Week 12 - ANOVA and Linear Models

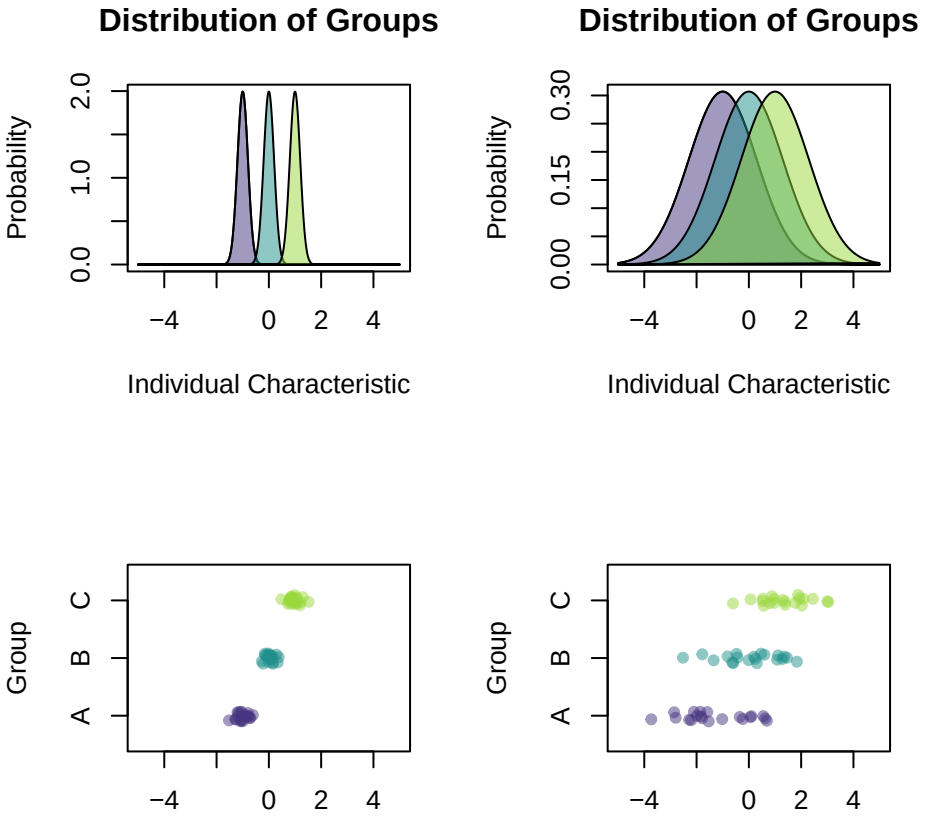
Nicholas Kortessis

April 9, 2025

ANOVA

In an ANOVA, variability among all the individuals in the group are partitioned into two components: variation attributable to the groups and variation attributable to something else other than the groups.

The figure below shows data that might be analyzed using ANOVA. The distributions of individual characteristics in three groups, labelled by color, are shown as well as samples from each group in the bottom panels.



In the example on the left, the within-group variance is small compared to the between-group variance. In the example on the right, the within-group variance is large compared to the between-group variance. You might suspect that you should be more confident that the groups differ in the left panel than in the right panel. This suspicion is likely based on the comparison of within and between group variances. ANOVA makes exactly this comparison. When variation between groups is large compared to variation within groups, ANOVA identifies strong evidence against the idea that the null hypothesis is true.

To figure out how ANOVA partitioning of variance works, remember the equation for variance:

$$\text{Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The variance is simply an average, but an average of what? The average is over "squared deviations", or simply "squares". The way to think about this is that we want to know how much each observation is different from the mean. That is just $x_i - \bar{X}$, but this comes with the problem of keeping track of the signs of deviations. Because deviations are in comparison to the mean, *some are above the mean and some are below the mean* (to convince yourself of this point, put some data points on a number line and try to construct an average where all the points are above or below the average). To not have to keep track of signs and to just measure *distance from the mean*, we take the square of these deviations $(X_i - \bar{X})^2$.

Sum of Squares (SS) are simply this, a measure of the **total amount of differences between data points and their mean**. ANOVA works by analyzing these sums of squares and partitioning them into different "buckets". If most of the deviations in the data come from the fact that the individuals are in different groups, then we have pretty good confidence that the groups are different. If most of the deviations in the data are between observations in the same group, we have pretty good evidence that the groups are not different.

The total sum of squares is:

$$SS_{\text{tot}} = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X})^2,$$

where X_{ij} is the data point representing the value of the jth individual of the jth group.

This total sum of squares has two components, called error sum of squares and treatment (or group) sums of squares. We can thus rewrite this as $SS_{\text{tot}} = SS_{\text{error}} + SS_{\text{trt}}$. The equations for the other sums of squares are

$$SS_{\text{err}} = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$$

for error sums of squares and

$$SS_{\text{trt}} = \sum_{i=1}^{k}\sum_{j=1}^{n}(\overline{X}_i - \overline{X})^2 = \sum_{i=1}^{k} n_i(\overline{X}_i - \overline{X})^2$$

for treatment (or group) sums of squares.

You'll notice that the Mean Squares (MS) are the SS divided by the degrees of freedom. This dividing by the degrees of freedom means that mean squares are just variances.

The mean squared treatment is:

$$MS_{\text{trt}} = \frac{SS_{\text{trt}}}{t-1}$$

The mean squared error is:

$$MS_{\text{err}} = \frac{SS_{\text{err}}}{N-t}$$

In an ANOVA, the MS between groups (MS_{trt}) is compared to the MS within groups (MS_{err}):

$$F = \frac{MS_{\text{trt}}}{MS_{\text{err}}}$$

Under the null hypothesis that the means of all the groups are the same, the ratio of variances follows an F distribution, which has two parameters: $df_1 = t - 1$, the treatment (or group) degrees of freedom, and $df_2 = N - t$, the error degrees of freedom. (Note that the total degrees of freedom is $df_1 + df_2 = t - 1 + N - t = N - 1$, just like in a normal estimate of the variance.)

To make a statement about the evidence against the null hypothesis, we compare our observed F value to an F distribution with these same degrees of freedom. If the F value looks likely, we take that as evidence that differences between groups are due to sampling variability. However, if we see that the F value is very unlikely (I.e., when it is large), we take that as evidence against the null hypothesis and conclude that it is unlikely that differences in the means by treatment occurred because of sampling variability.

ANOVA with Simple Designs: One-Way ANOVA

ANOVA partitions variance among groups, but how the groups are constructed matters. The simplest possible design is one where all the groups are of equal interest and are unique compared to all others. Such ANOVAs are called "One-Way" or "One-Factor" ANOVA.

An example comes from an experiment where scientists looked at the effect of light treatment on circadian rhythms in humans. The scientists implemented treatments in which light was shown in a person's eyes or on the back of their knees, and a control for a total of 3 treatments. They then measured the shift in the circadian rhythm.

And yes, we do possess photoreceptors on the back of our knees:

http://www.independent.co.uk/news/science-knees-hold-clue-to-human-body-clock-1138865.html

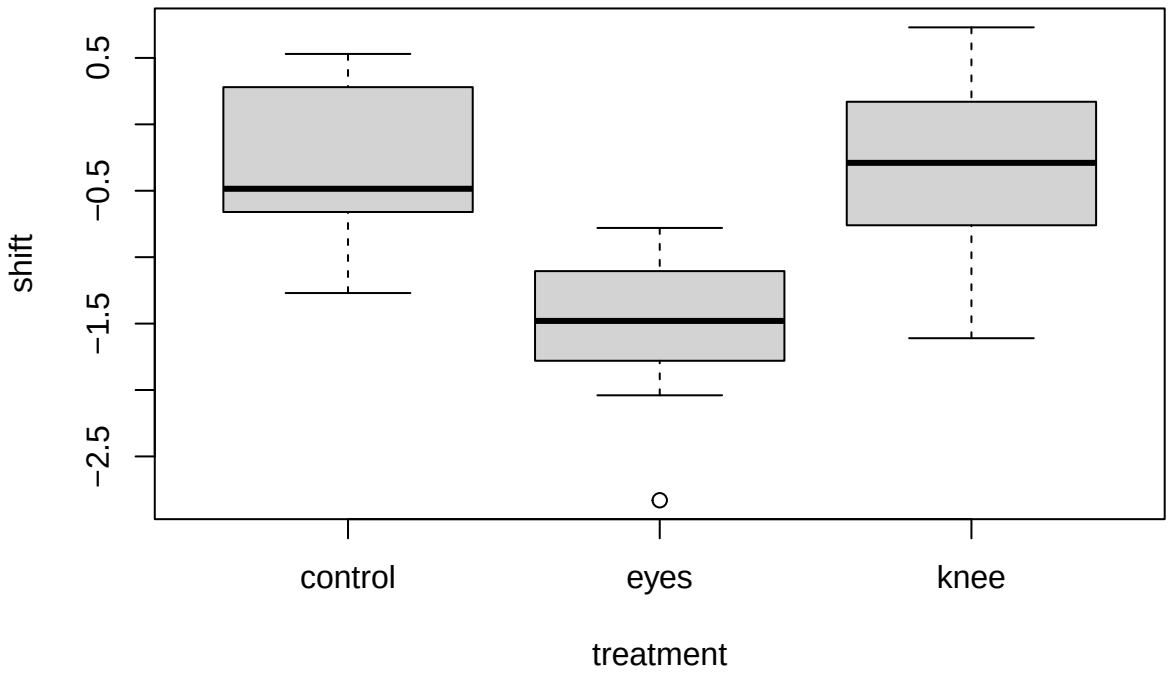
First, load the "knees" data and take a look at it.

```
df.k<-read.csv(file = "knees.csv")
str(df.k)
```

```
## 'data.frame':    22 obs. of  2 variables:
##  $ treatment: chr  "control" "control" "control" "control" ...
##  $ shift    : num  0.53 0.36 0.2 -0.37 -0.6 -0.64 -0.68 -1.27 0.73 0.31 ...
```

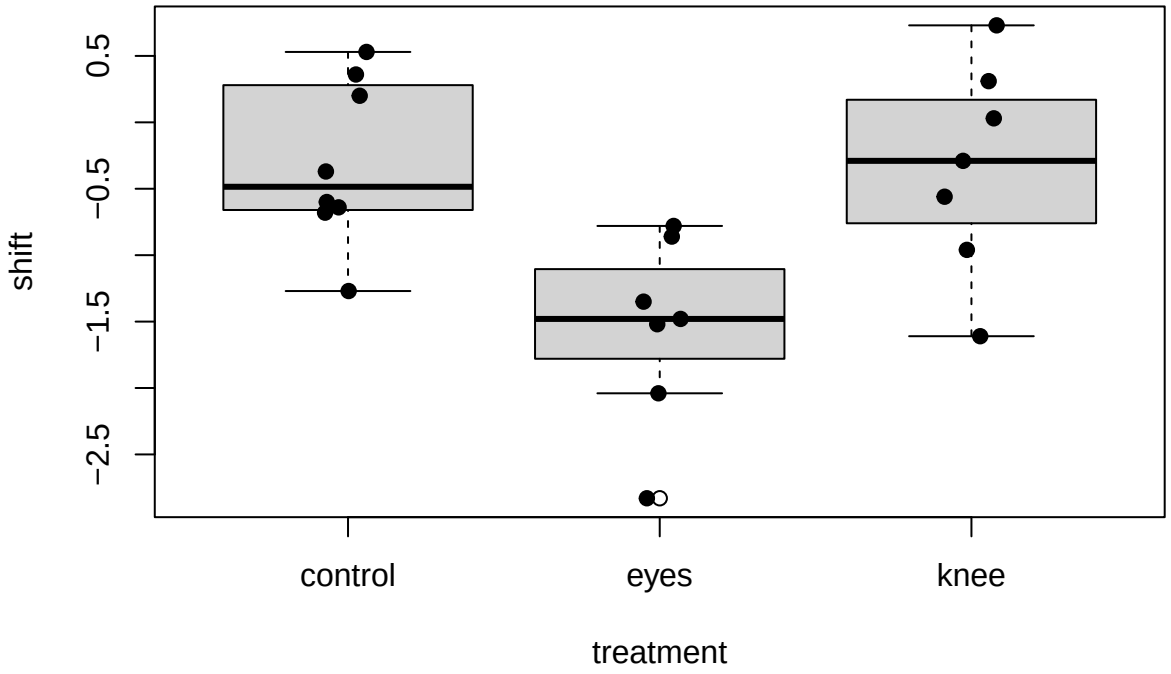
Checkpoint 1: What are the names of each treatment and how many individuals are in each group? Before doing any statistical analysis, make sure you have a good handle on the data. Be sure to look at it first. Use boxplots to make a quick and dirty figure.

```
boxplot(shift ~ treatment, data = df.k)
```



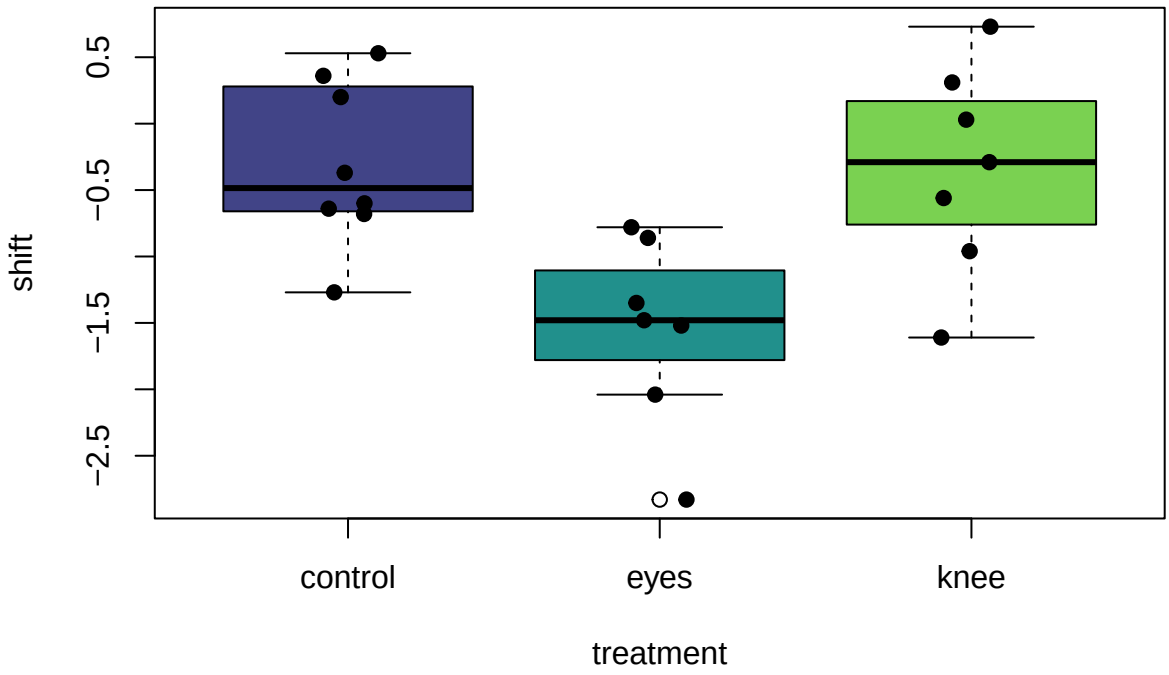
And we could add the data points on if we wanted.

```
boxplot(shift ~ treatment, data = df.k)
stripchart(subset(df.k, subset = treatment == 'control')$shift, at = 1, add = T,
           vertical = T, pch = 19, method = 'jitter')
stripchart(subset(df.k, subset = treatment == 'eyes')$shift, at = 2, add = T,
           vertical = T, pch = 19, method = 'jitter')
stripchart(subset(df.k, subset = treatment == 'knee')$shift, at = 3, add = T,
           vertical = T, pch = 19, method = 'jitter')
```



If you like color, we could try color coding the groups using the `viridis` colormap (to use it, you will need the package `viridisLite`. Download the package if you don't have it.

```
library(viridisLite)
cols <- viridis(3, begin = 0.2, end = 0.8)
boxplot(shift ~ treatment, data = df.k, col = cols)
stripchart(subset(df.k, subset = treatment == 'control')$shift, at = 1, add = T,
           vertical = T, pch = 19, method = 'jitter')
stripchart(subset(df.k, subset = treatment == 'eyes')$shift, at = 2, add = T,
           vertical = T, pch = 19, method = 'jitter')
stripchart(subset(df.k, subset = treatment == 'knee')$shift, at = 3, add = T,
           vertical = T, pch = 19, method = 'jitter')
```



Looks like pretty good evidence that the groups are different. But let's run a test to find out.

The first step in doing ANOVA is to make a linear model. A linear model is a general way to do a large class of statistical model fitting and it fits important parameters necessary to calculate sums of squares, mean squares, and F statistics. Let's fit the model.

```
m1 <- lm(shift~treatment, data=df.k)
```

This reads as a model where `shift` is a function of `treatment`, meaning that we want to know how shift differs for different values of treatment. Because `treatment` is a column with group names as factors, this asks whether how the average values of `shift` change for each `treatment` value.

Once you've got a linear model, we can ask for a one-way ANOVA using either `summary(aov())` or `Anova()`. They both do the same thing in this context. Try both, but to do use `Anova()` you will need the package `car`. Install it and load it.

```
library(car)
```

```
## Loading required package: carData
```

```
aov(m1)
```

```
## Call:
##    aov(formula = m1)
##
## Terms:
##                   treatment Residuals
## Sum of Squares   7.224492  9.415345
## Deg. of Freedom         2        19
##
## Residual standard error: 0.7039492
## Estimated effects may be unbalanced
```

```
summary(aov(m1))
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## treatment    2  7.224   3.612   7.289 0.00447 **
```

```
## Residuals   19  9.415   0.496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(m1)
```

```
## Anova Table (Type II tests)
##
## Response: shift
##           Sum Sq Df F value   Pr(>F)
## treatment 7.2245  2  7.2894 0.004472 **
## Residuals 9.4153 19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These each contain the most important parts of ANOVA: sums of squares and degrees of freedom.

Checkpoint 2: What is the estimate of variation that is attributable to the groups and what is the estimate of variation that is attributable to something other than groups?

The F-distribution

This analysis estimates sums of squares and means squares for you to compute an F value. It does what is listed in the equations above. In the end, this information about partitioning of variance is summarized with an F value, which is the test statistic for an ANOVA. Like all hypothesis testing, a test-statistic needs to be compared against a distribution of the test statistic under the null hypothesis. The null hypothesis here is that the groups all have the same mean, and as such, the variance between groups should be small compared with the variance within groups. Stated differently, MS_{group} should be smaller than MS_{error} and so F should typically be less than 1.

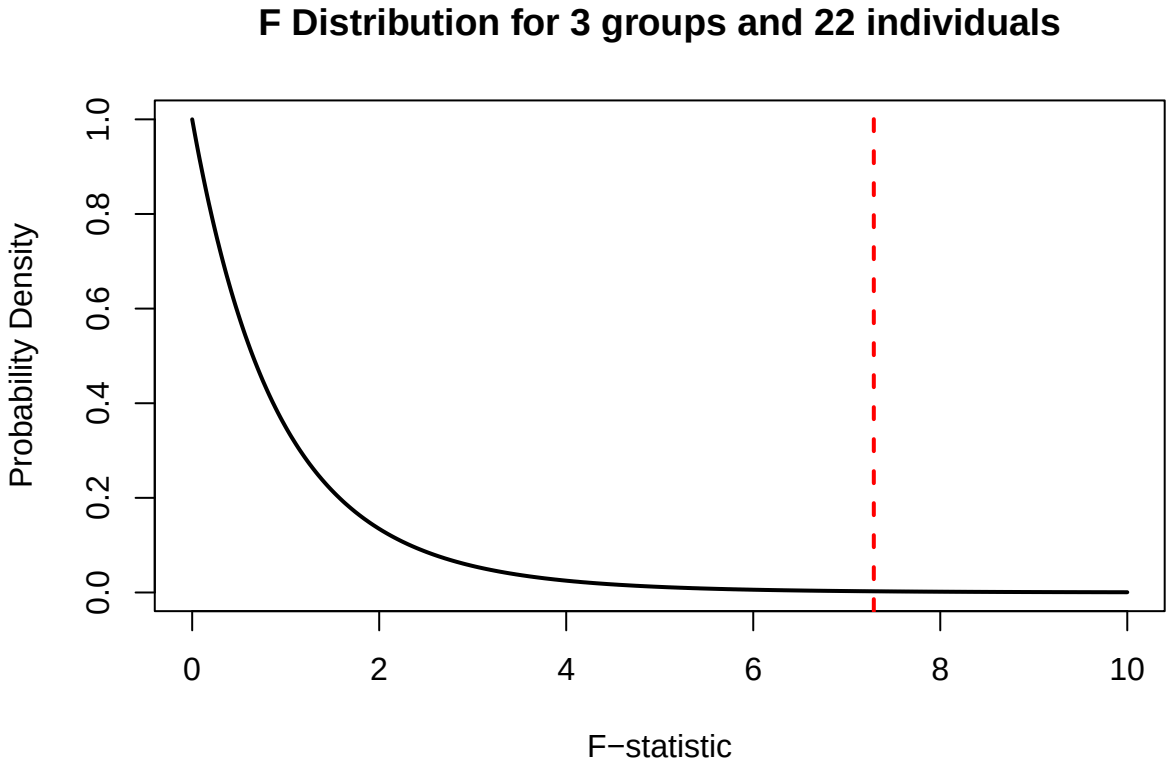
The probability distribution for the F statistic under the null is called, well, an F distribution. This distribution has two parameters: group degrees of freedom and error degrees of freedom.

These two parameters are estimated by ANOVA and are listed in the ANOVA table. Here, the "treatment" df is 2 because we had three groups and so df_{group} = number of groups $- 1 = 3 - 1 = 2$. The "error" degrees of freedom is 19, which comes from the fact that error degrees of freedom are the number of data points - the number of groups = 22 - 3 = 19.

We can make a null distribution using the function `df`, which gives the distribution for the F-distribution. We'll also put our calculated F-value on the plot.

```
F.values <- seq(from = 0, to = 10, length = 1000)
F.prob <- df(F.values, df1 = 2, df2 = 19)
plot(F.values, F.prob, xlab = 'F-statistic', ylab = 'Probability Density',
     main = 'F Distribution for 3 groups and 22 individuals',
     type = 'l', lwd = 2)

#Store ANOVA results
m1.ANOVA <- Anova(m1)
# Pull out and plot the F-value
abline(v = m1.ANOVA$'F value', lty = 2, col = 'red', lwd = 2)
```



You can see this outcome is particularly unlikely under the null hypothesis that the groups are the same. Since larger values of F constitute more extreme differences from the null hypothesis, the p-value in this case is just the probability that F is larger than the F-value we calculated. That is,

$$Pr(F > 7.289)$$

This is, of course, 1 minus the probability of everything to the left, which we can calculate using the cdf of the F distribution. Let's do that.

```
(p.value <- 1-pf(m1.ANOVA$'F value', df1 = 2, df2 = 19))
```

```
## [1] 0.004472271          NA
```

That matches the p-value from our ANOVA table exactly.

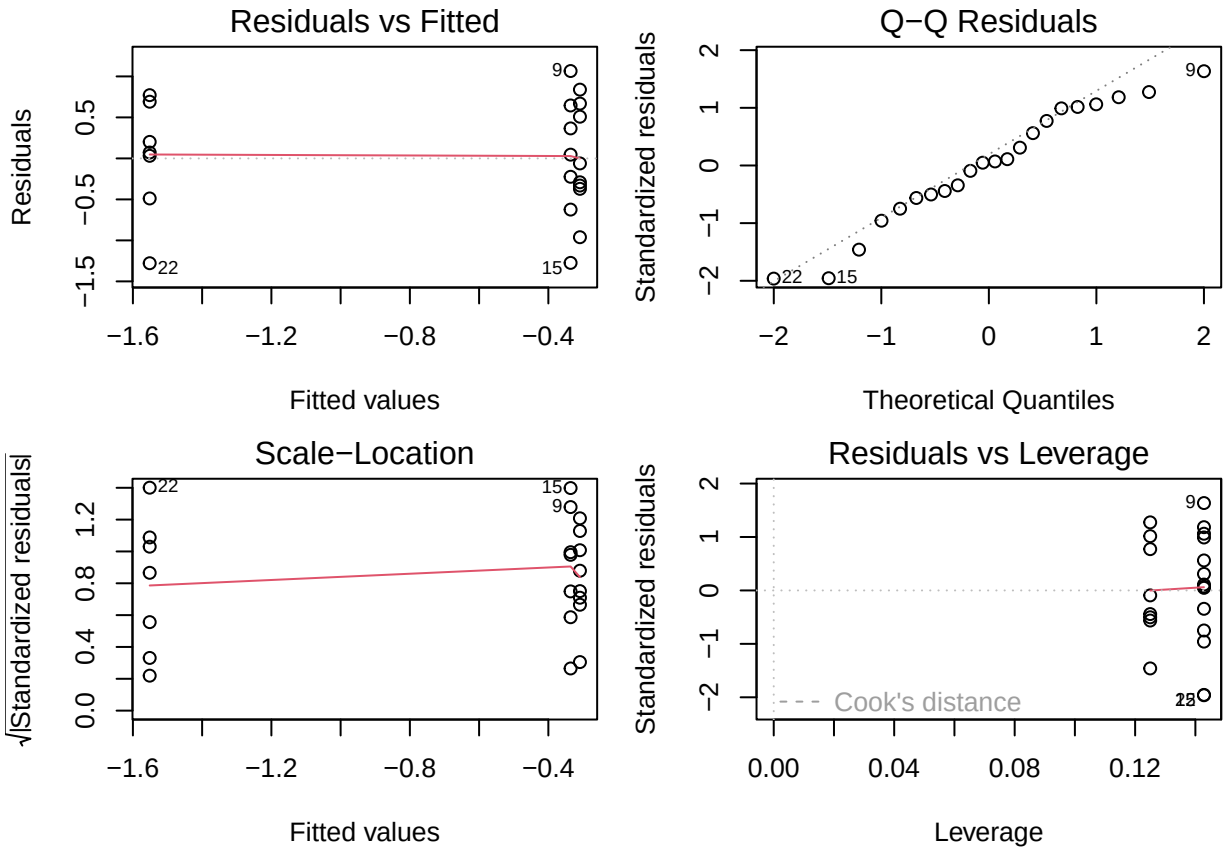
The next step in doing an ANOVA is to check your assumptions: does your model fit the data?

ANOVA makes a few technical assumptions. They can be summarized as that the residuals are normally distributed with mean zero and with equal variance. Equal variance is typically termed "**homoskedasticity**". If you don't have equal variances, then the groups are said to be "**heteroskedastic**".

So you need to do all the typical checks for normality that we have done before, but now you need to do them on the ***residuals***, which are the differences of each data point from the group average.

Luckily, R has an easy way to see these things. We will make 4 plots to check this.

```
par(mfrow=c(2,2), mar=c(4,4,2,1))
plot(m1)
```



Here are the four graphs:

- ***Residuals vs. Fitted***
  - **Residuals** are the distance of each data point from the fitted model. For ANOVA, the fitted data points don't matter much, so just focus on the residuals, which are the y-values.
  - The residuals plot removes any group differences and looks at the distribution of data within groups. ANOVA assumes normal residuals. If the data are normal, the residuals should fall equally on both sides of the fitted value (i.e., are symmetric). Visually, the red line should have its intercept near Residuals = 0 and have no slope or curvature.
- ***Normal Q-Q***
  - A Q-Q plot takes your sample data, sorts it in ascending order, and then plots that against the quantiles of a theoretical distribution. The points should lie along a straight line, not be banana-shaped or S-shaped. Note that we are assuming a normal distribution but a Q-Q plot can be made for any distribution using the *qqplot()* command.
- ***Scale-Location***
  - This is like a positive-value version of the 1st graph, and is good for detecting non-constancy of variance (heteroscedasticity), which shows up as a triangular scatter. The red line should be flat as well.
- ***Residuals vs Leverage***
  - Leverage measures the extent to which a point is highly influential on the parameter estimates of the model. The red line should be flat. Points outside the dotted red line (Cook's distance) are particularly influential. Ask yourself whether these points might be from a coding error or are real. If real, ask whether you inferences about the entire dataset rely on this (or these) individual data point(s). If so, try to figure out what that means relative to your question. Is there something special about this point? Was it collected in a different location or time? These questions can often provide insight into meaningful variation in your dataset.
  - One way to deal with datasets that have high leverage points is to run what is called a "Jackknife"

analysis in which you sequentially leave out every point in the dataset, rerun your analysis, and record the entity of interest (p-value, effect size, presence of an interaction). You then look at the distribution of this entity across all points being left out and determine whether your inferences are contingent on one (or a few) data points. If so, you should be skeptical about the robustness of the results. We won't do jackknife analysis (or any other 'leave one out' analysis) in this class, but you should be aware of their existence.

Checkpoint 3: Does our model match the assumptions well?

Consequences of violating assumptions

Unequal variance (heteroskedasticity) can be problematic in ANOVA. Heteroskedasticity alters the assumptions underlying the F-test and may cause the P value to be over- or underestimated. Most researchers cope with heterogeneous variances through transformations, commonly a logarithmic or root transformation for residuals that funnel out or a reciprocal transformation for residuals that funnel in. Importantly, moderate violation of homoskedasticity can be ignored in balanced ANOVA designs (those with equal numbers of replicates for each treatment), because the bias in the P value is small.

Failure to meet the normality assumption is usually of minimal concern in ANOVA, unless the errors are highly skewed. The F-tests used in ANOVA tend to be robust to non-normal errors, except when an experiment is highly unbalanced, although power may be reduced by non-normality. Moreover, parameter estimates from regression analyses are robust to non-normality, except when the non-normality is due to outliers.

Solutions to violation of assumptions

If your data violate assumptions of ANOVA, you have three viable options.

1. Transform the data. Try a few, fit the model, and look at the diagnostic plots again. If the Q-Q plots, the residual plots and scale-location plots look good, then you are ok. If not, try another transformation.
2. Run a randomization test. A randomization test is like bootstrapping, but where you shuffle the group labels while keeping the other data the same. This randomizes the group labels and so mimics the null hypothesis that the groups have the same mean. You randomize, calculate an F-value, and then store this F-value. Do that 10,000 times to get your own distribution of F-values under the null. The F-value from your data can be compared with this distribution. The null distribution you bootstrapped includes any violation of assumptions and so you don't need to worry about them anymore.
3. Try a different method. If neither of those work, you may need to move to more complex statistics, including generalized linear modeling or Bayesian analysis.

Post-hoc tests After you have run your ANOVA, perhaps you find that you can reject the null hypothesis. But this just says at least one of the means is different from the others. Often you want to know which ones. For data with k groups, there are $k(k-1)/2$ possible pairwise comparisons between groups and multiple comparisons raise your chance of Type I error. You'll need a method to account inflated Type I error. There are many approaches, and still much debate in the scientific literature about the proper approach for different questions.

Here are two main approaches that are generally agreed work well:

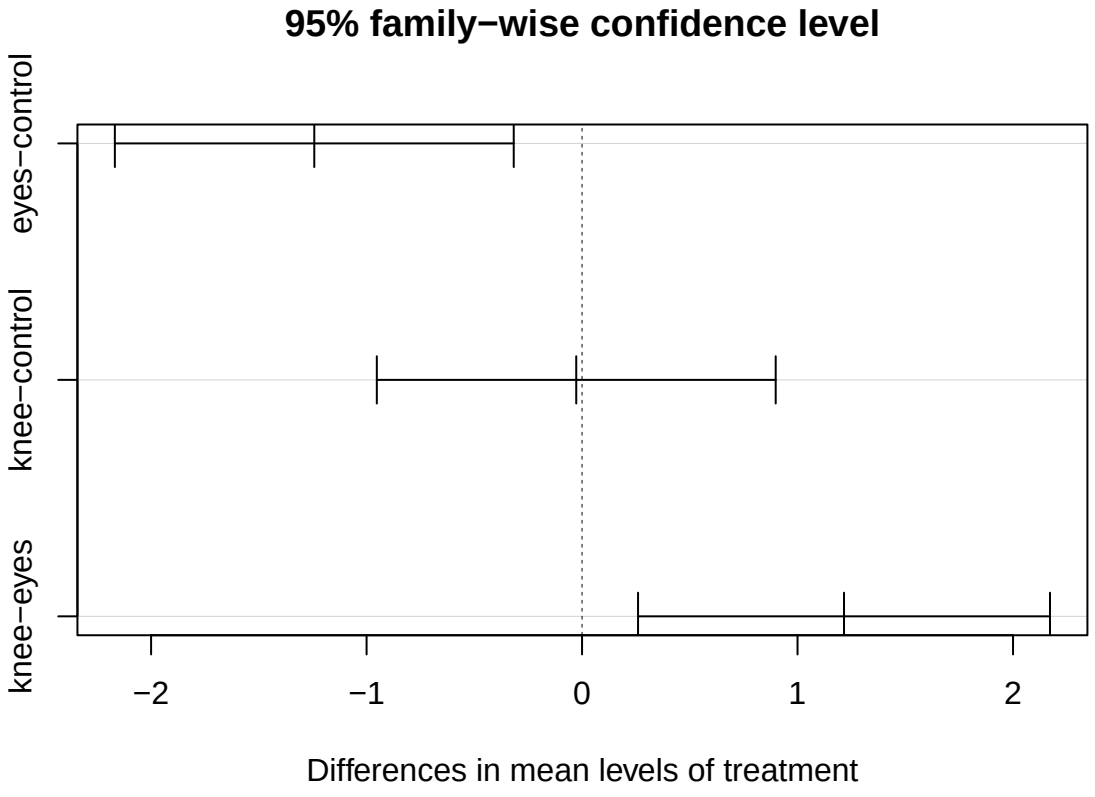
- A Tukey post-hoc test using `TukeyHSD` This method creates a set of confidence intervals on each factor level combination, and then compares them to each other. The intervals are based on the range of sample means rather than individual pairwise comparisons, and thus inherently corrects for multiple tests. This also has a nice plot option. I strongly encourage this because it provides confidence intervals.
- A Benjamini-Hotchberg correction using `pairwise.t.test` and the argument `p.adjust.method="BH"` for Benjamini-Hotchberg. This is one tests differences between groups and provides appropriate p-values to deal with multiple-testing problems.

We know that there is a significant effect of treatment, but which groups differ from the others? Is "eyes" different from both "control" and "knees", or just one of them? We need to do a post-hoc test to find out.

```
?TukeyHSD
### Unfortunately, TukeyHSD only works with the "aov" command:
  TukeyHSD(aov(m1))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = m1)
##
## $treatment
##                     diff        lwr        upr     p adj
## eyes-control -1.24267857 -2.1682364 -0.3171207 0.0078656
## knee-control -0.02696429 -0.9525222  0.8985936 0.9969851
## knee-eyes     1.21571429  0.2598022  2.1716263 0.0116776
```

```
  par(mfcol = c(1,1))
  plot(TukeyHSD(aov(m1)))
```




```
pairwise.t.test(df.k$shift, df.k$treatment, p.adjust.method="BH")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df.k$shift and df.k$treatment
##
##      control eyes
## eyes 0.0066  -
## knee 0.9418  0.0066
```

```
##
## P value adjustment method: BH
```

Checkpoint 3: Which groups are different from the others?

Checkpoint 4: Do the two methods give the same answers?

Checkpoint 5: What do you conclude from this study?

ANOVA with more complicated designs: Two-factor (or more) ANOVAs

Two factor ANOVAs include two different treatments that are given factorially. An example we discussed in class is one where herbivores and nutrients are manipulated and grassland productivity is measured. The factors here are the herbivores and nutrients. Factors have levels, which represent the number of manipulations of the factor.

For the example from class, each factor had two levels.

- Herbivores were either **removed** or left as a **control**
- Nutrients were either **added** or left as a **control**

When we looked at all possible combinations of these two factors at two levels, we had four combinations:

1. Herbivore **control**, nutrient **control**
2. Herbivore **control**, nutrient **addition**
3. Herbivore **removal**, nutrient **control**
4. Herbivore **removal**, nutrient **addition**

The goal with this design is to figure out the **average effect of each factor**, but also to identify **if there is an interaction between factors.**

These can be estimated with the notation

$$\text{Productivity} \sim \text{Herbivore} * \text{Nutrients}$$

The `*` notation says to find all main effects of each factor and their interaction. If you want to *only* look for main effects, you can use the notation

$$\text{Productivity} \sim \text{Herbivore} + \text{Nutrients}$$

The kinds of comparison you want are estimated with the function `lm()`. This specifies the kinds of contrasts you want to do. When you put that into `aov()` or `Anova()`, it estimates particular effects as specified by the notation.

ANOVA is great at estimating these effects and does so seamlessly when your design is balanced. Balanced designs happen when all groups have the same number of individuals.

Let's take a look at this with an example data set with snails.

This experiment measured snail growth (mm/day) as a response to temperature (two levels: 10C and 15C) and salinity (two levels: 20ppm and 30ppm). We are interested in the effects of these treatments (and their interaction) on snail growth.

Checkpoint 6: Load the data set "snailgrowth_bal.csv" The treatments are in temperature and salinity values. R will treat these as numbers unless we tell it different. Let's make them factors for use in ANOVA and take a look at a basic plot.

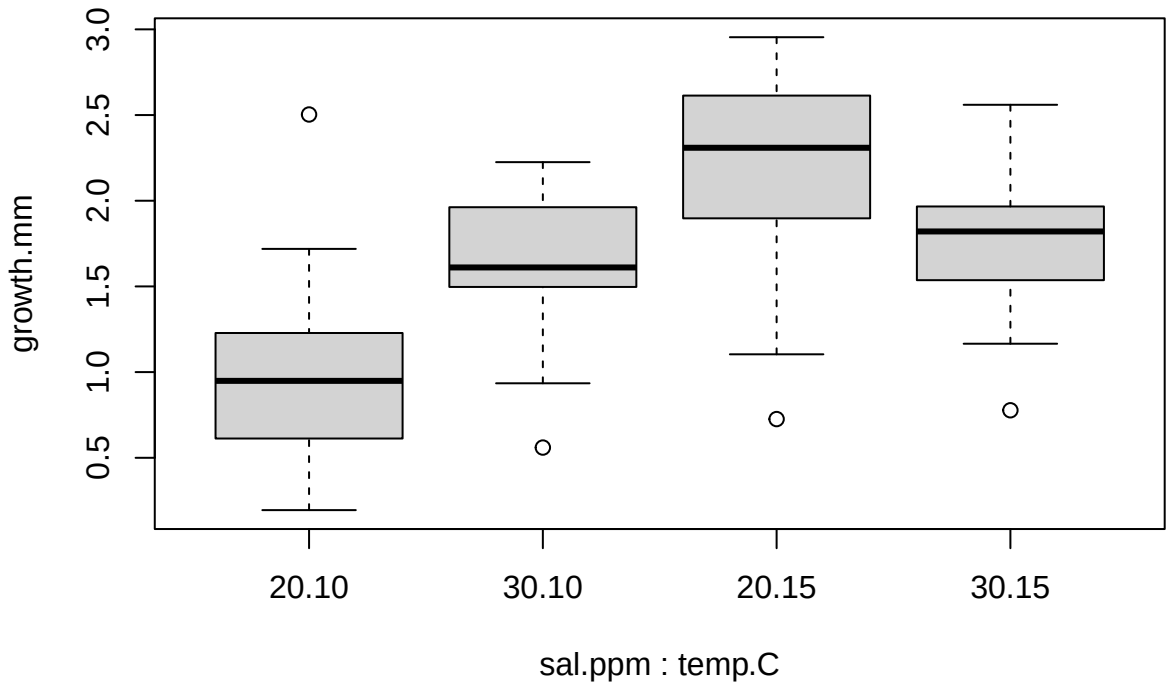
```
str(df.g)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ temp.C   : int  10 10 10 10 10 10 10 10 10 10 ...
##  $ sal.ppm  : int  20 20 20 20 20 20 20 20 20 20 ...
##  $ growth.mm: num  0.49 0.859 0.312 0.195 0.564 ...
```

```
df.g$temp.C <- as.factor(df.g$temp.C)
df.g$sal.ppm <- as.factor(df.g$sal.ppm)
str(df.g)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ temp.C   : Factor w/ 2 levels "10","15": 1 1 1 1 1 1 1 1 1 1 ...
##  $ sal.ppm  : Factor w/ 2 levels "20","30": 1 1 1 1 1 1 1 1 1 1 ...
##  $ growth.mm: num  0.49 0.859 0.312 0.195 0.564 ...
```

```
# Basic plot
par(mfrow=c(1,1))
boxplot(growth.mm~sal.ppm*temp.C, data=df.g)
```



This design is balanced. You can check it this way.

```
tapply(df.g$growth.mm, list(df.g$sal.ppm, df.g$temp.C), length)
```

```
##    10 15
## 20 15 15
## 30 15 15
```

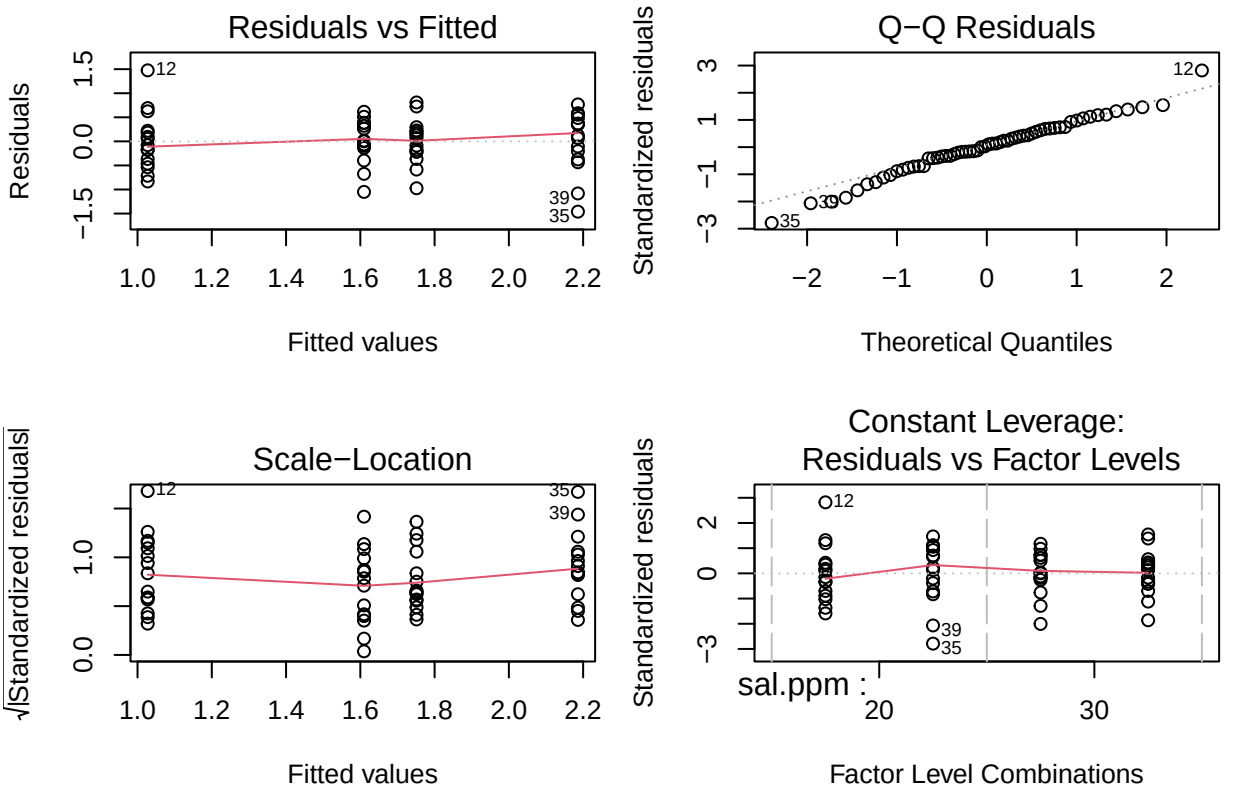
This means there are 15 individuals in each treatment.

Now we fit a linear model specifying that we want to look at main effects of each factor and their interaction.

```
# Make a lm
  m2 <- lm(growth.mm~sal.ppm*temp.C, data=df.g)
```

We should check the assumptions of normal residuals and homoskedastic residuals among groups.

```
# Check assumptions
  par(mfrow=c(2,2), mar=c(4,4,4,1))
  plot(m2)
```



Looks pretty good. Now we are ready to run the ANOVA and ask for estimates of main effects of temperature and salinity as well as a potential interaction.

```
# Do 2-WAY ANOVA
# One way
summary(aov(m2))
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## sal.ppm         1  0.082   0.082    0.28 0.598957
## temp.C          1  6.339   6.339   21.59 2.09e-05 ***
## sal.ppm:temp.C  1  3.879   3.879   13.21 0.000605 ***
## Residuals      56 16.441   0.294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Another way
Anova(m2)
```

```
## Anova Table (Type II tests)
##
## Response: growth.mm
##                 Sum Sq Df F value    Pr(>F)
```

```
## sal.ppm         0.0821  1  0.2797 0.5989575
## temp.C          6.3385  1 21.5904 2.092e-05 ***
## sal.ppm:temp.C  3.8794  1 13.2142 0.0006051 ***
## Residuals      16.4405 56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You can see here that there are now F values for each factor as well as their interaction. These are F values for specific contrasts.

Main Effects

This model shows a small F-value for salinity in the row `sal.ppm`, suggesting weak evidence against the null hypothesis of no average effect of salinity.

This model also shows a very large F-value for temperature in the row `temp.C`, suggesting evidence highly inconsistent with the hypothesis that temperature has no effect on snail growth.

Interaction Effects

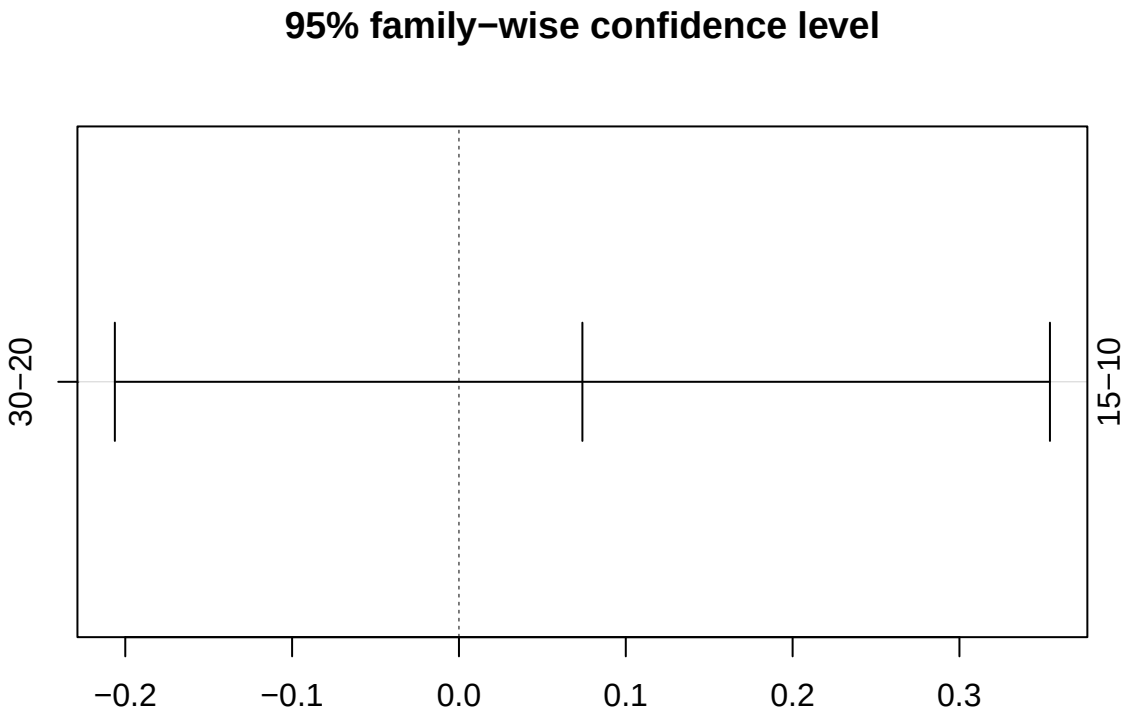
The model also estimates an interaction between temperature and salinity with a pretty large F-value (in the row `sal.ppm:temp.C`. This suggests that the effect of temperature on snail growth depends on salinity.

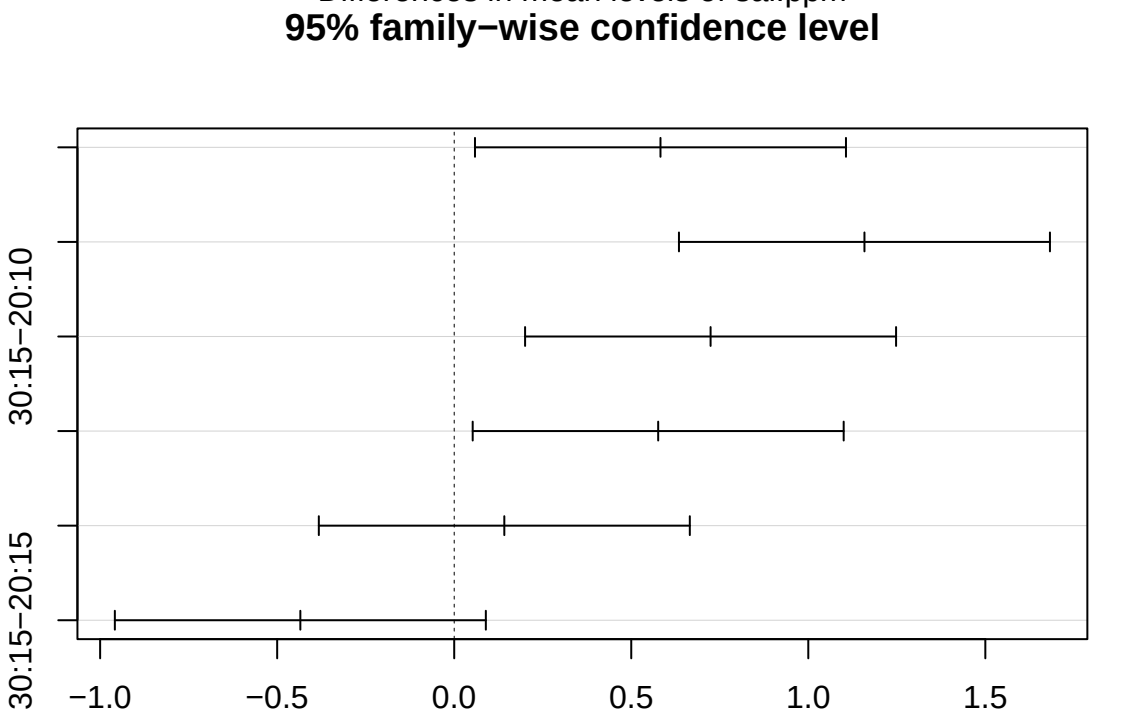
If you want to see the magnitude of these effects, make a plot and do post-hoc tests.

```
### Post-hoc comparisons
  TukeyHSD(aov(m2))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = m2)
##
## $sal.ppm
##             diff       lwr       upr     p adj
## 30-20 0.07399476 -0.206259 0.3542486 0.5989575
##
## $temp.C
##            diff       lwr      upr    p adj
## 15-10 0.6500522 0.3697984 0.930306 2.09e-05
##
## $`sal.ppm:temp.C`
##                    diff         lwr        upr     p adj
## 30:10-20:10   0.5825499  0.05866869 1.10643120 0.0236428
## 20:15-20:10   1.1586074  0.63472611 1.68248862 0.0000015
## 30:15-20:10   0.7240469  0.20016569 1.24792820 0.0030625
## 20:15-30:10   0.5760574  0.05217617 1.09993868 0.0257665
## 30:15-30:10   0.1414970 -0.38238425 0.66537826 0.8906609
## 30:15-20:15 -0.4345604 -0.95844168 0.08932083 0.1367726
```

```
  plot(TukeyHSD(aov(m2)))
```






```
group <- paste( df.g$sal.ppm,df.g$temp.C, sep=".")
pairwise.t.test(df.g$growth.mm, group, p.adjust.method = "bonferroni")
```

```
##
```

```
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df.g$growth.mm and group
##
##       20.10   20.15  30.10
## 20.15 1.6e-06 -      -
## 30.10 0.0282  0.0309 -
## 30.15 0.0034  0.1933 1.0000
##
## P value adjustment method: bonferroni
```

```
tapply(df.g$growth.mm, group, mean)
```

```
##    20.10    20.15    30.10    30.15
## 1.027380 2.185987 1.609930 1.751427
```

```
#    20.10    20.15    30.10    30.15
# 1.132264 2.150755 1.648887 1.867930
#    a         b       c         c
# Conclude:   At 10C, growth increased with salinity, but at 15C, growth decreased with salinity. At 30
```

Checkpoint 7: What are the conclusions of the snail growth study?

A complication: More than one way to partition variances with unbalanced designs

However, for unbalanced designs with 2 or more factors, the answer depends on how the SS are calculated. There are three ways, illustrated here for the model

$$y \sim A * B$$

- Type I
  - **the SS depends on whether A or B is first in the model equation**. Assuming A is first:
  - calculate SS for factor A first: SS(A)
  - calculate SS for factor B, while controlling for the main effect of factor A: SS(B | A)
  - test SS for interaction A:B, while controlling for the main effects of factors and and B: SS(AB | B, A)
  - implemented in `aov()` or `anova()`
- Type II
  - **order doesn't matter whether A or B is first in the model equation**
  - calculate SS for factor A, while controlling for the effect of factor B: SS(A | B)
  - calculate SS for factor B, while controlling for the effect of factor A: SS(B | A)
  - This **assumes there is no significant interaction**. What should you do before using Type II SS?
  - implemented `Anova(type=2)` from the `car` package
- Type III
  - **order doesn't matter whether A or B is first in the model**
  - calculate SS for factor A, while controlling for the effect of factor B and the interaction: SS(A | B, A:B)
  - calculate SS for factor B, while controlling for the effect of factor A and the interaction: SS(B | A, A:B)
  - implemented `Anova(type=3)` from the `car` package. Note that you need to **specify contrasts** before using this.

Recommended approach to a 2-way, unbalanced ANOVA

- Test for an interaction first
- When there is no interaction, Type 2 is a more powerful test than Type 3 sum of squares
- When there is an interaction, Type 3 is a valid approach. BUT, it is often not interesting to interpret a main effect if interactions are present (generally speaking, if a significant interaction is present, the main effects should not be further analysed). In that case, make your analysis and interpretations based on *figures of the data* rather than the coefficients themselves.