STAT 516 Lec 06

Two-way factorial design (balanced)

Karl Gregory

2024-03-21

Tensile strength example from Kuehl (2000)

Table 6.3 Tensile strength (psi) of asphaltic concrete specimens for two aggregate types with each of four compaction methods

	Compaction Method			
		Kneading		
Aggregate Type	*Static*	*Regular*	*Low*	*Very Low*
Basalt	68	126	93	56
	63	128	101	59
	65	133	98	57
Silicious	71	107	63	40
	66	110	60	41
	66	116	59	44

Source: A. M. Al-Marshed (1981), Compaction effects on asphaltic concrete durability. M.S. thesis, Civil Engineering, University of Arizona.

```
y <- c(68,63,65,71,66,66,126,128,133,107,110,116,
       93,101,98,63,60,59,56,59,57,40,41,44)
agg <- as.factor(rep(c(rep("B",3),rep("S",3)),4))
comp <- as.factor(c(rep("st",6),rep("r",6),
                    rep("l",6),rep("vl",6)))
```

The two-way factorial experimental design

- Two factors of interest.
- Each factor comprehends a set of treatments, called its levels.
- EUs randomly assigned to treatments.
- Each treatment is a unique combinations of factor levels.
- If Factor A has a levels and Factor B has b levels.
- There are ab treatment groups.

We want to make inferences about:

1. The effects of each factor.
2. The interactions between the two factors.
3. Various differences in treatment group means.

Discuss: Give factors and their levels in the tensile strength experiment.

Main effects and interactions

A main effect is an effect of a factor which does not depend on the level of the other factor.

An interaction is any dependence on the effect of one factor on the level of the other factor.

Two-way treatment effects model

Suppose the responses arise as

$$Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}$$

for $i = 1, \dots, a$, $j = 1, \dots, b$, and $k = 1, \dots, n_{ij}$, where

- Y_{ijk} is the response for EU k under level i of A and level j of B.
- μ represents a baseline or overall mean
- The τ_i are the main effects for Factor A.
- The γ_j are the main effects for Factor B.
- The $(\tau\gamma)_{ij}$ are the interaction effects between A and B.
- The ε_{ijk} are Normal$(0, \sigma^2)$ error terms.

Assume for now

1. a balanced design, i.e. $n_{ij} = n$ for all i, j
2. with replication, i.e. $n \geq 2$.

Parameter constraints in the treatment effects model

Treatment effects model has $a+b+ab+1$ parameters for ab means...

To identify the parameters uniquely, we impose one of these constraints:

1. To give μ a baseline interpretation, set

$$\tau_a = 0, \quad \gamma_b = 0, \quad \text{and} \quad (\tau\gamma)_{aj} = (\tau\gamma)_{ib} = 0 \text{ for all } i, j.$$

2. To give μ an overall mean interpretation, set

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

Exercise: Let $a = 2$ and $b = 3$. Can tabulate the response means as:

	1	2	3
1	$\mu+\tau_1+\gamma_1+(\tau\gamma)_{11}$	$\mu+\tau_1+\gamma_2+(\tau\gamma)_{12}$	$\mu+\tau_1+\gamma_3+(\tau\gamma)_{13}$
2	$\mu+\tau_2+\gamma_1+(\tau\gamma)_{21}$	$\mu+\tau_2+\gamma_2+(\tau\gamma)_{22}$	$\mu+\tau_2+\gamma_3+(\tau\gamma)_{23}$

Rewrite the table under the μ*-as-baseline* constraints

Cell means model representation

Assume

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

for $i = 1, \dots, a$, $j = 1, \dots, b$, and $k = 1, \dots, n_{ij}$, where

- Y_{ijk} is the response for EU k under level i of A and level j of B.
- μ_{ij} is the response mean under level i of A and level j of B.
- The ε_{ijk} are Normal$(0, \sigma^2)$ error terms.

Define the marginal means

$$\bar{\mu}_{i.} = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij}, \quad i = 1, \dots, a \quad \text{and} \quad \bar{\mu}_{.j} = \frac{1}{a}\sum_{i=1}^{a}\mu_{ij}, \quad j = 1, \dots, b.$$

and the overall mean $\bar{\mu}_{..} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij}$.

Hypotheses of interest in the two-way factorial experiment

1. H_0: Factor A has no main effect.

$$H_0\colon\ \bar{\mu}_{1.} = \cdots = \bar{\mu}_{a.}.$$

2. H_0: Factor B has no main effect.

$$H_0\colon\ \bar{\mu}_{.1} = \cdots = \bar{\mu}_{.b}.$$

3. H_0: There is no interaction between Factor A and Factor B.

$$H_0\colon\ \mu_{ij} = \bar{\mu}_{i.} + \bar{\mu}_{.j} - \bar{\mu}_{..} \quad \text{for all } i, j.$$

Example: Let $a = 2$ and $b = 3$. Can tabulate the response means as:

	1	2	3
1	μ_{11}	μ_{12}	μ_{13}
2	μ_{21}	μ_{22}	μ_{23}

Make sense of the hypotheses of no main effects and of no interaction.

Goals in two-way factorial experiments

In the two-way treatment effects model

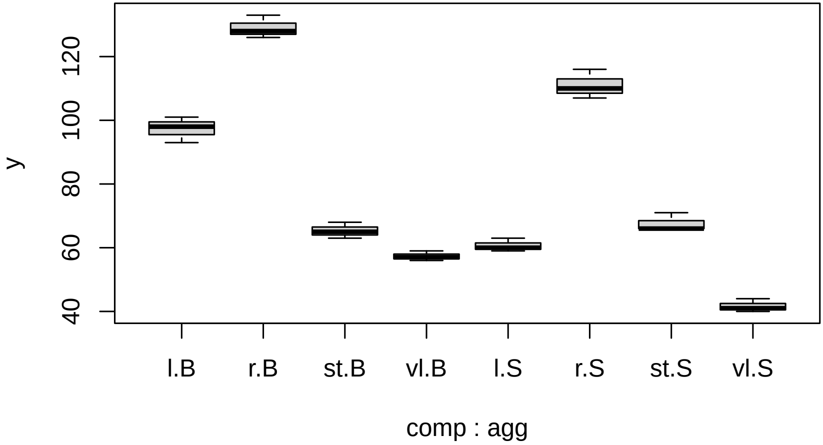
$$Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk},$$

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

1. Visualize the data.
2. Estimate the parameters μ, τ_i, γ_j, $(\tau\gamma)_{ij}$.
3. Estimate the error term variance σ^2.
4. Check whether the model assumptions are satisfied.
5. Decompose the variation in the Y_{ij} into its sources.
6. Test whether there is *any* effect of the factors on the response.
7. Test for main effects and interaction effects.
8. Do multiple comparisons.

Tensile strength data (cont)

```
boxplot(y ~ comp:agg)
```



Tensile example (cont)

Use `summary()` on the `lm()` function output.

```
lm_out <- lm(y ~ agg + comp + agg:comp)
summary(lm_out)
```

```
Call:
lm(formula = y ~ agg + comp + agg:comp)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3333 -1.6667 -0.6667  2.3333  5.0000 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   97.333      1.780  54.697  < 2e-16 ***
aggS         -36.667      2.517 -14.570 1.18e-10 ***
compr         31.667      2.517  12.583 1.03e-09 ***
compst       -32.000      2.517 -12.716 8.85e-10 ***
compvl       -40.000      2.517 -15.894 3.20e-11 ***
aggS:compr    18.667      3.559   5.245 8.01e-05 ***
aggS:compst   39.000      3.559  10.958 7.58e-09 ***
aggS:compvl   21.000      3.559   5.900 2.24e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.082 on 16 degrees of freedom
Multiple R-squared:  0.9921,    Adjusted R-squared:  0.9887 
F-statistic: 287.6 on 7 and 16 DF,  p-value: 1.305e-15
```

Exercise: See how the parameter estimates build the treatment means.

```
aggregate(y, by = list(agg, comp), FUN = mean)
```

```
  Group.1 Group.2         x
1       B       l  97.33333
2       S       l  60.66667
3       B       r 129.00000
4       S       r 111.00000
5       B      st  65.33333
6       S      st  67.66667
7       B      vl  57.33333
8       S      vl  41.66667
```

The fitted values

Define the treatment group means

$$\bar{Y}_{ij.} = \frac{1}{n} \sum_{k=1}^{n} Y_{ijk} \quad \text{for } i = 1, \ldots, a, \quad j = 1, \ldots, b.$$

Then the

- fitted values are the treatment group means, i.e. $\hat{Y}_{ijk} = \bar{Y}_{ij.}$.
- residuals are the deviations from the group means $\varepsilon_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$.

In the cell means model, we estimate μ_{ij} with $\hat{\mu}_{ij} = \bar{Y}_{ij.}\ \forall ij$.

Estimating the error term variance

An unbiased estimator of the error term variance σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{ab(n-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{ij.})^2,$$

which is the sum of the squared residuals divided by $ab(n-1)$.

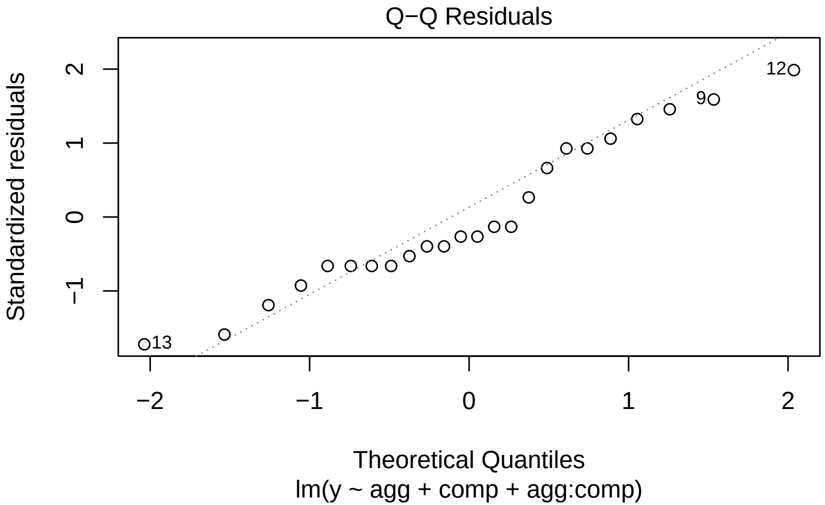
Checking model assumptions

Validity of the methods in these slides depends on these assumptions:

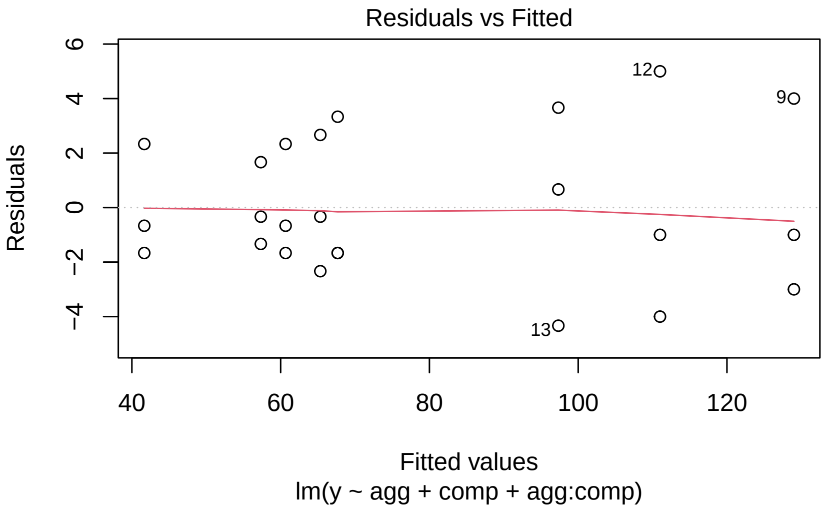
1. The responses are normally distributed around the treatment means (Check QQ plot of residuals).
2. The response has the same variance in all treatment groups (Check residuals vs fitted values plot).
3. The response values are independent of each other (No way to check; must trust experimental design).

Tensile example (cont)

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



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



Sums of squares in the two-way factorial experiment

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

- Total sum of squares: $\mathrm{SS}_{\mathrm{Tot}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{...})^2$
- Treatments sum of squares: $\mathrm{SS}_{\mathrm{Trt}} = \sum_{i=1}^{a} \sum_{j=1}^{b} n(\bar{Y}_{ij.} - \bar{Y}_{...})^2$
- Error sum of squares: $\mathrm{SS}_{\mathrm{Error}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{ij.})^2$

We have $\mathrm{SS}_{\mathrm{Tot}} = \mathrm{SS}_{\mathrm{Trt}} + \mathrm{SS}_{\mathrm{Error}}$.

We again define $R^2 = \dfrac{\mathrm{SS}_{\mathrm{Trt}}}{\mathrm{SS}_{\mathrm{Tot}}}$.

Sampling distributions of our sums of squares

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

- $\mathrm{SS}_{\mathrm{Tot}} / \sigma^2 \sim \chi^2_{abn-1}(\phi_{\mathrm{Tot}})$
- $\mathrm{SS}_{\mathrm{Trt}} / \sigma^2 \sim \chi^2_{ab-1}(\phi_{\mathrm{Trt}})$
- $\mathrm{SS}_{\mathrm{Error}} / \sigma^2 \sim \chi^2_{ab(n-1)}$,

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

Mean squares in the two-way factorial experiment

Dividing $\mathrm{SS}_{\mathrm{Trt}}$ and $\mathrm{SS}_{\mathrm{Error}}$ by their dfs, we define:

- Treatments mean square: $\mathrm{MS}_{\mathrm{Trt}} = \dfrac{\mathrm{SS}_{\mathrm{Trt}}}{ab-1}$
- Error mean square: $\mathrm{MS}_{\mathrm{Error}} = \dfrac{\mathrm{SS}_{\mathrm{Error}}}{ab(n-1)}$

The ratio $F_{\mathrm{stat}} = \dfrac{\mathrm{MS}_{\mathrm{Trt}}}{\mathrm{MS}_{\mathrm{Error}}}$ has an F distribution.

The Analysis of Variance (ANOVA) table

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

Source	Df	SS	MS	F value	p-value
Treatments	$ab-1$	$\mathrm{SS}_{\mathrm{Trt}}$	$\mathrm{MS}_{\mathrm{Trt}}$	F_{stat}	$P(F > F_{\mathrm{stat}})$
Error	$ab(n-1)$	$\mathrm{SS}_{\mathrm{Error}}$	$\mathrm{MS}_{\mathrm{Error}}$		
Total	$abn-1$	$\mathrm{SS}_{\mathrm{Tot}}$			

In the table $F_{\mathrm{stat}} = \dfrac{\mathrm{MS}_{\mathrm{Trt}}}{\mathrm{MS}_{\mathrm{Error}}}$.

The p-value is based on $F \sim F_{ab-1,ab(n-1)}$.

The `summary()` function on the output of `lm()` gives these values.

Exercise: Fill in the ANOVA table for the tensile strength data.

```r
a <- 2
b <- 4
n <- 3
yhat <- predict(lm_out)
ehat <- y - yhat
SSE <- sum(ehat^2)
MSE <- SSE / (a*b*(n-1))
SSR <- sum((yhat - mean(y))^2)
MSR <- SSR / (a*b - 1)
SST <- sum((y - mean(y))^2)
Fstat <- MSR / MSE
pval <- 1 - pf(Fstat,a*b-1,a*b*(n-1))
```

Source	Df	SS	MS	F value	p-value
Treatments	7	19122.50	2731.79	287.5564	0.0000
Error	16	152.00	9.50		
Total	23	19274.50			

Overall F test for any kind of effect

Using `summary()` on the `lm()` output prints the overall F test result.

The overall F test in the two-way treatment effects model tests

H_0: All the μ_{ij} are the same.

H_1: The μ_{ij} are not all the same.

The null says neither factor has any effect on the mean response.

The alternate says at least one of the factors has some effect.

Exercise: Interpret overall F test result for the tensile strength data.

Further decomposition of treatments sum of squares

Define main effect and interaction sums of squares $\mathrm{SS_A}$, $\mathrm{SS_B}$, and $\mathrm{SS_{AB}}$:

- $\mathrm{SS_A} = bn\sum_{i=1}^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2$
- $\mathrm{SS_B} = an\sum_{j=1}^{b}(\bar{Y}_{.j.} - \bar{Y}_{...})^2$
- $\mathrm{SS_{AB}} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij.} - (\bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}))^2$

In the balanced design we have

$$\mathrm{SS_{Trt}} = \mathrm{SS_A} + \mathrm{SS_B} + \mathrm{SS_{AB}}\,.$$

Distributions of main effect and interaction SS

For some noncentrality parameters ϕ_A, ϕ_B, and ϕ_{AB}, we have

- $\mathrm{SS_A}/\sigma^2 \sim \chi^2_{a-1}(\phi_A)$
- $\mathrm{SS_B}/\sigma^2 \sim \chi^2_{b-1}(\phi_B)$
- $\mathrm{SS_{AB}}/\sigma^2 \sim \chi^2_{(a-1)(b-1)}(\phi_{AB})$.

Define the corresponding mean squares

$$\mathrm{MS_A} = \frac{\mathrm{SS_A}}{a-1}, \quad \mathrm{MS_B} = \frac{\mathrm{SS_B}}{b-1}, \quad \mathrm{MS_{AB}} = \frac{\mathrm{SS_{AB}}}{(a-1)(b-1)}.$$

Full ANOVA table for balanced two-way factorial design

Source	Df	SS	MS	F value
A	$a-1$	$\mathrm{SS_A}$	$\mathrm{MS_A}$	$F_\mathrm{A} = \mathrm{MS_A} \,/\, \mathrm{MS_{Error}}$
B	$b-1$	$\mathrm{SS_B}$	$\mathrm{MS_B}$	$F_\mathrm{B} = \mathrm{MS_B} \,/\, \mathrm{MS_{Error}}$
AB	$(a-1)(b-1)$	$\mathrm{SS_{AB}}$	$\mathrm{MS_{AB}}$	$F_\mathrm{AB} = \mathrm{MS_{AB}} \,/\, \mathrm{MS_{Error}}$
Error	$ab(n-1)$	$\mathrm{SS_{Error}}$	$\mathrm{MS_{Error}}$	
Total	$abn-1$	$\mathrm{SS_{Tot}}$		

1. Reject H_0: no Factor A main effect if $F_\mathrm{A} > F_{a-1,ab(n-1),\alpha}$.
2. Reject H_0: no Factor B main effect if $F_\mathrm{B} > F_{b-1,ab(n-1),\alpha}$.
3. Reject H_0: no A and B interaction if $F_\mathrm{AB} > F_{(a-1)(b-1),ab(n-1),\alpha}$.

Tensile strength data (cont)

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

```
anova(lm(y ~ agg + comp + agg:comp))
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
agg        1   1734  1734.0 182.526 3.628e-10 ***
comp       3  16244  5414.5 569.947 < 2.2e-16 ***
agg:comp   3   1145   381.7  40.175 1.124e-07 ***
Residuals 16    152     9.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Important: `anova()` function only appropriate for a balanced design.

Interaction is significant. Now what?

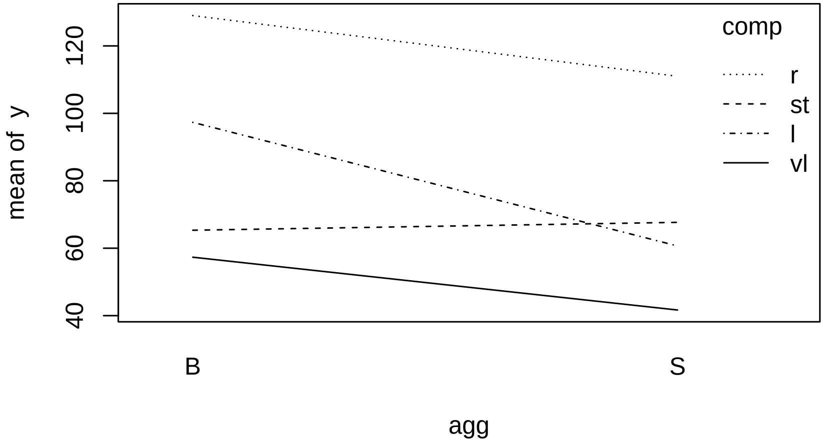
When you find a significant interaction:

1. Make interaction plots (next slides).
2. Be very cautious about interpreting main effects, even when these are statistically significant.

Tensile strength data (cont)

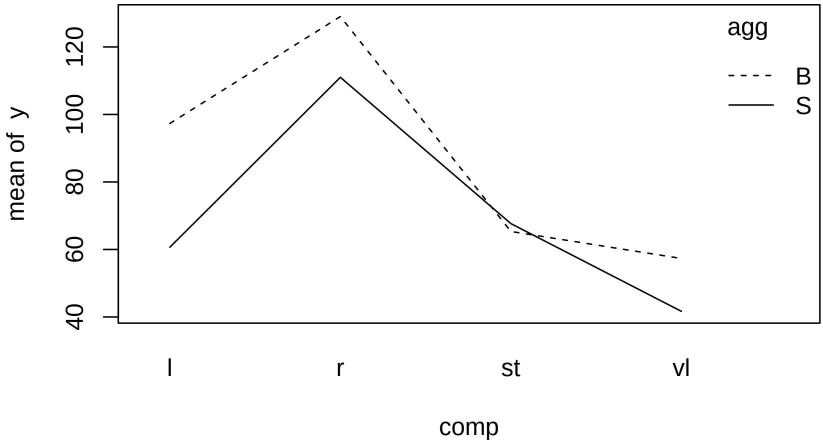
Use the `interaction.plot()` function to visualize an interaction:

```
interaction.plot(agg,comp,y)
```



Interactions appear as crossing lines or differing slopes.

```r
interaction.plot(comp, agg, y)
```



Estimates of cell and marginal means in the balanced case

The estimators of the cell and marginal means are given by

- $\hat{\mu}_{ij} = \bar{Y}_{ij.}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$
- $\hat{\bar{\mu}}_{i.} = \bar{Y}_{i..}$, $i = 1, \ldots, a$.
- $\hat{\bar{\mu}}_{.j} = \bar{Y}_{.j.}$, $j = 1, \ldots, b$.

We estimate $\hat{\bar{\mu}}_{i.}$ with $\bar{Y}_{i..}$ (and $\hat{\bar{\mu}}_{.j}$ with $\bar{Y}_{.j.}$) only when $n_{ij} = n \ \forall ij$.

Some CI formulas (without familywise adjustment)

These CI formulas are for the balanced design $n_{ij} = n \ \forall ij$.

Target	$(1-\alpha)100\%$ confidence interval
μ_{ij}	$\bar{Y}_{ij.} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}}$
$\mu_{ij} - \mu_{i'j'}$	$\bar{Y}_{ij.} - \bar{Y}_{i'j'.} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{2}{n}}$
$\bar{\mu}_{i.}$	$\bar{Y}_{i..} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{1}{bn}}$
$\bar{\mu}_{.j}$	$\bar{Y}_{.j.} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{1}{an}}$
$\bar{\mu}_{i.} - \bar{\mu}_{i'.}$	$\bar{Y}_{i..} - \bar{Y}_{i'..} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{2}{bn}}$
$\bar{\mu}_{.j} - \bar{\mu}_{.j'}$	$\bar{Y}_{.j.} - \bar{Y}_{.j'.} \pm t_{ab(n-1),\alpha/2}\hat{\sigma}\sqrt{\frac{2}{an}}$

In the above $\hat{\sigma} = \sqrt{\mathrm{MS}_{\mathrm{Error}}}$.

Comparing means at all factor level combinations

- Tukey's for comparing all pairs among μ_{ij}, $i = 1, \dots, a$, $j = 1, \dots, b$:

$$\bar{Y}_{ij.} - \bar{Y}_{i'j'.} \pm q_{ab,ab(n-1),\alpha} \; \hat{\sigma} \frac{1}{\sqrt{n}}, \quad (i,j) \neq (i',j').$$

- Dunnett's for comparing all means μ_{ij} to a baseline μ_{ab}:

$$\bar{Y}_{ij.} - \bar{Y}_{ab.} \pm d_{ab,ab(n-1),\alpha} \; \hat{\sigma} \sqrt{\frac{2}{n}}, \quad (i,j) \neq (a,b).$$

Use $\hat{\sigma} = \sqrt{\mathrm{MS}_{\mathrm{Error}}}$.

These may make more comparisons than are of interest...

Tensile strength data (cont)

```
TukeyHSD(aov(lm(y ~ agg:comp)))
```

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

Fit: aov(formula = lm(y ~ agg:comp))

$`agg:comp`
                diff        lwr         upr     p adj
S:l-B:l   -36.666667 -45.379555 -27.9537785 0.0000000
B:r-B:l    31.666667  22.953778  40.3795548 0.0000000
S:r-B:l    13.666667   4.953778  22.3795548 0.0011210
B:st-B:l  -32.000000 -40.712888 -23.2871118 0.0000000
S:st-B:l  -29.666667 -38.379555 -20.9537785 0.0000001
B:vl-B:l  -40.000000 -48.712888 -31.2871118 0.0000000
S:vl-B:l  -55.666667 -64.379555 -46.9537785 0.0000000
B:r-S:l    68.333333  59.620445  77.0462215 0.0000000
S:r-S:l    50.333333  41.620445  59.0462215 0.0000000
B:st-S:l    4.666667  -4.046222  13.3795548 0.5964603
S:st-S:l    7.000000  -1.712888  15.7128882 0.1678762
B:vl-S:l   -3.333333 -12.046222   5.3795548 0.8765993
S:vl-S:l  -19.000000 -27.712888 -10.2871118 0.0000257
S:r-B:r   -18.000000 -26.712888  -9.2871118 0.0000501
B:st-B:r  -63.666667 -72.379555 -54.9537785 0.0000000
S:st-B:r  -61.333333 -70.046222 -52.6204452 0.0000000
B:vl-B:r  -71.666667 -80.379555 -62.9537785 0.0000000
S:vl-B:r  -87.333333 -96.046222 -78.6204452 0.0000000
B:st-S:r  -45.666667 -54.379555 -36.9537785 0.0000000
S:st-S:r  -43.333333 -52.046222 -34.6204452 0.0000000
B:vl-S:r  -53.666667 -62.379555 -44.9537785 0.0000000
S:vl-S:r  -69.333333 -78.046222 -60.6204452 0.0000000
S:st-B:st   2.333333  -6.379555  11.0462215 0.9785200
B:vl-B:st  -8.000000 -16.712888   0.7128882 0.0842128
S:vl-B:st -23.666667 -32.379555 -14.9537785 0.0000015
B:vl-S:st -10.333333 -19.046222  -1.6204452 0.0145554
S:vl-S:st -26.000000 -34.712888 -17.2871118 0.0000004
S:vl-B:vl -15.666667 -24.379555  -6.9537785 0.0002561
```

Easiest way to do Dunnett's is to convert the design to a one-way:

```
agg_comp <- as.factor(paste(agg,comp,sep="_"))
levels(agg_comp)
```

```
[1] "B_l"  "B_r"  "B_st" "B_vl" "S_l"  "S_r"  "S_st" "S_vl"
```

```
library(DescTools)
DunnettTest(y ~ agg_comp, control = "B_st", conf.level = 0.95)
```

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

$B_st
                diff     lwr.ci      upr.ci    pval
B_l-B_st   32.000000  24.643829  39.3561715 1.1e-10 ***
B_r-B_st   63.666667  56.310495  71.0228381 < 2e-16 ***
B_vl-B_st  -8.000000 -15.356171  -0.6438285  0.0305 *
S_l-B_st   -4.666667 -12.022838   2.6895048  0.3265
S_r-B_st   45.666667  38.310495  53.0228381 < 2e-16 ***
S_st-B_st   2.333333  -5.022838   9.6895048  0.8881
S_vl-B_st -23.666667 -31.022838 -16.3104952 5.0e-08 ***

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

Comparing factor level means at fixed level of other factor

Fix Factor A at level i and compare means across factor B:

- Tukey's for comparing all pairs among $\bar{\mu}_{i1}, \ldots, \bar{\mu}_{ib}$:

$$\bar{Y}_{ij.} - \bar{Y}_{ij'.} \pm q_{b,ab(n-1),\alpha}\, \hat{\sigma} \frac{1}{\sqrt{n}}, \quad 1 \leq j < j' \leq b.$$

- Dunnett's for comparing $\bar{\mu}_{ij}, \ldots, \bar{\mu}_{i,b-1}$ to baseline μ_{ib}:

$$\bar{Y}_{ij.} - \bar{Y}_{ib.} \pm d_{b,ab(n-1),\alpha}\, \hat{\sigma} \sqrt{\frac{2}{n}}, \quad j = 1, \ldots, b-1.$$

To do this at *all* levels $i = 1, \ldots, a$, divide α by a (à la Bonferroni)!

Likewise if comparing means of factor B at fixed levels of factor A.

Tensile strength data (cont)

For each aggregate type, compare all pairs of compaction method means.

Ensure that the familywise coverage probability is at least 0.95.

Use $q_{4,16,0.05/2}$ = `qtukey(1-0.05/2,4,16)` = 4.54763.

```
y11. <- mean(y[agg == "B" & comp == "l"])
y12. <- mean(y[agg == "B" & comp == "r"])
y13. <- mean(y[agg == "B" & comp == "st"])
y14. <- mean(y[agg == "B" & comp == "vl"])

y21. <- mean(y[agg == "S" & comp == "l"])
y22. <- mean(y[agg == "S" & comp == "r"])
y23. <- mean(y[agg == "S" & comp == "st"])
y24. <- mean(y[agg == "S" & comp == "vl"])

alpha <- 0.05
me <- qtukey(1-alpha/a,b,a*b*(n-1)) * sqrt(MSE) / sqrt(n)
```

```r
ttab <- rbind(c(y11. - y12. - me, y11. - y12. + me),
              c(y11. - y13. - me, y11. - y13. + me),
              c(y11. - y14. - me, y11. - y14. + me),
              c(y12. - y13. - me, y12. - y13. + me),
              c(y12. - y14. - me, y12. - y14. + me),
              c(y13. - y14. - me, y13. - y14. + me),
              c(y21. - y22. - me, y21. - y22. + me),
              c(y21. - y23. - me, y21. - y23. + me),
              c(y21. - y24. - me, y21. - y24. + me),
              c(y22. - y23. - me, y22. - y23. + me),
              c(y22. - y24. - me, y22. - y24. + me),
              c(y23. - y24. - me, y23. - y24. + me))
rownames(ttab) <- c("B:l-r","B:l-st","B:l-vl","B:r-st","B:r-vl","B:st-vl",
                   "S:l-r","S:l-st","S:l-vl","S:r-st","S:r-vl","S:st-vl")
colnames(ttab) <- c("lower","upper")
```

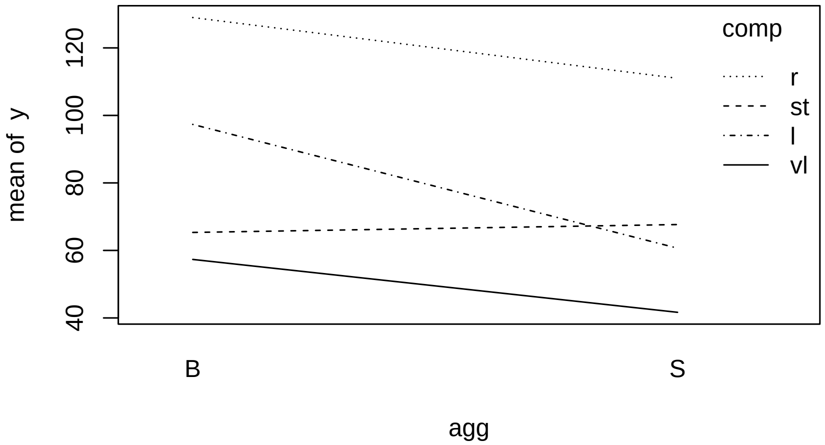
Tukey/Bonferroni-adjusted confidence intervals comparing all pairs of compaction methods for each aggregate type:

```
ttab
```

```
               lower      upper
B:l-r   -39.75923299 -23.574100
B:l-st   23.90743367  40.092566
B:l-vl   31.90743367  48.092566
B:r-st   55.57410034  71.759233
B:r-vl   63.57410034  79.759233
B:st-vl  -0.09256633  16.092566
S:l-r   -58.42589966 -42.240767
S:l-st  -15.09256633   1.092566
S:l-vl   10.90743367  27.092566
S:r-st   35.24076701  51.425900
S:r-vl   61.24076701  77.425900
S:st-vl  17.90743367  34.092566
```

Identify which differences are *not* significant in the interaction plot:

```
interaction.plot(agg,comp,y)
```



Tensile strength data (cont)

For only the basalt aggregate type, compare all compaction method means to that of the static method.

Use $d_{4,16,0.05} = 2.59$ from the table on the next slide.

```
me <-  2.59 * sqrt(MSE) * sqrt(2/n)
dtab <- rbind(c(y11. - y13. - me, y11. - y13. + me),
              c(y12. - y13. - me, y12. - y13. + me),
              c(y14. - y13. - me, y14. - y13. + me))
rownames(dtab) <- c("B:l-st","B:r-st","B:vl-st")
colnames(dtab) <- c("lower","upper")
```

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

		T = Number of Groups Counting Both Treatments and Control						
Error df	**Two-sided α**	**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

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

Figure 1: Table A.5 from Mohr, Wilson, and Freund (2021)

Dunnett's comparison of compaction method means to the static method when the aggregate type is basalt:

```
dtab
```

```
            lower     upper
B:l-st   25.48198 38.518024
B:r-st   57.14864 70.184690
B:vl-st -14.51802 -1.481976
```

Interaction *not* significant. Then what?

If the interaction is not significant:

1. We can interpret main effects.
2. We can make meaningful comparisons among marginal means.

Comparing marginal means in the absence of interaction

For making comparisons among the marginal means of Factor A:

- Tukey's for comparing all pairs among $\bar{\mu}_{1.}, \dots, \bar{\mu}_{a.}$:

$$\bar{Y}_{i..} - \bar{Y}_{i'..} \pm q_{a,ab(n-1),\alpha}\ \hat{\sigma}\frac{1}{\sqrt{bn}}, \quad 1 \le i < i' \le a.$$

- Dunnett's for comparing $\bar{\mu}_{1.}, \dots, \bar{\mu}_{a-1.}$ to a control mean $\bar{\mu}_{a.}$:

$$\bar{Y}_{i..} - \bar{Y}_{a..} \pm d_{a,ab(n-1),\alpha}\ \hat{\sigma}\sqrt{\frac{2}{bn}}, \quad i = 1, \dots, a-1.$$

Still use $\hat{\sigma} = \sqrt{\mathrm{MS}_{\mathrm{Error}}}$.

Do likewise for making comparisons among $\bar{\mu}_{.1}, \dots, \bar{\mu}_{.b}$ of Factor B.

Serum glucose example from Kuehl (2000)

Two methods for measuring serum glucose level at three glucose levels.

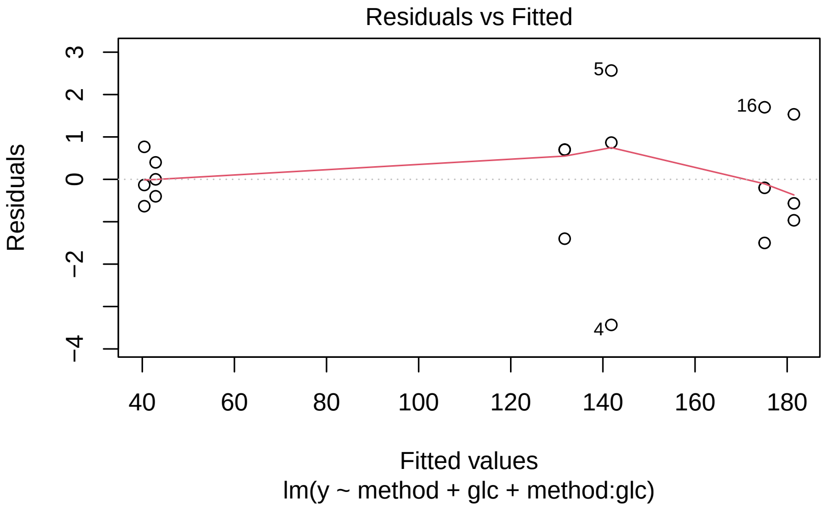
	Method 1			*Method 2*		
Glucose Level	*1*	*2*	*3*	*1*	*2*	*3*
	42.5	138.4	180.9	39.8	132.4	176.8
	43.3	144.4	180.5	40.3	132.4	173.6
	42.9	142.7	183.0	41.2	130.3	174.9

Source: Dr. J. Anderson, Beckman Instruments Inc.

```
y <- c(42.5,43.3,42.9,138.4,144.4,142.7,180.9,180.5,183.0,
       39.8,40.3,41.2,132.4,132.4,130.3,176.8,173.6,174.9)
method <- as.factor(c(1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2))
glc <- as.factor(c(1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3))
```

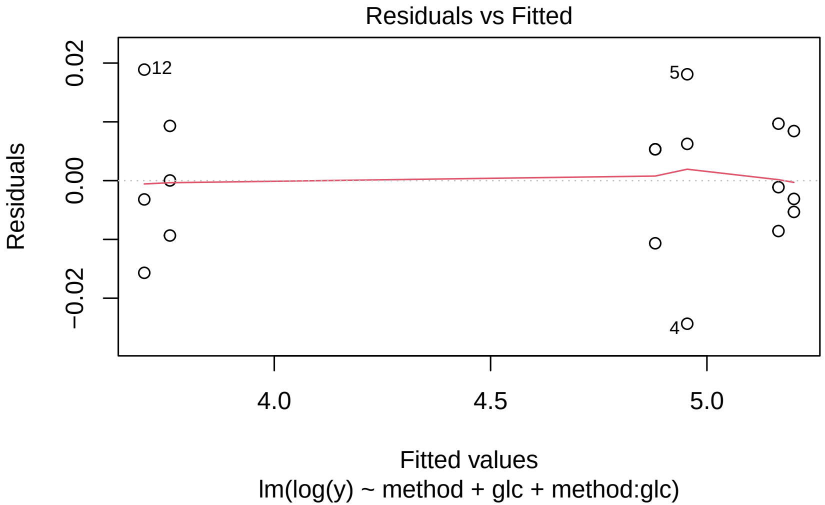
- Is there an interaction between the method and the glucose level?
- If not, can we describe the main effect of the method?

```r
lm_glc <- lm(y ~ method + glc + method:glc)
plot(lm_glc,which = 1)
```



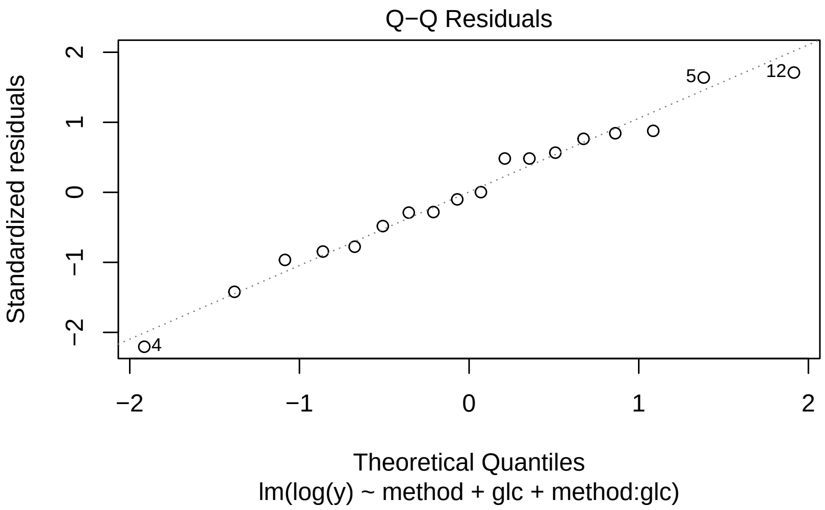
Variance appears smaller at lower glucose level. Try using $\log(Y_{ijk})$.

```r
lm2_glc <- lm(log(y) ~ method + glc + method:glc)
plot(lm2_glc,which = 1)
```



This looks better.

```r
plot(lm2_glc,which = 2)
```



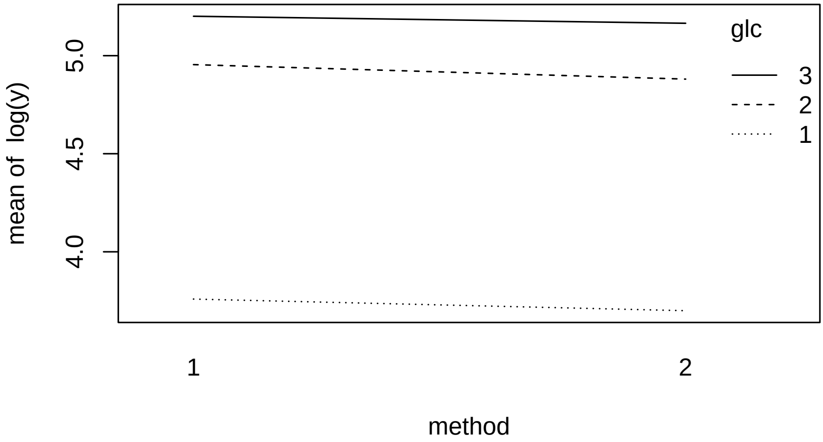
Normality check looks okay.

```r
anova(lm2_glc)
```

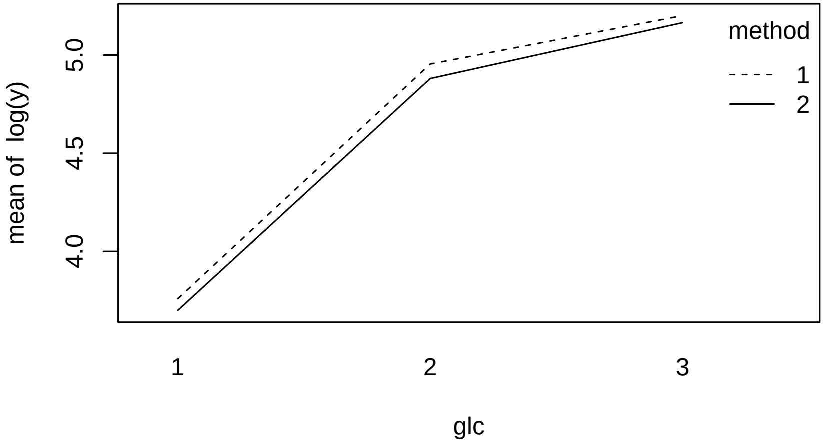
```
Analysis of Variance Table

Response: log(y)
           Df Sum Sq Mean Sq    F value    Pr(>F)    
method      1 0.0143  0.0143    78.1091 1.337e-06 ***
glc         2 7.1935  3.5967 19670.4837 < 2.2e-16 ***
method:glc  2 0.0011  0.0006     3.0574    0.0845 .  
Residuals  12 0.0022  0.0002                         
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is only weak evidence of interaction. Check interaction plot.

```r
interaction.plot(method,glc,log(y))
```



```r
interaction.plot(glc,method,log(y))
```



It appears safe to ignore the interaction and report on main effects.

Call the method Factor A; build a CI for $\bar{\mu}_{1.} - \bar{\mu}_{2.}$ (just one comparison).

Since $a = 2$, $b = 3$, and $n = 3$, use $\bar{Y}_{1..} - \bar{Y}_{2..} \pm q_{2,2\cdot 3(3-1),0.05}\ \hat{\sigma}\frac{1}{\sqrt{3\cdot 3}}$.

```
a <- 2
b <- 3
n <- 3
alpha <- 0.05
y1.. <- mean(log(y[method == 1])) # remember we are using log(y)
y2.. <- mean(log(y[method == 2]))
MSE <- sum(lm2_glc$residuals^2) / (a*b*(n-1))
me <- qtukey(1-alpha,a,a*b*(n-1)) * sqrt(MSE) / sqrt(n*b)
lo <- y1.. - y2.. - me
up <- y1.. - y2.. + me
c(lo,up)
```

```
[1] 0.04244809 0.07022545
```

Since $a = 2$, $q_{a,ab(n-1),\alpha} = \sqrt{2}\cdot t_{ab(n-1),\alpha/2}$, so it is just a t-interval.

Possible workflow for factorial experiments

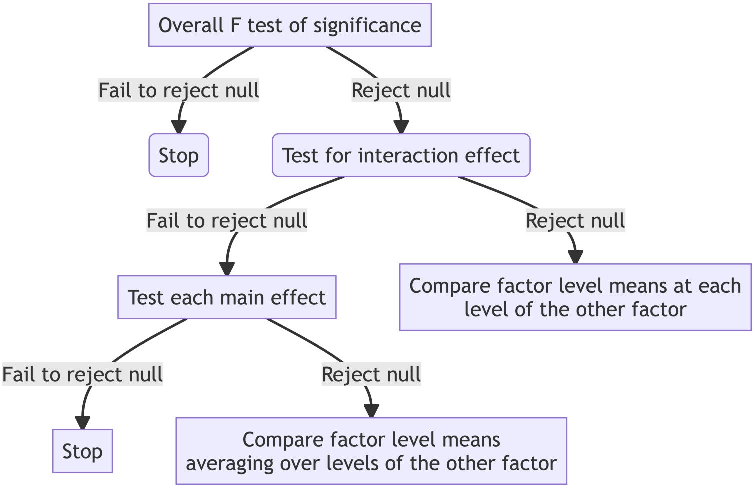


Figure 2: A justifiable workflow for analyzing factorial experiment data.

References

Kuehl, R. O. 2000. *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury/Thomson Learning.

Mohr, Donna L, William J Wilson, and Rudolf J Freund. 2021. *Statistical Methods*. Academic Press.