from baseline parameterization:

```
# use as.factor() to designate brand as a "factor"  
lm_out <- lm(score ~ as.factor(brand), data = rust)  
lm_out
```

deviations from baseline parameterization:

Call:

```
lm(formula = score ~ as.factor(brand), data = rust)
```

Coefficients:

(Intercept) as.factor(brand)2 as.factor(brand)3 as.factor(brand)4
43.14 46.30 24.81 -2.67

```
# see how these are obtained from the group means  
aggregate(rust$score, by = list(rust$brand), FUN = mean)
```

Group.1	x
1	1 43.14
2	2 89.44
3	3 67.95
4	4 40.47

$$\bar{y}_2$$

$$\bar{y}_2 - \bar{y}_1$$
$$89.44 - 43.14$$

$$\bar{y}_3 - \bar{y}_1$$
$$67.95 - 43.14$$

$$\bar{y}_4 - \bar{y}_1$$
$$40.47 - 43.14$$

Estimation of the error term variance σ^2

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \varepsilon_{ij} \sim \text{Normal}(0, \sigma^2)$$

As in linear regression, define the

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{Y}_i$$

▶ fitted values \hat{Y}_{ij} as $\hat{Y}_{ij} = \bar{Y}_i$ for $j = 1, \dots, n_i$, and the

▶ residuals $\hat{\varepsilon}_{ij}$ as $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_i$.

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_i$$

for $j = 1, \dots, n_i$, $i = 1, \dots, a$.

Then an unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{N - a} \sum_{i=1}^a \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^2 = \frac{1}{N - a} \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2.$$

need to estimate a treatment means to get residuals

Divide by $N - a$ since the N residuals depend on a estimated quantities...

$$N = n_1 + \dots + n_a = \text{total sample size}$$

Rust inhibitors example (cont)

```
tab <- cbind(rust$brand,rust$score,lm_out$fitted.values,lm_out$residuals)
colnames(tab) <- c("brand","score","Fitted value","Residual")
head(tab,n = 13)
```

	Y_{ij}	$\hat{Y}_{ij} = \bar{Y}_i$	$Y_{ij} - \bar{Y}_i$	
brand	score	Fitted value	Residual	
1	1	43.9	43.14	0.76
2	1	39.0	43.14	-4.14
3	1	46.7	43.14	3.56
4	1	43.8	43.14	0.66
5	1	44.2	43.14	1.06
6	1	47.7	43.14	4.56
7	1	43.6	43.14	0.46
8	1	38.9	43.14	-4.24
9	1	43.6	43.14	0.46
10	1	40.0	43.14	-3.14
11	2	89.8	89.44	0.36
12	2	87.1	89.44	-2.34
13	2	92.7	89.44	3.26

Handwritten notes: $43.9 - 43.14$ next to the first residual; "average" with an arrow pointing to the fitted values for brand 1; a downward arrow next to the residual for brand 2.

$\hat{\sigma}^2 =$ `sum(lm_out$residuals^2) / (nrow(rust) - 4)`
`sgsqhat`

Handwritten annotations: $\hat{\epsilon}_{ij}$ under the residuals; N under the denominator; a downward arrow next to the `- 4`.

`[1] 6.139833` = $\hat{\sigma}^2$

An upward arrow points to the output value.

The value of $\hat{\sigma}$ is printed in the `summary()` output:

```
summary(lm_out)
```

Call:

```
lm(formula = score ~ as.factor(brand), data = rust)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-4.270 -1.597  0.395  1.275  4.730
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    43.1400    0.7836  55.056 <2e-16 ***
as.factor(brand)2  46.3000    1.1081  41.782 <2e-16 ***
as.factor(brand)3  24.8100    1.1081  22.389 <2e-16 ***
as.factor(brand)4  -2.6700    1.1081  -2.409  0.0212 *
```

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

Residual standard error: 2.478 ^{$\hat{\sigma}$} on 36 degrees of freedom

Multiple R-squared: 0.9863, Adjusted R-squared: 0.9852

F-statistic: 866.1 on 3 and 36 DF, p-value: < 2.2e-16

Sums of squares in the one-way ANOVA model

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

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij} \leftarrow \text{ignore the groups.}$$

As in linear regression we decompose the variation in the Y_{ij} by defining:

▶ Total sum of squares: $SS_{\text{Tot}} = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$

▶ Treatment sum of squares: $SS_{\text{Trt}} = \sum_{i=1}^a n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$

▶ Error sum of squares: $SS_{\text{Error}} = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$

sum of squared residuals

We have $SS_{\text{Tot}} = SS_{\text{Trt}} + SS_{\text{Error}}$.

Note that SS_{Trt} is computed just like SS_{Reg} in linear regression.

We again define $R^2 = \frac{SS_{\text{Trt}}}{SS_{\text{Tot}}}$.

$$SS_{\text{Reg}} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\hat{Y}_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^a n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

= $\bar{Y}_{i.}$, the mean of treatment group i

Sampling distributions of our sums of squares

The SS, appropriately scaled, follow chi-square distributions:

▶ $SS_{\text{Tot}} / \sigma^2 \sim \chi_{N-1}^2(\phi_{\text{Tot}})$

▶ $SS_{\text{Trt}} / \sigma^2 \sim \chi_{a-1}^2(\phi_{\text{Trt}})$

▶ $SS_{\text{Error}} / \sigma^2 \sim \chi_{N-a}^2$

where ϕ_{Tot} and ϕ_{Trt} are noncentrality parameters.

The mean squares in the one-way ANOVA model

Dividing SS_{Trt} and SS_{Error} by their dfs, we define:

▶ Treatment mean square: $MS_{\text{Trt}} = \frac{SS_{\text{Trt}}}{a - 1}$ $\leftarrow a-1$ degrees of freedom.

▶ Error mean square: $MS_{\text{Error}} = \frac{SS_{\text{Error}}}{N - a} = \hat{\sigma}^2$

The ratio $F_{\text{stat}} = \frac{MS_{\text{Trt}}}{MS_{\text{Error}}}$ has an F distribution.

The Analysis of Variance (ANOVA) table

$$SS_{TRT} = \sum_{i=1}^a n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 =$$

sum of a quantities after estimating 1 overall mean.

We often present the SS, df, and MS values in a table like this:

Source	Df	SS	MS	F value	p-value
Treatment	$a - 1$	SS_{Trt}	MS_{Trt}	F_{stat}	$P(F > F_{stat})$
Error	$N - a$	SS_{Error}	MS_{Error}		
Total	$N - 1$	SS_{Tot}			

sum of N residuals obtained after estimating a treatment means

In the table $F_{stat} = \frac{MS_{Trt}}{MS_{Error}}$.

The p-value is based on $F \sim F_{a-1, N-a}$.

$$N = n_1 + \dots + n_a$$

$$a = 4$$

$$N = 40$$

$$N - a = 40 - 4 = 36$$

Rust inhibitors example (cont)

Obtain the ANOVA table with the `anova()` function on the `lm()` output.

```
anova(lm_out)
```

Analysis of Variance Table

Response: score

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(brand)	3	15954	5317.8	866.12	< 2.2e-16 ***
Residuals	36	221	6.1		

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

$4 - 1$ (4 brands - 1)

$$\frac{5317.8}{6.1} = 866.12$$

$$MS_{Error} = \hat{\sigma}^2$$

$$36 = 40 - 4$$

↑ 4 brands

$$N = n_1 + n_2 + n_3 + n_4 = 10 + 10 + 10 + 10$$

Testing whether there is any difference in treatment means

Given $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $j = 1, \dots, n_i$, $i = 1, \dots, a$, we wish to test

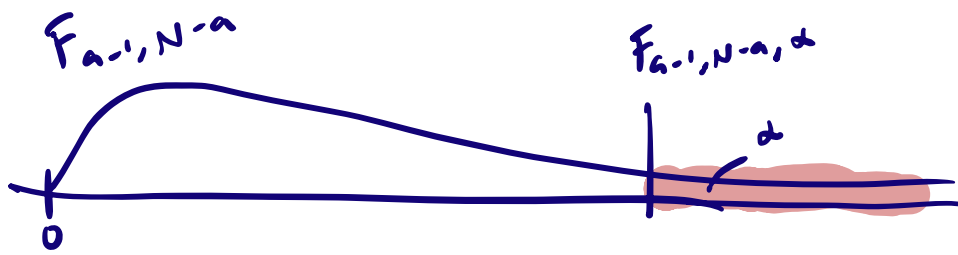
$H_0: \tau_i = 0$ for all i versus H_1 : At least one τ_i is nonzero.

none of the treatment means differs from the baseline.

We use the overall F test of significance:

1. Compute $F_{\text{stat}} = \frac{MS_{\text{Trt}}}{MS_{\text{Error}}}$ Under $H_0 \sim F_{a-1, N-a}$
2. Reject H_0 at α if $F_{\text{stat}} > F_{a-1, N-a, \alpha}$.
3. Obtain p-value as $P(F > F_{\text{stat}})$, where $F \sim F_{a-1, N-a}$.

The value of F_{stat} and the p-value are printed in the `summary()` output.



Interpretation of F statistic

$$F_{\text{stat}} = \frac{MS_{\text{TRT}}}{MS_{\text{Error}}} = \frac{SS_{\text{TRT}} / (a-1)}{SS_{\text{Error}} / (N-a)} = \frac{\sum_{i=1}^a n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 / (a-1)}{\sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 / (N-a)}$$

Variation between treatment means.

σ^2

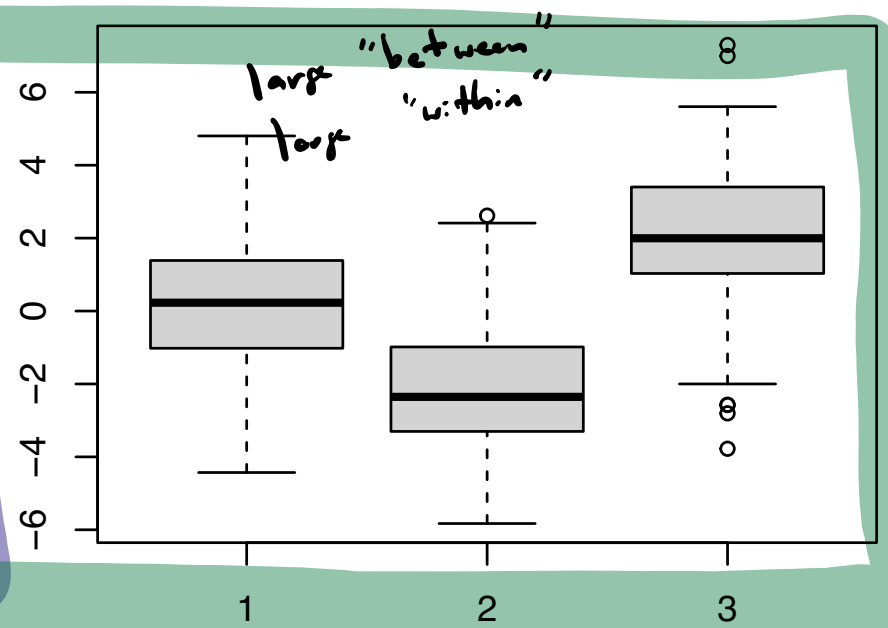
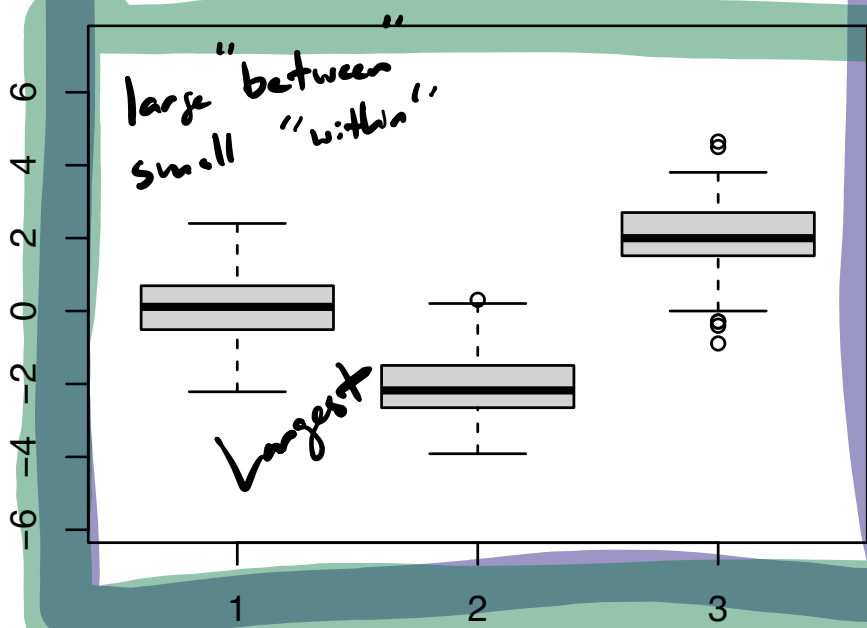
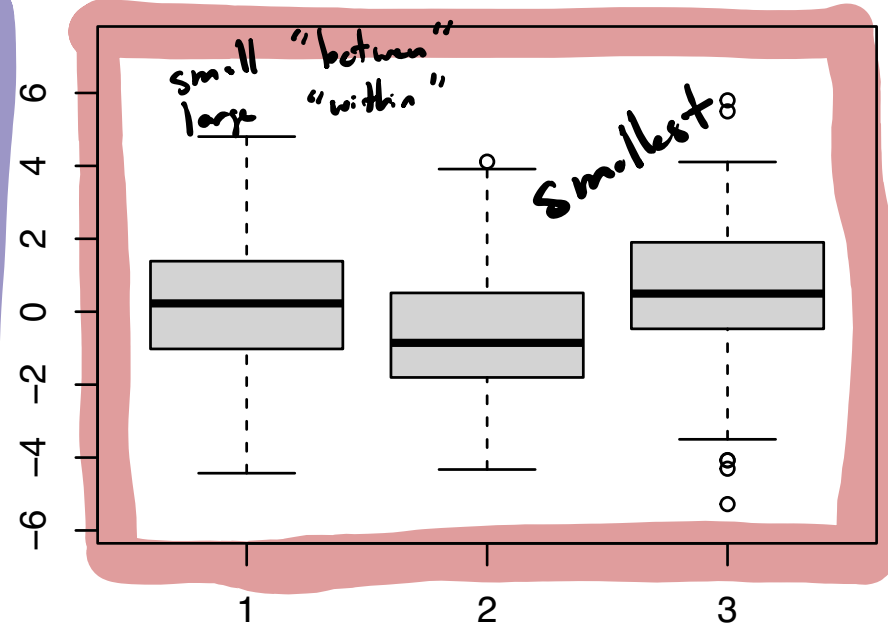
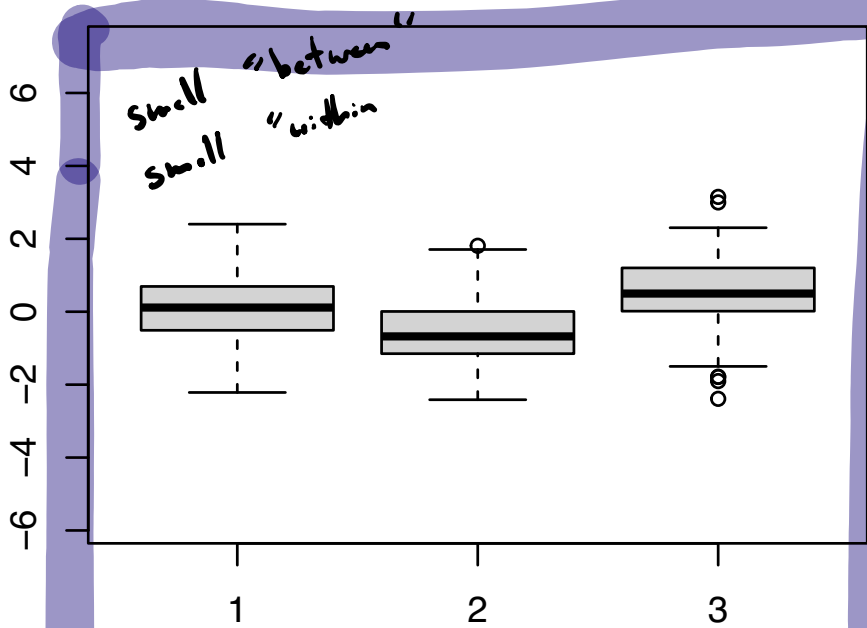
Note that F_{stat} is a ratio of the form $\frac{\text{Between treatment variation}}{\text{Within treatment variation}}$.

H_0 : All treatment means are equal.

Larger F_{stat} carries more evidence against H_0 .

Exercise: For which data set will the F-statistic be largest/smallest?

$q=3$



$$F_{\text{stat}} = \frac{MS_{\text{Treat}}}{MS_{\text{Error}}} = \frac{\text{"Between"}}{\text{"Within"}}$$

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{N} \sum_{i=1}^a n_i \bar{Y}_{i.} = \frac{10}{40} \sum_{i=1}^4 \bar{Y}_{i.}$$

Exercise: Compute F_{stat} for the rust data using the summary info:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

group	replicates	mean	standard deviation
1	10	43.14	3.00
2	10	89.44	2.22
3	10	67.95	2.17
4	10	40.47	2.44

$$a = 4$$

$$n_1 = n_2 = n_3 = n_4 = 10$$

$$N = 40$$

sample variance in treat group i .

Hint: $SS_{\text{Error}} = \sum_{i=1}^a (n_i - 1) S_i^2$, where $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$

$$(10 - 1) S_i^2 = \sum_{j=1}^{10} (Y_{ij} - \bar{Y}_{i.})^2$$

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{40 - 4}$$

$$= \frac{1}{36} \sum_{i=1}^4 \sum_{j=1}^{10} (Y_{ij} - \bar{Y}_{i.})^2 = \frac{1}{36} \sum_{i=1}^4 (10 - 1) S_i^2$$

$$= \frac{1}{36} (10-1) \sum_{i=1}^4 s_i^2 = \frac{9}{36} ((3.00)^2 + (2.22)^2 + (2.17)^2 + (2.49)^2)$$

$$= 6.15$$

$$MS_{\text{Error}} = 6.15$$

$$MS_{\text{TFT}} = \frac{\sum_{i=1}^4 n_i (\bar{y}_{i.} - \bar{y}_{..})^2}{4-1} = \frac{1}{4-1} \sum_{i=1}^4 10 (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \frac{1}{4-1} 10 \left((43.14 - 60.25)^2 + (89.44 - 60.25)^2 + (67.95 - 60.25)^2 + (40.47 - 60.25)^2 \right)$$

$$= 5317.8$$

$$\bar{y}_{..} = \frac{1}{4} (43.14 + 89.44 + 67.95 + 40.47) = 60.25$$

$$F_{\text{stat}} = \frac{5317.8}{6.15} \approx 866 \dots$$

Post-hoc comparisons of means

↑ after F test rejects $H_0: \mu_1 = \dots = \mu_a$.

μ_1 vs μ_2
 μ_2 vs μ_3 ...

- ▶ If we reject $H_0: \mu_1 = \dots = \mu_a$, then we may wish to compare means.
- ▶ Call such comparisons post-hoc as they follow the F-test.
- ▶ We may wish to compare several pairs of means, which is like testing several hypotheses at once.
- ▶ When several hypotheses are tested at once, the familywise Type I error rate is the probability that *any* Type I error is committed.
- ▶ We discuss two methods for post-hoc comparisons of means which control the familywise Type I error rate.

Suppose we want to test these two sets of hypotheses:

$$H_0: \mu_1 = \mu_2 \quad \text{vs} \quad H_1: \mu_1 \neq \mu_2$$

$$H_0: \mu_2 = \mu_3 \quad \text{vs} \quad H_1: \mu_2 \neq \mu_3$$

Comparing all pairs of means

μ_1, \dots, μ_a . Compare μ_i and μ_j for all $i \neq j$.
 Test $H_0: \mu_i = \mu_j$ vs $H_1: \mu_i \neq \mu_j$ for $i \neq j$.

- ▶ We want to build a CI for $\mu_i - \mu_j$ for all pairs $i \neq j$.
- ▶ Suppose the design is balanced, i.e. $n_i = n$ for all $i = 1, \dots, a$.
- ▶ If we build for all $i \neq j$ the ordinary $(1 - \alpha) \times 100\%$ CIs

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm t_{a(n-1), \alpha/2} \hat{\sigma} \sqrt{2/n},$$

← ordinary two-sample CI for a difference in means.

each one will cover its target with probability $1 - \alpha$.

- ▶ We want *simultaneous* coverage with probability $1 - \alpha$.
- ▶ I.e., we want the familywise coverage of all the intervals to be $1 - \alpha$.

a groups

$$\binom{a}{2} = \frac{a!}{2!(a-2)!}$$

$a = 4$

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (2 \cdot 1)} = 6.$$

Multiple comparisons of means with Tukey's HSD

HSD: Honest significant difference.

- ▶ Suppose the design is balanced, i.e. $n_i = n$ for all $i = 1, \dots, a$.
- ▶ Suppose we could find the value $q_{a, a(n-1), \alpha}$ such that

$$P \left(\max_{i \neq j} \left\{ \frac{|\bar{Y}_{i.} - \bar{Y}_{j.}| - (\mu_i - \mu_j)|}{\hat{\sigma} / \sqrt{n}} \right\} \leq q_{a, a(n-1), \alpha} \right) = 1 - \alpha.$$

- ▶ Then with probability $1 - \alpha$ the CIs

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm q_{a, a(n-1), \alpha} \hat{\sigma} / \sqrt{n}$$

$\sqrt{MS_{Error}}$

will simultaneously cover the targets $\mu_i - \mu_j$ for all $i \neq j$. **Show!**

- ▶ Tukey made tables of the values $q_{a, a(n-1), \alpha}$.
- ▶ Can use the simultaneous intervals to sort/compare the means.

$q_{a, a(n-1), \alpha}$
 \uparrow # treatment groups
 $a(n-1) = N - a = df$ for SS_{Error}

$$P\left(\max_{i \neq j} \left\{ \frac{|(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - (\mu_i - \mu_j)|}{\hat{\sigma} / \sqrt{n}} \leq \delta_\alpha \right\} \right) = 1 - \alpha$$

"

$$P\left(\bigcap_{i \neq j} \left\{ \frac{|(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - (\mu_i - \mu_j)|}{\hat{\sigma} / \sqrt{n}} \leq \delta_\alpha \right\} \right) = 1 - \alpha$$

"

$$P\left(\bigcap_{i \neq j} \left\{ \mu_i - \mu_j \in \left[\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm \delta_\alpha \frac{\hat{\sigma}}{\sqrt{n}} \right] \right\} \right) = 1 - \alpha$$

So the intervals

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm \delta_\alpha \frac{\hat{\sigma}}{\sqrt{n}}, \quad i \neq j$$

simultaneously over $\mu_i - \mu_j$, $i \neq j$, with prob $1 - \alpha$.

Table A.6 Critical Values of the Studentized Range, for Tukey's HSD.

Error df	Two-sided α	Number of Groups							
		2	3	4	5	6	7	8	
5	0.05	3.64	4.6	5.22	5.67	6.03	6.33	6.58	
5	0.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	
6	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	
6	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	
7	0.01	4.95	5.92	6.54	7.00	7.37	7.68	7.94	
8	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	
8	0.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	
9	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	
9	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	
10	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	
11	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	
12	0.01	4.32	5.05	5.50	5.84	6.1	6.32	6.51	
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	
13	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	
14	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	
15	0.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	
15	0.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	
16	0.01	4.13	4.79	5.19	5.49	5.72	5.91	6.08	
17	0.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	
17	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	
18	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	
19	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	
20	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	
25	0.05	2.91	3.52	3.89	4.15	4.36	4.53	4.67	
25	0.01	3.94	4.53	4.88	5.14	5.35	5.51	5.65	
30	0.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	
30	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	
40	0.01	3.82	4.37	4.69	4.93	5.11	5.26	5.39	
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	
60	0.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	

Table produced using the SAS System using function PROBMCR(SRANGE', 1 - α , df, T).

Rust:
 $a = 4$ treatments
 $n = 10$ replicates in each group
 $\alpha = 0.05$

$a(n-1) = N - a$

$f_{\alpha, a(n-1), d}$
 "
 $f_{4, 4(10-1), 0.05}$
 "
 $f_{4, 36, 0.05}$

$\bar{y}_i - \bar{y}_j \pm 3.85 \frac{\hat{\sigma}}{\sqrt{n}}$

$\hat{\sigma} = \sqrt{MSE}$

smaller error df

3.809 falls between

Figure 1: Table A.6 from Mohr, Wilson, and Freund (2021)

Rust inhibitors example (cont)

For the rust data we have $n = 10$ and $a = 4$.

At $\alpha = 0.05$ we have $q_{a,a(n-1),\alpha} = q_{4,36,0.05} \approx 3.85$ from table.

Obtain exact value with `qtukey(.95, 4, 36) = 3.8087984`.

Build the Tukey HSD CI for $\mu_2 - \mu_1$.

a $a(n-1) = N - a$

```
MSE <- sum(lm_out$residuals^2) / (40 - 4)
```

```
y1bar <- mean(rust$score[rust$brand == 1])
```

```
y2bar <- mean(rust$score[rust$brand == 2])
```

```
lo21 <- y2bar - y1bar - qtukey(.95, 4, 36) * sqrt(MSE) / sqrt(10)
```

```
up21 <- y2bar - y1bar + qtukey(.95, 4, 36) * sqrt(MSE) / sqrt(10)
```

```
c(lo21, up21)
```

$\hat{\sigma}^2$
 y_1
 y_2

$n = 10$

```
[1] 43.31554 49.28446
```



Rust inhibitors example (cont)

Use TukeyHSD() on aov() output to obtain the simultaneous CIs.

```
# must use the aov() function instead of the lm() function
aov_out <- aov(score ~ as.factor(brand), data = rust)
TukeyHSD(aov_out)
```

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = score ~ as.factor(brand), data = rust)

```
$`as.factor(brand)`
      diff      lwr      upr      p adj
2-1  46.30  43.315536  49.2844635 0.0000000
3-1  24.81  21.825536  27.7944635 0.0000000
4-1  -2.67  -5.654464   0.3144635 0.0933303
3-2 -21.49 -24.474464 -18.5055365 0.0000000
4-2 -48.97 -51.954464 -45.9855365 0.0000000
4-3 -27.48 -30.464464 -24.4955365 0.0000000
```

$\mu_2 - \mu_1$
 $\mu_3 - \mu_1$
 $\mu_4 - \mu_1$
 $\mu_3 - \mu_2$
...

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

can trust these comparisons.

Dunnett's, for comparison

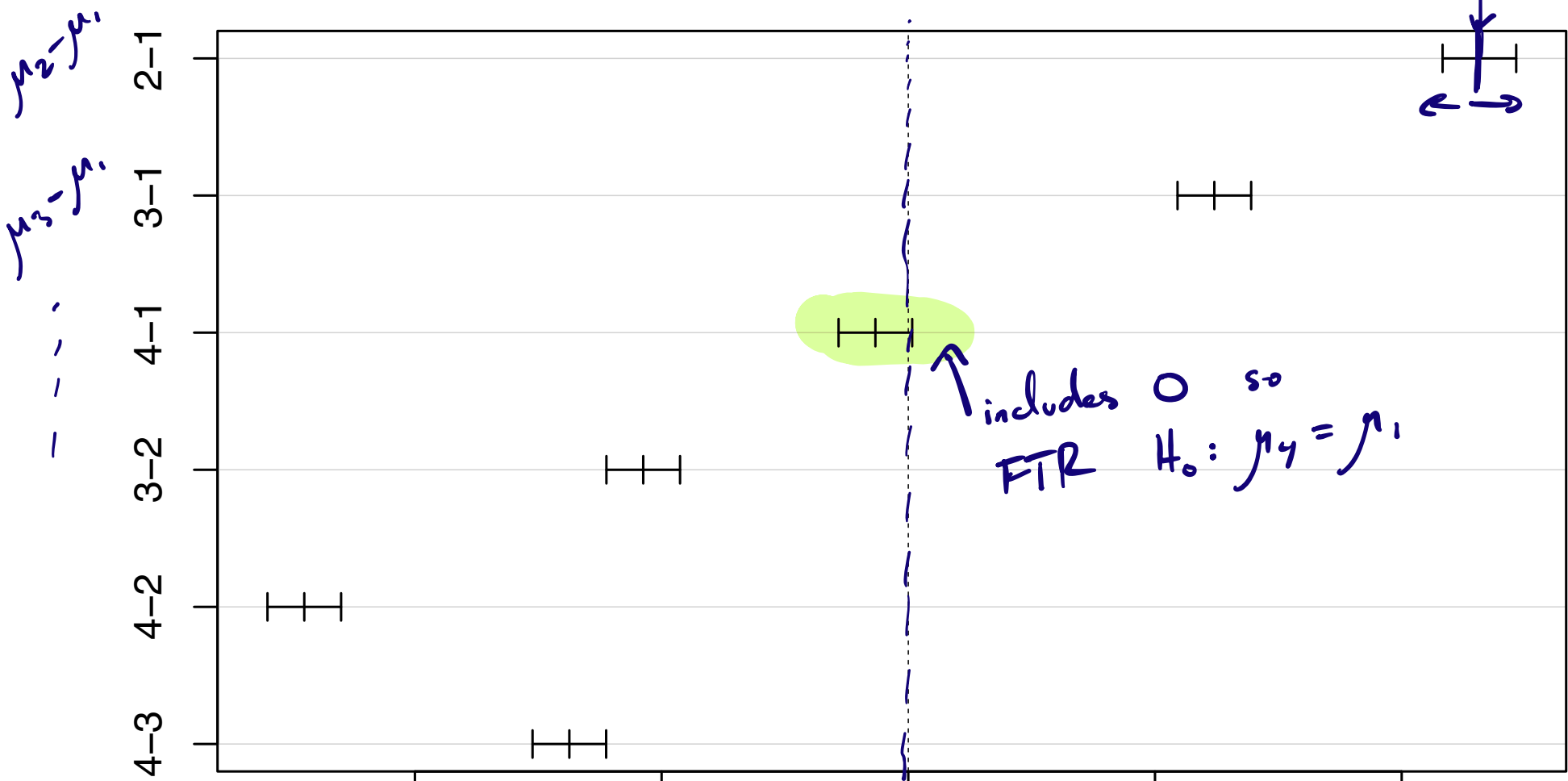
$\mu_2 - \mu_1$ | lwr.ci | upr.ci
43.582516 49.017484

↑ slightly narrower.

```
plot(TukeyHSD(aov_out))
```

95% family-wise confidence level

$\bar{y}_{2.} - \bar{y}_{1.}$



Differences in mean levels of as.factor(brand)

Comparison of treatments with a baseline treatment

Tukey's is for comparing all pairs of means: $\binom{a}{2}$ pairs.

Let treatment a is baseline. Compare μ_i to μ_a , $i=1, \dots, a-1$.

- ▶ It may be that not all pairwise comparisons are of interest.
- ▶ Then Tukey's method is too conservative (CIs wider than necessary).
- ▶ Say we want to compare all treatments to a "baseline" treatment.
- ▶ Build CIs for $\mu_j - \mu_a$, $j = 1, \dots, a - 1$, a the baseline treatment.
- ▶ This is $a - 1$ CIs instead of $\binom{a}{2}$ CIs.
- ▶ Can use Dunnett's method.

Dunnett's method for comparisons with a baseline

▶ Assume $n_i = n$ for all i (balanced case).

▶ Given a value $d_{n,a(n-1),\alpha}$ such that

$$P \left(\max_{1 \leq i \leq a-1} \left| \frac{(\bar{Y}_i - \bar{Y}_{a.}) - (\mu_i - \mu_a)}{\hat{\sigma} \sqrt{2/n}} \right| \leq d_{n,a(n-1),\alpha} \right) = 1 - \alpha,$$

with probability $1 - \alpha$ the CIs

$$\bar{Y}_i - \bar{Y}_{a.} \pm d_{n,a(n-1),\alpha} \hat{\sigma} \sqrt{2/n}$$

will simultaneously cover the targets $\mu_i - \mu_a$ for all $i = 1, \dots, a - 1$.

▶ Dunnett made tables of the values $d_{n,a(n-1),\alpha}$.

▶ Cannot sort the means after Dunnett's.

Post:
 $a = 4$
 $n = 10$
 $i = 1, 2, 3, 4$
 Say Brand 1 is the "control".
 Build CIs for
 $\mu_2 - \mu_1$
 $\mu_3 - \mu_1$
 $\mu_4 - \mu_1$

$a(n-1) = N - a$

Table A.5 Critical Values for Dunnett's Two-Sided Test of Treatments versus Control.

Error df	Two-sided α	Number of Groups Counting Both Treatments and Control							
		2	3	4	5	6	7	8	
5	0.05	2.57	3.03	3.29	3.48	3.62	3.73	3.82	
5	0.01	4.03	4.63	4.97	5.22	5.41	5.56	5.68	
6	0.05	2.45	2.86	3.10	3.26	3.39	3.49	3.57	
6	0.01	3.71	4.21	4.51	4.71	4.87	5.00	5.10	
7	0.05	2.36	2.75	2.97	3.12	3.24	3.33	3.41	
7	0.01	3.50	3.95	4.21	4.39	4.53	4.64	4.74	
8	0.05	2.31	2.67	2.88	3.02	3.13	3.22	3.29	
8	0.01	3.36	3.77	4.00	4.17	4.29	4.40	4.48	
9	0.05	2.26	2.61	2.81	2.95	3.05	3.14	3.20	
9	0.01	3.25	3.63	3.85	4.01	4.12	4.22	4.30	
10	0.05	2.23	2.57	2.76	2.89	2.99	3.07	3.14	
10	0.01	3.17	3.53	3.74	3.88	3.99	4.08	4.16	
11	0.05	2.20	2.53	2.72	2.84	2.94	3.02	3.08	
11	0.01	3.11	3.45	3.65	3.79	3.89	3.98	4.05	
12	0.05	2.18	2.50	2.68	2.81	2.90	2.98	3.04	
12	0.01	3.05	3.39	3.58	3.71	3.81	3.89	3.96	
13	0.05	2.16	2.48	2.65	2.78	2.87	2.94	3.00	
13	0.01	3.01	3.33	3.52	3.65	3.74	3.82	3.89	
14	0.05	2.14	2.46	2.63	2.75	2.84	2.91	2.97	
14	0.01	2.98	3.29	3.47	3.59	3.69	3.76	3.83	
15	0.05	2.13	2.44	2.61	2.73	2.82	2.89	2.95	
15	0.01	2.95	3.25	3.43	3.55	3.64	3.71	3.78	
16	0.05	2.12	2.42	2.59	2.71	2.80	2.87	2.92	
16	0.01	2.92	3.22	3.39	3.51	3.60	3.67	3.73	
17	0.05	2.11	2.41	2.58	2.69	2.78	2.85	2.90	
17	0.01	2.90	3.19	3.36	3.47	3.56	3.63	3.69	
18	0.05	2.10	2.40	2.56	2.68	2.76	2.83	2.89	
18	0.01	2.88	3.17	3.33	3.44	3.53	3.60	3.66	
19	0.05	2.09	2.39	2.55	2.66	2.75	2.81	2.87	
19	0.01	2.86	3.15	3.31	3.42	3.50	3.57	3.63	
20	0.05	2.09	2.38	2.54	2.65	2.73	2.80	2.86	
20	0.01	2.85	3.13	3.29	3.40	3.48	3.55	3.60	
25	0.05	2.06	2.34	2.50	2.61	2.69	2.75	2.81	
25	0.01	2.79	3.06	3.21	3.31	3.39	3.45	3.51	
30	0.05	2.04	2.32	2.47	2.58	2.66	2.72	2.77	
30	0.01	2.75	3.01	3.15	3.25	3.33	3.39	3.44	
40	0.05	2.02	2.29	2.44	2.54	2.62	2.68	2.73	
40	0.01	2.70	2.95	3.09	3.19	3.26	3.32	3.37	
60	0.05	2.00	2.27	2.41	2.51	2.58	2.64	2.69	
60	0.01	2.66	2.90	3.03	3.12	3.19	3.25	3.29	

$\alpha = 0.05$
 $\bar{y}_{.j} - \bar{y}_{.1} \pm 2.47 \hat{\sigma} \sqrt{\frac{1}{n}}$
 $d_{a, a(n-1), \alpha}$
 $d_{4, 36, 0.05}$
 $\hat{\sigma} = \sqrt{MSE}$
 take larger

This table produced from the SAS System using function PROBMC('DUNNETT2', ..., 1 - α , df, k), where $k = T - 1$.

Figure 2: Table A.5 from Mohr, Wilson, and Freund (2021)

Rust inhibitor data (cont)

For the rust data we have $n = 10$ and $a = 4$.

At $\alpha = 0.05$ we have $d_{a,a(n-1),\alpha} = d_{4,36,0.05} \approx 2.47$

Use value 2.47 in the table (should be close).

Treat Brand 1 as the baseline and make comparisons with Dunnett's.

$\mu_2 - \mu_1$

```
y1bar <- mean(rust$score[rust$brand == 1])
y2bar <- mean(rust$score[rust$brand == 2])
lo21 <- y2bar - y1bar - 2.47 * sqrt(MSE) * sqrt(2/10)
up21 <- y2bar - y1bar + 2.47 * sqrt(MSE) * sqrt(2/10)
c(lo21, up21)
```

[1] 43.59615 49.00385

Rust inhibitor data (cont)

Use `DunnettTest()` from R package DescTools.

```
library(DescTools) # first time run install.packages("DescTools")
Dunnett_out <- DunnettTest(score ~ as.factor(brand), data = rust, control = "1")
Dunnett_out
```

specify control/
baseline treatment.

Dunnett's test for comparing several treatments with a control :
95% family-wise confidence level

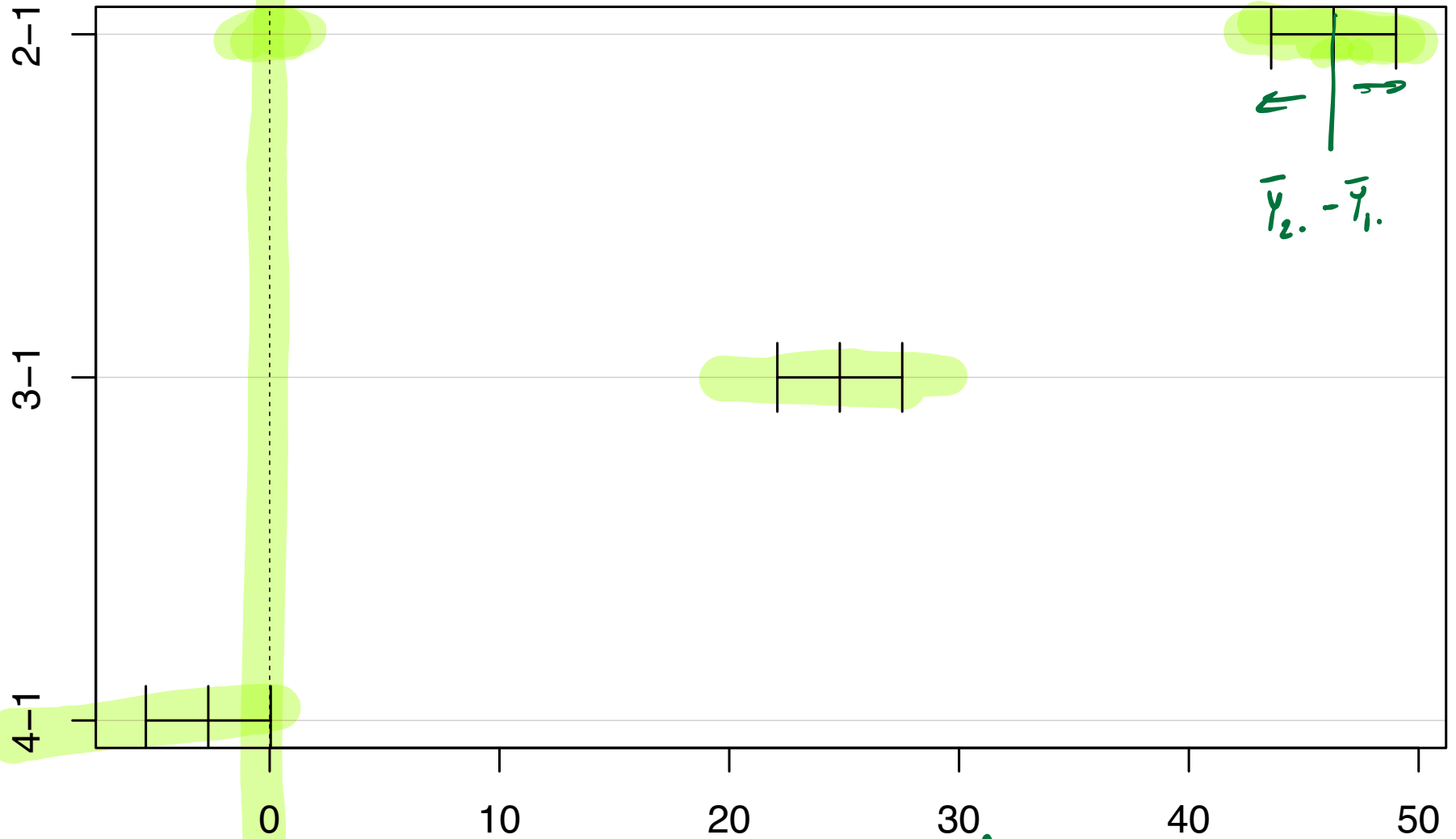
```
$`1`
      diff   lwr.ci   upr.ci   pval
 $\mu_2 - \mu_1$  2-1 46.30 43.582516 49.017484 <2e-16 ***
 $\mu_3 - \mu_1$  3-1 24.81 22.092516 27.527484 <2e-16 ***
 $\mu_4 - \mu_1$  4-1 -2.67 -5.387484  0.047484 0.0549 .
```

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

plot(Dunnett_out)

95% family-wise confidence level

(43.58, 49.02)



Differences in mean levels μ_1

Dunnett's vs Tukey's

- ▶ Tukey's is for comparisons between all pairs of means.
- ▶ Dunnett's is for comparison of means with a baseline.
- ▶ So Tukey's must make greater adjustments to control the familywise Type I error.
- ▶ Tukey intervals will be wider than Dunnett intervals.
- ▶ Tukey's allows you to sort the means, while Dunnett's does not.

▶ Both methods (as presented here) require a balanced design,
i.e. $n_i = n$ for $i = 1, \dots, a$.

Modifications for unbalanced designs exist, but are not straightforward to implement.

Bonferroni correction - For any set of comparisons.

Tukey's and Dunnett's and Bonferroni are all methods for multiple comparisons, i.e. building multiple C.I.s.

If building B CIs you can ALWAYS use the Bonferroni correction:

- ▶ Build each CI ordinarily, but use α/B instead of α .
- ▶ Ensures simultaneous coverage of all CIs with probability $\geq 1 - \alpha$.
- ▶ True probability of simultaneous coverage may be greater.
- ▶ Bonferroni-corrected CIs will be wider than Dunnett's and wider than Tukey's if used for making those same comparisons.
- ▶ Use when we do not know how to adjust for multiple comparisons.

95%:

$$\alpha = 0.05$$

$$B = 10$$

→

use

$$\alpha^* = \frac{0.05}{10} = 0.005.$$

makes each interval wider.

$$a = 4 : \text{Make } \binom{a}{2} = \binom{4}{2} = 6 \text{ pairwise comparisons. } \Rightarrow \text{Bonferroni says use } \frac{\alpha}{6}.$$

Rust inhibitor data (cont)

$$40 - 4 = 36$$

↑
N

$$\bar{y}_3 - \bar{y}_4 \pm t_{36, \underbrace{(0.05/2)}_{\text{Bonferroni}}} / 2 \hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{n}}$$

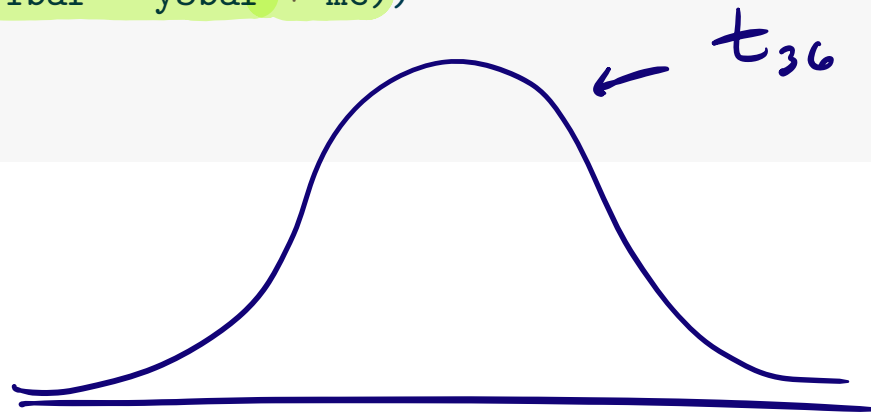
$\hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{n}}$
 $\sqrt{MSE_{Error}}$

Compare Brand 3 to 4 and Brand 1 to 3, using the Bonferroni correction to control the familywise error rate.

```

y1bar <- mean(rust$score[rust$brand == 1])
y3bar <- mean(rust$score[rust$brand == 3])
y4bar <- mean(rust$score[rust$brand == 4])
alpha <- 0.05
B <- 2 = # C.I.s
me <- qt(1 - (alpha/B)/2, a*(n-1)) * sqrt(MSE) * sqrt(2/n)
tab <- rbind(c(y3bar - y4bar - me, y3bar - y4bar + me),
             c(y1bar - y3bar - me, y1bar - y3bar + me))
rownames(tab) <- c("3-4", "1-3")
colnames(tab) <- c("lower", "upper")
tab
    
```

	lower	upper
3-4	24.888	30.072
1-3	-27.402	-22.218



Checking model assumptions

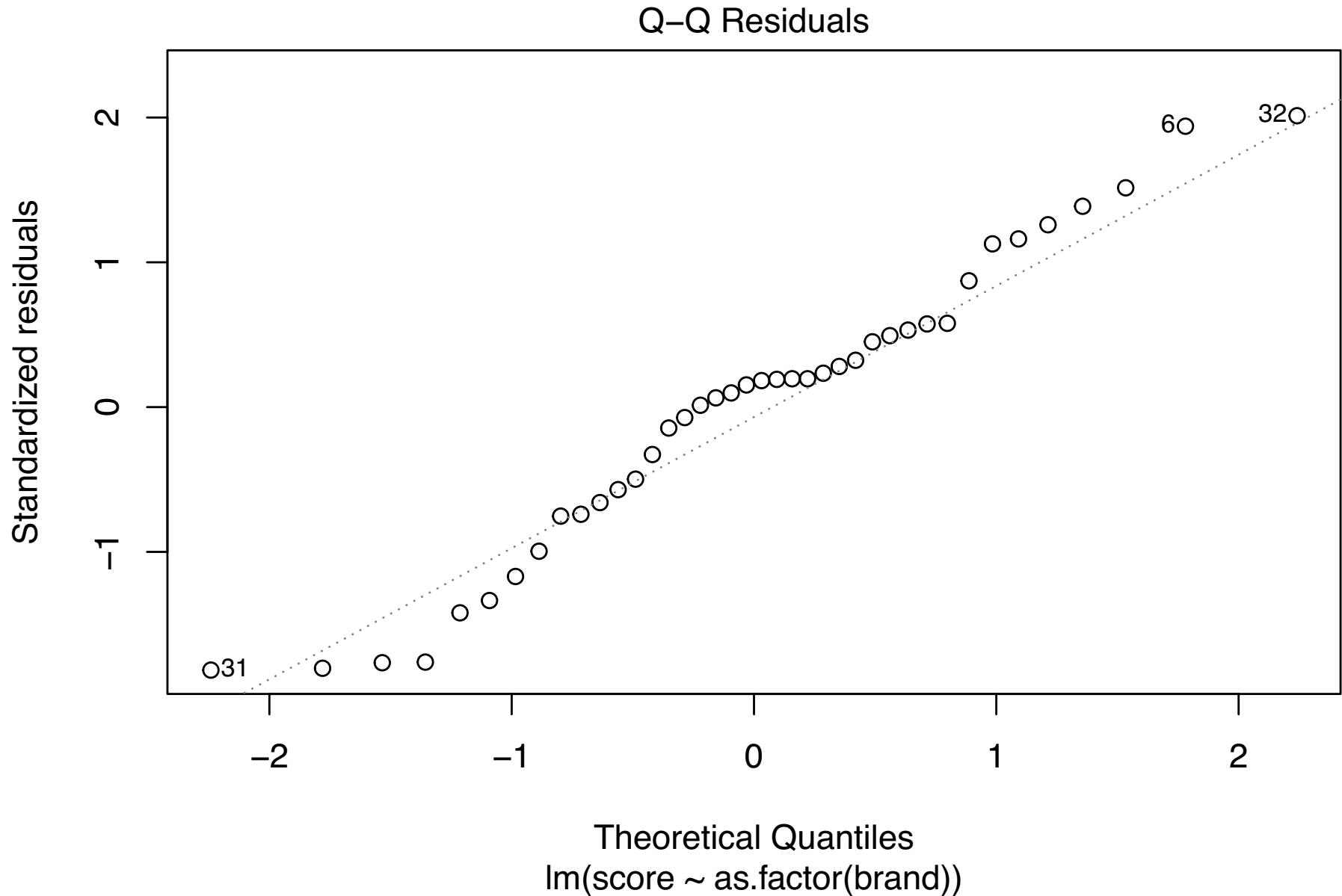
$$Y_{ij} = \underbrace{\mu + \tau_i}_{\text{treatment mean}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Validity of the foregoing analyses depends on these assumptions:

1. The responses are normally distributed around the treatment means (Check QQ plot of residuals).
2. The response has the same variance in all treatment groups (Check residuals vs fitted values plot).
3. The response values are independent of each other (No way to check; must trust experimental design).

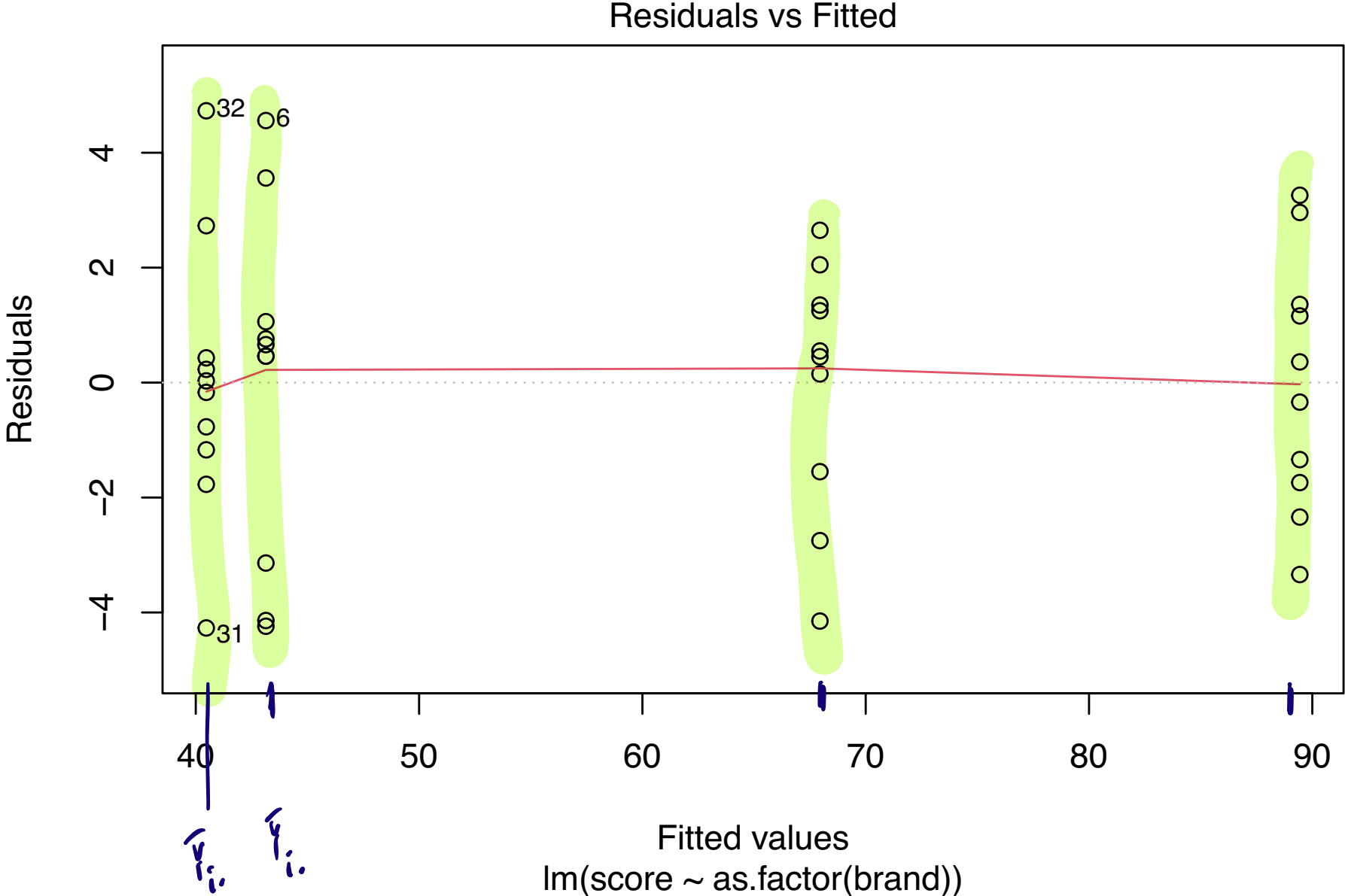
Rust inhibitors example (cont)

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



Rust inhibitors example (cont)

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



Perception of slope example

WILL NOT TEST
ABOUT LEVENE'S
TEST.

Do axis re-scalings affect how we perceive an x-y relationship?

For a single data set with data pairs (X_i, Y_i) , with $X_i \sim \text{Normal}(0, 1)$ and $Y_i = \text{Normal}(X_i, 1)$ for $i = 1, \dots, 50$, three scatterplot treatments were constructed:

1. "Control" used x and y plotting limits given by the range of the data.
2. "X" extended the x-limits by 1.5 in each direction.
3. "Y" extended the y-limits by 1.5 in each direction.

Each student in a class was randomly assigned a scatterplot and told to draw with a ruler the best-fitting line through the data. The slope of each student-drawn line was measured and recorded as the response.

Is the response mean the same in the three treatment groups?

An artifact from each treatment group:

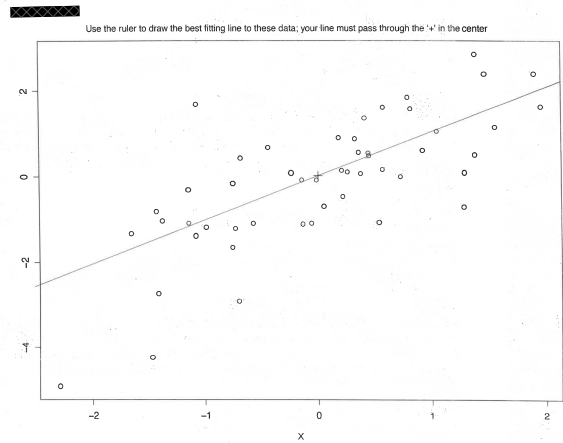


Figure 3: "Control"

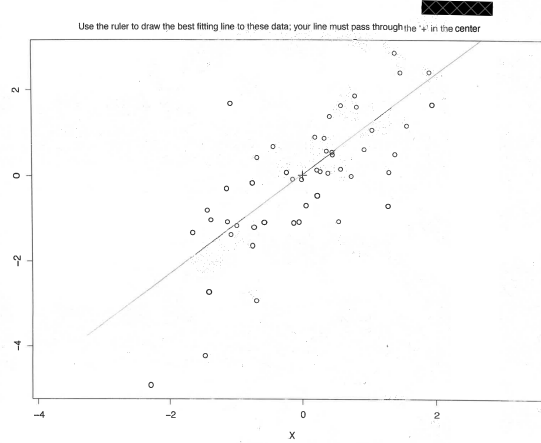


Figure 4: "X"

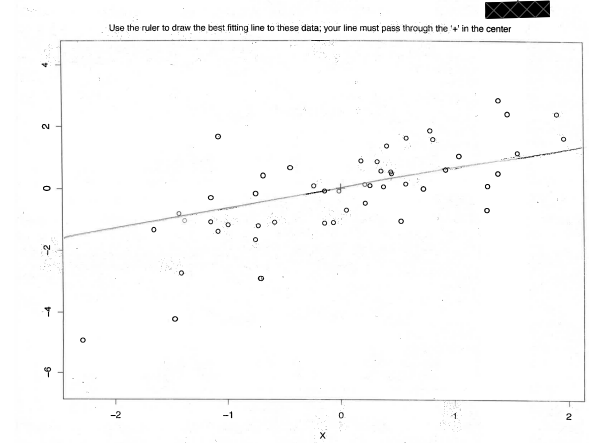
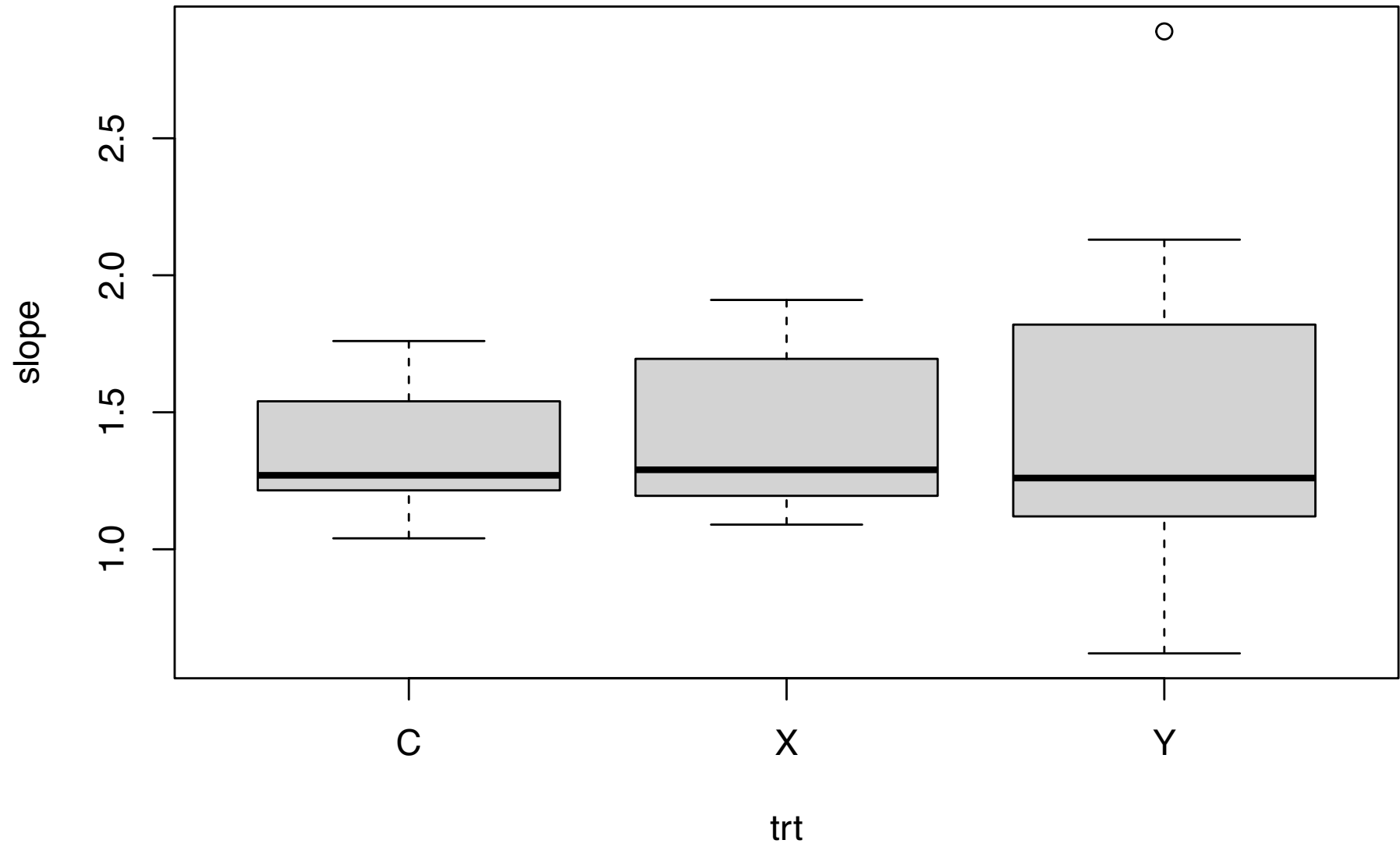


Figure 5: "Y"

```
slope <- c(1.23,1.80,1.81,1.29,2.89,1.58,0.99,1.24,  
          1.26,1.57,1.27,1.19,1.82,1.76,1.91,1.25,  
          1.09,1.29,1.12,1.51,2.13,1.16,0.62,1.04)  
trt <- c("X","Y","X","X","Y","X","Y","C",  
        "Y","C","C","C","Y","C","X","Y",  
        "X","X","Y","C","Y","X","Y","C")
```



```
boxplot(slope ~ trt)
```



```
lm_slope <- lm(slope ~ as.factor(trt))
summary(lm_slope)
```

Call:

```
lm(formula = slope ~ as.factor(trt))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9222	-0.2847	-0.1293	0.2628	1.3478

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.36857	0.18161	7.536	2.12e-07 ***
as.factor(trt)X	0.05143	0.24868	0.207	0.838
as.factor(trt)Y	0.17365	0.24215	0.717	0.481

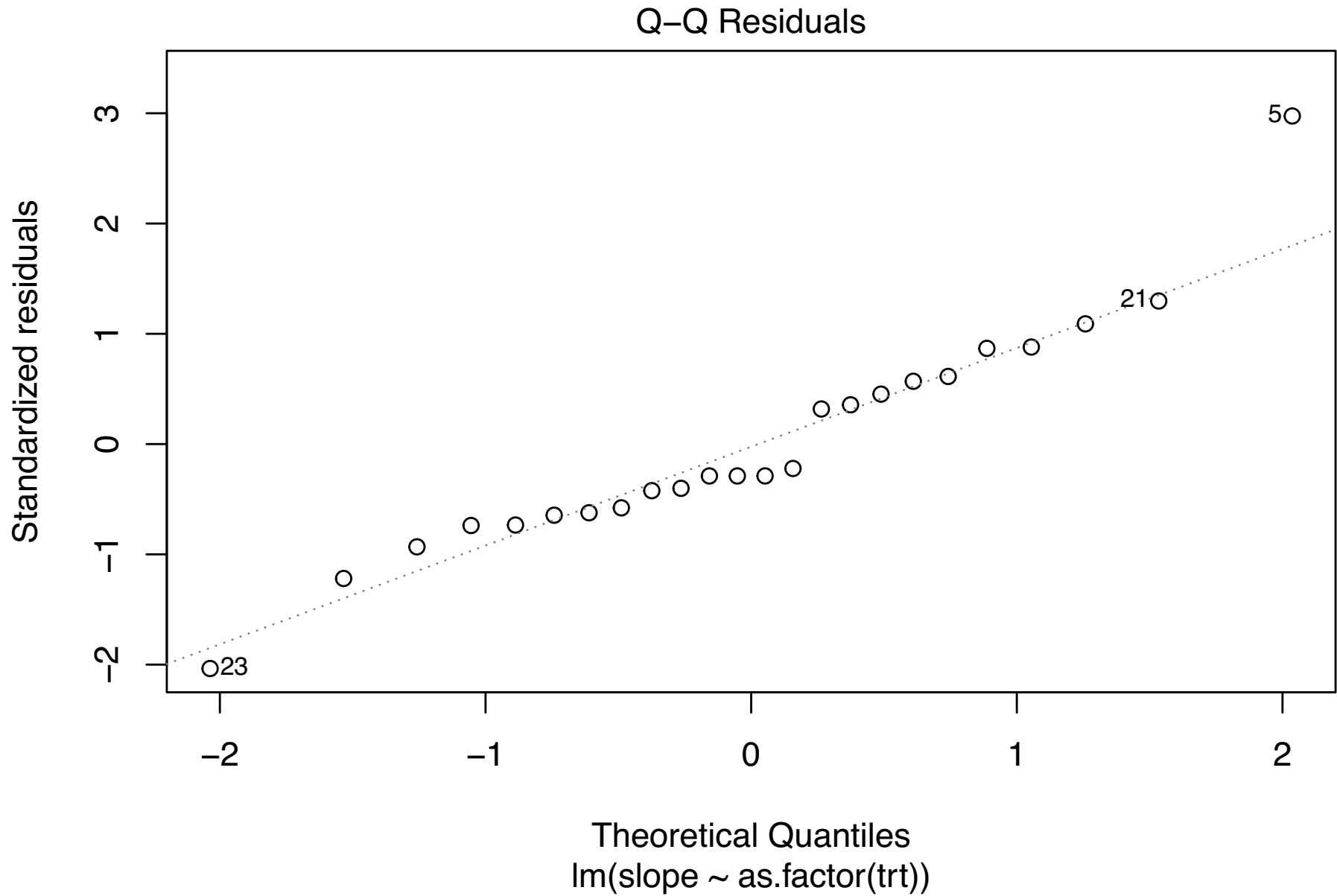
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4805 on 21 degrees of freedom

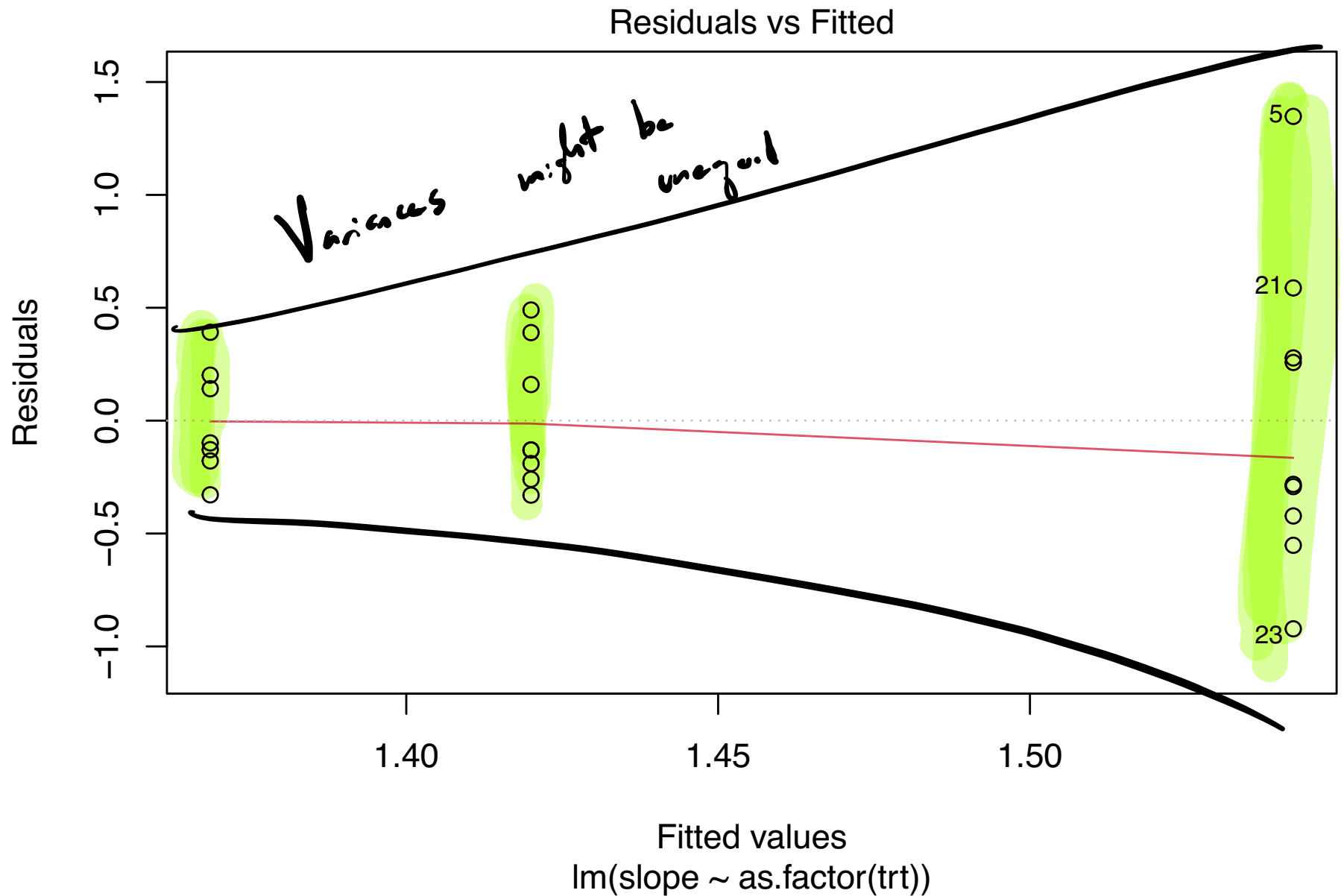
Multiple R-squared: 0.02614, Adjusted R-squared: -0.06661

F-statistic: 0.2818 on 2 and 21 DF, p-value: 0.7572

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



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



Levene's test for equality of variances

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

Checks if the mean magnitude of the residuals is equal across groups:

1. Obtain the residuals $\hat{\varepsilon}_{ij}$ from the one-way ANOVA model.
2. Treat the absolute values $|\hat{\varepsilon}_{ij}|$ of the residuals as new responses.
3. Test for equal means of the new responses with the F test.

So, do the ordinary F-test with the $|\hat{\varepsilon}_{ij}|$ as the responses.

Perception of slope example (cont)

Perform Levene's test:

```
ehat <- lm_slope$residuals  
lm_levne <- lm(abs(ehat) ~ as.factor(trt))  
summary(lm_levne)
```

$|\hat{\epsilon}_i|$

Call:

```
lm(formula = abs(ehat) ~ as.factor(trt))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.29136	-0.12769	-0.04980	0.08219	0.79864

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.20980	0.09352	2.243	0.0358 *
as.factor(trt)X	0.05020	0.12805	0.392	0.6990
as.factor(trt)Y	0.33934	0.12469	2.721	0.0128 *

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

Residual standard error: 0.2474 on 21 degrees of freedom

Multiple R-squared: 0.303, Adjusted R-squared: 0.2367

F-statistic: 4.565 on 2 and 21 DF, p-value: 0.02258

Reject H_0 at $\alpha = 0.05$

Can also use the `leveneTest()` function in the R package `car`.

```
library(car)
leveneTest(slope~as.factor(trt),center = mean)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
```

```
      Df F value Pr(>F)
group  2  4.5652 0.02258 *
      21
```

```
---
```

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

We conclude that the variances are *not* equal across treatment groups.

References

- Kutner, Michael H, Christopher J Nachtsheim, John Neter, and William Li. 2005. *Applied Linear Statistical Models*. McGraw-hill.
- Mohr, Donna L, William J Wilson, and Rudolf J Freund. 2021. *Statistical Methods*. Academic Press.