

Chapter 6 Two-way ANOVA

```
# Load my usual packages
library(tidyverse) # ggplot2, dplyr, tidyr
library(ggfortify) # autoplot() for lm objects
library(emmeans)  # pairwise contrasts stuff
```

6.1 Review of 1-way ANOVA

Given a categorical covariate (which I will call a factor) with I levels, we are interested in fitting the model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$, μ is the overall mean, and τ_i are the offset of factor level i from μ . Unfortunately this model is not identifiable because I could add a constant (say 5) to μ and subtract that same constant from each of the τ_i values and the group mean $\mu + \tau_i$ would not change. There are two easy restrictions we could make to make the model identifiable:

1. Set $\mu = 0$. In this case, τ_i represents the expected value of an observation in group level i . We call this the “cell means” representation.
2. Set $\tau_1 = 0$. Then μ represents the expected value of treatment 1, and the τ_i values will represent the offsets from group 1. The group or level that we set to be zero is then referred to as the reference group. We can call this the “offset from reference” model.

We will be interested in testing the null and alternative hypotheses

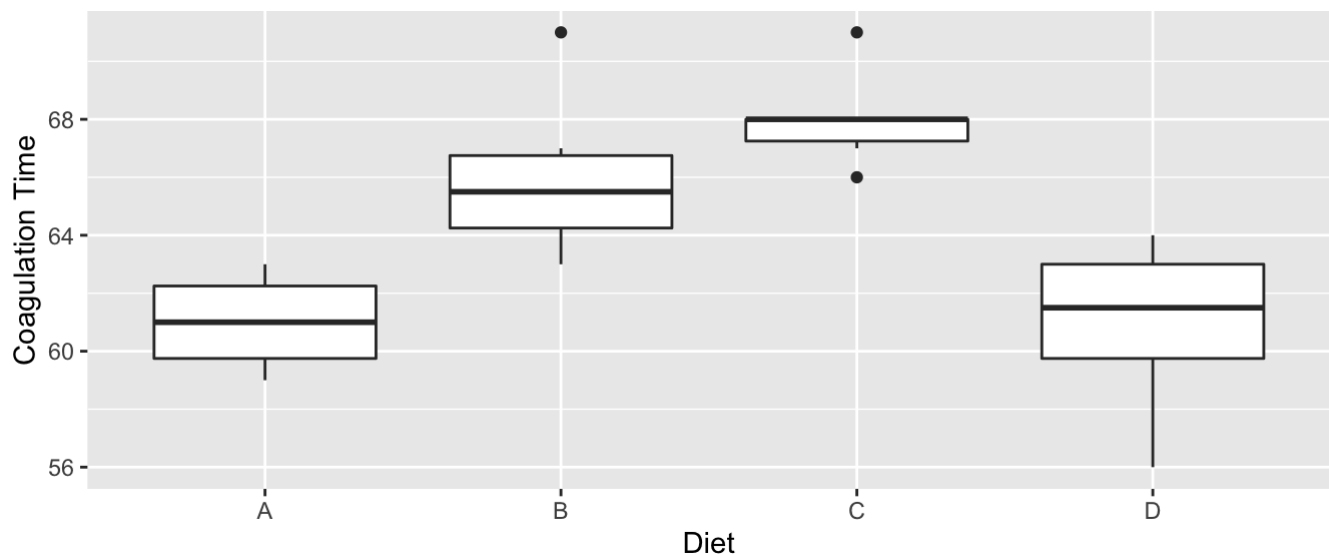
$$H_0 : y_{ij} = \mu + \epsilon_{ij}$$

$$H_a : y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

6.1.1 An Example

We look at a dataset that comes from the study of blood coagulation times: 24 animals were randomly assigned to four different diets and the samples were taken in a random order. The diets are denoted as A, B, C , and D and the response of interest is the amount of time it takes for the blood to coagulate.

```
data('coagulation', package='faraway')
ggplot(coagulation, aes(x=diet, y=coag)) +
  geom_boxplot() +
  labs( x='Diet', y='Coagulation Time' )
```



Just by looking at the graph, we expect to see that diets A and D are similar while B and C are different from A and D and possibly from each other, too. We first fit the offset model.

```
m <- lm(coag ~ diet, data=coagulation)
summary(m)
```

```
##
## Call:
## lm(formula = coag ~ diet, data = coagulation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.00  -1.25   0.00   1.25   5.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.100e+01  1.183e+00  51.554 < 2e-16 ***
## dietB       5.000e+00  1.528e+00   3.273 0.003803 **
## dietC       7.000e+00  1.528e+00   4.583 0.000181 ***
## dietD      2.991e-15  1.449e+00   0.000 1.000000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.366 on 20 degrees of freedom
## Multiple R-squared:  0.6706, Adjusted R-squared:  0.6212
## F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

Notice that diet A is the reference level and it has a mean of 61. Diet B has an offset from A of 5, etc. From the very small F-statistic, we conclude that simple model

$$y_{ij} = \mu + \epsilon_{ij}$$

is not sufficient to describe the data.

6.1.2 Degrees of Freedom

Throughout the previous example, the degrees of freedom that are reported keeps changed depending on what models we are comparing. The simple model we are considering is

$$y_{ij} \sim \mu + \epsilon_{ij}$$

which has 1 parameter that defines the expected value versus

$$y_{ij} \sim \mu + \tau_i + \epsilon_{ij}$$

where there really are only 4 parameters that define the expected value because $\tau_1 = 0$. In general, the larger model is only adding $I - 1$ terms to the model where I is the number of levels of the factor of interest.

6.1.3 Pairwise Comparisons

After detecting differences in the factor levels, we are often interested in which factor levels are different from which. Often we are interested in comparing the mean of level i with the mean of level j .

As usual we let the vector of parameter estimates be $\hat{\beta}$ then the contrast of interest can be written as

$$\mathbf{c}^T \hat{\beta} \pm t_{n-p}^{1-\alpha/2} \hat{\sigma} \sqrt{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}$$

for some vector \mathbf{c} .

Unfortunately this interval does not take into account the multiple comparisons issue (i.e. we are making $I(I - 1)/2$ contrasts if our factor has I levels). To account for this, we will not use a quantile from a t-distribution, but from Tukey's studentized range distribution $q_{n,n-I}$ divided by $\sqrt{2}$. The intervals we will use are:

$$\mathbf{c}^T \hat{\beta} \pm \frac{q_{n,n-I}^{1-\alpha/2}}{\sqrt{2}} \hat{\sigma} \sqrt{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}$$

There are several ways to make R calculate this interval, but the easiest is to use the `emmeans` package. This package computes the above intervals which are commonly known as Tukey's Honestly Significant Differences.

```
m <- lm(coag ~ diet, data=coagulation) # use the lm() function as usual
emmeans(m, specs= pairwise~diet) %>%
  summary(level=0.90)
```

```
## $emmeans
##   diet emmean      SE df lower.CL upper.CL
##   A         61 1.183 20     59.0     63.0
##   B         66 0.966 20     64.3     67.7
##   C         68 0.966 20     66.3     69.7
##   D         61 0.837 20     59.6     62.4
##
## Confidence level used: 0.9
##
## $contrasts
##   contrast estimate      SE df t.ratio p.value
##   A - B          -5 1.53 20  -3.273  0.0183
##   A - C          -7 1.53 20  -4.583  0.0010
##   A - D           0 1.45 20   0.000  1.0000
##   B - C          -2 1.37 20  -1.464  0.4766
##   B - D           5 1.28 20   3.912  0.0044
##   C - D           7 1.28 20   5.477  0.0001
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

Here we see that diets *A* and *D* are similar to each other, but different than *B* and *C* and that *B* and *C* are not statistically different from each other at the 0.10 level.

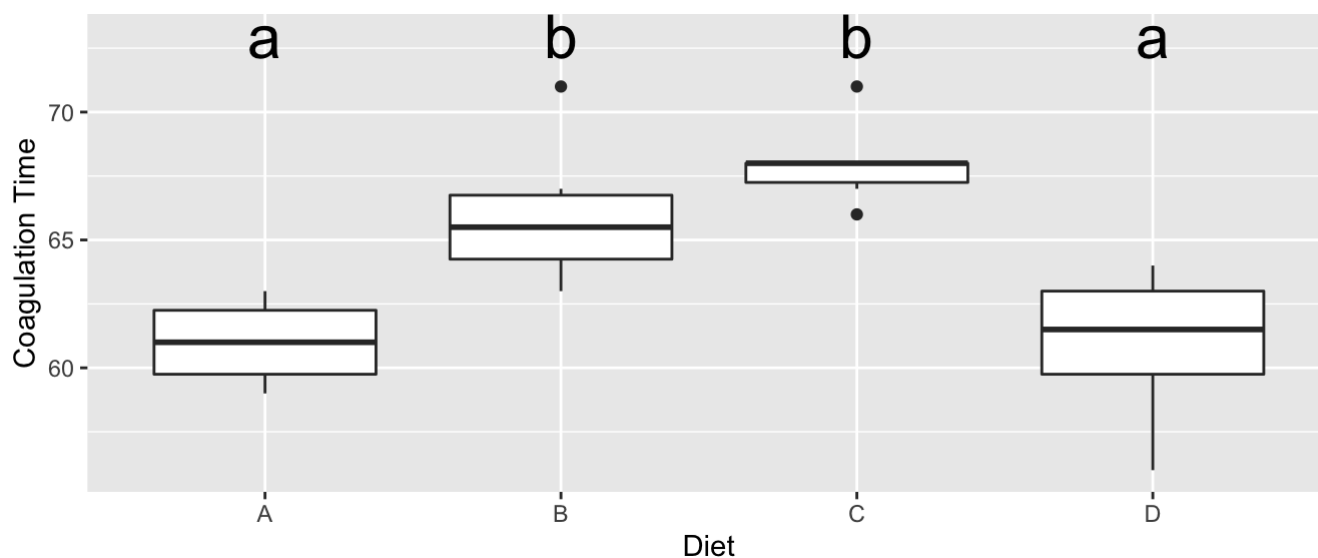
Often I want to produce the “Compact Letter Display” which identifies which groups are significantly different. Unfortunately this leads to a somewhat binary decision of “is statistically significant” or “is NOT statistically significant”, but we should at least know how to do this calculation.

```
LetterData <- emmeans(m, specs= ~ diet) %>% # cld() will freak out if you have p
  multcomp::cld(Letters=letters, level=0.95) %>%
  mutate(.group = str_remove_all(.group, '\\s')) %>% # remove the spaces
  mutate( y = 73 ) # height to place the letters at.
LetterData
```

```
## diet emmean SE df lower.CL upper.CL .group y
## A 61 1.183 20 58.5 63.5 a 73
## D 61 0.837 20 59.3 62.7 a 73
## B 66 0.966 20 64.0 68.0 b 73
## C 68 0.966 20 66.0 70.0 b 73
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

I can easily add these to my boxplot with the following:

```
ggplot(coagulation, aes(x=diet, y=coag)) +
  geom_boxplot() +
  labs( x='Diet', y='Coagulation Time' ) +
  geom_text( data=LetterData, aes(x=diet, y=y, label=.group), size=8 )
```



6.2 Two-Way ANOVA

Given a response that is predicted by two different categorical variables. Suppose we denote the levels of the first factor as α_i and has I levels. The second factor has levels β_j and has J levels. As usual we let $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$, and we wish to fit the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

which has the main effects of each covariate or possibly the model with the interaction

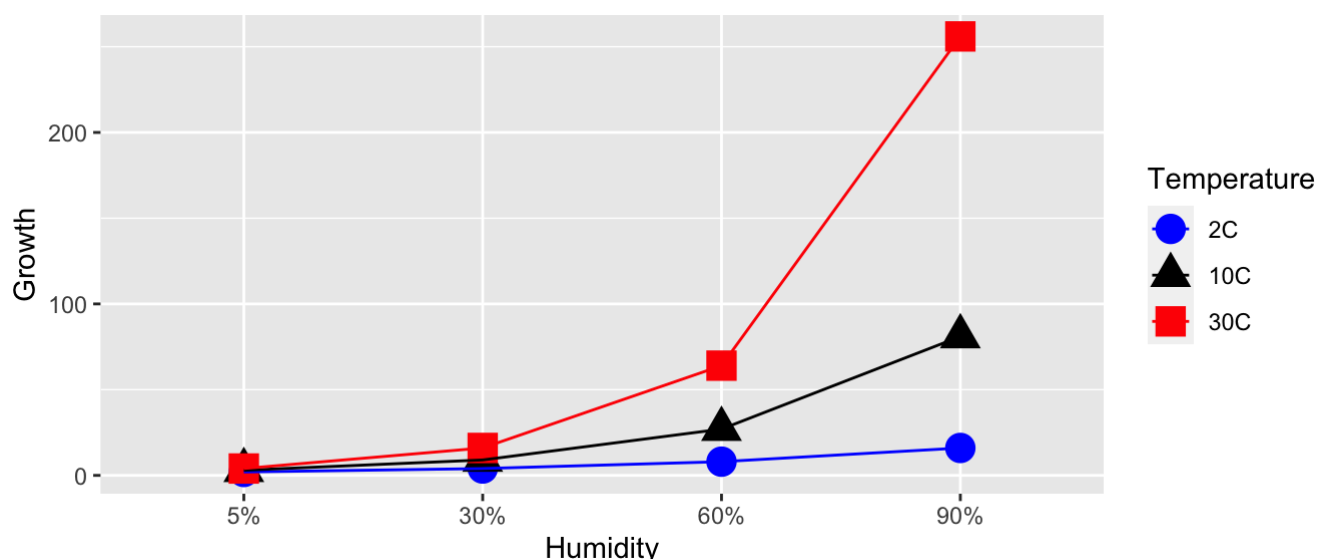
$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

To consider what an interaction term might mean consider the role of temperature and humidity on the amount of fungal growth. You might expect to see data similar to this (where the numbers represent some sort of measure of fungal growth):

		5%	30%	60%	90%
	2C	2	4	8	16
Temperature	10C	3	9	27	81
	30C	4	16	64	256

In this case we see that increased humidity increases the amount of fungal growth, but the amount of increase depends on the temperature. At 2 C, the increase is humidity increases are significant, but at 10 C the increases are larger, and at 30 C the increases are larger yet. The effect of changing from one humidity level to the next *depends on which temperature level we are at*. This change in effect of humidity is an interaction effect. A memorable example is that chocolate by itself is good. Strawberries by themselves are also good. But the combination of chocolate and strawberries is a delight greater than the sum of the individual treats.

We can look at a graph of the Humidity and Temperature vs the Response and see the effect of increasing humidity changes based on the temperature level. Just as in the ANCOVA model, the interaction manifested itself in non-parallel slopes, the interaction manifests itself in non-parallel slopes when I connect the dots across the factor levels.



Unfortunately the presence of a significant interaction term in the model makes interpretation difficult, but examining the interaction plots can be quite helpful in understanding the effects. Notice in this example, we 3 levels of temperature and 4 levels of humidity for a total of 12 different possible treatment combinations. In general I will refer to these combinations as cells.

6.3 Orthogonality

When designing an experiment, I want to make sure than none of my covariates are confounded with each other and I'd also like for them to not be correlated. Consider the following three experimental designs, where the number in each bin is the number of subjects of that type. I am interested in testing 2 different drugs and studying its effect on heart disease within the gender groups.

Design 1	Males	Females	Design 2	Males	Females
Treatment A	0	10	Treatment A	1	9
Treatment B	6	0	Treatment B	5	1
Design 3	Males	Females	Design 4	Males	Females
Treatment A	3	5	Treatment A	4	4
Treatment B	3	5	Treatment B	4	4

1. This design is very bad. Because we have no males taking drug 1, and no females taking drug 2, we can't say if any observed differences are due to the effect of drug 1 versus 2, or gender. When this situation happens, we say that the gender effect is confounded with the drug effect.
2. This design is not much better. Because we only have one observation in the Male-Drug 1 group, any inference we make about the effect of drug 1 on males is based on one observation. In general that is a bad idea.

3. Design 3 is better than the previous 2 because it evenly distributes the males and females among the two drug categories. However, it seems wasteful to have more females than males because estimating average of the male groups, I only have 6 observations while I have 10 females.
4. This is the ideal design, with equal numbers of observations in each gender-drug group.

Designs 3 and 4 are good because the correlation among my predictors is 0. In design 1, the drug covariate is perfectly correlated to the gender covariate. The correlation is less in design 2, but is zero in designs 3 and 4. We could show this by calculating the design matrix for each design and calculating the correlation coefficients between each of pairs of columns.

Having an orthogonal design with equal numbers of observations in each group has many nice ramifications. Most importantly, with an orthogonal design, the interpretation of parameter is not dependent on what other factors are in the model. Balanced designs are also usually optimal in the sense that the variances of $\hat{\beta}$ are as small as possible given the number of observations we have (barring any other *a priori* information).

6.4 Main Effects Model

In the one factor ANOVA case, the additional degrees of freedom used by adding a factor with I levels was $I - 1$. In the case that we consider two factors with the first factor having I levels and the second factor having J levels, then model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

adds $(I - 1) + (J - 1)$ parameters to the model because both $\alpha_1 = \beta_1 = 0$.

- The intercept term, μ is the reference point for all the other parameters. This is the expected value for an observation in the first level of factor 1 and the first level of factor two.
- α_i is the amount you expect the response to increase when changing from factor 1 level 1, to factor 1 level i (while the second factor is held constant).
- β_j is the amount you expect the response to increase when changing from factor 2 level 1 to factor 2 level j (while the first factor is held constant).

Referring back to the fungus example, let the α_i values be associated with changes in humidity and β_j values be associated with changes in temperature levels. Then the expected value of each treatment combination is

	5%	30%	60%	90%
2C	$\mu + 0 + 0$	$\mu + \alpha_2 + 0$	$\mu + \alpha_3 + 0$	$\mu + \alpha_4 + 0$
10C	$\mu + 0 + \beta_2$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_3 + \beta_2$	$\mu + \alpha_4 + \beta_2$
30C	$\mu + 0 + \beta_3$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_3 + \beta_3$	$\mu + \alpha_4 + \beta_3$

6.4.1 Example - Fruit Trees

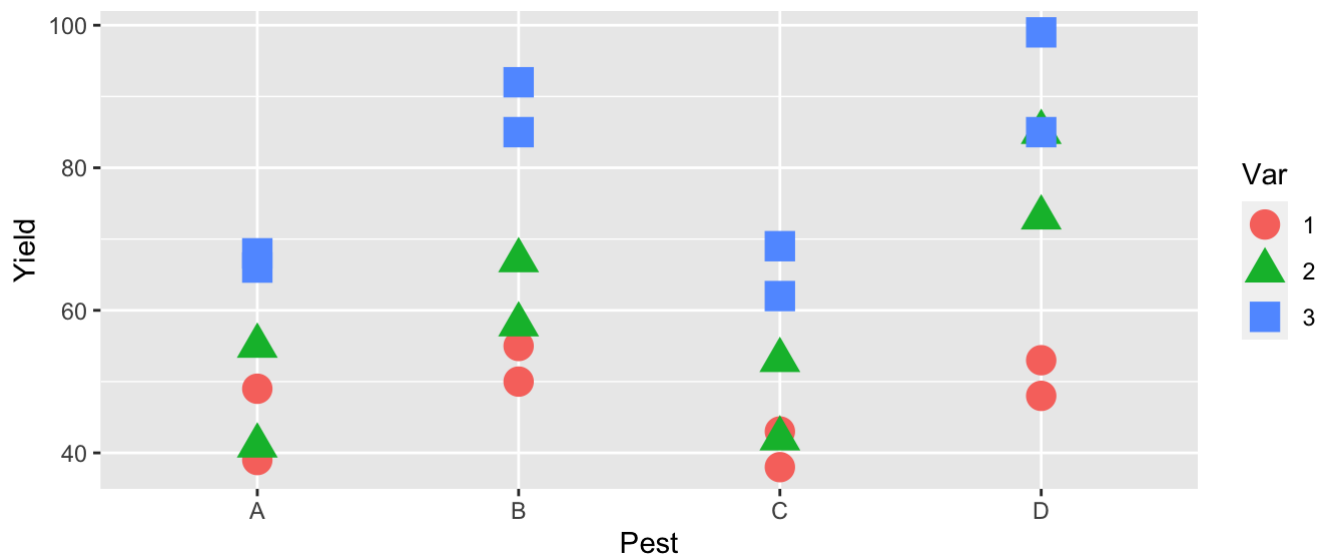
An experiment was conducted to determine the effects of four different pesticides on the yield of fruit from three different varieties of a citrus tree. Eight trees of each variety were randomly selected from an orchard. The four pesticides were randomly assigned to two trees of each variety and applications were made according to recommended levels. Yields of fruit (in bushels) were obtained after the test period.

Critically notice that we have equal number of observations for each treatment combination.

```
# Typing the data in by hand because I got this example from a really old text b
Pesticide <- factor(c('A','B','C','D'))
Variety <- factor(c('1','2','3'))
fruit <- data.frame( expand.grid(rep=1:2, Pest=Pesticide, Var=Variety) )
fruit$Yield <- c(49,39,50,55,43,38,53,48,55,41,67,58,53,42,85,73,66,68,85,92,69,6
```

The first thing to do (as always) is to look at our data

```
ggplot(fruit, aes(x=Pest, color=Var, y=Yield, shape=Var)) +
  geom_point(size=5)
```



The first thing we notice is that pesticides B and D seem to be better than the others and that variety 3 seems to be the best producer. The effect of pesticide treatment seems consistent between varieties, so we don't expect that the interaction effect will be significant.

```
m1 <- lm(Yield ~ Var, data=fruit)
m2 <- lm(Yield ~ Pest, data=fruit)
m3 <- lm(Yield ~ Var + Pest, data=fruit)
summary(m1)$coef %>% round(digits=3)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	46.875	4.359	10.754	0.000
## Var2	12.375	6.164	2.008	0.058
## Var3	31.375	6.164	5.090	0.000

```
summary(m2)$coef %>% round(digits=3)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	53.000	6.429	8.243	0.000
## PestB	14.833	9.093	1.631	0.118
## PestC	-1.833	9.093	-0.202	0.842
## PestD	20.833	9.093	2.291	0.033

```
summary(m3)$coef %>% round(digits=3)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	38.417	3.660	10.497	0.000
## Var2	12.375	3.660	3.381	0.003
## Var3	31.375	3.660	8.573	0.000
## PestB	14.833	4.226	3.510	0.003
## PestC	-1.833	4.226	-0.434	0.670
## PestD	20.833	4.226	4.930	0.000

Notice that the affects for Variety and Pesticide are the same *whether or not the other is in the model*. This is due to the orthogonal design of the experiment and makes it much easier to interpret the main effects of Variety and Pesticide.

6.4.2 ANOVA Table

Most statistical software will produce an analysis of variance table when fitting a two-way ANOVA. This table is very similar to the analysis of variance table we have seen in the one-way ANOVA, but has several rows which correspond to the additional factors added to the model.

Consider the two-way ANOVA with factors A and B which have levels I and J discrete levels respectively. For convenience let RSS_1 is the residual sum of squares of the intercept-only model, and RSS_A be the residual sum of squares for the model with just the main effect of factor A , and RSS_{A+B} be the residual sum of squares of the model with both main effects. Finally assume that we have a total of n observations. The ANOVA table for this model is as follows:

Source	df	Sum of Sq (SS)	Mean Sq
A	$df_A = I - 1$	$SS_A = RSS_1 - RSS_A$	$MS_A = SS_A/df_A$
B	$df_B = J - 1$	$SS_B = RSS_A - RSS_{A+B}$	$MS_B = SS_B/df_B$
Error	$df_e = n - I - J + 1$	RSS_{A+B}	$MSE = RSS_{A+B}/df_e$

Note, if the table is cut off, you can change decrease your font size and have it all show up...

This arrangement of the ANOVA table is referred to as “Type I” sum of squares.

We can examine this table in the fruit trees example using the `anova()` command but just passing a single model.

```

m4 <- lm(Yield ~ Var + Pest, data=fruit)
anova( m4 )

## Analysis of Variance Table
##
## Response: Yield
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Var        2 3996.1  1998.04   37.292 3.969e-07 ***
## Pest       3 2227.5   742.49   13.858 6.310e-05 ***
## Residuals 18  964.4    53.58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

We might think that this is the same as fitting three nested models and running an F-test on each successive pairs of models, but it isn't. While both will give the same Sums of Squares, the F statistics are different because the MSE of the complex model is different. In particular, the F-statistics are larger and thus the p-values are smaller for detecting significant effects.

```

m1 <- lm(Yield ~ 1, data=fruit)
m2 <- lm(Yield ~ Var, data=fruit)
m3 <- lm(Yield ~ Var + Pest, data=fruit)
anova( m1, m2 ) # Notice the F-statistic here is different than previous

```

```

## Analysis of Variance Table
##
## Model 1: Yield ~ 1
## Model 2: Yield ~ Var
##    Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      23 7188.0
## 2      21 3191.9  2    3996.1 13.146 0.0001987 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

anova( m2, m3 ) # This F-statistic matches what we saw previously

```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ Var
## Model 2: Yield ~ Var + Pest
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      21 3191.9
## 2      18  964.4  3    2227.5 13.858 6.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6.4.3 Estimating Contrasts

As in the one-way ANOVA, we are interested in which factor levels differ. For example, we might suspect that it makes sense to group pesticides B and D together and claim that they are better than the group of A and C.

Just as we did in the one-way ANOVA model, this is such a common thing to do that there is an easy way to do this, using `emmeans`.

```
m3 <- lm(Yield ~ Var + Pest, data=fruit)
emmeans(m3, spec=pairwise~Var)
```

```
## $emmeans
##   Var emmean   SE df lower.CL upper.CL
##   1     46.9 2.59 18     41.4     52.3
##   2     59.2 2.59 18     53.8     64.7
##   3     78.2 2.59 18     72.8     83.7
##
## Results are averaged over the levels of: Pest
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate   SE df t.ratio p.value
##   1 - 2         -12.4 3.66 18 -3.381  0.0089
##   1 - 3         -31.4 3.66 18 -8.573  <.0001
##   2 - 3         -19.0 3.66 18 -5.191  0.0002
##
## Results are averaged over the levels of: Pest
## P value adjustment: tukey method for comparing a family of 3 estimates

emmeans(m3, spec=pairwise~Pest)
```

```
## $emmeans
##   Pest emmean   SE df lower.CL upper.CL
##   A      53.0 2.99 18    46.7    59.3
##   B      67.8 2.99 18    61.6    74.1
##   C      51.2 2.99 18    44.9    57.4
##   D      73.8 2.99 18    67.6    80.1
##
## Results are averaged over the levels of: Var
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate   SE df t.ratio p.value
##   A - B      -14.83 4.23 18  -3.510  0.0122
##   A - C        1.83 4.23 18   0.434  0.9719
##   A - D     -20.83 4.23 18  -4.930  0.0006
##   B - C      16.67 4.23 18   3.944  0.0048
##   B - D      -6.00 4.23 18  -1.420  0.5038
##   C - D     -22.67 4.23 18  -5.364  0.0002
##
## Results are averaged over the levels of: Var
## P value adjustment: tukey method for comparing a family of 4 estimates
```

These outputs are nice and they show the main effects of variety and pesticide. Similar to the 1-way ANOVA, we also want to be able to calculate the compact letter display.

```
m3 <- lm(Yield ~ Var + Pest, data=fruit)
emmeans(m3, spec= ~Var) %>% multcomp::cld(Letters=letters)
```



```
## Var emmean SE df lower.CL upper.CL .group
## 1 46.9 2.59 18 41.4 52.3 a
## 2 59.2 2.59 18 53.8 64.7 b
## 3 78.2 2.59 18 72.8 83.7 c
##
## Results are averaged over the levels of: Pest
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
```

```
emmeans(m3, spec= ~Pest) %>% multcomp::cld(Letters=letters)
```

```
## Pest emmean SE df lower.CL upper.CL .group
## C 51.2 2.99 18 44.9 57.4 a
## A 53.0 2.99 18 46.7 59.3 a
## B 67.8 2.99 18 61.6 74.1 b
## D 73.8 2.99 18 67.6 80.1 b
##
## Results are averaged over the levels of: Var
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

So we see that each variety is significantly different from all the others and among the pesticides, A and C are indistinguishable as are B and D , but there is a difference between the A, C and B, D groups.

6.5 Interaction Model

When the model contains the interaction of the two factors, our model is written as

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Interpreting effects effects can be very tricky. Under the interaction, the effect of changing from factor 1 level 1 to factor 1 level i depends on what level of factor 2 is. In essence, we are fitting a model that allows each of the $I \times J$ cells in my model to vary independently. As

such, the model has a total of $I \times J$ parameters but because the model without interactions had $1 + (I - 1) + (J - 1)$ terms in it, the interaction is adding df_{AB} parameters. We can solve for this via:

$$\begin{aligned} I \times J &= 1 + (I - 1) + (J - 1) + df_{AB} \\ I \times J &= I + J - 1 + df_{AB} \\ IJ - I - J &= -1 + df_{AB} \\ I(J - 1) - J &= -1 + df_{AB} \\ I(J - 1) - J + 1 &= df_{AB} \\ I(J - 1) - (J - 1) &= df_{AB} \\ (I - 1)(J - 1) &= df_{AB} \end{aligned}$$

This makes sense because the first factor added $(I - 1)$ columns to the design matrix and an interaction with a continuous covariate just multiplied the columns of the factor by the single column of the continuous covariate. Creating an interaction of two factors multiplies each column of the first factor by all the columns defined by the second factor.

The expected value of the ij combination is $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$. Returning to our fungus example, the expected means for each treatment under the model with main effects and the interaction is

	5%	30%	60%	90%
2C	$\mu + 0 + 0 + 0$	$\mu + \alpha_2 + 0 + 0$	$\mu + \alpha_3 + 0 + 0$	$\mu + \alpha_4 + 0 +$
10C	$\mu + 0 + \beta_2 + 0$	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}$	$\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}$	$\mu + \alpha_4 + \beta_2 +$
30C	$\mu + 0 + \beta_3 + 0$	$\mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23}$	$\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{33}$	$\mu + \alpha_4 + \beta_2 +$

Notice that we have added $6 = 3 \cdot 2 = (4 - 1)(3 - 1) = (I - 1)(J - 1)$ interaction parameters $(\alpha\beta)_{ij}$ to the main effects only model. The interaction model has $p = 12$ parameters, one for each cell in my treatment array.

In general it is hard to interpret the meaning of α_i , β_j , and $(\alpha\beta)_{ij}$ and the best way to make sense of them is to look at the interaction plots.

6.5.1 ANOVA Table

Most statistical software will produce an analysis of variance table when fitting a two-way ANOVA. This table is very similar to the analysis of variance table we have seen in the one-way ANOVA, but has several rows which correspond to the additional factors added to the model.

Consider the two-way ANOVA with factors A and B which have levels I and J discrete levels respectively. For convenience let RSS_1 be the residual sum of squares of the intercept-only model, and RSS_A be the residual sum of squares for the model with just the main effect of factor A . Likewise RSS_{A+B} and RSS_{A*B} shall be the residual sum of squares of the model with just the main effects and the model with main effects and the interaction. Finally assume that we have a total of n observations. The ANOVA table for this model is as follows:

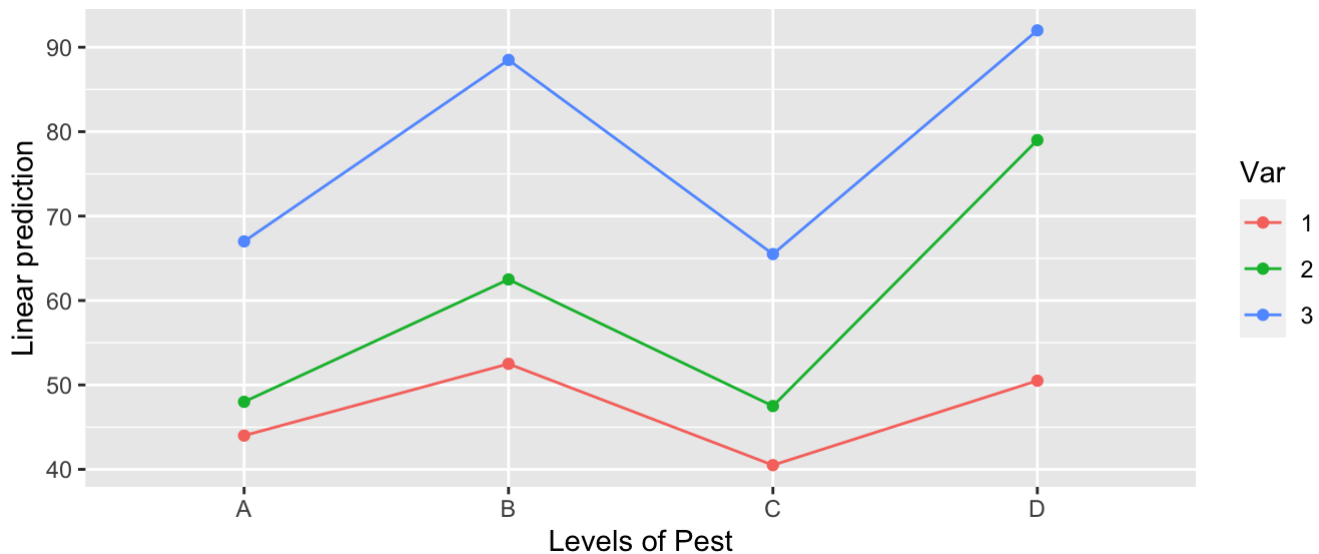
	df	Sum Sq (SS)	MS
A	$df_A = I - 1$	$SS_A = RSS_1 - RSS_A$	$MS_A = SS_A/df_A$
B	$df_B = J - 1$	$SS_B = RSS_A - RSS_{A+B}$	$MS_B = SS_B/df_B$
AB	$df_{AB} = (I - 1)(J - 1)$	$SS_{A*B} = RSS_{A+B} - RSS_{A*B}$	$MS_{AB} = SS_{AB}/df_{AB}$
Error	$df_\epsilon = n - IJ$	RSS_{A*B}	$MSE = RSS_{A*B}/df$

This arrangement of the ANOVA table is referred to as “Type I” sum of squares. Type III sums of squares are the difference between the full interaction model and the model removing each parameter group, even when it doesn’t make sense. For example in the Type III table, $SS_A = RSS_{B+A:B} - RSS_{A*B}$. There is an intermediate form of the sums of squares called Type II, that when removing a main effect also removes the higher order interaction. In the case of balanced (orthogonal) designs, there is no difference between the different types, but for non-balanced designs, the numbers will change. To access these other types of sums of squares, use the `Anova()` function in the package `car`.

6.5.2 Example - Fruit Trees (continued)

We next consider whether or not to include the interaction term to the fruit tree model. We fit the model with the interaction and then graph the results.

```
# Create the Interaction Plot using emmeans package. IP stands for interaction plot
m4 <- lm(Yield ~ Var * Pest, data=fruit)
emmip( m4, Var ~ Pest ) # color is LHS of the formula
```

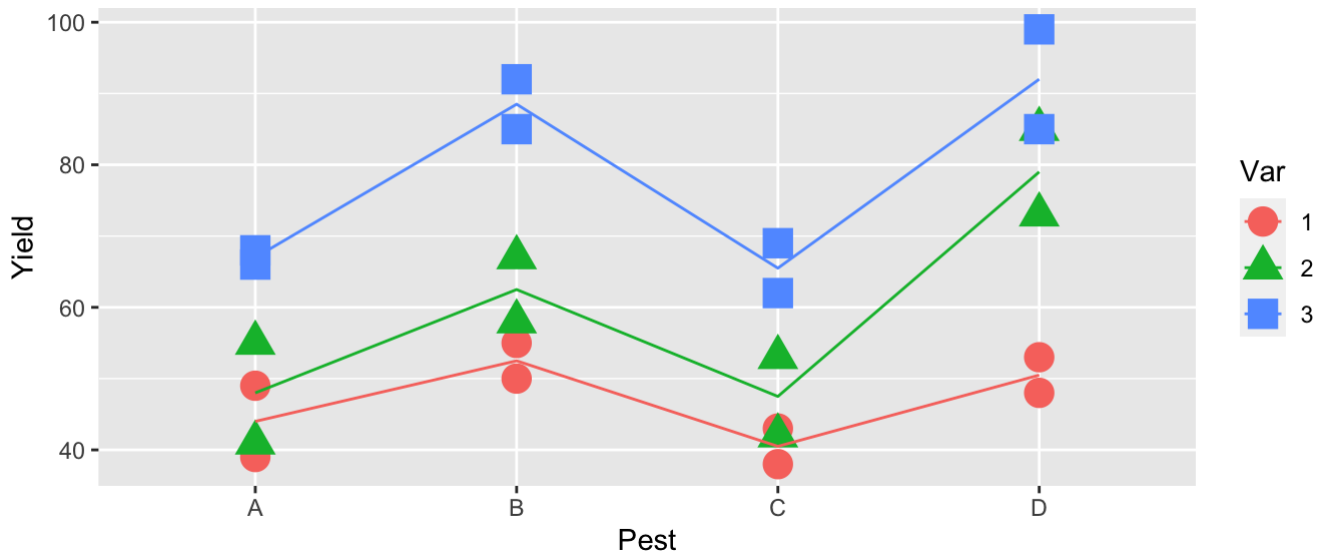


Create the interaction plot by hand

```
m4 <- lm(Yield ~ Var * Pest, data=fruit)
```

```
fruit$y.hat <- predict(m4)
```

```
ggplot(fruit, aes(x=Pest, color=Var, shape=Var, y=Yield)) +  
  geom_point(size=5) +  
  geom_line(aes(y=y.hat, x=as.integer(Pest)))
```



All of the line segments are close to parallel so, we don't expect the interaction to be significant.

```
anova( m4 )
```

```
## Analysis of Variance Table
##
## Response: Yield
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Var           2 3996.1  1998.04  47.2443 2.048e-06 ***
## Pest          3 2227.5   742.49  17.5563 0.0001098 ***
## Var:Pest       6  456.9    76.15   1.8007 0.1816844
## Residuals    12  507.5    42.29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

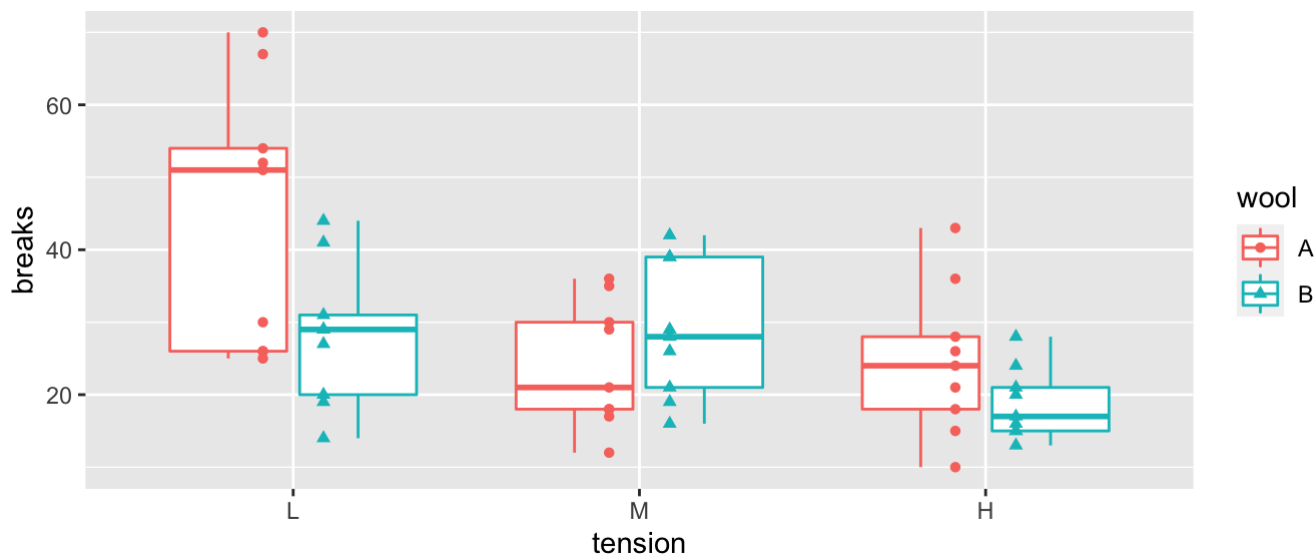
Examining the ANOVA table, we see that the interaction effect is not significant and we will stay with simpler model `Yield~Var+Pest` .

6.5.3 Example - Warpbreaks

This data set looks at the number of breaks that occur in two different types of wool under three different levels of tension (low, medium, and high). The fewer number of breaks, the better.

As always, the first thing we do is look at the data. In this case, it looks like the number of breaks decreases with increasing tension and perhaps wool B has fewer breaks than wool A.

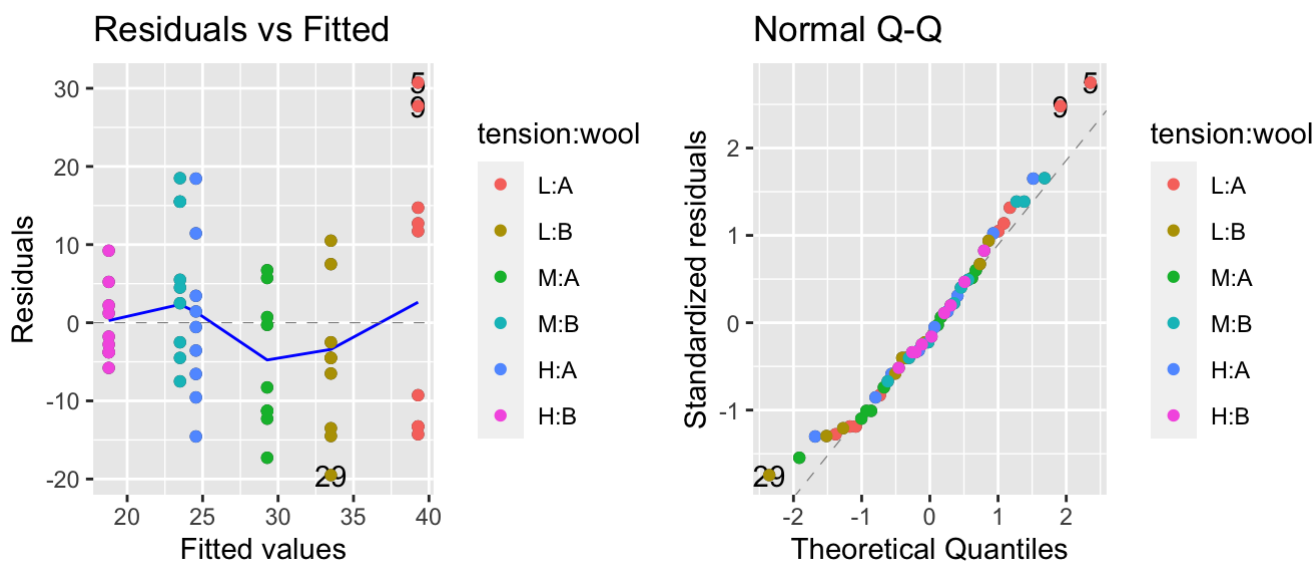
```
data(warpbreaks)
ggplot(warpbreaks, aes(x=tension, y=breaks, color=wool, shape=wool), size=2) +
  geom_boxplot() +
  geom_point(position=position_dodge(width=.35)) # offset the wool groups
```



We next fit our linear model and examine the diagnostic plots.

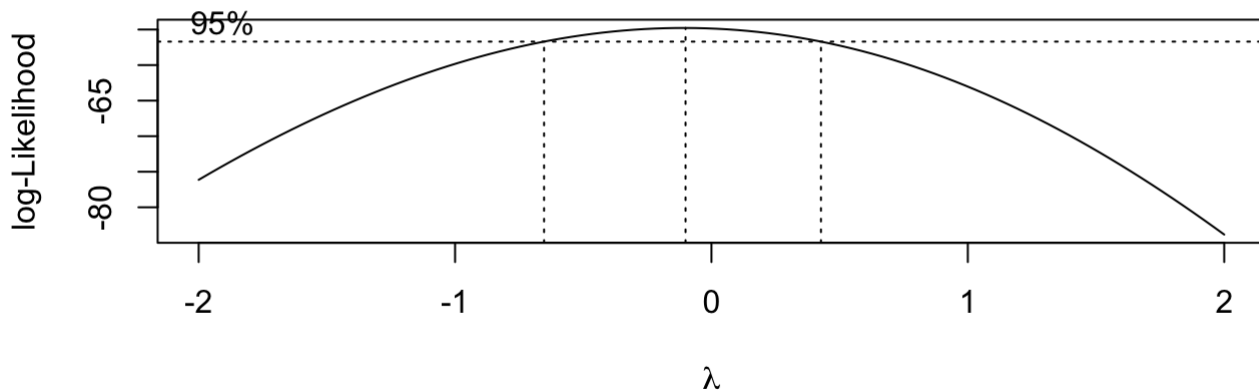
```
model <- lm(breaks ~ tension + wool, data=warpbreaks)
autoplot(model, which=c(1,2)) + geom_point(aes(color=tension:wool))

## Warning: `arrange()` is deprecated as of dplyr 0.7.0.
## Please use `arrange()` instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```



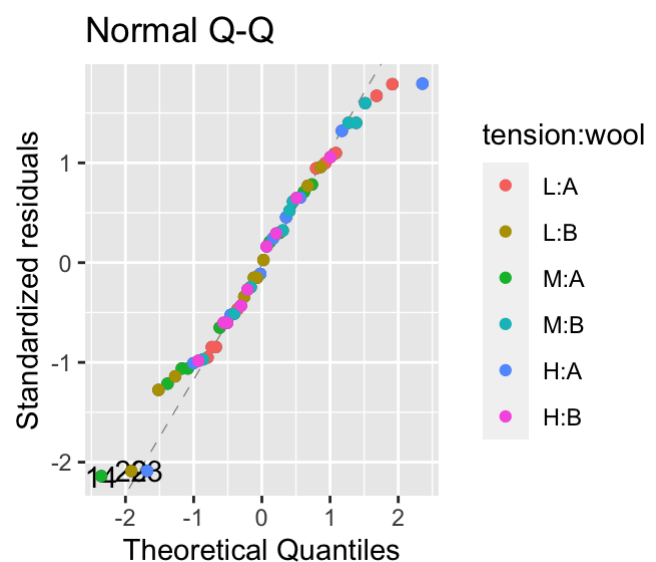
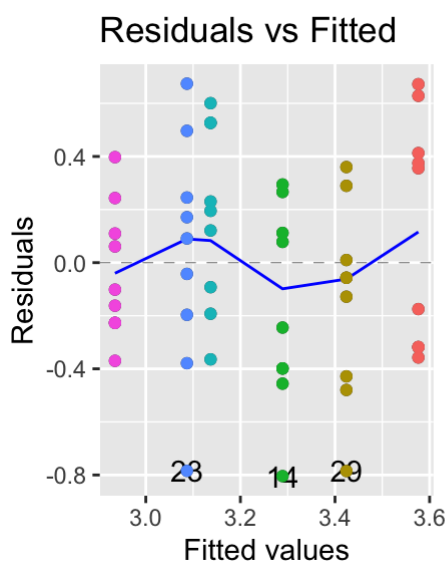
The residuals vs fitted values plot is a little worrisome and appears to be an issue with non-constant variance, but the normality assumption looks good. We'll check for a Box-Cox transformation next.

MASS::boxcox(model)



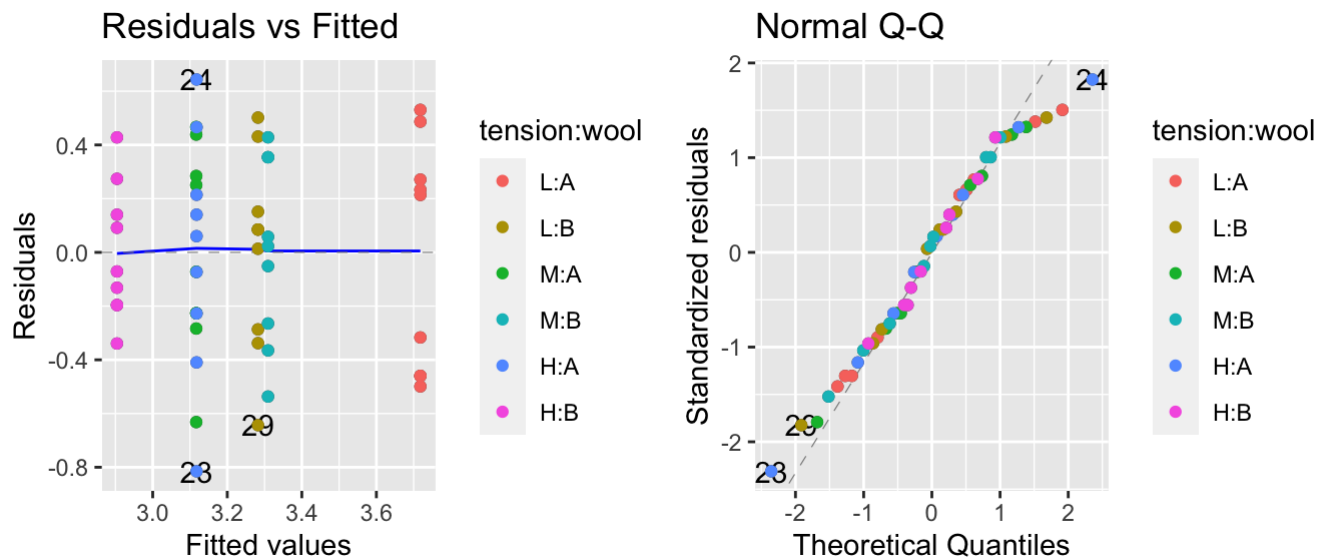
This suggests we should make a log transformation, though because the confidence interval is quite wide we might consider if the increased difficulty in interpretation makes sufficient progress towards making the data meet the model assumptions. The diagnostic plots of the resulting model look better for the constant variance assumption, but the normality is now a worse off. Because the Central Limit Theorem helps deal with the normality question, I'd rather stabilize the variance at the cost of the normality.

```
model.1 <- lm(log(breaks) ~ tension + wool, data=warpbreaks)
autoplot(model.1, which=c(1,2)) + geom_point(aes(color=tension:wool))
```



Next we'll fit the interaction model and check the diagnostic plots. The diagnostic plots look good and this appears to be a legitimate model.

```
model.2 <- lm(log(breaks) ~ tension * wool, data=warpbreaks)
autoplot(model.2, which=c(1,2)) + geom_point(aes(color=tension:wool))
```



Then we'll do an F-test to see if it is a better model than the main effects model. The p-value is marginally significant, so we'll keep the interaction in the model, but recognize that it is a weak interaction.

```
anova(model.1, model.2) # explicitly look model1 vs model2
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: log(breaks) ~ tension + wool
```

```
## Model 2: log(breaks) ~ tension * wool
```

```
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
```

```
## 1      50 7.6270
```

```
## 2      48 6.7138  2   0.91315 3.2642 0.04686 *
```

```
## ---
```

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

```
anova(model.2) # table of sequentially added terms in model 2
```

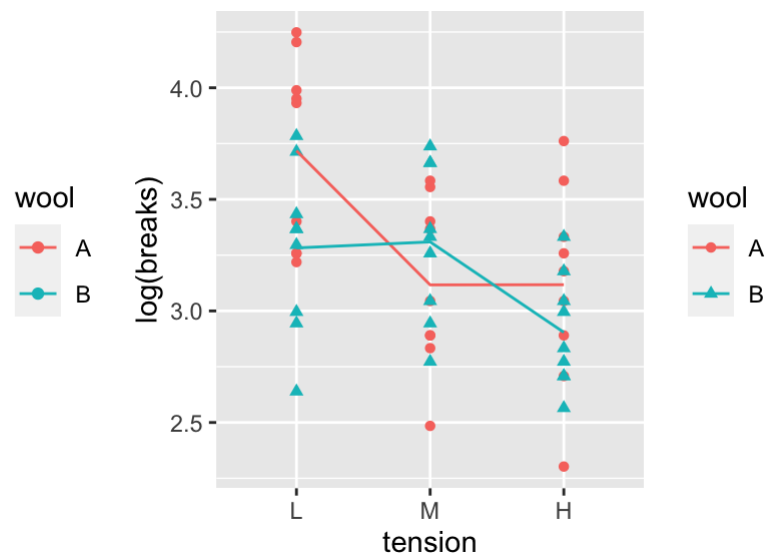
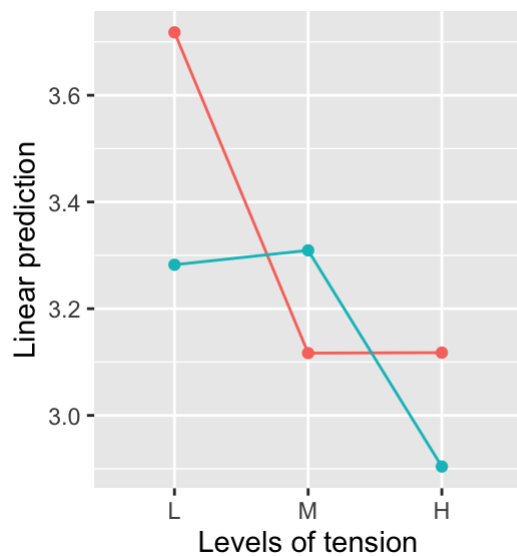


```
## Analysis of Variance Table
##
## Response: log(breaks)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## tension      2  2.1762  1.08808    7.7792 0.001185 **
## wool          1  0.3125  0.31253    2.2344 0.141511
## tension:wool  2  0.9131  0.45657    3.2642 0.046863 *
## Residuals    48  6.7138  0.13987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next we look at the effect of the interaction and the easiest way to do this is to look at the interaction plot. The `emmeans::emmip()` just shows the mean of each treatment combination, while the plot I made by hand shows the mean of each treatment combination along with the raw data.

```
A <- emmip(model.2, wool ~ tension) # LHS is color, RHS is the x-axis variable

B <- warpbreaks %>%
  mutate( logy.hat = predict(model.2) ) %>%
  ggplot(aes(x=tension, y=log(breaks), color=wool, shape=wool)) +
  geom_point() +
  geom_line(aes(y=logy.hat, x=as.integer(tension))) # make tension continuous to
cowplot::plot_grid(A,B) # Plot these version of the interaction plot side-by-side
```



We can see that it appears that wool A has a decrease in breaks between low and medium tension, while wool B has a decrease in breaks between medium and high. It is actually quite difficult to see this interaction when we examine the model coefficients.

```
summary(model.2)
```

```
##
## Call:
## lm(formula = log(breaks) ~ tension * wool, data = warpbreaks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81504 -0.27885  0.04042  0.27319  0.64358
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.7179      0.1247  29.824 < 2e-16 ***
## tensionM        -0.6012      0.1763  -3.410  0.00133 **
## tensionH        -0.6003      0.1763  -3.405  0.00134 **
## woolB           -0.4356      0.1763  -2.471  0.01709 *
## tensionM:woolB   0.6281      0.2493   2.519  0.01514 *
## tensionH:woolB   0.2221      0.2493   0.891  0.37749
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.374 on 48 degrees of freedom
## Multiple R-squared:  0.3363, Adjusted R-squared:  0.2672
## F-statistic: 4.864 on 5 and 48 DF,  p-value: 0.001116
```

To test if there is a statistically significant difference between medium and high tensions for wool type B, we really need to test the following hypothesis:

$$H_0 : (\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) - (\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}) = 0$$

$$H_a : (\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) - (\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}) \neq 0$$

This test reduces to testing if $\alpha_2 - \alpha_3 + (\alpha\beta)_{22} - (\alpha\beta)_{23} = 0$. Calculating this difference from the estimated values of the summery table we have $-.6012 + .6003 + .6281 - .2221 = 0.4051$, we don't know if that is significantly different than zero.

In the main effects model, we were able to read off the necessary test using `emmeans`. Fortunately, we can do the same thing here. In this case, we'll look at the interactions piece of the `emmeans` command. In this case, we find the test H:B - M:B in the last row of the interactions.

```

emmeans(model.2, specs= pairwise~tension*wool)

## $emmeans
##   tension wool emmean      SE df lower.CL upper.CL
##   L       A      3.72 0.125 48      3.47      3.97
##   M       A      3.12 0.125 48      2.87      3.37
##   H       A      3.12 0.125 48      2.87      3.37
##   L       B      3.28 0.125 48      3.03      3.53
##   M       B      3.31 0.125 48      3.06      3.56
##   H       B      2.90 0.125 48      2.65      3.15
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##   contrast      estimate      SE df t.ratio p.value
##   L A - M A    0.601196 0.176 48   3.410 0.0158
##   L A - H A    0.600323 0.176 48   3.405 0.0160
##   L A - L B    0.435567 0.176 48   2.471 0.1535
##   L A - M B    0.408619 0.176 48   2.318 0.2071
##   L A - H B    0.813794 0.176 48   4.616 0.0004
##   M A - H A   -0.000873 0.176 48  -0.005 1.0000
##   M A - L B   -0.165629 0.176 48  -0.939 0.9341
##   M A - M B   -0.192577 0.176 48  -1.092 0.8821
##   M A - H B    0.212598 0.176 48   1.206 0.8319
##   H A - L B   -0.164756 0.176 48  -0.935 0.9355
##   H A - M B   -0.191704 0.176 48  -1.087 0.8840
##   H A - H B    0.213471 0.176 48   1.211 0.8295
##   L B - M B   -0.026948 0.176 48  -0.153 1.0000
##   L B - H B    0.378227 0.176 48   2.145 0.2823
##   M B - H B    0.405175 0.176 48   2.298 0.2149
##
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 6 estimates

```

The last call to `emmeans` gives us all the pairwise tests comparing the cell means. If we don't want to wade through all the other pairwise contrasts we could do the following:

```

# If I want to not wade through all those contrasts and just grab the
# contrasts for wool type 'B' and tensions 'M' and 'H'
emmeans(model.2, pairwise~tension*wool, at=list(wool='B', tension=c('M','H')))

## $emmeans
##   tension wool emmean    SE df lower.CL upper.CL
##   M        B      3.31 0.125 48     3.06     3.56
##   H        B      2.90 0.125 48     2.65     3.15
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate    SE df t.ratio p.value
##   M B - H B      0.405 0.176 48  2.298   0.0260
##
## Results are given on the log (not the response) scale.

```

What would happen if we just looked at the main effects? In the case where our experiment is balanced with equal numbers of observations in each treatment cell, we can interpret these differences as follows. Knowing that each cell in our table has a different estimated mean, we could consider the average of all the type A cells as the typical wool A. Likewise we could average all the cell means for the wool B cells. Then we could look at the difference between those two averages. In the balanced design, this is equivalent to removing the tension term from the model and just looking at the difference between the average log number of breaks.

```
emmeans(model.2, specs= pairwise~tension )
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##   tension emmean      SE df lower.CL upper.CL
##   L           3.50 0.0882 48     3.32     3.68
##   M           3.21 0.0882 48     3.04     3.39
##   H           3.01 0.0882 48     2.83     3.19
##
## Results are averaged over the levels of: wool
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate      SE df t.ratio p.value
##   L - M          0.287 0.125 48  2.303   0.0649
##   L - H          0.489 0.125 48  3.925   0.0008
##   M - H          0.202 0.125 48  1.622   0.2465
##
## Results are averaged over the levels of: wool
## Results are given on the log (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates

emmeans(model.2, specs= pairwise~wool )

## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##   wool emmean    SE df lower.CL upper.CL
##   A      3.32 0.072 48     3.17     3.46
##   B      3.17 0.072 48     3.02     3.31
##
## Results are averaged over the levels of: tension
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate    SE df t.ratio p.value
##   A - B      0.152 0.102 48  1.495   0.1415
##
## Results are averaged over the levels of: tension
## Results are given on the log (not the response) scale.
```

Using `emmeans`, we can see the wool effect difference between types B and A is -0.1522 . We can calculate the mean number of log breaks for each wool type and take the difference by the following:

```
warpbreaks %>%
  group_by(wool) %>%
  summarise( wool_means = mean(log(breaks)) ) %>%
  summarise( diff(wool_means) )

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 1 x 1
##   `diff(wool_means)`
##   <dbl>
## 1      -0.152
```

In the unbalanced case taking the average of the cell means produces a different answer than taking the average of the data. The `emmeans` package chooses to take the average of the cell means.

6.6 Exercises

As we have developed much of the necessary theory, we are moving into exercises that emphasize modeling decisions and interpretation. As such, grading of exercises will move to emphasizing interpretation and justification of why modeling decisions were made. For any R output produced, be certain to discuss what is important. Feel free to not show exploratory work, but please comment that you considered. Furthermore, as students progress, necessary analysis steps will not be listed and students will be expected to appropriately perform the necessary work and comment as appropriate. These are leading to, at the end of the course, students being given a dataset and, with no prompts as to what is an appropriate analysis, told to model the data appropriately.

1. In the `faraway` package, the data set `rats` has data on a gruesome experiment that examined the time till death of 48 rats when they were subjected to three different types of poison administered in four different manners (which they called treatments). We are interested in assessing which poison works the fastest as well as which administration method is most effective.
 - a. The response variable `time` needs to be transformed. To see this, we'll examine the diagnostic plots from the interaction model and there is clearly a problem with non-constant variance. We'll look at the Box-Cox family of transformations and see that y^{-1} is a reasonable transformation.

```
data('rats', package='faraway')
model <- lm( time ~ poison * treat, data=rats)
# lindia::gg_diagnose(model) # All the plots...
lindia::gg_diagnose(model, plot.all=FALSE)[[4]] # just resid vs fitted
lindia::gg_boxcox(model)
rats <- rats %>% mutate( speed = time^(-1) )
```

- b. Fit the interaction model using the transformed response. Create a graph of data and the predicted values. Visually assess if you think the interaction is significant.
 - c. Perform an appropriate statistical test to see if the interaction is statistically significant.
 - d. What do you conclude about the poisons and treatment (application) types?
2. In the `faraway` package, the dataset `butterfat` has information about the the percent of the milk was butterfat (more is better) taken from $n = 100$ cows. There are 5 different

breeds of cows and 2 different ages. We are interested in assessing if Age and Breed affect the butterfat content

- a. Graph the data. Do you think an interaction model is justified?
- b. Perform an appropriate set of tests to select a model for predicting Butterfat .
- c. Discuss your findings.

3. In the `faraway` package, the dataset `alfalfa` has information from a study that examined the effect of seed inoculum, irrigation, and shade on alfalfa yield. This data has $n = 25$ observations.

- a. Examine the help file for this dataset. Graph the data. What effects seem significant?
- b. Consider the main effects model with all three predictor variables. Which effects are significant? *Using the model you ultimately select, examine the diagnostic plots. These all look fine, but it is useful to see examples where everything is ok.*
- c. Consider the model with `shade` and `inoculum` and the interaction between the two. Examine the anova table. Why does R complain that the fit is perfect? *Hint: Think about the degrees of freedom of the model compared to the sample size.*
- d. Discuss your findings and the limitations of your investigation based on data.