Two-Way ANOVA

SDS 328M

University of Texas at Austin

Announcements

- TODAY: Two-way ANOVA and Interactions (!)
- But first we are going to review one-way ANOVA
- Then, we are going to see how to think about two-way ANOVA
- In the process, we will build intuition for what interactions are!
- Finally, we will do a full example with all steps
- At the end of the slides, I'll post extra practice problems!

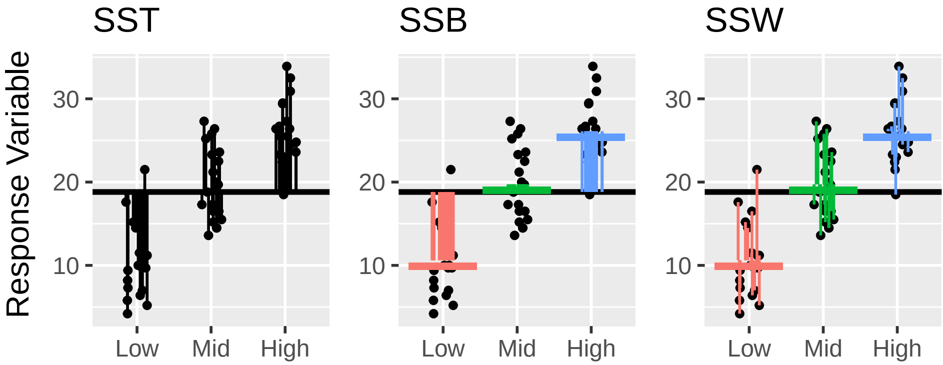
SDS ANOVA App!

https://utsds.shinyapps.io/ANOVAVariance/

Set group means to $90, 100, 100$ ($n = 25$ in each group)

- What happens to the F-stat and p-value when you increase the population standard deviation? $F\downarrow, p\uparrow$
- What happens when you decrease it? $F\uparrow, p\downarrow$
- What happens when means are further apart? $F\uparrow, p\downarrow$
- Closer together? $F\downarrow, p\uparrow$

One-Way ANOVA Review



$$SS_{Total} = SS_{Between} + SS_{Within}$$

(total variation) = (grp means around overall mean) + (obs around grp means)

$$\sum_{i}^{N}(y_{ik} - \bar{y}_{\bullet\bullet})^2 = \sum_{k}^{K} n_k(\bar{y}_{\bullet k} - \bar{y}_{\bullet\bullet})^2 + \sum_{k}^{K}\sum_{i}^{n_k}(y_{ik} - \bar{y}_{\bullet k})^2$$

ANOVA Review: F Statistic

- $H_0 : \mu_{Low} = \mu_{Mid} = \mu_{High}$
- H_A : at least one of these means differs from the others
- H_0 in words: true mean of response variable is the same for all 3 groups
- To test this H_0, we use the F statistic!
- Ratio of *variance explained* (MS_B) to *variance unexplained* (MS_W)

$$F = \frac{\textit{Variation betweeen groups}}{\textit{Variation within groups}} = \frac{SS_B/(K-1)}{SS_W/(N-K)} = \frac{\mathbf{MS_B}}{\mathbf{MS_W}}$$

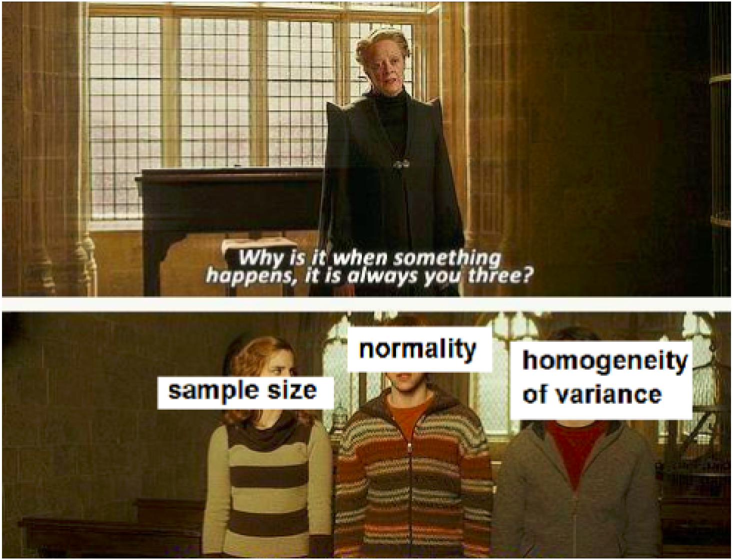
- If F is large, variation between groups large relative to variation within
  - Group means spread out (numerator) relative to the noise (denominator)
  - If large enough, reject null hypothesis!
- If F is small, little variation between groups and/or lots of noise within
  - Group means close together; true means probably not different!

ANOVA Assumptions

Usual suspects:

- Random sample and independent observations (always!)
- Independent samples (groups)
- Normal distribution (in each group) or large sample in each (25+)
  - Can eyeball a histogram/boxplot/qqplot of response in each group
  - Equivalently, make one plot of the residuals (instead of a plot for each group)!
  - Could also run Shapiro-Wilk on residuals
- Equal variance (variance/sd of each group is the same)
  - Eyeball boxplots for each group or run Levene's test
  - Good rule of thumb: *sd* should not be more than $2\times$ bigger in one group than another

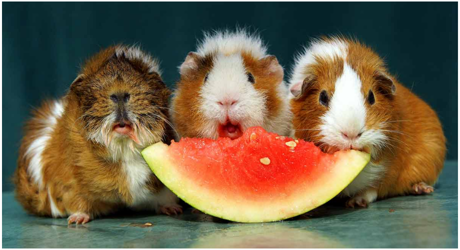
ANOVA Assumptions



One-Way ANOVA Table

Source	SS	Deg. of Freedom (df)	Mean Squares (MS)	F
Between Grps	SS_B	$K-1$	$MS_B = \frac{SS_B}{K-1}$	$\frac{MS_B}{MS_W}$
Within Grps	SS_W	$N-K$	$MS_W = \frac{SS_W}{N-K}$	
Total	SS_T	$N-1$		

Let's do a quick example



- How is guinea pig tooth growth influenced by vitamin C dosage?
- Three dosages (0.5 mg, 1.0 mg, and 2.0 mg), 20 guineas per condition!

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
```

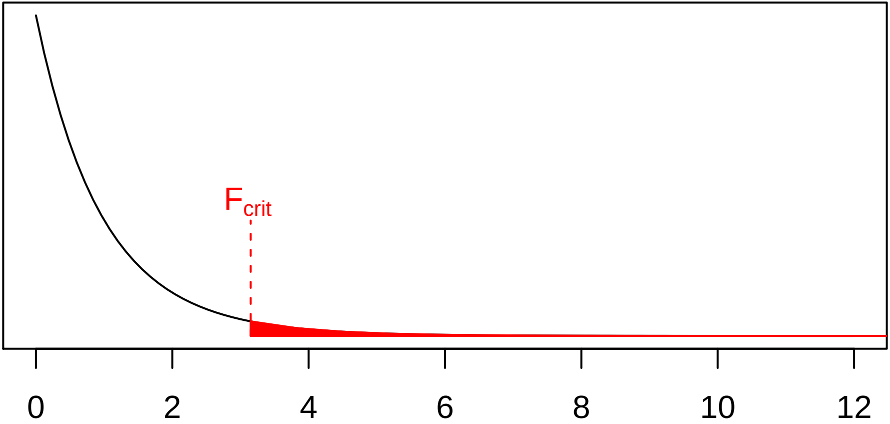
Let's do a quick example

```r
library(car) #need this for the Anova() function
fit <- lm(len ~ dose, data=ToothGrowth)
Anova(fit)
```

```
...
## Response: len
##           Sum Sq Df F value    Pr(>F)
## dose      2426.4  2  67.416 9.533e-16 ***
## Residuals 1025.8 57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

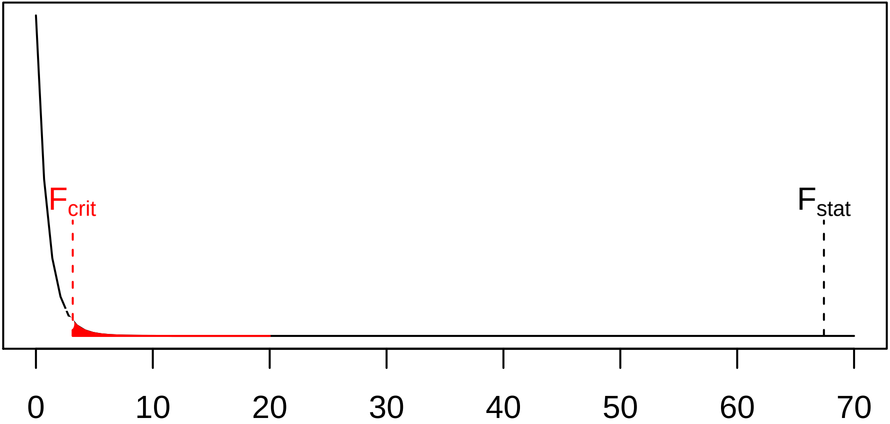
ANOVA: F Distribution

- $F_{crit} = 3.15 \rightarrow$ F-table $(num.\ df = 2,\ denom.\ df = 57,\ \alpha = .05)$
- Note that F can only be positive!
- $F_{stat} = 67.42$



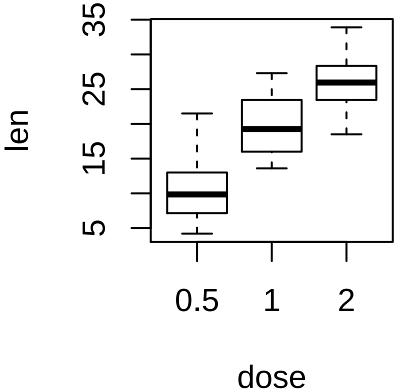
ANOVA: F Distribution

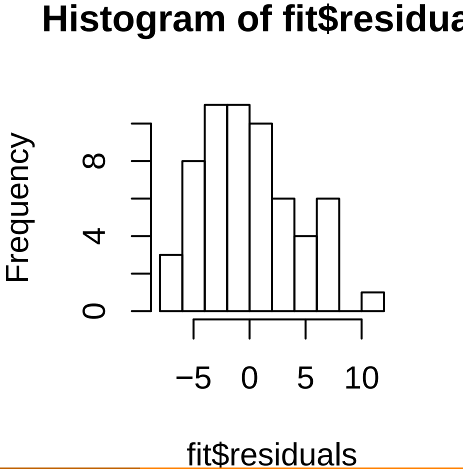
- $F_{crit} = 3.15 \rightarrow$ F-table $(num.\ df = 2,\ denom.\ df = 57,\ \alpha = .05)$
- Note that F can only be positive!
- $F_{stat} = 67.42$



Test Assumptions: Eyeball It Method

```
par(mfrow=c(1,2))
boxplot(len~dose,data=ToothGrowth) #shape/spread look OK
hist(fit$residuals) #residuals look mound-shaped!
```





Testing Assumptions: Look At P-Value Method

- Test normality: H_0 : sample is from a normally distributed population

```
shapiro.test(fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit$residuals
## W = 0.96731, p-value = 0.1076
```

- Test homoskedasticity: H_0 : samples from pops with equal variance

```
leveneTest(len~dose,data=ToothGrowth)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.6457 0.5281
##       57
```

Post-hoc t tests to see which groups differ

```r
# Note: adjust="none" does not correct p-values
# Instead, let's do boneferroni by hand!
library(emmeans)
emmeans(fit,pairwise~dose, adjust="none")$contrasts
```

```
##  contrast estimate   SE df t.ratio p.value
##  0.5 - 1     -9.13 1.34 57  -6.806 <.0001
##  0.5 - 2    -15.49 1.34 57 -11.551 <.0001
##  1 - 2       -6.37 1.34 57  -4.745 <.0001
```

- Three t tests, so probability of at least one type I error is $1 - .95^3 = .143$
- To keep overall Type I error rate at .05, set $\alpha = .05/3 = .01667$ (Bonferroni correction)
- Even still, using this conservative α, everything is significant!

Post Hoc Tests

How much of an effect was there? Effect size!

- How much of variability in growth can we attribute to dosage?
  $\rightarrow$ How much of **total** variability due to differences **between** groups?
- Identical to R^2 but usually called η^2 in ANOVA contexts
- $\eta^2 = \frac{\textit{Between-groups SS}}{\textit{Total SS}} = \frac{SS_B}{SS_W + SS_B} = \frac{SS_B}{SS_T}$
- Represents proportion of total variation due to differences in dosage
- $\eta^2 = \frac{2426.4}{2426.4 + 1025.8} = \frac{2426.4}{3452.2} = .703$
- "70.3% of the variation in tooth growth is attributable to dosage"

ANOVA Write-Up

After completing a one-way ANOVA test, we reject the null hypothesis and conclude that mean tooth growth does differ by vitamin C dosage. We conclude this since our F-statistic was more extreme than the critical value, $F(2, 57) = 67.42$, $F_{critical} = 3.15$, $p =< .0001$. A large effect was found, as 70.3% of the variation in tooth growth was accounted for by dosage differences. Post-hoc tests were performed to determine which dosage levels differed in mean tooth growth: t tests with bonferroni correction (adjusted $\alpha = .05/3 = .016667$) revealed that all groups were significantly different: guinea pigs receiving a dosage of 2 mg had significantly greater mean tooth growth than those receiving a dosage of 0.5 mg ($t = 12.802$, $p =< .0001$) or a dosage of 1.0 mg ($t = 5.259$, $p =< .0001$), and guinea pigs receiving a dosage of 1.0 mg had significantly greater tooth growth than those receiving a dosage of 0.5 mg ($t = 7.543$, $p < .0001$)

Two-Way ANOVA!

- Simultaneously tests effects of two categorical explanatory variables ("factors") on a numeric response variable

- Test the effect of Factor A controlling for Factor B (and vice versa)
- Test interaction between Factors A and B

- Each factor/interaction has its own set of **hypotheses**
- Each factor/interaction has its own **F-statistic**
- Each factor/interaction has its own **p-value**

- An interaction tests if the effect of one factor on the response differs across levels of the other factor (this will make sense shortly)

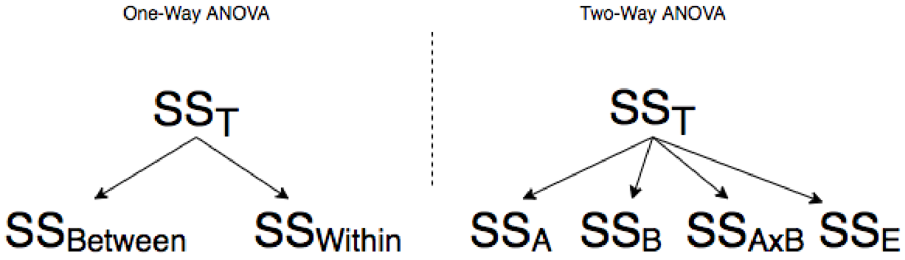
Assumptions

1. Random sample, independent observations
2. Each group within each factor is roughly normal (or $n > 25$)
3. Each group within each factor has equal variance

- To assess normality, look at boxplots (or histograms/qq-plots) for each group *of each factor!*
  - Or just make a histogram of your model's residuals!

To assess homoskedasticity, Levene's test, boxplots, or look at sample sds ($\forall i, j : s_i < 2s_j$)

Partitioning the Total Sum of Squares (SS_T)



- One-Way ANOVA, **two** sources of variation:
  (between-groups, within-groups)
- Two-Way ANOVA, **four** sources:
  (factor A, factor B, A$\times$B interaction, within-cell/error)
- For a two-way ANOVA, $SS_T = SS_A + SS_B + SS_{A\times B} + SS_E$
- If you know four of these, you can solve for the other!

Two-Way ANOVA table (FULL, WITH INTERACTION)

Source	SS	df	MS	F
Factor A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Factor B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
A×B	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{A\times B}}{MS_E}$
Error	SS_E	$N-ab$	$MS_E = \frac{SS_E}{N-ab}$	
Total	SS_T	$N-1$		

- Notice! All F-ratios have the same denominator (MSE)
- You will have a significant effect of Factor A controlling for Factor B

Two-Way ANOVA table (NO INTERACTION)

Source	SS	df	MS	F
Factor A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Factor B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Error	SS_E	$N-a-b+1$	$MS_E = \frac{SS_E}{N-a-b+1}$	
Total	SS_T	$N-1$		

- Notice! All F-ratios have the same denominator (MSE)
- You will have a significant effect of Factor A controlling for Factor B

You will NOT have to calculate these sums-of-squares by hand!

(You'll mostly be running 2-way ANOVAs and interpreting output.)

However, I'm still going to show you how in case you are curious!

How to Think about Two-Way ANOVA

	B_1	B_2	$\ldots$	B_b	
A_1	$y_{1,1,1}, \ldots, y_{1,1,r}$	$y_{1,2,1}, \ldots, y_{1,2,r}$	$\ldots$	$y_{1,b,1}, \ldots, y_{1,3,r}$	$\bar{y}_{1\bullet\bullet}$
A_2	$y_{2,1,1}, \ldots, y_{2,1,r}$	$y_{2,2,1}, \ldots, y_{2,2,r}$	$\ldots$	$y_{2,b,1}, \ldots, y_{2,3,r}$	$\bar{y}_{2\bullet\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_a	$y_{a,1,1}, \ldots, y_{2,1,r}$	$y_{2,2,1}, \ldots, y_{2,2,r}$	$\ldots$	$y_{2,3,1}, \ldots, y_{2,3,r}$	$\bar{y}_{a\bullet\bullet}$
	$\bar{y}_{\bullet 1\bullet}$	$\bar{y}_{\bullet 2\bullet}$	$\ldots$	$\bar{y}_{\bullet b\bullet}$	$\bar{y}_{\bullet\bullet\bullet}$

- y is the numeric response variable; A and B are categorical explanatory variables ("Factors")
- There are a levels of factor A (indexed by i)
- There are b levels of factor B (indexed by j)
- Thus, there are $a \times b$ cells (indexed by k)
- Each cell contains r observations
- Thus, there are a total of $N = a \times b \times r$ observations!

- Observation y_{ijk} is the k^{th} observation in row i and column j
- $\bar{y}_{\bullet\bullet\bullet}$ = grand mean; $\bar{y}_{i\bullet\bullet}$ = row i mean; $\bar{y}_{\bullet j\bullet}$ = column j mean
- $\bar{y}_{ij\bullet}$ = cell ij mean

Two-Way ANOVA Scenario

- You want to see whether detergent type (A or B) or water temperature (Cold, Warm, Hot) affects how clean your clothes get (numeric, 1-15 rating).
- You wash 24 loads, 4 in each of the 6 conditions (A:cold, A:warm, A:hot, B:cold, B:warm, B:hot): The cleanness ratings for each load are shown in their respective cells below

	Detergent A	*Detergent B*	
Cold	$4, 5, 6, 5$	$6, 6, 4, 4$	$\bar{y}_{1\bullet\bullet}$
Warm	$7, 9, 8, 12$	$13, 15, 12, 12$	$\bar{y}_{2\bullet\bullet}$
Hot	$10, 12, 11, 9$	$12, 13, 10, 13$	$\bar{y}_{3\bullet\bullet}$
	$\bar{y}_{\bullet 1\bullet}$	$\bar{y}_{\bullet 2\bullet}$	$\bar{y}_{\bullet\bullet\bullet}$

- Here, $a = 3$, $b = 2$, $r = 4$, so $N = 3 \times 2 \times 4 = 24$

Two-Way ANOVA

Computing the row and column means shows you average cleanness for each detergent and temperature

	Detergent A	*Detergent B*	
Cold	$4, 5, 6, 5$	$6, 6, 4, 4$	$\bar{y}_{1\bullet\bullet} = 5$
Warm	$7, 9, 8, 12$	$13, 15, 12, 12$	$\bar{y}_{2\bullet\bullet} = 11$
Hot	$10, 12, 11, 9$	$12, 13, 10, 13$	$\bar{y}_{3\bullet\bullet} = 11.25$
	$\bar{y}_{\bullet 1\bullet} = 8.167$	$\bar{y}_{\bullet 2\bullet} = 10$	$\bar{y}_{\bullet\bullet\bullet} = 9.083$

How to Think about Two-Way ANOVA

	B_1	B_2	$\ldots$	B_b	
A_1	$y_{1,1,1},\ldots,y_{1,1,r}$	$y_{1,2,1},\ldots,y_{1,2,r}$	$\ldots$	$y_{1,b,1},\ldots,y_{1,3,r}$	$\bar{y}_{1\bullet\bullet}$
A_2	$y_{2,1,1},\ldots,y_{2,1,r}$	$y_{2,2,1},\ldots,y_{2,2,r}$	$\ldots$	$y_{2,b,1},\ldots,y_{2,3,r}$	$\bar{y}_{2\bullet\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_a	$y_{a,1,1},\ldots,y_{2,1,r}$	$y_{2,2,1},\ldots,y_{2,2,r}$	$\ldots$	$y_{2,3,1},\ldots,y_{2,3,r}$	$\bar{y}_{a\bullet\bullet}$
	$\bar{y}_{\bullet 1\bullet}$	$\bar{y}_{\bullet 2\bullet}$	$\ldots$	$\bar{y}_{\bullet b\bullet}$	$\bar{y}_{\bullet\bullet\bullet}$

- $\mathbf{SS_A} = rb\sum_i^a(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the means of Factor A from the grand mean"
- $\mathbf{SS_B} = ra\sum_j^b(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the means of Factor B from the grand mean"
- $\mathbf{SS_{A\times B}} = r\sum_i^a\sum_j^b \overbrace{((\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}))^2}^{(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2}$
  "Deviations of the cell means from the grand mean, minus the effects of factors A and B"
- $\mathbf{SS_E} = \sum_i^a\sum_j^b\sum_k^r(y_{ijk} - \bar{y}_{ij\bullet})^2$
  "Deviations of the observations from their cell means" (These are the residuals!)
- $\mathbf{SS_T} = \sum_i^a\sum_j^b\sum_k^r(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the observations from the grand mean"

How to Think about Two-Way ANOVA

	B_1	B_2	$\ldots$	B_b	
A_1	$\bar{y}_{1,1\bullet}$	$\bar{y}_{1,2\bullet}$	$\ldots$	$\bar{y}_{1,b\bullet}$	$\bar{y}_{1\bullet\bullet}$
A_2	$\bar{y}_{2,1\bullet}$	$\bar{y}_{2,2\bullet}$	$\ldots$	$\bar{y}_{2,b\bullet}$	$\bar{y}_{2\bullet\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_a	$\bar{y}_{a,1\bullet}$	$\bar{y}_{a,2\bullet}$	$\ldots$	$\bar{y}_{a,b\bullet}$	$\bar{y}_{a\bullet\bullet}$
	$\bar{y}_{\bullet 1\bullet}$	$\bar{y}_{\bullet 2\bullet}$	$\ldots$	$\bar{y}_{\bullet b\bullet}$	$\bar{y}_{\bullet\bullet\bullet}$

- $\mathbf{SS_A} = rb\sum_i^a(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the means of Factor A from the grand mean"
- $\mathbf{SS_B} = ra\sum_j^b(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the means of Factor B from the grand mean"
- $\mathbf{SS_{AB}} = r\sum_i^a\sum_j^b \overbrace{((\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}))^2}^{(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2}$
  "Deviations of the cell means from the grand mean, minus the effects of factors A and B"
- $\mathbf{SS_E} = \sum_i^a\sum_j^b\sum_k^r(y_{ijk} - \bar{y}_{ij\bullet})^2$
  "Deviations of the observations from their cell means" (these are the residuals!)
- $\mathbf{SS_T} = \sum_i^a\sum_j^b\sum_k^r(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2$
  "Deviations of the observations from the grand mean"

Two-Way ANOVA Illustration

- Does wash quality (1-15) differ by detergent type?
- Does wash quality (1-15) differ by water temperature?
- Is there an interaction?

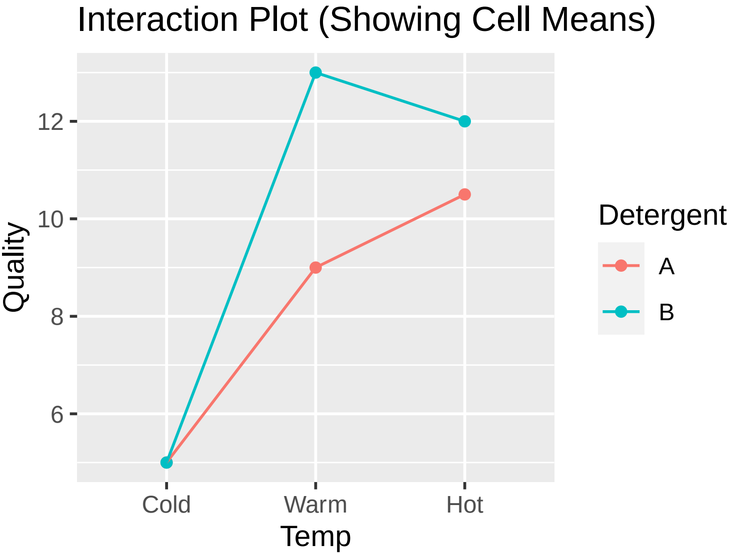
	Detergent A	*Detergent B*	
Cold	$4, 5, 6, 5$	$6, 6, 4, 4$	$\bar{y}_{1\bullet\bullet} = 5$
Warm	$7, 9, 8, 12$	$13, 15, 12, 12$	$\bar{y}_{2\bullet\bullet} = 11$
Hot	$10, 12, 11, 9$	$12, 13, 10, 13$	$\bar{y}_{3\bullet\bullet} = 11.25$
	$\bar{y}_{\bullet 1\bullet} = 8.167$	$\bar{y}_{\bullet 2\bullet} = 10$	$\bar{y}_{\bullet\bullet\bullet} = 9.083$

Two-Way ANOVA Illustration

- Does wash quality (1-15) differ by detergent type?
- Does wash quality (1-15) differ by water temperature?
- Is there an interaction? What would that mean?

	Detergent A	*Detergent B*	
Cold	$\bar{y}_{1,1\bullet} = 5$	$\bar{y}_{1,2\bullet} = 5$	$\bar{y}_{1\bullet\bullet} = 5$
Warm	$\bar{y}_{2,1\bullet} = 9$	$\bar{y}_{2,2\bullet} = 13$	$\bar{y}_{2\bullet\bullet} = 11$
Hot	$\bar{y}_{3,1\bullet} = 10.5$	$\bar{y}_{3,2\bullet} = 12$	$\bar{y}_{3\bullet\bullet} = 11.25$
	$\bar{y}_{\bullet 1\bullet} = 8.167$	$\bar{y}_{\bullet 2\bullet} = 10$	$\bar{y}_{\bullet\bullet\bullet} = 9.083$

Two-Way ANOVA Illustration



Sums of Squares from R

```
...
## Response: Quality
##                   Sum Sq Df F value   Pr(>F)
## Detergent         20.167  1  9.8108 0.005758 **
## Temp             200.333  2 48.7297 5.44e-08 ***
## Detergent:Temp    16.333  2  3.9730 0.037224 *
## Residuals         37.000 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

- We will do a full worked example later with a write-up: this output is just to get you thinking!

TOTALLY EXTRA: YOU'LL NEVER HAVE TO DO THESE BY HAND!

$$\mathbf{SS}_{\mathbf{Detergent}} = rb\sum_{i}^{a}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$$

$$= 4 * 3 * ((8.167 - 9.083)^2 + (10 - 9.083)^2) = \mathbf{20.16}$$

$$\mathbf{SS}_{\mathbf{Temperature}} = ra\sum_{i}^{b}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$$

$$= 4 * 2 * ((5 - 9.083)^2 + (11 - 9.083)^2 + (11.25 - 9.083)^2) = \mathbf{200.33}$$

$$\mathbf{SS}_{\mathbf{Interaction}} = r\sum_{i}^{a}\sum_{j}^{b}\overbrace{((\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}))^2}^{(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2}$$

$$= 4 * (((5 - 9.083) - (8.167 - 9.083) - (5 - 9.083))^2 +$$
$$((9 - 9.083) - (8.167 - 9.083) - (11 - 9.083))^2 +$$
$$((10.5 - 9.083) - (8.167 - 9.083) - (11.25 - 9.083))^2 +$$
$$((5 - 9.083) - (10 - 9.083) - (5 - 9.083))^2 +$$
$$((13 - 9.083) - (10 - 9.083) - (11 - 9.083))^2 +$$
$$((12 - 9.083) - (10 - 9.083) - (11.25 - 9.083))^2) = \mathbf{16.333}$$

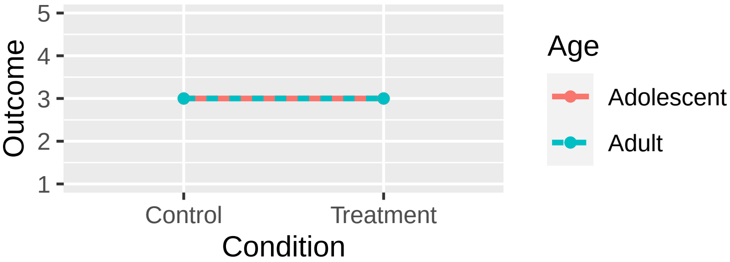
Two-Way ANOVA: 8 possibilities

- No effects
- Main effect of A
- Main effect of B
- Main effect of A and B
- A$\times$B Interaction, No main effects
- A$\times$B Interaction, Main effect of A
- A$\times$B Interaction, Main effect of B
- A$\times$B Interaction, Main effects of A and B

1. No Interactions, No Main Effects

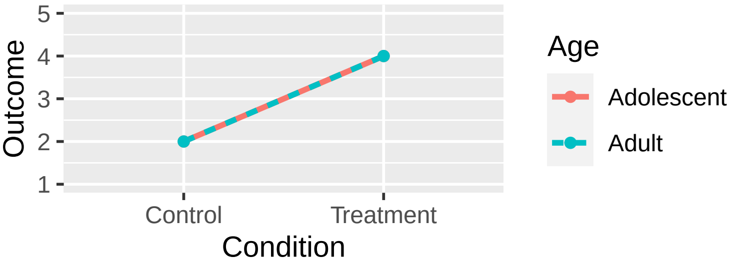
- No differences in cell means or group means

Condition	Adolescent	Adult	
Control	3	3	**3**
Treatment	3	3	**3**
	3	**3**	**3**



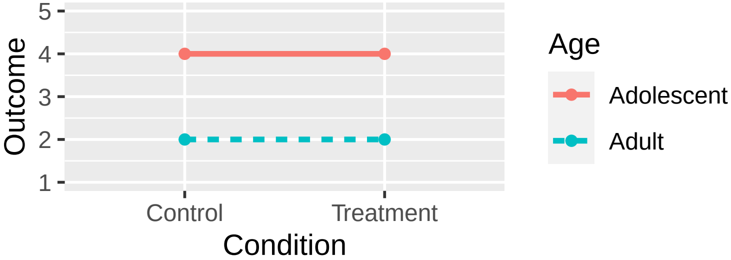
2. No Interaction, Main Effect of Treatment

Condition	Adolescent	Adult	
Control	2	2	**2**
Treatment	4	4	**4**
	3	**3**	**3**



3. No Interaction, Main Effect of Age

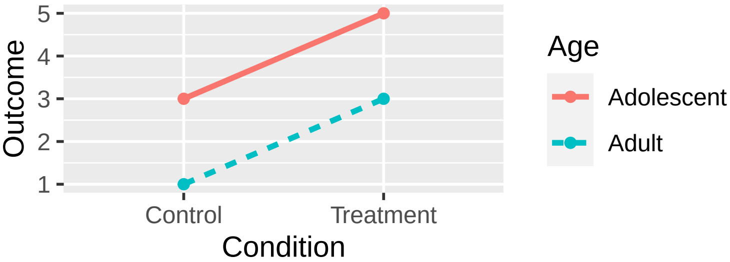
Condition	Adolescent	Adult	
Control	4	2	**3**
Treatment	4	2	**3**
	4	**2**	**3**



- Notice that each dot corresponds to a cell mean!

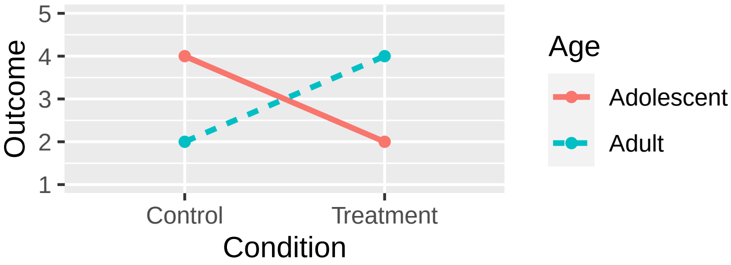
4. No Interaction, Two Main Effects

Condition	Adolescent	Adult	
Control	3	1	**2**
Treatment	5	3	**4**
	4	**2**	**3**



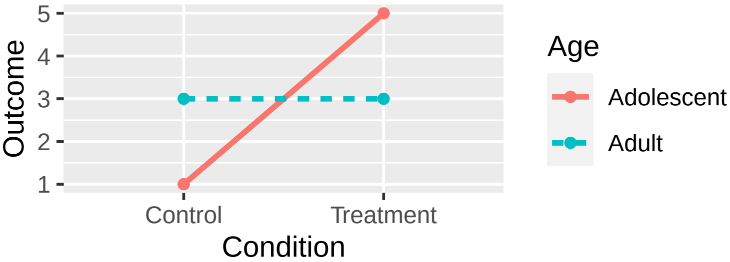
5. Interaction, No Main Effects

Condition	Adolescent	Adult	
Control	4	2	**3**
Treatment	2	4	**3**
	3	**3**	**3**



6. Interaction and Main Effect of Treatment

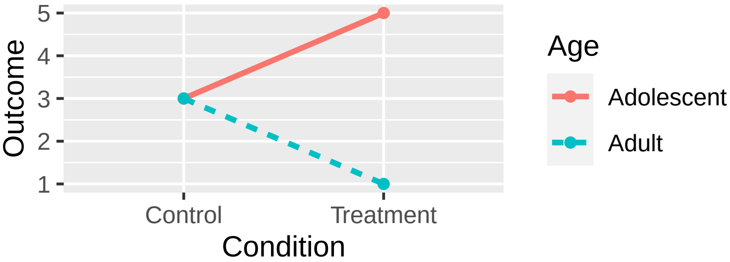
Condition	Adolescent	Adult	
Control	1	3	**2**
Treatment	5	3	**4**
	3	**3**	**3**



7. Interaction and Main Effect of Age

- Interaction is what's important, not main effect.

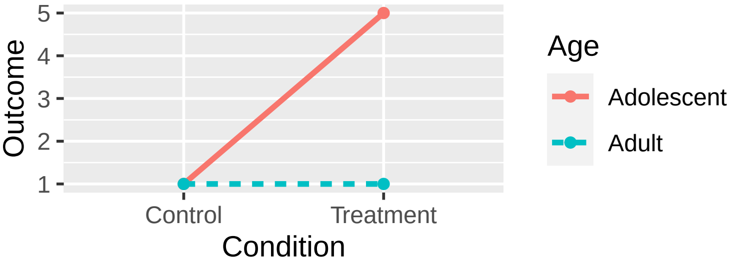
Condition	Adolescent	Adult	
Control	3	3	**3**
Treatment	5	1	**3**
	4	**2**	**3**



8. Interaction and Both Main Effects

- Interaction is what's important, not main effect.

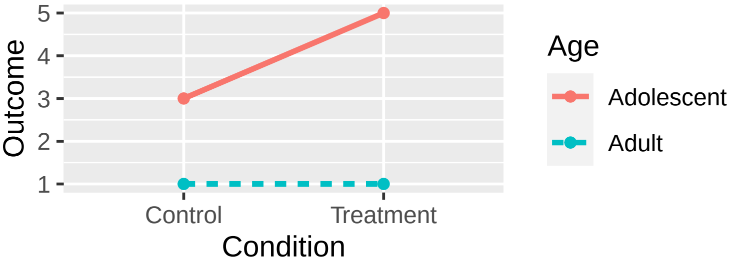
Condition	Adolescent	Adult	
Control	1	1	**1**
Treatment	5	1	**3**
	3	**1**	**2**

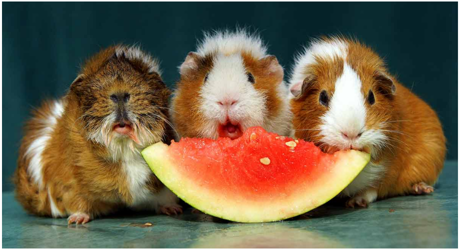


8. Interaction and Both Main Effects (II)

- Main effect of age holds for both conditions, so it is interpretable

Condition	Adolescent	Adult	
Control	3	1	**2**
Treatment	5	1	**3**
	4	**1**	**2.5**





- How is guinea pig tooth growth influenced by vitamin C dosage?
- Is it better to give them vitamin C tablets or orange juice?
- Does the effect of dosage depend on which supplement type is used?

```
#this dataset comes with R
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##       0.5  1  2
##   OJ   10 10 10
##   VC   10 10 10
```

```r
head(ToothGrowth,20)
```

```
##     len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
## 11 16.5   VC    1
## 12 16.5   VC    1
## 13 15.2   VC    1
## 14 17.3   VC    1
## 15 22.5   VC    1
## 16 17.3   VC    1
## 17 13.6   VC    1
## 18 14.5   VC    1
## 19 18.8   VC    1
## 20 15.5   VC    1
```

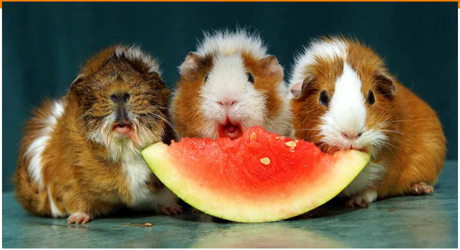
```r
head(tail(ToothGrowth,40),20)
```

```
##     len supp dose
## 21 23.6   VC    2
## 22 18.5   VC    2
## 23 33.9   VC    2
## 24 25.5   VC    2
## 25 26.4   VC    2
## 26 32.5   VC    2
## 27 26.7   VC    2
## 28 21.5   VC    2
## 29 23.3   VC    2
## 30 29.5   VC    2
## 31 15.2   OJ  0.5
## 32 21.5   OJ  0.5
## 33 17.6   OJ  0.5
## 34  9.7   OJ  0.5
## 35 14.5   OJ  0.5
## 36 10.0   OJ  0.5
## 37  8.2   OJ  0.5
## 38  9.4   OJ  0.5
## 39 16.5   OJ  0.5
## 40  9.7   OJ  0.5
```

```r
tail(ToothGrowth,20)
```

```
##     len supp dose
## 41 19.7   OJ    1
## 42 23.3   OJ    1
## 43 23.6   OJ    1
## 44 26.4   OJ    1
## 45 20.0   OJ    1
## 46 25.2   OJ    1
## 47 25.8   OJ    1
## 48 21.2   OJ    1
## 49 14.5   OJ    1
## 50 27.3   OJ    1
## 51 25.5   OJ    2
## 52 26.4   OJ    2
## 53 22.4   OJ    2
## 54 24.5   OJ    2
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

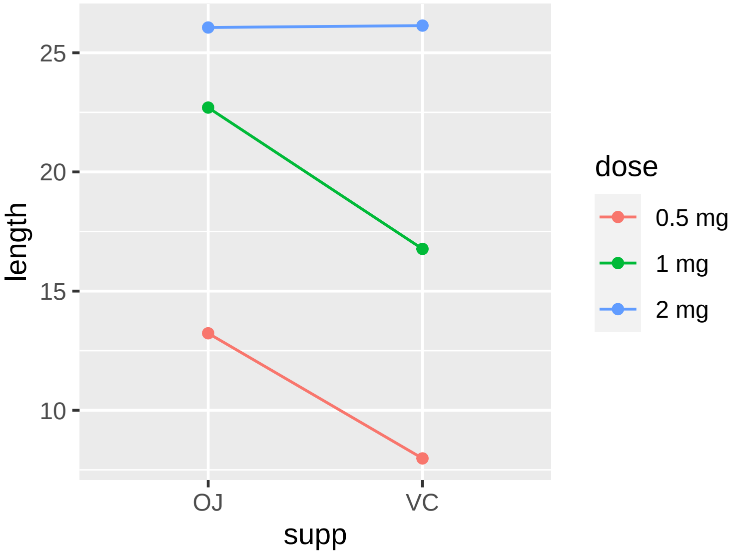
Example



supp	0.5 mg	1 mg	2 mg	
OJ	13.23	22.7	26.06	**20.663**
VC	7.98	16.77	26.14	**16.963**
	10.605	**19.735**	**26.1**	**18.813**

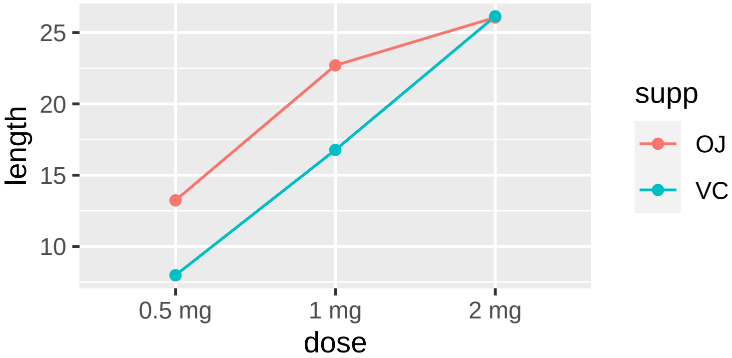
- What's the mean tooth length on the 0.5 mg dose? 10.605
- What's the mean tooth length overall? 18.813
- What's the mean tooth length for OJ and 2 mg? 26.06

Plot of the cell means (interaction plot)



Plot of the cell means (interaction plot)

Or plot them like this:



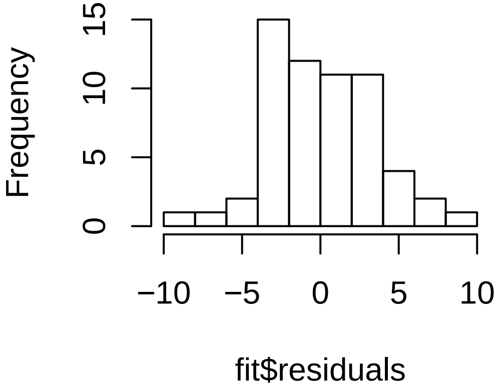
Hypotheses

1. H_0 : Controlling for dose, no effect of supplement type: $\mu_{OJ} = \mu_{VC}$
   H_A : Effect of dose: at least group mean one differs
2. H_0 : Controlling for supplement type, no effect of dose:
   $\mu_{0.5mg} = \mu_{1mg} = \mu_{2mg}$
   H_A : Effect of dose: at least group mean one differs
3. H_0 : No *dose* $\times$ *supplement* interaction effect
   H_A : There is a significant interaction between dose and supplement type on tooth growth

Assumptions

```
fit<-lm(len ~ dose * supp, data=ToothGrowth) #fit full model
```

```
#or eyeball normality: looks fine!
hist(fit$residuals, main="")
```




```
#or do formal tests
#normality is OK (p > .05)
shapiro.test(fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit$residuals
## W = 0.98499, p-value = 0.6694
```

```
#homoskedasticity is OK (p > .05)
leveneTest(fit)
```

```
## Levene's Test for Homogeneity
##       Df F value Pr(>F)
## group  5  1.7086 0.1484
##       54
```

Sums of Squares for Supplement Type (OJ vs VC)

$$
\begin{aligned}
SS_{supp} &= rb \sum_{i}^{2} (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2 \\
&= 10 * 3 * \sum_{i}^{2} (\bar{y}_{i\bullet\bullet} - 18.813)^2 \\
&= 10 * 3 * \left((20.663 - 18.813)^2 + (16.963 - 18.813)^2 \right) \\
&= 205.35
\end{aligned}
$$

Example: One-Way ANOVA

- Overall effect of supplement type:

```
fit<-lm(len ~ supp, data=ToothGrowth)
Anova(fit)
```

```
...
## Response: len
##           Sum Sq Df F value  Pr(>F)
## supp       205.4  1  3.6683 0.06039 .
## Residuals 3246.9 58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

Sums of Squares for Dose (0.5, 1, or 2)

$$
\begin{aligned}
SS_{dose} &= ra \sum_{j}^{3} (\bar{y}_{\bullet j \bullet} - \bar{y}_{\bullet\bullet\bullet})^2 \\
&= 10 \times 2 \times \sum_{j}^{3} (\bar{y}_{\bullet j \bullet} - 18.813)^2 \\
&= 20 \times \left((10.605 - 18.813)^2 + (19.735 - 18.813)^2 + (26.1 - 18.813)^2\right) \\
&= 2426.434
\end{aligned}
$$

Example: One-Way ANOVA

- Overall effect of dose:

```
fit<-lm(len ~ dose, data=ToothGrowth)
Anova(fit)
```

```
...
## Response: len
##            Sum Sq Df F value    Pr(>F)
## dose       2426.4  2  67.416 9.533e-16 ***
## Residuals 1025.8 57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

Example: Two-Way ANOVA, no Interaction

- Effect of supplement type, *controlling for dose* (it's significant now!)
- Effect of dose, *controlling for supplement type*

```
fit<-lm(len ~ supp + dose, data=ToothGrowth)
Anova(fit)
```

```
...
## Response: len
##            Sum Sq Df F value    Pr(>F)
## supp       205.35  1  14.017 0.0004293 ***
## dose      2426.43  2  82.811 < 2.2e-16 ***
## Residuals  820.43 56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

Calculation of Dose $\times$ Supplement Interaction Effect

Shortcut:

- $SS_{cells} = r\sum_i^a \sum_j^b (\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$ (deviations of cell means around grand mean)
- $SS_{A\times B} = SS_{cells} - SS_A - SS_B$

$$
\begin{aligned}
SS_{cells} &= 10\sum_i^2 \sum_j^3 (\bar{y}_{ij\bullet} - 18.813)^2 \\
&= 10 \times ((13.23 - 18.813)^2 + (22.7 - 18.813)^2 + (26.06 - 18.813)^2 + \\
&\quad ...(7.98 - 18.813)^2 + (16.77 - 18.813)^2 + (26.14 - 18.813)^2) \\
&= 10 \times (31.170 + 15.109 + 52.519 + 4.174 + 53.685) \\
&= 2740.103
\end{aligned}
$$

- $SS_{supp\times dose} = SS_{cells} - SS_{supp} - SS_{dose}$
- $SS_{supp\times dose} = 2740.103 - 205.35 - 2426.434 = \mathbf{108.32}$

Example: Two-Way ANOVA with Interaction

- Effect of *supplement*, controlling for dose
- Effect of *dose*, controlling for supplement
- Effect of *dose* $\times$ *supp*

```
fit<-lm(len ~ supp + dose + supp:dose, data=ToothGrowth)
Anova(fit)
```

```
...
## Response: len
##            Sum Sq Df F value    Pr(>F)
## supp       205.35  1  15.572 0.0002312 ***
## dose      2426.43  2  92.000 < 2.2e-16 ***
## supp:dose  108.32  2   4.107 0.0218603 *
## Residuals  712.11 54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

```r
#Here's a quick way to get all of the means
model.tables(aov(fit), type = "means")
```

```
## Tables of means
## Grand mean
##
## 18.81333
##
##  supp
## supp
##     OJ     VC
## 20.663 16.963
##
##  dose
## dose
## 0.5 mg   1 mg   2 mg
## 10.605 19.735 26.100
##
##  supp:dose
##     dose
## supp 0.5 mg 1 mg  2 mg
##   OJ 13.23  22.70 26.06
##   VC  7.98  16.77 26.14
```

ANOVA table for Guinea Pig example

Source	SS	df	MS	F
dose	2426.43	2	$\frac{2426.43}{2} = 1213.215$	$\frac{1213.215}{13.18} = 92.05$
supp	205.35	1	$\frac{205.35}{1} = 205.35$	$\frac{205.35}{13.18} = 15.58$
dose×supp	108.32	2	$\frac{108.32}{2} = 54.16$	$\frac{54.16}{13.18} = 4.11$
error	712.11	54	$\frac{712.11}{54} = 13.18$	
total	3452.21	59		

- Critical value for main effect of *dose*: $F^*_{(2,54)} = 3.16$
- Critical value for main effect of *supplement*: $F^*_{(1,54)} = 4.02$
- Critical value for interaction effect: $F^*_{(2,54)} = 3.16$
- **Decision?** Reject 'em all, but *the interaction is the main event!*

R^2 (η^2)

$$\begin{aligned} R^2 &= 1 - \frac{SS_E}{SS_T} \\ &= 1 - \frac{712.11}{205.35 + 2426.43 + 108.32 + 712.11} \\ &= 1 - \frac{712.11}{3452.21} \\ &= .794 \end{aligned}$$

```
summary(fit)$r.sq
```

```
## [1] 0.7937246
```

WARNING!

- The sum-of-squares hand-calculations will work fine when the data is balanced (equal number of observations in each cell)
- However, in the real world, much data is unbalanced (i.e., unequal cell sizes)
- This imbalance complicates sum-of-squares calculations!
- To get the correct calculations, always run your model in R with type 3 sums-of-squares like this (ignore the intercept entirely)

ALWAYS RUN YOUR 2-WAY ANOVA LIKE THIS:

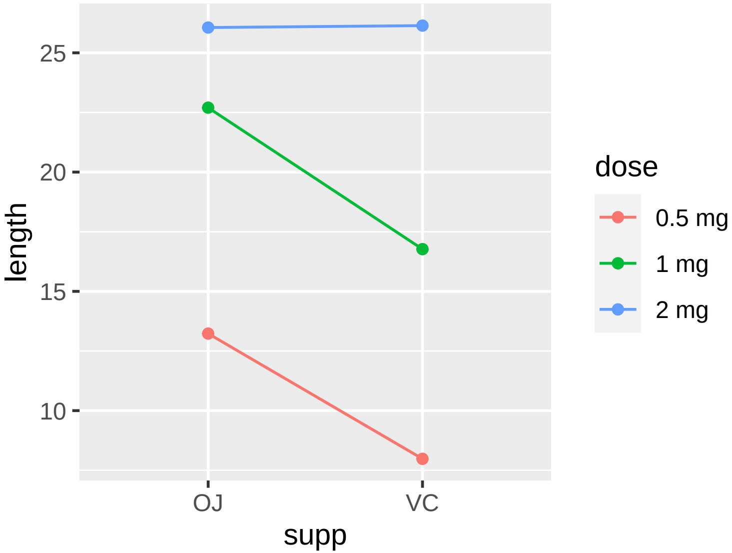
```
options(contrasts = c("contr.sum","contr.poly"))
fit<-lm(len ~ supp*dose, data=ToothGrowth)
Anova(fit, type=3)
```

```
...
## Response: len
##              Sum Sq Df  F value    Pr(>F)
## (Intercept) 21236.5  1 1610.393 < 2.2e-16 ***
## supp          205.4  1   15.572 0.0002312 ***
## dose         2426.4  2   92.000 < 2.2e-16 ***
## supp:dose     108.3  2    4.107 0.0218603 *
## Residuals     712.1 54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

Two-Way ANOVA write-up

A two-way ANOVA was performed to assess the effect of dosage and supplement type on guinea pig tooth growth. There were two significant main effects: While controlling for supplement type, there was a significant effect of dose ($F(2, 54) = 92.00$, $p < .0001$), and while controlling for dose, there was a significant effect of supplement type ($F(1, 54) = 15.57$, $p = .0002$). However, these were qualified by a significant interaction, $F(2, 54) = 4.11$, $p = .022$. The effect of dosage on tooth-growth depends on which supplement type is taken! Vitamin C dose, supplement type, and their interaction account for 79.4% of the variability in guinea pig tooth growth!

Which cell means are significantly different?



Post Hoc t Tests (only do comparisons of interest!)

```r
emmeans(fit,pairwise~dose|supp, adjust="none")$contrasts
```

```
## supp = OJ:
##  contrast        estimate   SE df t.ratio p.value
##  0.5 mg - 1 mg      -9.47 1.62 54  -5.831 <.0001
##  0.5 mg - 2 mg     -12.83 1.62 54  -7.900 <.0001
##  1 mg - 2 mg        -3.36 1.62 54  -2.069 0.0434
##
## supp = VC:
##  contrast        estimate   SE df t.ratio p.value
##  0.5 mg - 1 mg      -8.79 1.62 54  -5.413 <.0001
##  0.5 mg - 2 mg     -18.16 1.62 54 -11.182 <.0001
##  1 mg - 2 mg        -9.37 1.62 54  -5.770 <.0001
```

```r
#Or, if you want ALL comparisons of ALL means, do
#emmeans(fit,pairwise~dose*supp, adjust="none")$contrasts
```

- Doing 6 tests, $\alpha = .05/6 = .0083$, so all sig. diff. except OJ 1mg vs. OJ 2mg

Course Packet Ch 15

PRACTICE

- One-Way ANOVA example (Ch 15A)
- Two-Way ANOVA example (Ch 15B)
    - 15B #1 is with NO interaction
    - 15B #3 is WITH an interaction!
- **Review these (and the next examples) if you get stuck!**

Full write-up for the laundry example

(In practice, you would state hypotheses and test assumptions first: this is just the write-up!)

```
...
## Response: Quality
##                  Sum Sq Df F value    Pr(>F)
## Detergent        20.167  1  9.8108 0.005758 **
## Temp            200.333  2 48.7297 5.44e-08 ***
## Detergent:Temp   16.333  2  3.9730 0.037224 *
## Residuals        37.000 18
...
```

A Two-Way ANOVA was conducted to assess the impact of detergent type (A vs. B) and temperature (cold, warm, hot) on clothes cleanness. While controlling for temperature, there is a main effect of detergent on cleanness, $F(1, 18) = 9.81$, $p = .006$. Further, controlling for detergent, there is a main effect of temperature on cleanness $F(2, 18) = 48.73$, $p < .0001$ However, these effects are qualified by a significant interaction, $F(2, 18) = 3.97$, $p = .04$: the effect of temperature differred between detergents. The model including reading temperature, detergent, and their interaction explained roughly $1 - (37/273.833) = 86.5\%$ of the variation in cleanness rating.

Another Full Example

- Do students learn to read better using Phonics or Whole Language instruction?
- Does the effect of curriculum type depend on the students' reading proficiency level?

```r
#some data
learndat<-data.frame(
  score=c(14,12,10,12,11,7,9,11,13,21,22,17,14,12,10,10,6,5),
  curr=factor(c(rep("PH",9),rep("WL",9))),
  prof=factor(rep(gl(3,3),2)))
```

Hypothesize

H_0 : Controlling for proficiency, Phonics and Whole-Language curricula are equally effective
H_A : Controlling for proficiency, Phonics and Whole-Language curricula are not equally effective

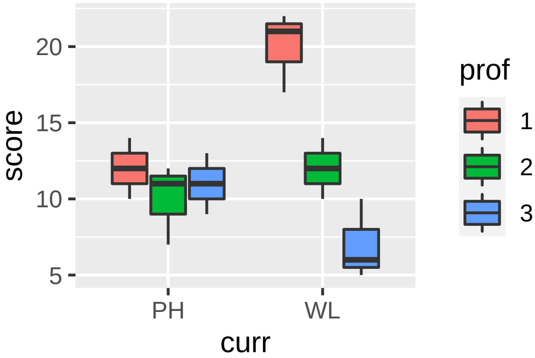
H_0 : There is no interaction between proficiency and curriculum on mean reading score

H_A : There is an interaction between proficiency and curriculum on mean reading score

(Could also test the effect of proficency, but it would be a waste of a test: we are already confident that more proficient readers would have higher reading scores on average)

Assumptions

```
library(car)
ggplot(learndat,aes(x = curr,y=score,fill=prof))+geom_boxplot()
```

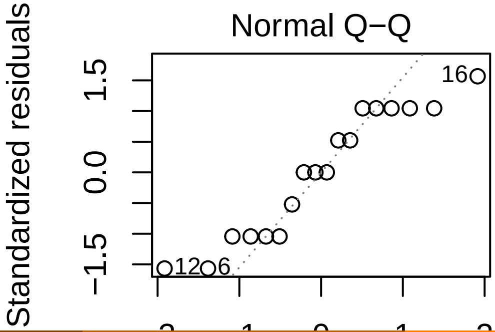


Assumptions continued

```r
#homoskedasticity good
leveneTest(score~curr*prof, data=learndat)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.0353 0.9992
##       12
```

```r
#normality not great: small sample. Let's proceed anyway
plot(lm(score~curr*prof,data=learndat),2)
```



Fill in ANOVA table to conduct F tests

Source	SS	df	MS	F
curr	18			
prof	156			
curr×prof				
error	66			
total	348			

Fill in ANOVA table to conduct F tests

Source	SS	df	MS	F
curr	18	1	$\frac{18}{1} = 18$	$\frac{18}{5.5} = 3.273$
prof	156	2	$\frac{156}{2} = 78$	
curr×prof	108	2	$\frac{108}{2} = 54$	$\frac{54}{5.5} = 9.818$
error	66	12	$\frac{66}{12} = 5.5$	
total	348	17		

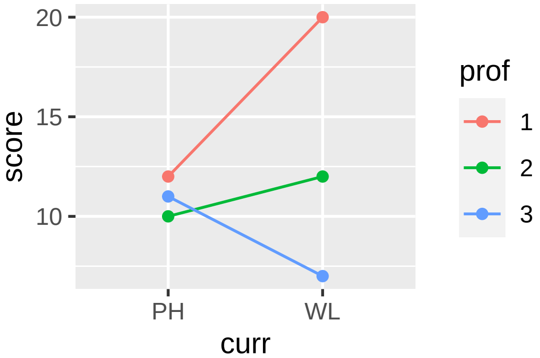
- Critical value for curriculum test: $F^*_{1,12} = 3.89$
- Critical value for interaction test: $F^*_{2,12} = 4.75$
- Decision?

Run it in R

```
learnfit<-lm(score~curr*prof,data=learndat)
Anova(learnfit)
```

```
...
## Response: score
##           Sum Sq Df F value    Pr(>F)
## curr          18  1  3.2727 0.0955444 .
## prof         156  2 14.1818 0.0006905 ***
## curr:prof    108  2  9.8182 0.0029783 **
## Residuals     66 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
...
```

R^2 and Interaction plot



$$R^2 = 1 - \frac{66}{348} = .8103$$

Writeup

While controlling for reading proficiency, there is no main effect of curriculum type on reading score, $F_(1, 12) = 3.273$, $p = .09$. However, this null is qualified by a significant interaction: the effect of curriculum differred by proficiency level such that proficient students did worse under the Whole-Language curriculum while the least proficient readers did the best under the Whole-Language curriculum, $F_{2,12} = 9.818$, $p < .0001$. The model including reading proficiency, curriculum, and their interaction explained roughly 81% of the variation in reading test score.

- If we had just run a one-way ANOVA, we would have no way of detecting this effect!
- Note that you don't need post-hoc tests (there are just two groups, so you automatically know they are different)

More Post Hoc Tests

Bonferroni

You want to test whether your tires are wearing down at different rates, so you sample 25 random locations on each tire and measure the tread. You run an ANOVA to see if there is a difference in average tread across the 4 wheels, and it is significant. How many post hoc tests do you need to run to find out which tires are different?

- How many unique groups of r from n distinct things: $\frac{n!}{(n-r)!r!}$
- How many unique groups of 2 from n distinct things: $\frac{n(n-1)}{2}$

$$\frac{n!}{(n-r)!r!} = \frac{4!}{2!2!} = \frac{4*3*2*1}{(2*1)(2*1)} = \frac{4*3}{2*1} = \frac{12}{2} = 6$$

- e.g., with groups A, B, C, D you have AB, AC, AD, BC, BD, CD

Post Hoc Test with Bonferroni Correction

Question

What adjusted α level (bonferroni) should you use to ensure overall $\alpha = .05$?

Question

Which tires differ significantly using the bonferroni correction? (p-values for each comparison shown below)

```
...
##
##    BL     BR      FL
## BR 0.1056 -       -
## FL 0.0020 5.6e-06 -
## FR 0.0897 0.0012  0.1471
##
## P value adjustment method: none
...
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  bonf$Response and bonf$Tire
##
##    BL     BR      FL
## BR 0.1056 -       -
## FL 0.0020 5.6e-06 -
## FR 0.0897 0.0012  0.1471
##
## P value adjustment method: none
```

Using bonferroni correction, $\alpha = .05/6 = .0083$

- BL different from FL ($p < .0083$)
- BR different from FL ($p < .0083$)
- FR different from BR ($p < .0083$)
- BL not different from BR ($p > .0083$)
- BL not different from FR ($p > .0083$)
- FR not different from FL ($p > .0083$)