ANOVA

We think of regression as having a quantitative response and (usually) quantitative predictors, while ANOVA has a quantitative response and *qualitative* predictors. We'll see later that ANOVA is a really a special case of regression. The regression framework can handle qualitative predictors and mixes of qualitative and quantitative predictors.

We'll spend some time in the ANOVA setting, where all predictors are qualitative, before going back to the general regression setting.

ANOVA

Often ANOVA arises in designed experiments, where the the experimenter decides certain conditions to be manipulated. A lot of concepts from ANOVA historically came from agricultural experiments, where different growing conditions were randomly assigned to different plots of land to see which farming techniques affected the yield of the crop. Variables that could be manipulated might include things like watering regimes and type of fertilizer used.

In a greenhouse, experimenters can also control things like temperature and humidity. Whether or not these are considered qualitative or quantitative can depend on the design of the experiment. In many cases, experiments designed as ANOVAs just use high and low values for variables that could be treated as quantitative, such as temperature.

ANOVA

There are different types of experimental designs that can be described, depending on how individual observations are assigned to different treatments (i.e., predictors). In many studies, some aspects are observational and some experimental. For example, in a medical study, subjects might be randomly assigned either a placebo or a control, where the age of the patient is just observed.

ANOVA

The phrase **completely randomized design** is used to refer to an experiment in which only one primary factor (i.e., treatment) is analyzed, and this factor or treatment is randomly assigned to each individual.

What exactly is meant by random might not be completely clear. The experimenter can decide how many times each treatment is given. For example, if an experiment has 3 levels (treatment A, treatment B, and placebo) for blood pressure, and there will be 30 subjects, then the experimenter can randomly choose 10 patients to receive treatment A, then randomly pick 10 patients from the remaining 20 to receive treatment B, then give the placebo to the remaining 10 subjects.

ANOVA

This randomization is more like distributing cards from a shuffled deck (where you sample without replacement) than rolling a die (sampling with replacement). This randomization could be accomplished in R as follows, assuming that patients are numbered with IDs 1 through 30:

```
> patient <- 1:30
> treatment <- c(rep("a",10),rep("b",10),rep("placebo",10))
> treatment <- sample(treatment,replace=F)
> mydata <- data.frame(cbind(patient,treatment))
> head(mydata)
  patient treatment
1       1         b
2       2   placebo
3       3   placebo
4       4         a
5       5   placebo
6       6         a
```

ANOVA

In this example, because the total sample size is a multiple of the number of groups (30 is a multiple of 3), you can have equal sample sizes in each group. This is called **balanced ANOVA** or a **balanced design**. This is usually preferable in terms of statistical power. In other words, if there are different means for the groups, then you have a higher probability of detecting those differences using equal sample sizes than using unequal sample sizes if other assumptions are met (such as equal variances).

This is one reason for using the sampling without replacement approach. If each of the 30 subjects was assigned each treatment independently (using `replace=T`), then it would be unlikely to get exactly 10 individuals assigned to each of the three possible treatments.

ANOVA

Although this example is considered an experiment (because of random assignment), differences in blood pressure could be due to a number of unmeasured variables such as age, sex/gender, initial blood pressure, genetic influences to responses to the drugs, differences in lifestyle etc. It is usually hoped that by doing random assignment, the different groups will be similar in terms of the distribution of age, sex, genetics, etc. for the different groups. This is more likely for large samples than for small samples. Alternately, the experimenters could try to make the population sampled from more uniform by only recruiting people from one sex, one age range, etc.

ANOVA

So far, we have described a **one-factor ANOVA**, which was also analyzed last semester. ANOVA can still be an appropriate analysis tool even if the factor isn't controlled via randomization, but as we get into more complex designs and more complicated variables, it will be useful to introduce concepts of design. In addition, we introduce some new notation to help generalize to more complex designs.

Thinking of the ANOVA model as a response and predictors, we can write the model as

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

Here y_{ij} refers to the jth individual in the ith treatment group. For the blood pressure example, we would have $j = 1, \ldots, 10$ and $i = 1, 23$. The quantity $y_{2,5}$ for example, would mean the 5th indvidual receiving treatment B, and $y_{3,2}$ would be the second individual receiving the placebo.

ANOVA

Based on the model, we assume that there is a mean value associated with each treamtent, μ_1, μ_2 and μ_3. An individual in group 2 has an expected blood pressure of μ_2, and $\varepsilon_{2,j}$ represents the deviation of the jth individual in group 2 from μ_2. As in regression, the values ε_{ij} is a residual, meaning the difference between the observed and expected values for the individual.

Although I have said that ANOVA is a special case of regression, the notation used here is a bit different from simple linear regression, where there is only one subscript for the response and one subscript for the residuals and predictor values.

ANOVA

Note that the null hypothesis for ANOVA is that

$$\mu_1 = \mu_2 = \cdots = \mu_I$$

were I is the number of groups. (For the blood pressure example, $I = 3$.)

If the null hypothesis is true, we can let $\mu = \mu_1 = \mu_2 = \cdots = \mu_I$, so that

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

ANOVA

Another way of thinking about the model is that there is an overall mean, μ, and each treatment might have a different effect:

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

Here $\mu_i = \mu + \alpha_i$, were the null hypothesis can be written as $\alpha_i = 0$ for each $i = 1, \ldots, I$.
We can also think of μ as the **Grand mean**

$$\mu = \frac{1}{I} \sum_{i=1}^{I} \mu_i$$

and $\alpha_i = \mu - \mu_i$ as the **treatment effect**.

ANOVA

Typically, computer programs will estimate μ, the grand mean, which serves as an intercept, plus $\alpha_i, \ldots, \alpha_{I-1}$, setting $\alpha_I = 0$. The reason for these is that you can't separately estimate μ and all of the α_i terms because there isn't a unique solution. If there are I groups, then only I means can be estimated.

Typically, the first or last group is used as the default. For example if placebo is the default, you might be looking at the effect of treatment A or treatment B in comparison to the placebo used as a baseline.

ANOVA: randomized block design

The completely randomized design assigned treatments randomly to each subject, treating the pool of subjects as coming from a single population or group. This works best when the subjects come from a fairly homogeneous pool.

If the subjects come from different groups but are fairly homogeneous within these groups, then it might make sense to use a randomized block design, where you estimate the effect of being in different groups. Blocks for medical patients could be based on say, sex, age category, or whether or not the person smokes. Subejcts within each block are then randomly assigned to the possible treatments.

ANOVA: randomized block design

In a randomized block design, you estimate effects contributed by each block as well as the treatment effects.

In agriculture, randomized block designs came about because there might be differences in soil fertility in different plots of land. Researchers wanted the effect of fertilizer (for example) to be estimated but needed to account for the fact that some plots of land might have had different types of soil, so that differences in crop yield depended on both the treatment (type of fertilizer) and block (type of soil). The desire is to account for the effect of the soil wen estimating the effect of the fertilizer.

Another example of a block effect would be different varieties of a species (such as different strains of corn). Individuals can also be considered as blocks when the same individual is exposed to different treatments.

ANOVA: randomized block design

As an example where blocking is done on individuals is a study from Beecher (1959) on treatments for itching. There were 10 patient volunteers, all male and between 20 and 30 years old. There were seven treatments — 5 drugs, a placebo, and no drug — to relieve itching. Each subject was given a different treatment on seven study days. The time ordering of the treatments was randomized across days.

Randomizing the time ordering is not part of the statistical analysis but is scientifically a good idea — this helps reduce any accident effect due to time ordering. Without ordering, it could be the patients become more or less sensitive to itching over time regardless of treatment.

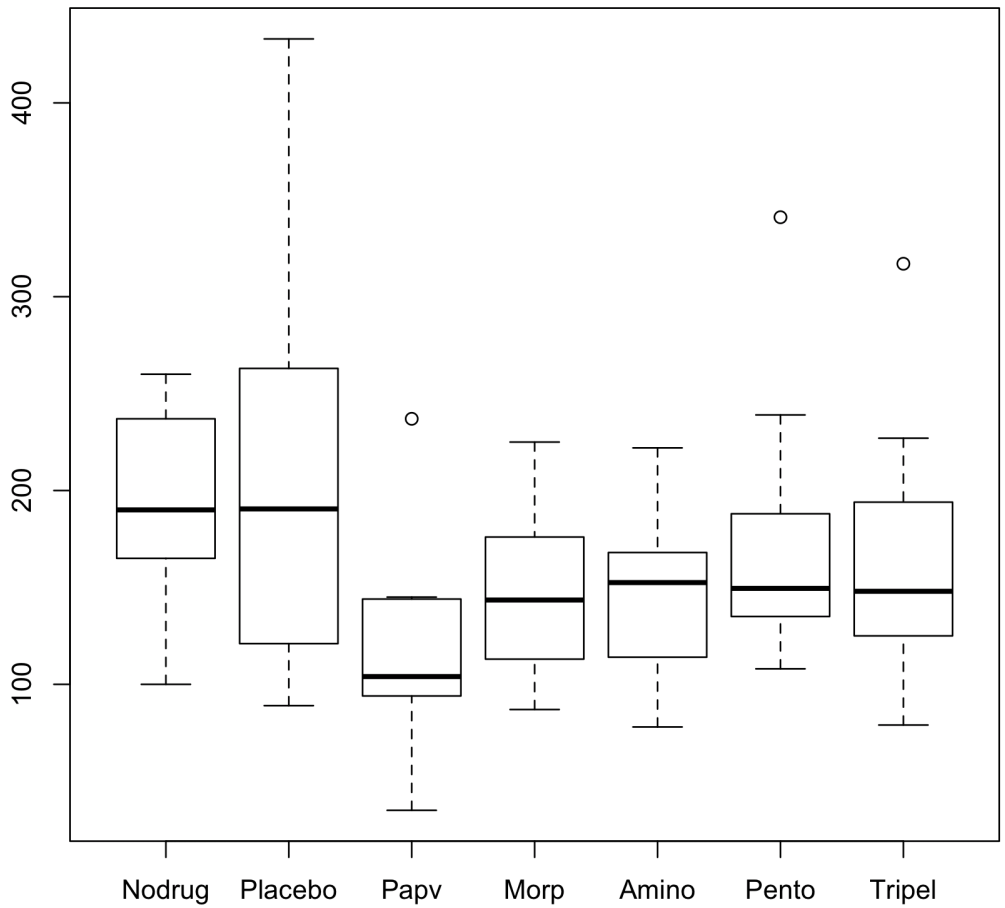
Except on the no-drug day, the subjects were given the treatment intravenously, and then itching was induced. on their forearms using an effective itch stimulus called cowage. The subjects recorded the duration of itching, in seconds. The data are given in the table below. From left to right the drugs are: papaverine, morphine, aminophylline, pentobarbitol, tripelenamine.

ANOVA: randomized block design

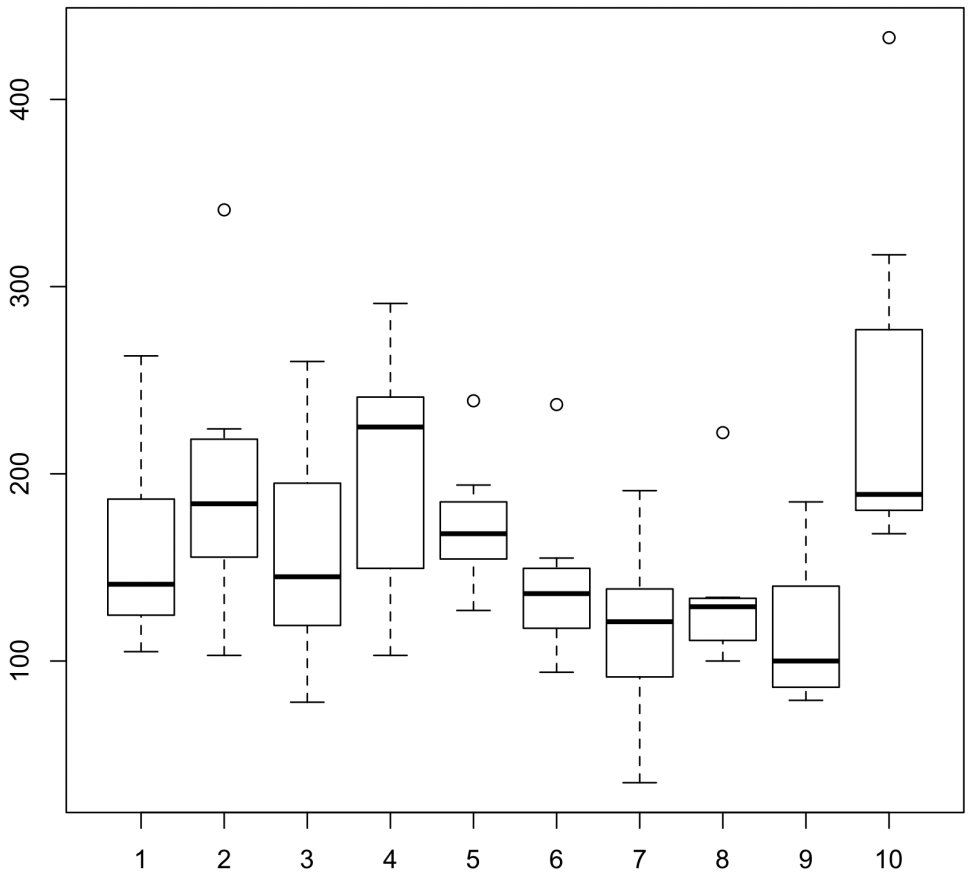
Here the subjejcts are treated as blocks because some subjects might have different mean levels of itchiness than others, and the effect of the treatments should have these differences accounted for.

Patient	Nodrug	Placebo	Papv	Morp	Amino	Pento	Tripel
1	174	263	105	199	141	108	141
2	224	213	103	143	168	341	184
3	260	231	145	113	78	159	125
4	255	291	103	225	164	135	227
5	165	168	144	176	127	239	194
6	237	121	94	144	114	136	155
7	191	137	35	87	96	140	121
8	100	102	133	120	222	134	129
9	115	89	83	100	165	185	79
10	189	433	237	173	168	188	317

ANOVA: itching data example



ANOVA: itching data example



ANOVA: randomized block design

Now to write the model, y_{ij} again represents the jth treatment for the ith block. The model is

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

Here each individual has their own mean. This might sound impossible to estimate because we only have one observation for each combination of block and treatment. However, we can also think of the model this way where $\mu_{ij} = \mu + \alpha_i + \beta_j$:

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

That is, μ is the grand mean, α_i is the effect of block i (i.e., subject i), and β_j is the effect of treatment j. Less formally

$$\text{Response} = \text{Grand mean} + \text{Treatment effect} + \text{Block effect}$$

ANOVA: randomized block design

The ANOVA table can be written as follows, where $\overline{y}_{..}$ is the mean of all observations, $\overline{y}_{i.}$ is the mean of

Source	*df*	*SS*	*MS*
Blocks	$I-1$	$J\sum_i(\overline{y}_{i.}-\overline{y}_{..})^2$	
Treatments	$J-1$	$I\sum_i(\overline{y}_{.j}-\overline{y}_{..})^2$	
Error	$(I-1)(J-1)$	$\sum_{ij}(y_{ij}-\overline{y}_{i.}-\overline{y}_{.j}+\overline{y}_{..})^2$	
Total	$IJ-1$	$\sum_{ij}(y_{ij}-\overline{y}_{..})^2$	

The *MS* (Mean square) column is filled in using SS/df for the same row. For the itching data set, there are $I = 10$ blocks and $J = 7$ treatments, so the total degrees of freedom is $70 - 1 = 69$.

ANOVA: randomized block design

Usually you are more interested in testing whether the treatment effects are 0 rather than whether the blocking effects are 0. In other words the hypothesis test of greatest interest is

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

However, mathematically and in the software, there are really just two types of effects (blocks and treatments), but the computer doesn't care which is which.

ANOVA: randomized block design

The formal hypothesis test is based on an F test using

$$F_{obs} = \frac{MS \text{ Treat}}{MS \text{ Error}}$$

Using $J-1$ numerator degrees of freedom and $(I-1)(J-1)$ denominator degrees of freedom. (Recall that for the F test, there are numerator and denominator degrees of freedom, so F distributions are indexed by two kinds of degrees of freedom).

Usually, the randomized block design is used when blocks are very different but observations within blocks would be very similar if the null hypothesis of no treatment effect is true. However, you could test

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

using an F test based on

$$F_{obs} = \frac{MS \text{ Blocks}}{MS \text{ Error}}$$

ANOVA: randomized block design

The estimates for μ, α_i and β_j are

$$\begin{aligned}\widehat{\mu} &= \overline{y}_{..} \\ \widehat{\alpha}_i &= \overline{y}_{i.} - \overline{y}_{..} \\ \widehat{\beta}_j &= \overline{y}_{.j} - \overline{y}_{..} \\ \widehat{\mu}_{ij} &= \widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j\end{aligned}$$

In other words, the estimated treatment effect (for a particular treatment) is the average response for that treatment minus the overall mean, and the estimated block effect (for a particular block) is the mean response in that block (i.e., for that patient) minus the overall mean.

ANOVA: randomized block design

The model can be fitted in R.

```
> x <- read.csv("itch.csv")
> head(x)
  Patient Nodrug Placebo Papv Morp Amino Pento Tripel
1       1    174     263  105  199   141   108    141
2       2    224     213  103  143   168   341    184
3       3    260     231  145  113    78   159    125
4       4    255     291  103  225   164   135    227
5       5    165     168  144  176   127   239    194
6       6    237     121   94  144   114   136    155
```

ANOVA: randomized block design

To analyze in R, the data should be in narrow format, with one column for the patient, one for the treatment, and one for the response.

```
> install.packages("reshape2")
> library(reshape2)
Warning message:
package ‘reshape2’ was built under R version 3.4.3
> R.Version()$version.string
[1] "R version 3.4.2 (2017-09-28)"
```

ANOVA: randomized block design

```
> itch.long &  <- melt(x
+                    , id.vars       = "Patient"
+                    , variable.name = "Treatment"
+                    , value.name    = "Seconds"
+ )
> head(itch.long)
  Patient Treatment Seconds
1       1    Nodrug     174
2       2    Nodrug     224
3       3    Nodrug     260
4       4    Nodrug     255
5       5    Nodrug     165
6       6    Nodrug     237
```

ANOVA: randomized block design

It is important to make the Patient ID a factor variable. Otherwise the patient ID is treated as quantitative!!

```
> itch.long$Patient <- factor(itch.long$Patient)
> attach(itch.long)
> model1 <- lm(Seconds ~ Patient + Treatment)
> library(car)
> Anova(model1,type=3)
Anova Table (Type III tests)

Response: Seconds
             Sum Sq Df F value    Pr(>F)
(Intercept) 155100  1 50.1133 3.065e-09 ***
Treatment    53013  6  2.8548  0.017303 *
Patient     103280  9  3.7078  0.001124 **
Residuals   167130 54
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

ANOVA: randomized block design

Based on the output, both the treatment and the patient are significant. To me, this suggests that it was important to take into account differences between patients. If we fit the model as a one-factor ANOVA (ignoring the effect of the individual patients), the evidence appears not as strong against the null hypothesis.

```
> model2 <- lm(Seconds ~ Treatment)
> Anova(model2,type=3)
Anova Table (Type III tests)

Response: Seconds
             Sum Sq Df F value    Pr(>F)
(Intercept) 364810  1 84.9935 2.709e-13 ***
Treatment    53013  6  2.0585   0.07082 .
Residuals   270409 63
---

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

ANOVA: randomized block design

Note that there are slight differences between type I, II, and III sums of squares. Type III sums of squares include an intercept term and is based on testing predictors in the context of other predictors. Type III is the default in SAS, while type II is the default in R. The two give equivalent p-values when the design is balanced (equal sample sizes in each combination of predictors).

```
> summary(model1)
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       188.286     26.598   7.079 3.07e-09 ***
TreatmentPlacebo   13.800     24.880   0.555  0.58141
TreatmentPapv     -72.800     24.880  -2.926  0.00501 **
TreatmentMorp     -43.000     24.880  -1.728  0.08965 .
TreatmentAmino    -46.700     24.880  -1.877  0.06592 .
TreatmentPento    -14.500     24.880  -0.583  0.56245
TreatmentTripel   -23.800     24.880  -0.957  0.34303
Patient2           35.000     29.737   1.177  0.24436
Patient3           -2.857     29.737  -0.096  0.92381
Patient4           38.429     29.737   1.292  0.20176
Patient5           11.714     29.737   0.394  0.69518
Patient6          -18.571     29.737  -0.625  0.53491
Patient7          -46.286     29.737  -1.557  0.12543
Patient8          -27.286     29.737  -0.918  0.36292
Patient9          -45.000     29.737  -1.513  0.13604
Patient10          82.000     29.737   2.758  0.00793 **
Multiple R-squared:  0.4832,Adjusted R-squared:  0.3397
F-statistic: 3.367 on 15 and 54 DF,  p-value: 0.00052
```

To interpret the output, the no drug treatment is a baseline, and Patient 1 is a baseline. So the estimate (fitted value) for patient 1 is the intercept, 188.286 seconds. Patient 1 actually recorded 174 seconds, so why isn't that the estimate? You can think of this.as meaning that if the experiement were performed again, you might expect patient 1 to itch for 174 seconds. The idea is that you are using information about that individual on the different treatments and other subjects on the no-drug treatment to have additional information to predict what patient 1 might experience on a new exposure to the treatment.

Based on the output, the fitted value for patient 10 on morphine is

$$188.286 - 43.000 + 82.000 = 227.286$$

seconds. Something to notice in the output is that every treatment other than placebo tended to reduce the itchiness, with papeverine as having te greatest expected reduction and also the strongest p-value.

In this sort of example, predicting the reduction in seconds is probably not as interesting as learning whether the treatments were different from each other, and which treatments were most effective.

From the linear model output, we also get an F test with a p-value, which is a p-value for testing whether both variables together (blocks and treatments) are significantly different from 0. This is usually not as interesting as testing whether just treatments are different from each other taking blocks into account.

In order to test which treatments are significantly different from no treatment, we should take into account that we are doing multiple comparisons. The package `multcomp` can be used to help do multiple comparisons.

```
> install.packages("multcomp")
> library(multcomp)
> comp.itch <- glht(aov(model1),linfct =
  mcp(Treatment = "Tukey"))
> summary(comp.itch)
```

	Estimate	Std. Error	t value	Pr(>|t|)	
Placebo - Nodrug == 0	13.80	24.88	0.555	0.9978	
Papv - Nodrug == 0	-72.80	24.88	-2.926	0.0697	.
Morp - Nodrug == 0	-43.00	24.88	-1.728	0.6005	
Amino - Nodrug == 0	-46.70	24.88	-1.877	0.5039	
Pento - Nodrug == 0	-14.50	24.88	-0.583	0.9971	
Tripel - Nodrug == 0	-23.80	24.88	-0.957	0.9610	
Papv - Placebo == 0	-86.60	24.88	-3.481	0.0165	*
Morp - Placebo == 0	-56.80	24.88	-2.283	0.2712	
Amino - Placebo == 0	-60.50	24.88	-2.432	0.2052	
Pento - Placebo == 0	-28.30	24.88	-1.137	0.9135	
Tripel - Placebo == 0	-37.60	24.88	-1.511	0.7370	
Morp - Papv == 0	29.80	24.88	1.198	0.8920	
Amino - Papv == 0	26.10	24.88	1.049	0.9398	
Pento - Papv == 0	58.30	24.88	2.343	0.2434	
Tripel - Papv == 0	49.00	24.88	1.969	0.4454	
Amino - Morp == 0	-3.70	24.88	-0.149	1.0000	
Pento - Morp == 0	28.50	24.88	1.146	0.9107	
Tripel - Morp == 0	19.20	24.88	0.772	0.9867	
Pento - Amino == 0	32.20	24.88	1.294	0.8516	
Tripel - Amino == 0	23.90	24.88	0.960	0.9676	

Based on the output, the only comparison that is statistically significant at the .05 level is papaverine versus placebo, and the second lowest adjusted p-value is for papaverine versus no drug. This suggests that there is some (but not overwhelming) evidence that this drug reduced itchiness.

Basically what the package is doing is similar to a t-test between the columns in the original data (not the long data set), but using the standard error that was obtained from the ANOVA instead of just the two columns in the data

```
> head(x)
  Patient Nodrug Placebo Papv Morp Amino Pento Tripel
1       1    174     263  105  199   141   108    141
2       2    224     213  103  143   168   341    184
3       3    260     231  145  113    78   159    125
4       4    255     291  103  225   164   135    227
5       5    165     168  144  176   127   239    194
6       6    237     121   94  144   114   136    155
> diff <- x$Placebo - x$Nodrug
> diff
 [1]   89  -11  -29   36    3 -116  -54    2  -26  244
> mean(diff)
[1] 13.8
> mean(diff)/24.88
[1] 0.5546624
```

The p-value for the Tukey multiple comparisons is based on the Tukey range distribution, which is similar to a t-test but results in different p-values.

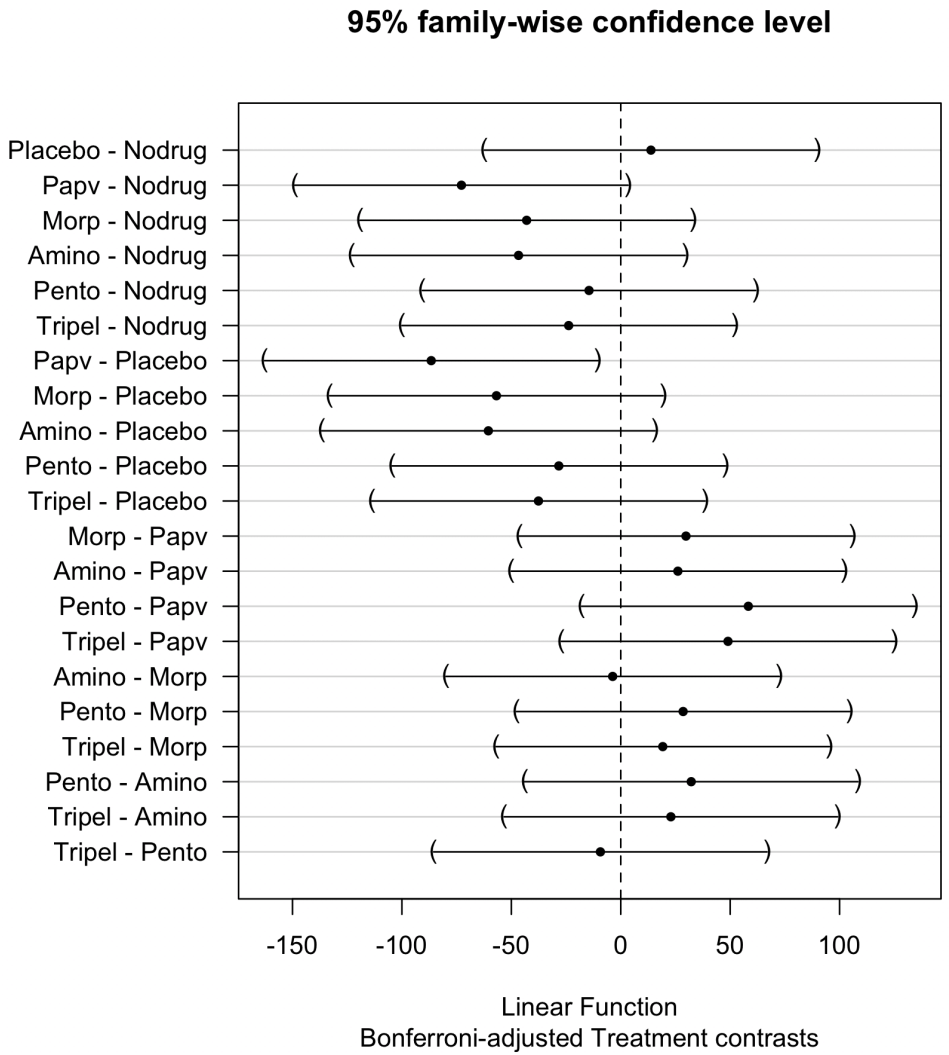
You could also do a Bonferroni correction instead.

```
summary(comp.itch, test = adjusted("bonferroni"))
```

	Estimate	Std. Error	t value	Pr(>|t|)	
Placebo - Nodrug == 0	13.80	24.88	0.555	1.000	
Papv - Nodrug == 0	-72.80	24.88	-2.926	0.105	
Morp - Nodrug == 0	-43.00	24.88	-1.728	1.000	
Amino - Nodrug == 0	-46.70	24.88	-1.877	1.000	
Pento - Nodrug == 0	-14.50	24.88	-0.583	1.000	
Tripel - Nodrug == 0	-23.80	24.88	-0.957	1.000	
Papv - Placebo == 0	-86.60	24.88	-3.481	0.021	*
Morp - Placebo == 0	-56.80	24.88	-2.283	0.554	
Amino - Placebo == 0	-60.50	24.88	-2.432	0.386	
Pento - Placebo == 0	-28.30	24.88	-1.137	1.000	
Tripel - Placebo == 0	-37.60	24.88	-1.511	1.000	
Morp - Papv == 0	29.80	24.88	1.198	1.000	
Amino - Papv == 0	26.10	24.88	1.049	1.000	
Pento - Papv == 0	58.30	24.88	2.343	0.479	
Tripel - Papv == 0	49.00	24.88	1.969	1.000	
Amino - Morp == 0	-3.70	24.88	-0.149	1.000	
Pento - Morp == 0	28.50	24.88	1.146	1.000	
Tripel - Morp == 0	19.20	24.88	0.772	1.000	
Pento - Amino == 0	32.20	24.88	1.294	1.000	
Tripel - Amino == 0	22.90	24.88	0.920	1.000	

You can also plot confidence intervals for the differences between treatments as follows.

```
# plot the summary
op <- par(no.readonly = TRUE) # the whole list of settable par's.
# make wider left margin to fit contrast labels
par(mar = c(5, 10, 4, 2) + 0.1) # order is c(bottom, left, top, r
# plot bonferroni-corrected difference intervals
plot(summary(comp.itch, test = adjusted("bonferroni"))
    , sub="Bonferroni-adjusted Treatment contrasts")
par(op) # reset plotting options
```


95% family-wise confidence level
Placebo - Nodrug
Papv - Nodrug
Morp - Nodrug
Amino - Nodrug
Pento - Nodrug
Tripel - Nodrug
Papv - Placebo
Morp - Placebo
Amino - Placebo
Pento - Placebo
Tripel - Placebo
Morp - Papv
Amino - Papv
Pento - Papv
Tripel - Papv
Amino - Morp
Pento - Morp
Tripel - Morp
Pento - Amino
Tripel - Amino
Tripel - Pento
-150
-100
-50
0
50
100
Linear Function
Bonferroni-adjusted Treatment contrasts

ANOVA: diagnostics

Part of an ANOVA or regression should ideally be diagnostic tests (although these are often not mentioned in scientific studies).

The assumptions of ANOVA needed to make p-values correct include that the response is normally distributed with the same variance and the same mean *for each combination of the predictors*.

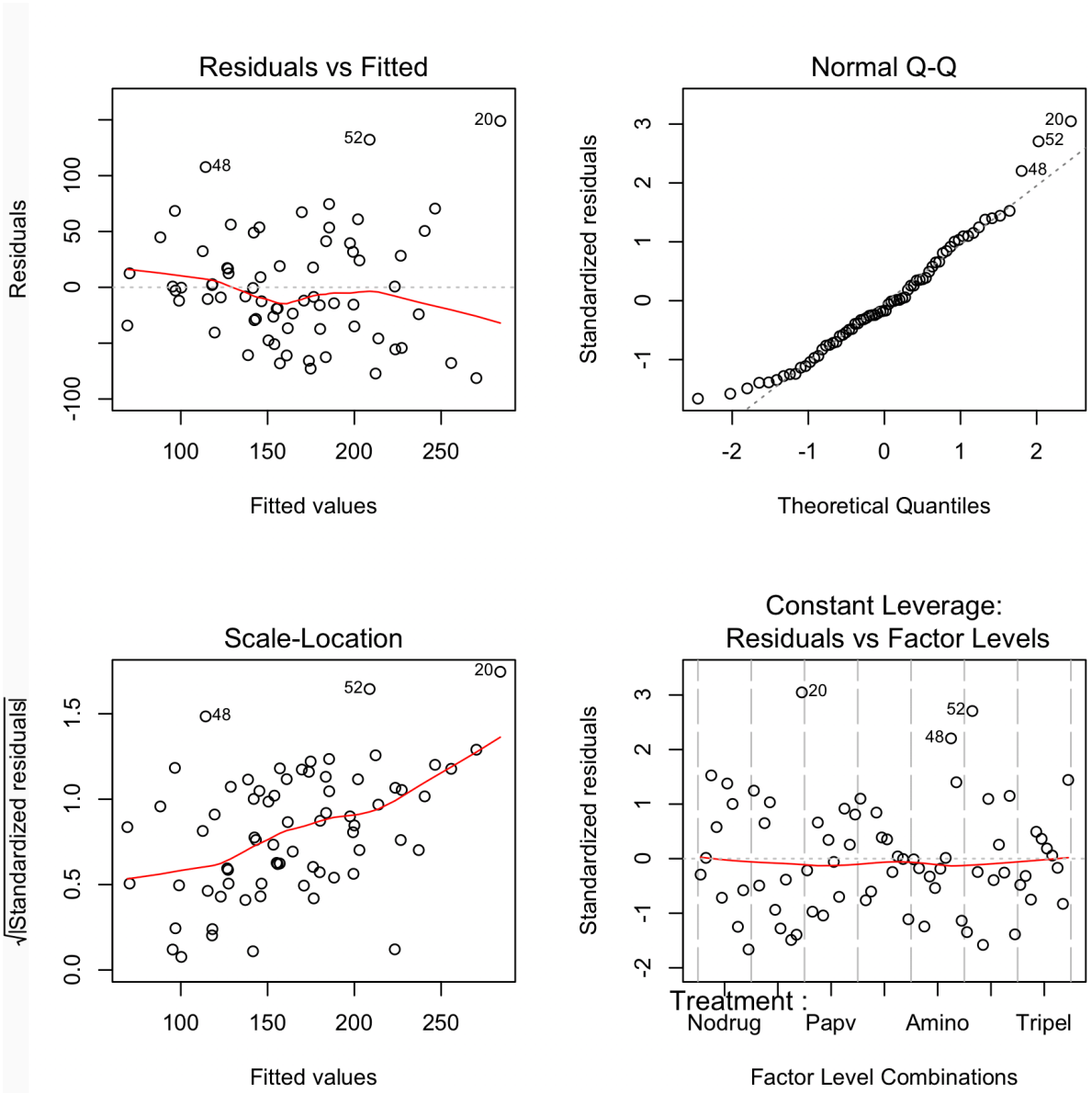
This also means that the residuals should be normally distributed with mean 0 and a common variance.

ANOVA: diagnostics

Typically, diagnostics are done visually and nor very formally, especially by examining residuals. Note that in the itchiness study, there is only one observation for each combination of predictors, so the normality would be impossible to assess looking at each combination of predictors separately. However, the residuals should all come from the same distribution, so there is still information in the residuals regarding the normality and constant variance assumptions.

If the `plot()` function is given saved model output, it will automatically generate diagnostic plots. For example

```
> par(mfrow=c(2,2))
> plot(model1)
```

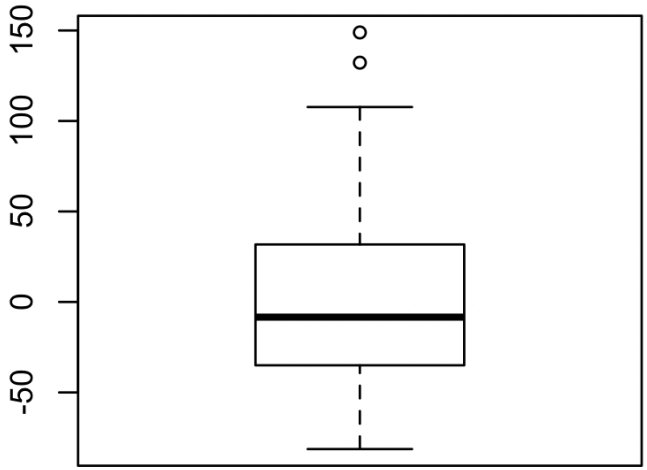

Residuals vs Fitted
Residuals
Fitted values
20
52
48
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
Factor Level Combinations
Treatment :
Nodrug
Papv
Amino
Tripel

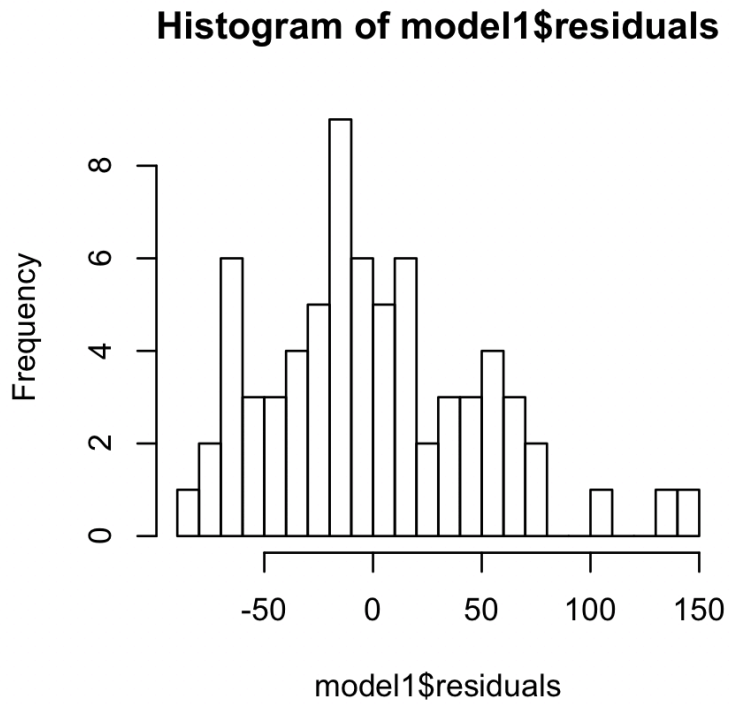
Things to look for in the diagnostic plots are that the residuals versus fitted values don't appear to have any pattern such as U shapes or funnel shapes, and that the QQ plot looks roughly straight. Here we see about three slight outliers.

The Residuals vs Fitted plot and the QQ plot are the most widely used. The Scale-Location plot should ideally be flat, and the leverage plot can find cases with unusual predictor values influencing the analysis, which is more useful in a regression setting with quantitative predictors.

Another approach is to plot the residuals using a histogram or boxplot. We see that the plots show some right-skew and possibly outliers, which is consistent with the QQ plot.

```
> boxplot(model1$residuals)
> histogram(model1$residuals,nclass=20)
> shapiro.test(model1$residuals)

Shapiro-Wilk normality test

data:  model1$residuals
W = 0.96345, p-value = 0.03895
```


-50
0
50
100
150


Histogram of model1$residuals
Frequency
0
2
4
6
8
-50
0
50
100
150
model1$residuals

ANOVA

A non-parametric alternative to ANOVA in this situation (ANOVA with one treatment, and one blocking variable, and no replication within treatment-block combinations), you can use the Friedman Test (named after economist Milton Friedman), which is similar to the Kruskal-Wallis test for one-way ANOVA. Here the values within each block are replaced by the ranks within that block. So patient 1 becomes

Patient	Nodrug	Placebo	Papv	Morp	Amino	Pento	Tripel
1	174	263	105	199	141	108	141

More general tests for dealing with ANOVA alternatives based on rank are called Durbin tests.

ANOVA

The Friedman test is implemented in R

```
> friedman.test(Seconds~Treatment | Patient,data=itch.long)

Friedman rank sum test

data:  Seconds and Treatment and Patient
Friedman chi-squared = 14.887, df = 6, p-value = 0.02115
```

This gives a similar result as the original ANOVA. Note that the syntax (and test) distinguishes blocks from treatments. Here you condition on the blocks (patients). If you swap Treatment and Patient variables, then you are testing whether patients differ from each other, controlling for type of medication. This would also result in a statistically significant test (p-value = .01).

ANOVA with two factors and replication

Generally, ANOVA can be run with more than two factors, some of which might be considered blocking variables (meaning we want to control for them), or we might be interested in all factors.

Often experiments are done with replication for different combinations of treatments. This is usually preferable to just having one observation for each combination (it is more data and allows better estimates of variability).

ANOVA with two factors and replication

For example, consider an experiment on beetles with four different insecticides and three different doses (low, medium, high). There are twelve combinations, and suppose each combination is replicated four times, with the survival time of the beetles recorded. This results in 48 observations.

For this data, the doses of high, medium, and low, are really ordinal (we don't know if they are equally spaced, for example, but they can be ranked), but the ANOVA will treat them as qualitative, like having three different brands without knowing the rankings. Time is measured in fractions of a 10 minute interval. (So 0.4 means 4 minutes.)

ANOVA with two factors and replication

```
dose insecticide t1 t2 t3 t4
1 A 0.31 0.45 0.46 0.43
1 B 0.82 1.10 0.88 0.72
1 C 0.43 0.45 0.63 0.76
1 D 0.45 0.71 0.66 0.62
2 A 0.36 0.29 0.40 0.23
2 B 0.92 0.61 0.49 1.24
2 C 0.44 0.35 0.31 0.40
2 D 0.56 1.02 0.71 0.38
3 A 0.22 0.21 0.18 0.23
3 B 0.30 0.37 0.38 0.29
3 C 0.23 0.25 0.24 0.22
3 D 0.30 0.36 0.31 0.33
```

ANOVA with two factors and replication

The original data has the doses listed as 1, 2, and 3. To do an ANOVA, and to not assume the doses are equally spaced, we should treat them as factor variables.

```
> x <- read.table("beetles",header=T)
> x$dose <- factor(x$dose, labels =
c("low", "medium", "high"))
> x
```

You should be careful here that the program assigns the correct labels to the observations.

ANOVA with two factors and replication

```
> x
     dose insecticide   t1   t2   t3   t4
1     low           A 0.31 0.45 0.46 0.43
2     low           B 0.82 1.10 0.88 0.72
3     low           C 0.43 0.45 0.63 0.76
4     low           D 0.45 0.71 0.66 0.62
5  medium           A 0.36 0.29 0.40 0.23
6  medium           B 0.92 0.61 0.49 1.24
7  medium           C 0.44 0.35 0.31 0.40
8  medium           D 0.56 1.02 0.71 0.38
9    high           A 0.22 0.21 0.18 0.23
10   high           B 0.30 0.37 0.38 0.29
11   high           C 0.23 0.25 0.24 0.22
12   high           D 0.30 0.36 0.31 0.33
```

ANOVA with two factors and replication

As usual, we need to reshape the data into the long format. Here the columns should be dose, insecticide, and replicate.

```
library(reshape2)
beetles.long <- melt(x
                , id.vars       = c("dose", "insecticide")
                , variable.name = "number"
                , value.name    = "hours10"
              )
str(beetles.long)
> str(beetles.long)
'data.frame': 48 obs. of  4 variables:
 $ dose       : Factor w/ 3 levels "low","medium",..: 1 1 1 1 2 2
 $ insecticide: Factor w/ 4 levels "A","B","C","D": 1 2 3 4 1 2 3
 $ number     : Factor w/ 4 levels "t1","t2","t3",..: 1 1 1 1 1 1
 $ hours10    : num  0.31 0.82 0.43 0.45 0.36 0.92 0.44 0.56 0.22
```

ANOVA with two factors and replication

```
> beetles.long
> head(beetles.long)
    dose insecticide number hours10
1    low           A     t1    0.31
2    low           B     t1    0.82
3    low           C     t1    0.43
4    low           D     t1    0.45
5 medium           A     t1    0.36
6 medium           B     t1    0.92
```

ANOVA with two factors and replication

Balanced ANOVA examples like this have an advantage in interpretation, which is that you can think about the average response for each combination of predictors, and that the average of say, all low doses is the average of the averages for each combination of low dose and insecticide.

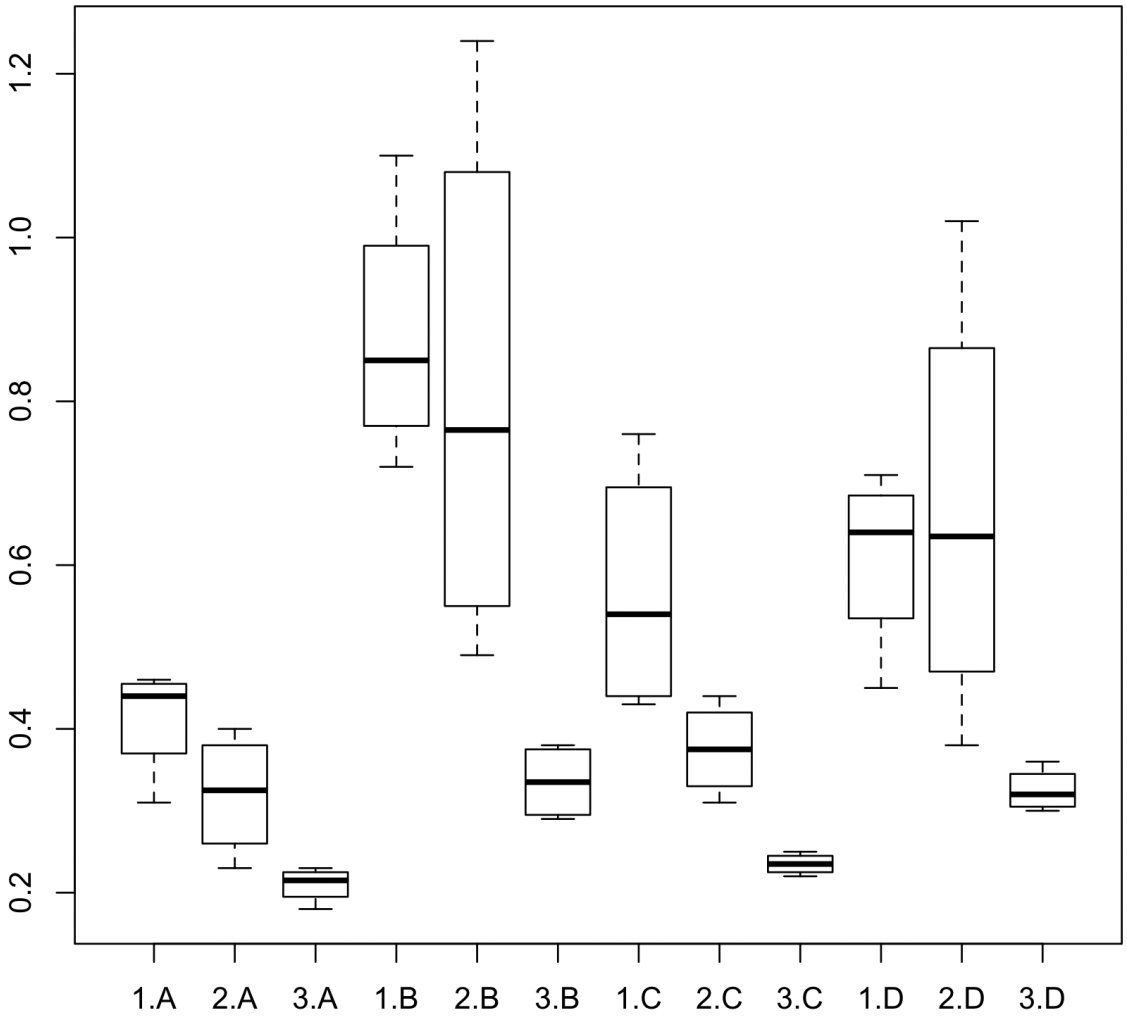
Cell Means	Dose			
Insecticide	1	2	3	Insect marg
A	0.413	0.320	0.210	0.314
B	0.880	0.815	0.335	0.677
C	0.568	0.375	0.235	0.393
D	0.610	0.668	0.325	0.534
Dose marg	0.618	0.544	0.277	0.480

ANOVA with two factors and replication

This is easier to interpret than the original data. Looking at the margins, the survival time was lowest for insecticides A and C. Higher doses also lead to lower survival times on average (without claiming statistical signficance here), but the survival times are not equally spaced—the difference in average survival times between doses 3 versus 2 is larger than for doses 2 versus 1 (again, not claiming any significance here).

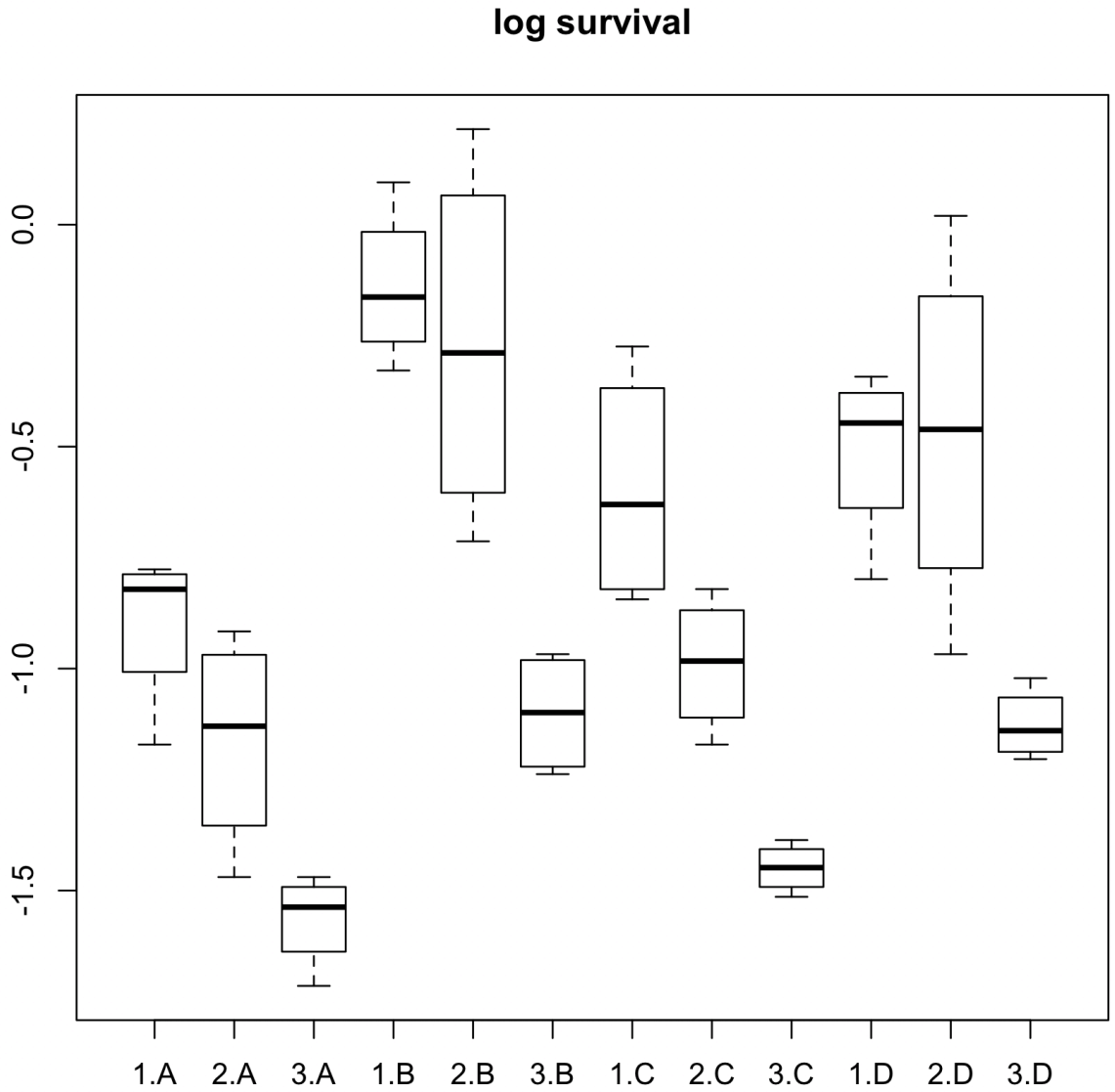
You can do boxplots for looking at the responses for combinations of predictors.

```
> boxplot(hours10 ~ dose + insecticide,data=beetles.long)
```


0.2
0.4
0.6
0.8
1.0
1.2
1.A
2.A
3.A
1.B
2.B
3.B
1.C
2.C
3.C
1.D
2.D
3.D

It looks like there are problems with the equal variances assumption! We'll proceed anyway to illustrate the ideas for two-way ANOVA.

To make the assumptions not so badly violated, one possibility is to transform the data, such as using log of the survival times. We'll analyze the data both ways.


log survival
0.0
-0.5
-1.0
-1.5
1.A
2.A
3.A
1.B
2.B
3.B
1.C
2.C
3.C
1.D
2.D
3.D

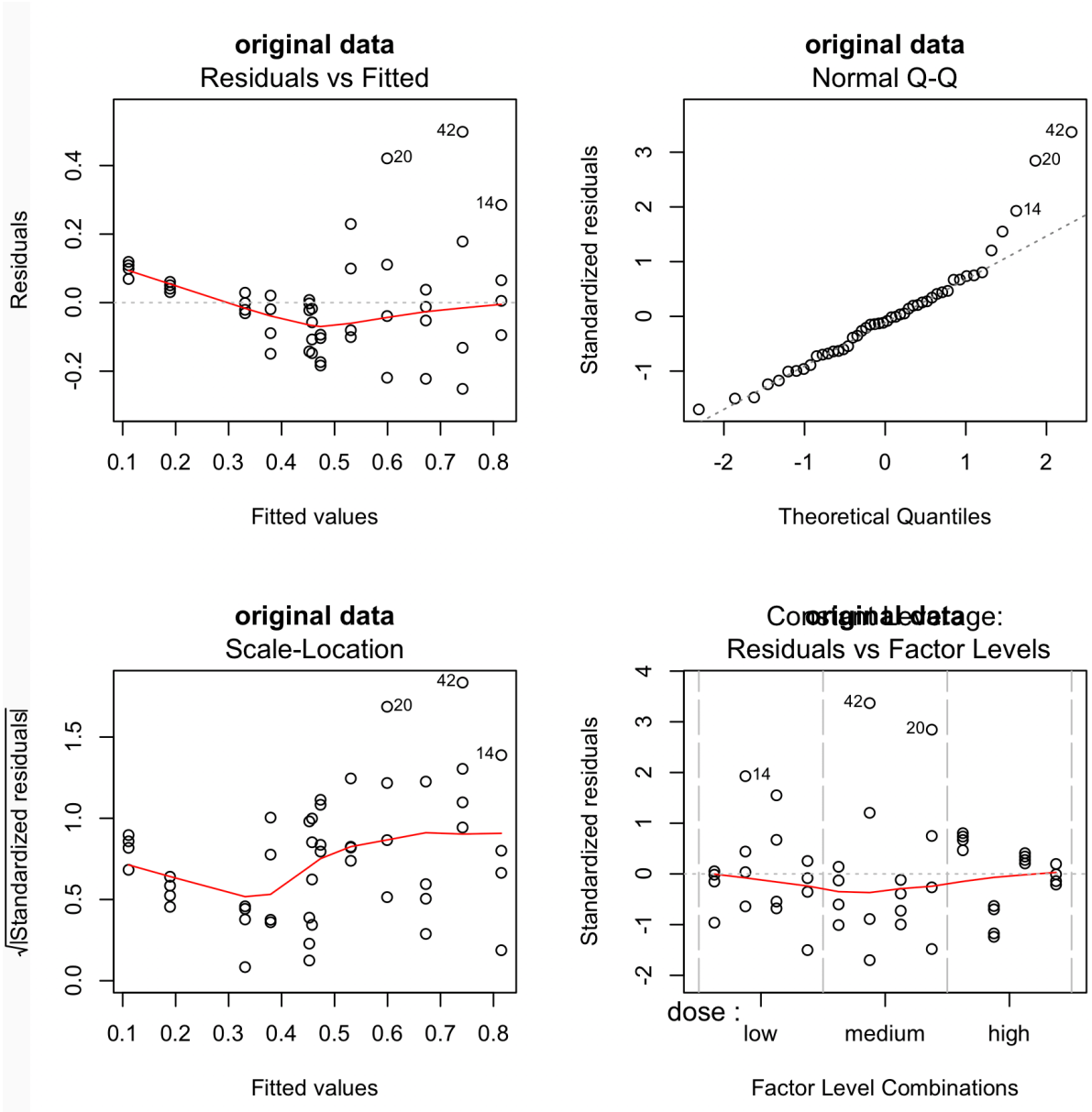
To analyze the data as a two-factor ANOVA, you can use the same code as for the randomized block design.

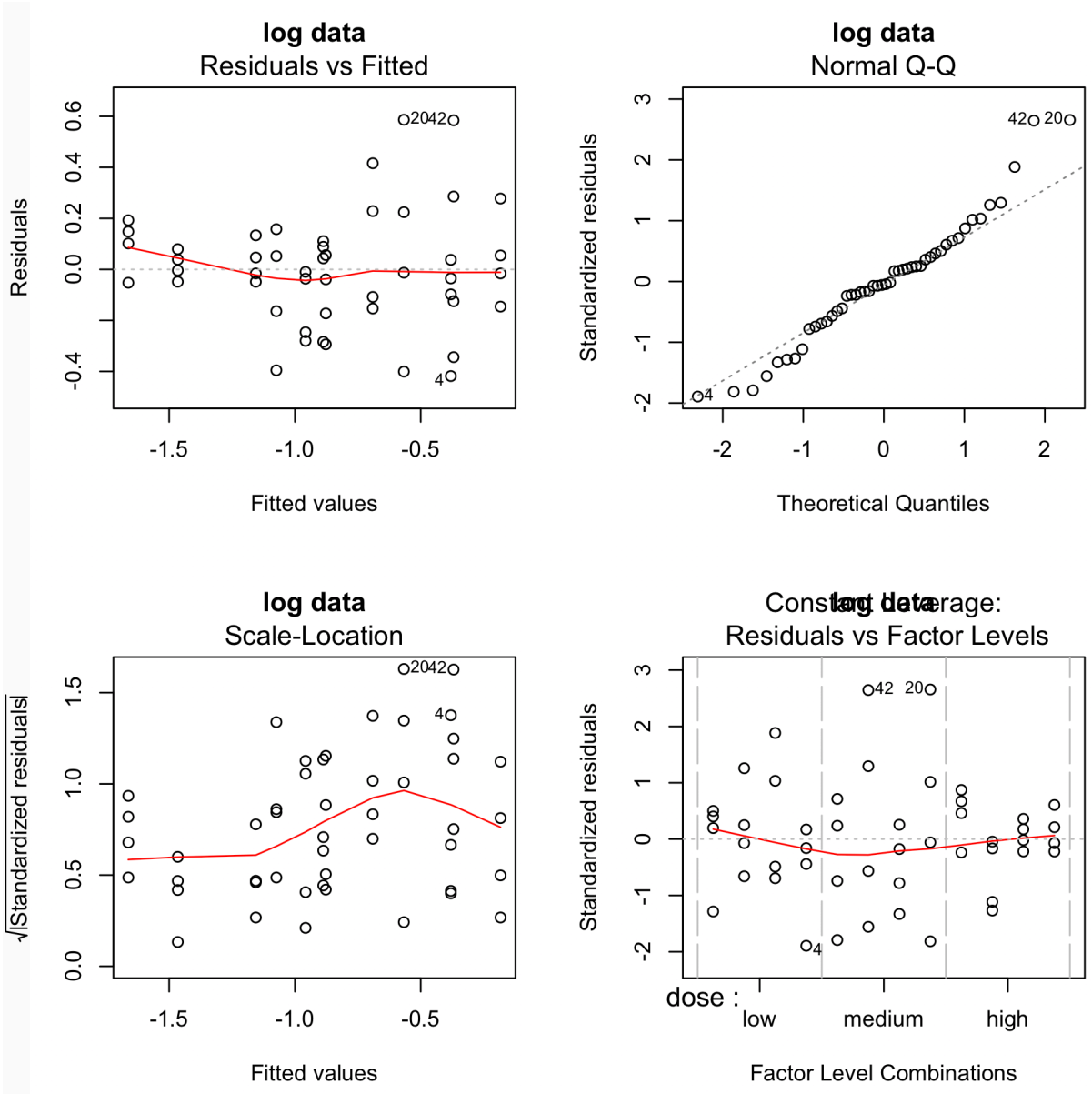
```
> attach(beetles.long)
> m1 <- lm(hours10 ~ dose + insecticide)
> library(car)
> Anova(m1,type=3)
Response: hours10
             Sum Sq Df F value    Pr(>F)
(Intercept) 1.63654  1  65.408 4.224e-10 ***
dose        1.03301  2  20.643 5.704e-07 ***
insecticide 0.92121  3  12.273 6.697e-06 ***
Residuals   1.05086 42
```

```
> m2 <- lm(log(hours10) ~ dose + insecticide)
> Anova(m2,type=3)
Anova Table (Type III tests)

Response: log(hours10)
             Sum Sq Df F value    Pr(>F)
(Intercept) 6.2985  1 112.941 1.768e-13 ***
dose        5.2375  2  46.958 1.948e-11 ***
insecticide 3.5572  3  21.262 1.560e-08 ***
Residuals   2.3423 42
---
```

Based on the boxplots, I would be more comfortable with using the log-transformed survival times, although it doesn't much change conclusions at this point. We can also look at diagnostic plots or tests of normality for the residuals. The log-transformed data is more consistent with normality assumptions.

```
> shapiro.test(m1$residuals)

Shapiro-Wilk normality test

data:  m1$residuals
W = 0.92242, p-value = 0.003622

> shapiro.test(m2$residuals)

Shapiro-Wilk normality test

data:  m2$residuals
W = 0.96408, p-value = 0.1475
```

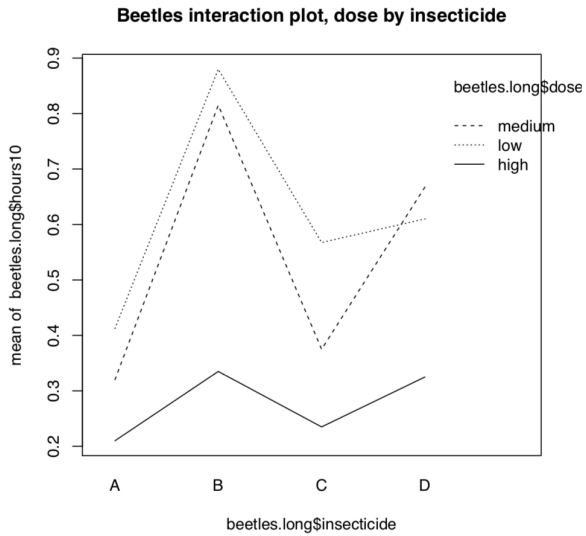

original data
Residuals vs Fitted
Residuals
Fitted values
42
20
14
original data
Normal Q-Q
Standardized residuals
Theoretical Quantiles
original data
Scale-Location
√|Standardized residuals|
Fitted values
original data
Residuals vs Factor Levels
Standardized residuals
dose :
low
medium
high
Factor Level Combinations


log data
Residuals vs Fitted
Residuals
Fitted values
log data
Normal Q-Q
Standardized residuals
Theoretical Quantiles
log data
Scale-Location
√|Standardized residuals|
Fitted values
log data
Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
dose :
low
medium
high
Factor Level Combinations

ANOVA with interaction

When there are two factors, it is possible that the effect of one factor depends on the value of the other factor. For this example, this could mean that the effect of the dose depends on the insecticide. To check for an interaction, you can use this code:

```
> m3 <- lm(hours10 ~ dose + insecticide + dose*insecticde)
> Anova(m3,type=3)
Anova Table (Type III tests)

Response: hours10
                  Sum Sq Df F value    Pr(>F)
(Intercept)      0.68063  1 30.6004 2.937e-06 ***
dose             0.08222  2  1.8482 0.1721570
insecticide      0.45395  3  6.8031 0.0009469 ***
dose:insecticide 0.25014  6  1.8743 0.1122506
Residuals        0.80072 36
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```




```
interaction.plot(beetles.long$dose, beetles.long$insecticide,
beetles.long$hours10 , main = "insecticide by dose")
interaction.plot(beetles.long$insecticide, beetles.long$dose,
beetles.long$hours10, main = "dose by insecticide")
```

ANOVA with interaction

The idea behind the plots is that we can see whether the effect of the insecticide depends on the dose, or similarly, whether the effect of the dose depends on the insecticide. For example, in the left plot on the previous slide, there is a rank ordering of insecticides based on survival times.

Here lower survival times means a more effective insecticide, and for each dose, we appear to have that insecticide A has the lowest survival time, followed by C, then followed by D, and finally B. If there were a strong interaction between dose and insecticide, you might find that one insecticide is the most effective at low doses, while another is the the most effective at higher doses. In this case, the rank ordering of insecticides doesn't change much.

ANOVA with interaction

A statistical test for interaction is testing whether the lines in the interaction plot are parallel, taking into account variability in the data. This does not necessarily mean that the lines are straight, but that the spacing in between the lines doesn't change significantly from level to level of the factor on the horizontal axis. An interaction can show up in the interaction plots either by curves crossing or by being significantly non-parallel.

Looking back at the table of cell means (slide 57), the idea is the differences between columns are similar, and the differences between rows are similar. For example, going from dose 1 to dose 2 (low to medium), the change in average survival for insecticide A is (0.413-0.320) = 0.093 (i.e., .93 minutes or 55 seconds), and the difference for insecticde B is (0.880 - 0.815) = 0.712 (i.e., 39 seconds). Given the variability in the data, the change going from low to medium doses is similar for insecticides A and B.

ANOVA with interaction

Often in a two-way ANOVA, you check for an interaction, and if the interaction is not signficiant, you don't consider it in a final model. This again is an issue for model selection, where you decide whether you prefer a model with an interaction term or with no interaction term. The interaction model can be written as

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

or in terms of means,

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Informally,

Response = Grand mean + F1 effect + F2 effect + F1-by-F2 interaction + residual

. The model with no interaction is called an additive model or main effects model, and is

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

ANOVA with interaction

The main effects and interaction effects can be estimated as follows:

$$
\begin{aligned}
\hat{\mu} &= \bar{y}_{\cdot\cdot} \quad \text{the estimated grand mean} \\
\hat{\alpha}_i &= \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \quad \text{the estimated F1 effect } i = 1, 2, \ldots, I \\
\hat{\beta}_j &= \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot} \quad \text{the estimated F2 effect } j = 1, 2, \ldots, J \\
\widehat{(\alpha\beta)}_{ij} &= \bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot} \quad \text{the estimated cell interaction}
\end{aligned}
$$

ANOVA with interaction

The ANOVA table (in the balanced case) is as follows:

Source	df	SS	MS
F1	$I-1$	$KJ\sum_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2$	MS F1=SS/df
F2	$J-1$	$KI\sum_j(\bar{y}_{\cdot j}-\bar{y}_{\cdot\cdot})^2$	MS F2=SS/df
Interaction	$(I-1)(J-1)$	$K\sum_{ij}(y_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y}_{\cdot\cdot})^2$	MS Inter=SS/df
Error	$IJ(K-1)$	$(K-1)\sum_{ij}s_{ij}^2$	MSE=SS/df
Total	$IJK-1$	$\sum_{ijk}(y_{ijk}-\bar{y}_{\cdot\cdot})^2$.	

ANOVA with interaction

Since there is no significant interaction for the insecticide example, we'll go ahead and look at comparing factors for significance, not using any interactions.

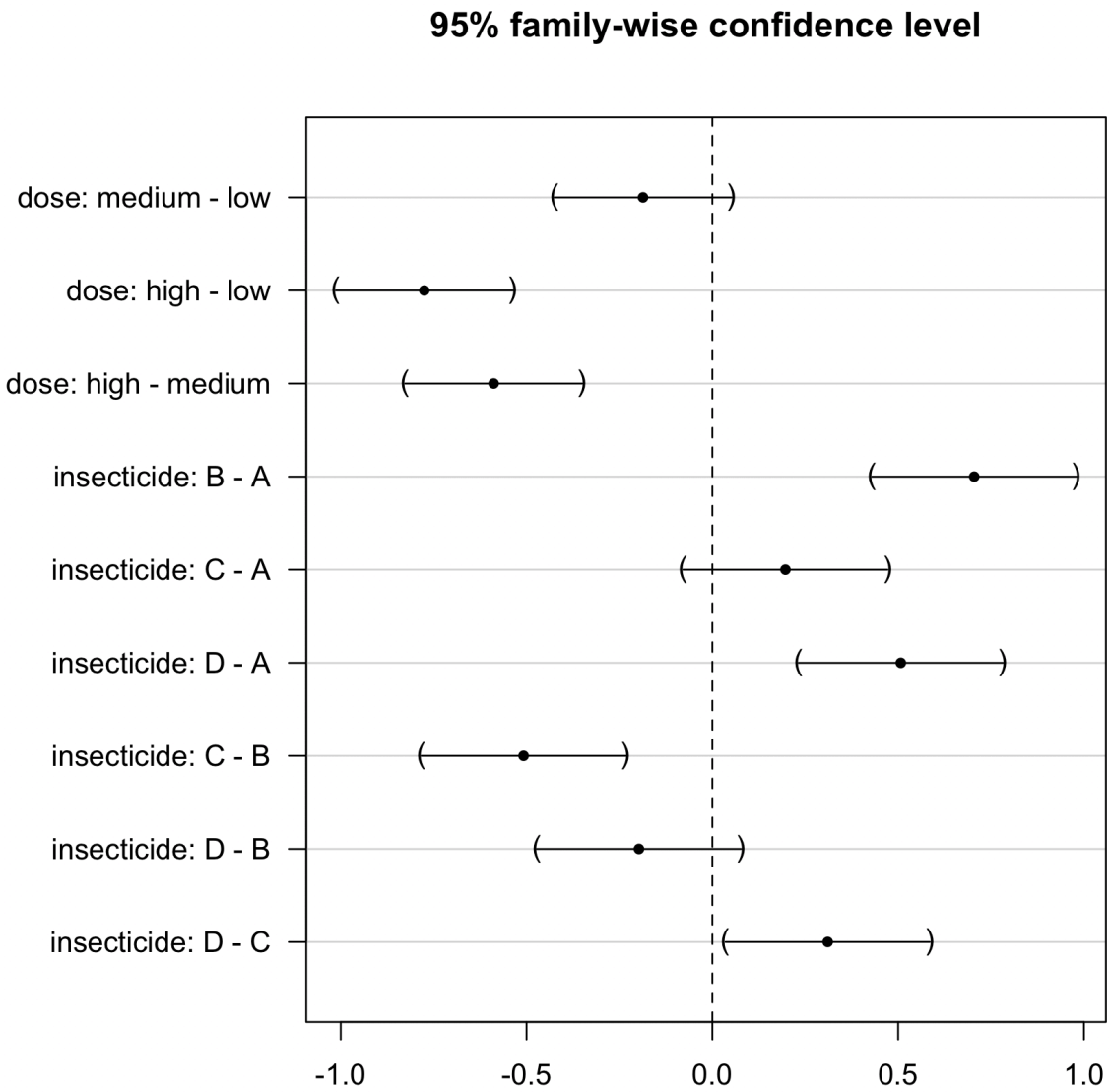
```
> library(multcomp)
> comparisons <- glht(aov(m2), linfct = mcp(dose = "Tukey",
insecticide="Tukey"))
> summary(comparisons)
Linear Hypotheses:
                           Estimate Std. Error t value Pr(>|t|)
dose: medium - low == 0   -0.18666    0.08349  -2.236   0.1941
dose: high - low == 0     -0.77515    0.08349  -9.284   <0.001 ***
dose: high - medium == 0  -0.58849    0.08349  -7.048   <0.001 ***
insecticide: B - A == 0    0.70465    0.09641   7.309   <0.001 ***
insecticide: C - A == 0    0.19671    0.09641   2.040   0.2786
insecticide: D - A == 0    0.50707    0.09641   5.260   <0.001 ***
insecticide: C - B == 0   -0.50795    0.09641  -5.269   <0.001 ***
insecticide: D - B == 0   -0.19759    0.09641  -2.049   0.2743
insecticide: D - C == 0    0.31036    0.09641   3.219   0.0195 *
```

ANOVA with interaction

Plotting the pairwise comparisons.

```
> op <- par(no.readonly = TRUE) # the whole list of settable par's
# make wider left margin to fit contrast labels
> par(mar = c(5, 10, 4, 2) + 0.1) # order is c(bottom, left, top,
> plot(summary(comparisons, test = adjusted("bonferroni"))
> par(op)
```

ANOVA with interaction



ANOVA with interaction

Several of the comparisons appear to be significant. Medium and low are not significantly different from each other, both are significantly different from high doses. For insecticides, the only pairwise comparisons that are not significant are A to C and B to D. To illustrate the grouping, it is helpful to present these in order of their marginal means (from the table on slide 57):

```
Doses:
1=Low  2=Med   3=High
0.618  0.544   0.276
-----------    -----

Insecticides:
B     D     C      A
0.677 0.534 0.393  0.314
-----------
      -----------
            -----------
```

ANOVA with interaction

We'll now try another two-factor ANOVA example in which the interaction term will be significant. For this example, the voltage of batter is measured at 3 different temperatures (50, 65, 80 degrees F), and using three different materials (metal plates) in the battery, just called 1, 2, and 3. Although the temperatures are equally spaced, we'll still analyze the data using a two-factor ANOVA.

```
battery <-
read.table("http://www.math.unm.edu/~james/STAT428/battery.txt",header=T)
```

ANOVA with interaction

The data has essentially the same structure as the beetle data, again with four replications for each combination of factors. To run this as an ANOVA, we'll convert the predictor variables to factors.

```
> battery$material <- factor(battery$material)
> battery$temp     <- factor(battery$temp)
> library(reshape2)
> battery.long <- melt(battery, id.vars =
 c("material","temp"), variable.name = "battery",
 value.name = "volt")
```

ANOVA with interaction

```
> battery
  material temp  v1  v2  v3  v4
1        1   50 130 155  74 180
2        1   65  34  40  80  75
3        1   80  20  70  82  58
4        2   50 150 188 159 126
5        2   65 136 122 106 115
6        2   80  25  70  58  45
7        3   50 138 110 168 160
8        3   65 174 120 150 139
9        3   80  96 104  82  60
```

ANOVA with interaction

```
> battery.long
   material temp battery volt
1         1   50      v1  130
2         1   65      v1   34
3         1   80      v1   20
4         2   50      v1  150
5         2   65      v1  136
6         2   80      v1   25
7         3   50      v1  138
8         3   65      v1  174
9         3   80      v1   96
```

ANOVA with interaction

```
> str(battery.long)
'data.frame': 36 obs. of  4 variables:
 $ material: Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 3 3
 $ temp    : Factor w/ 3 levels "50","65","80": 1 2 3 1 2 3
 $ battery : Factor w/ 4 levels "v1","v2","v3",..: 1 1 1 1 1
 $ volt    : int  130 34 20 150 136 25 138 174 96 155 ...
```

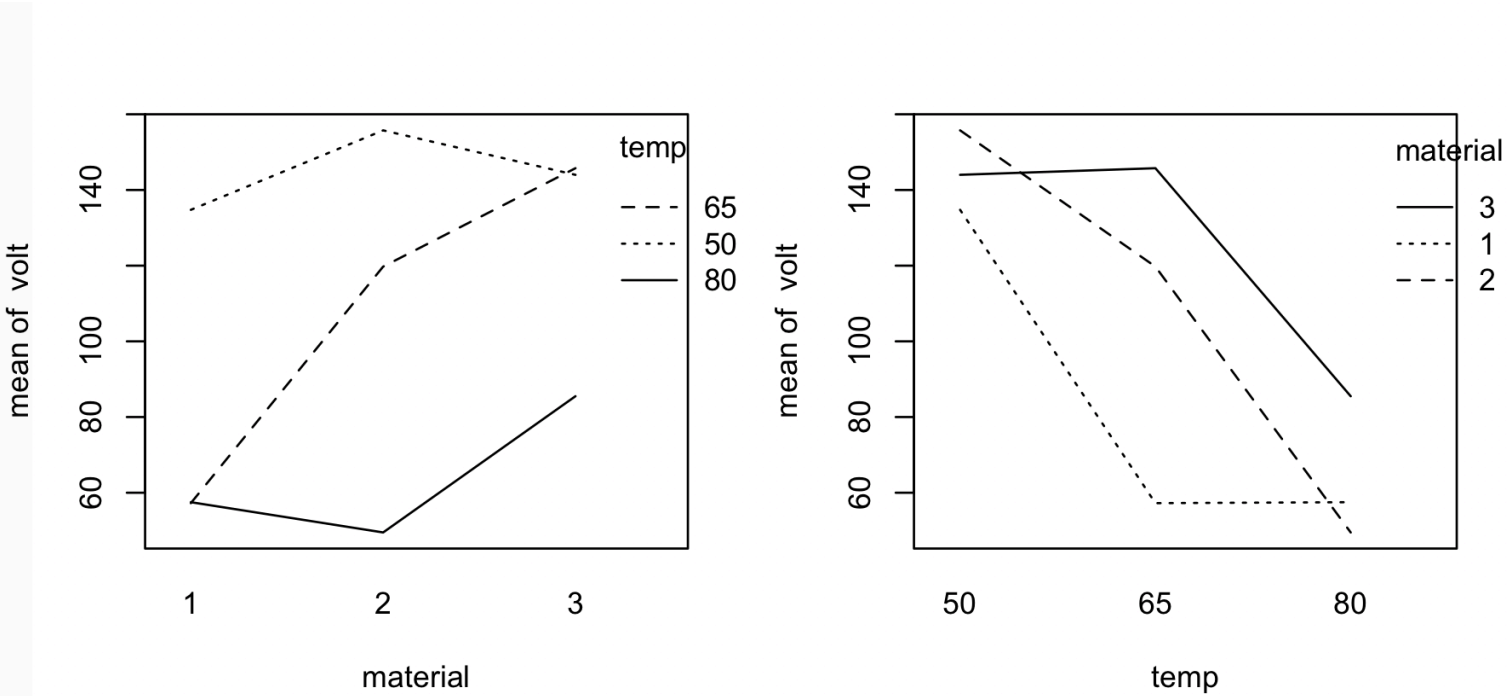
ANOVA with interaction

```
> library(car)
> attach(battery.long)
> m1 <- lm(volt ~ materail + temp + material*temp)
> # equivalently,
> m1 <- lm(volt ~ materail*temp)
> Anova(m1,type=3)
Anova Table (Type III tests)

Response: volt
              Sum Sq Df  F value    Pr(>F)
(Intercept)    72630  1 107.5664 6.456e-11 ***
material         886  2   0.6562 0.5268904
temp           15965  2  11.8223 0.0002052 ***
material:temp   9614  4   3.5595 0.0186112 *
Residuals      18231 27
```

ANOVA with interaction

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         134.75      12.99  10.371 6.46e-11 ***
material2            21.00      18.37   1.143 0.263107
material3             9.25      18.37   0.503 0.618747
temp65              -77.50      18.37  -4.218 0.000248 ***
temp80              -77.25      18.37  -4.204 0.000257 ***
material2:temp65     41.50      25.98   1.597 0.121886
material3:temp65     79.25      25.98   3.050 0.005083 **
material2:temp80    -29.00      25.98  -1.116 0.274242
material3:temp80     18.75      25.98   0.722 0.476759
```

```
> par(mfrow=c(2,2))
> interaction.plot(material,temp,volt)
> interaction.plot(temp,material,volt)
```



```
> by(volt,battery.long[,c(1,2)],mean)
material: 1
temp: 50
[1] 134.75
------------------------------------------------------------
material: 2
temp: 50
[1] 155.75
------------------------------------------------------------
material: 3
temp: 50
[1] 144
------------------------------------------------------------
material: 1
temp: 65
[1] 57.25
```

From the interaction plots, we see that as the temperature increases, the voltage tends to decrease for all three materials. However, for material 3, there is very little change from 50 to 65 degrees, and a big decrease from 65 to 80. For material 1, there is a large change in voltage from 50 to 65 degrees, and very little change from 65 to 80.

This suggests that the effect of temperature depends on the material, and similarly, the effect of the material depends on the temperature.

Note that in the original model, the test for the main effect for material doesn't appear significant, but because the interaction is significant, you can't conclude that the materials are not significantly affecting the voltage.

Note that if the model is made with the interaction term removed, then both material and temperature are significant. The p-value for material isn't significant only when the interaction is in the model.

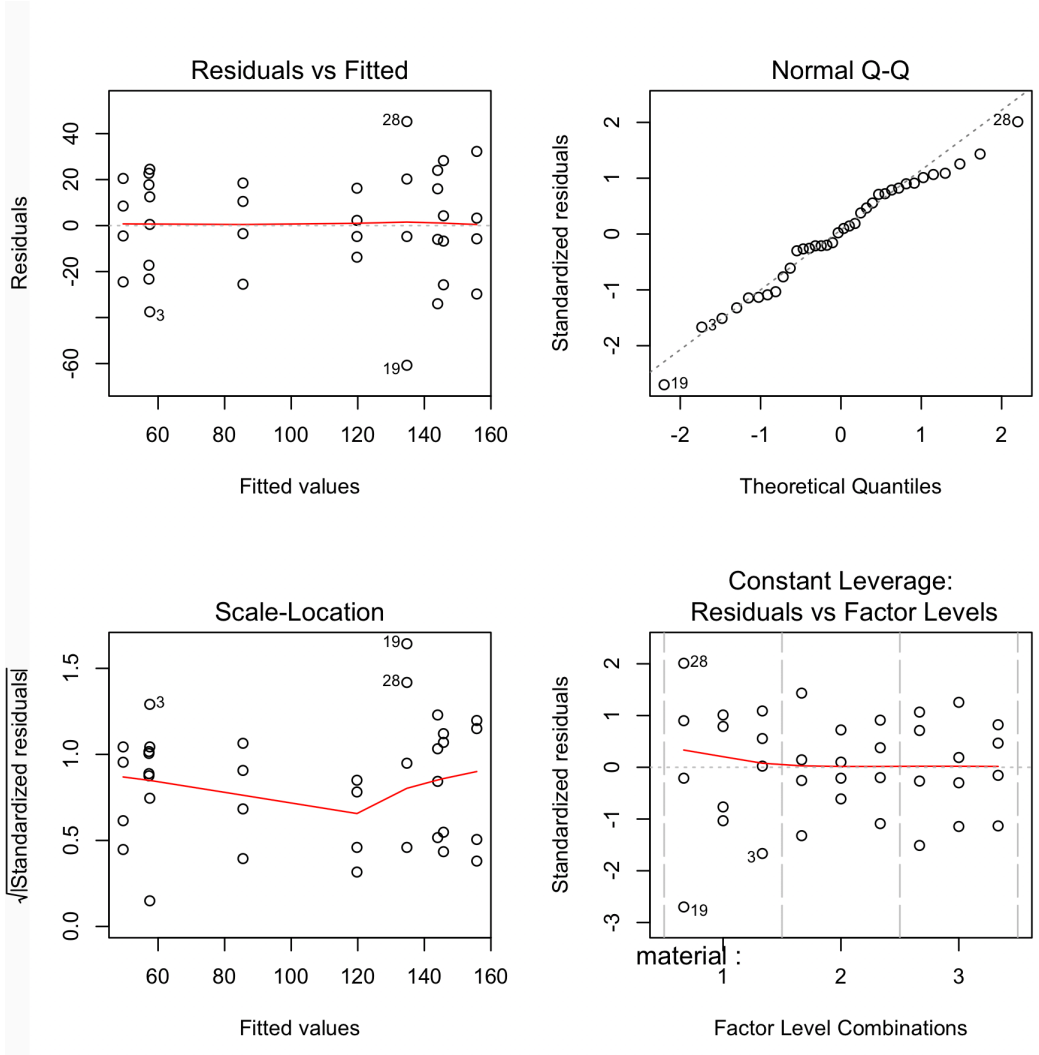
To give an example of what the model predicts for combinations of material and temperature, consider predicting the voltage for material 2 at 65 degrees. This would be

$$134.75 + 21.00 - 77.25 + 41.50 = 120$$

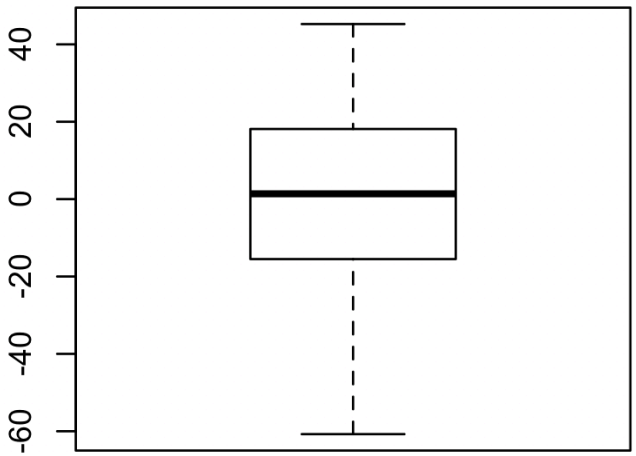
For material 1, temperature 80, the predicted voltage is

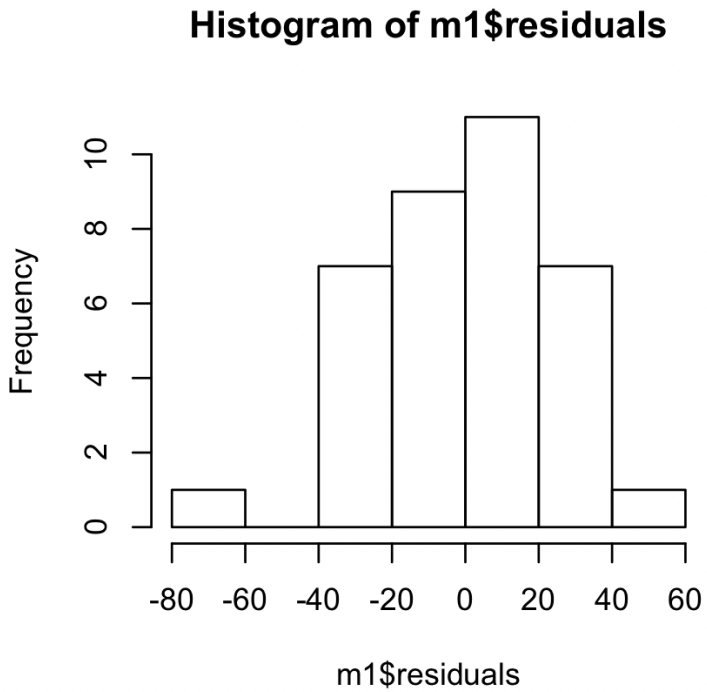
$$134.75 - 77.25 = 57.5$$

ANOVA with interaction

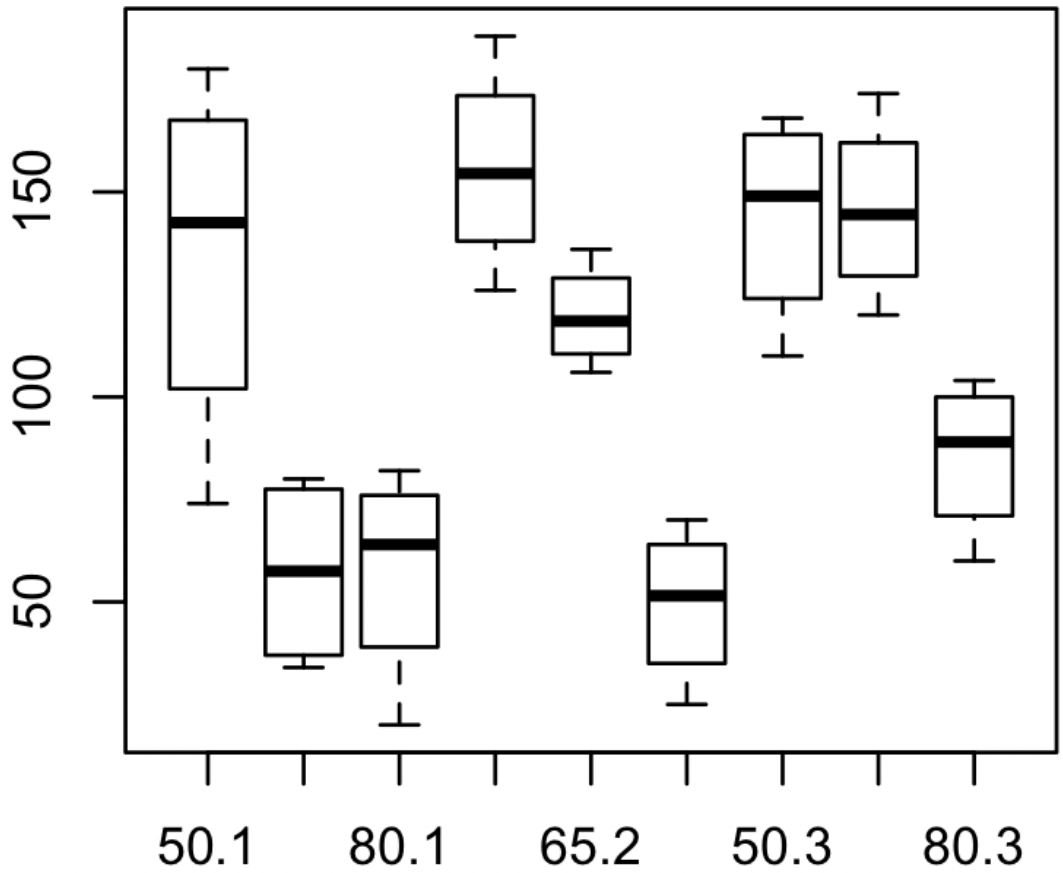


ANOVA with interaction





ANOVA with interaction



Battery example as a regression

Here we'll redo the battery example, treating temperature as quantitative instead of as a factor variable.

```
> x <- read.table("battery.txt",header=T)
> x
  material temp  v1  v2  v3  v4
1        1   50 130 155  74 180
2        1   65  34  40  80  75
3        1   80  20  70  82  58
4        2   50 150 188 159 126
> str(x)
'data.frame': 9 obs. of  6 variables:
 $ material: int  1 1 1 2 2 2 3 3 3
 $ temp    : int  50 65 80 50 65 80 50 65 80
 $ v1      : int  130 34 20 150 136 25 138 174 96
 $ v2      : int  155 40 70 188 122 70 110 120 104
 $ v3      : int  74 80 82 159 106 58 168 150 82
 $ v4      : int  180 75 58 126 115 45 160 139 60
```

Battery example as a regression

```
> x$material <- factor(x$material)
> library(reshape2)
> x2 <- melt(x,id.vars=c("material","temp"),
variable.name="number",value.name="volt")
> str(x2)
'data.frame': 36 obs. of  4 variables:
 $ material: Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 3 3 3 1
 $ temp    : int  50 65 80 50 65 80 50 65 80 50 ...
 $ number  : Factor w/ 4 levels "v1","v2","v3",..: 1 1 1 1 1 1 1 1
 $ volt    : int  130 34 20 150 136 25 138 174 96 155 ...
```

Battery example as a regression

```
> m4 <- lm(volt ~ material*temp)
> Anova(m4,type=3)
Anova Table (Type III tests)

Response: volt
              Sum Sq Df F value    Pr(>F)
(Intercept)  25825.8  1 30.2581 5.667e-06 ***
material      2093.3  2  1.2263 0.3076614
temp         11935.1  1 13.9835 0.0007772 ***
material:temp 2315.1  2  1.3562 0.2729805
Residuals    25605.5 30
```

```
m5:
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    250.5417    45.5469   5.501 5.67e-06 ***
material2       88.0000    64.4130   1.366 0.182036
material3        1.2917    64.4130   0.020 0.984134
temp            -2.5750     0.6886  -3.739 0.000777 ***
material2:temp  -0.9667     0.9738  -0.993 0.328825
material3:temp   0.6250     0.9738   0.642 0.525881
```

Battery example as a regression

```
> Anova(m5,type=3)
Anova Table (Type III tests)

Response: volt
             Sum Sq Df F value    Pr(>F)
(Intercept)  76854  1 88.0826 1.046e-10 ***
material     10684  2  6.1223  0.005606 **
temp         39043  1 44.7471 1.499e-07 ***
Residuals    27921 32
```

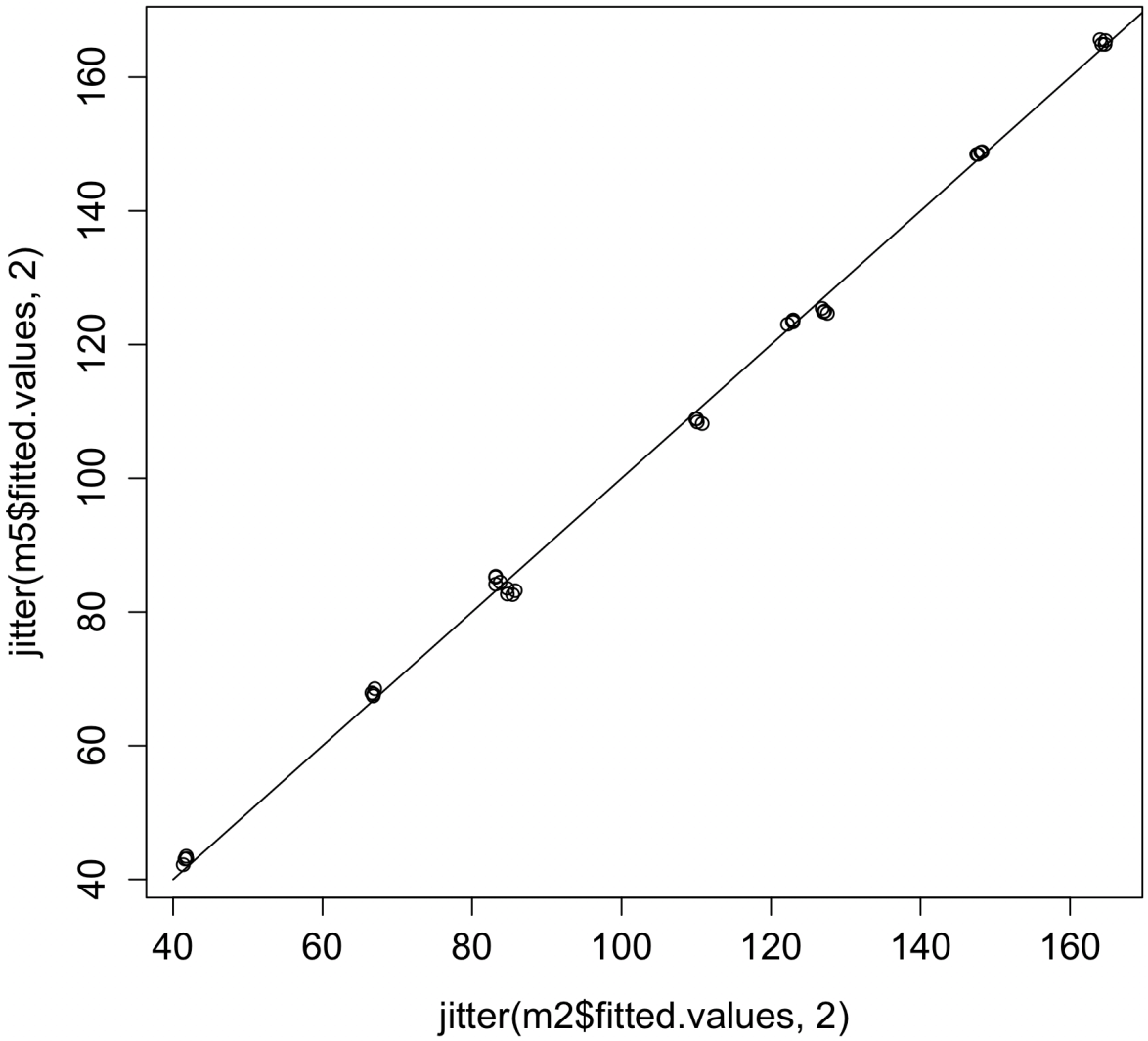
Battery example as a regression

```
> summary(m5)

Call:
lm(formula = volt ~ material + temp)

Residuals:
    Min      1Q  Median      3Q     Max
-55.417 -22.708   1.667  16.188  56.500

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  257.944     27.484   9.385 1.05e-10 ***
material2     25.167     12.059   2.087  0.04494 *
material3     41.917     12.059   3.476  0.00149 **
temp          -2.689      0.402  -6.689 1.50e-07 ***
```

Battery example as a regression

If we compare the models wit no interaction using the ANOVA versus the regression, we see that they make similar predictions. The models are different and make slightly different predictions, but the predictions are fairly similar, and are highly correlated.

```
> plot(jitter(m2$fitted.values,2),jitter(m5$fitted.values,2),
cex.lab=1.3,cex.axis=1.3)
> lines(c(40,180),c(40,180))
```


jitter(m5$fitted.values, 2)
jitter(m2$fitted.values, 2)
40
60
80
100
120
140
160

Battery example as a regression

Another way to compare the ANOVA versus regression models is by looking at the estimated effects and standard deviations:

```
m5:
(Intercept)  257.944     27.484   9.385 1.05e-10 ***
material2     25.167     12.059   2.087  0.04494 *
material3     41.917     12.059   3.476  0.00149 **
temp          -2.689      0.402  -6.689 1.50e-07 ***

m2:
(Intercept)   122.47      11.17  10.965 3.39e-12 ***
material2      25.17      12.24   2.057  0.04819 *
material3      41.92      12.24   3.426  0.00175 **
temp65        -37.25      12.24  -3.044  0.00472 **
temp80        -80.67      12.24  -6.593 2.30e-07 ***
```