

Agenda for Week 4 (Tuesday, Jan 26)

Week 4 Hour 1

AnOVA review.

Week 4 Hour 2

Multiple Testing

Tukey's HSD (Honestly Significant Difference).

Week 4 Hour 3 (Thursday)

Two-way AnOVA.

AnOVa Review

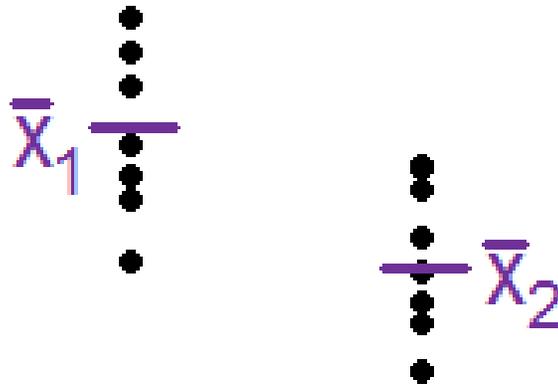
AnOVa stands for Analysis Of Variance.

Sometimes it's called ANOVA, and sometimes AOV.

ANOVA is the natural extension of the two-sample t-test.

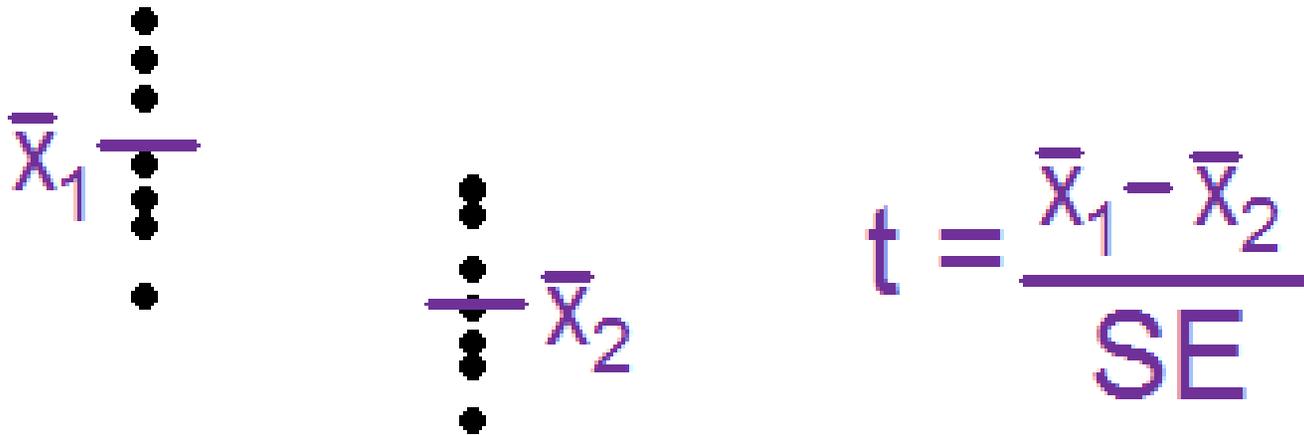
Two-sample T: Are the means of these two groups the same?

ANOVA: Are the means of ALL of these groups the same?



With the two-sample t-test, we did this by taking a sample mean from each group and comparing the difference to the ***standard error of the difference.***

Our method of testing whether this difference was significant (in other words, testing the null hypothesis that the difference between the true means was **zero**) was the t-score.


$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$$

It was always $t = (\text{difference}) / \text{Standard Error}$.

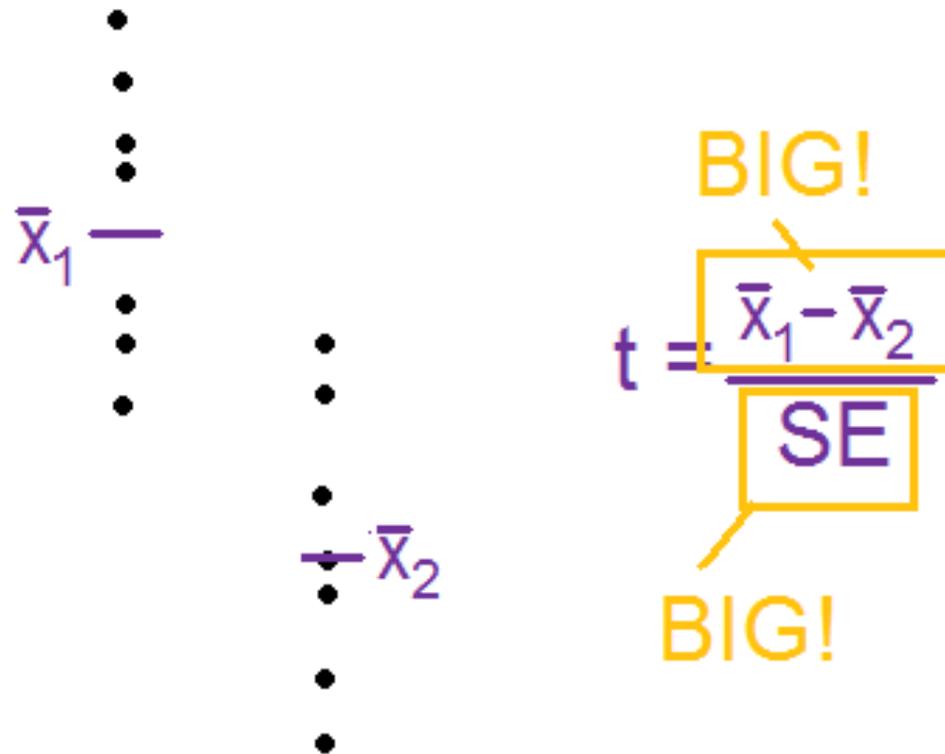
The definition of standard error depended on the details (paired/independent, pooled/non-pooled standard deviation)

If the difference was bigger, the t-score was bigger and we more often rejected the null hypothesis.

The diagram illustrates the components of a t-score. On the left, there are two vertical columns of dots representing data points. The top column is labeled \bar{X}_1 and the bottom column is labeled \bar{X}_2 . To the right, the t-score formula is shown: $t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$. The numerator, $\bar{X}_1 - \bar{X}_2$, is enclosed in a yellow box, and the word "BIG!" is written in yellow above it with an arrow pointing to the box. The denominator, SE, is written in purple below the box.

It's easier to say a difference is real when the sample mean difference is larger. (Easier to detect larger effects)

If there was more scatter between the points within a group, the standard error got bigger, and we more often failed to reject the null hypothesis.



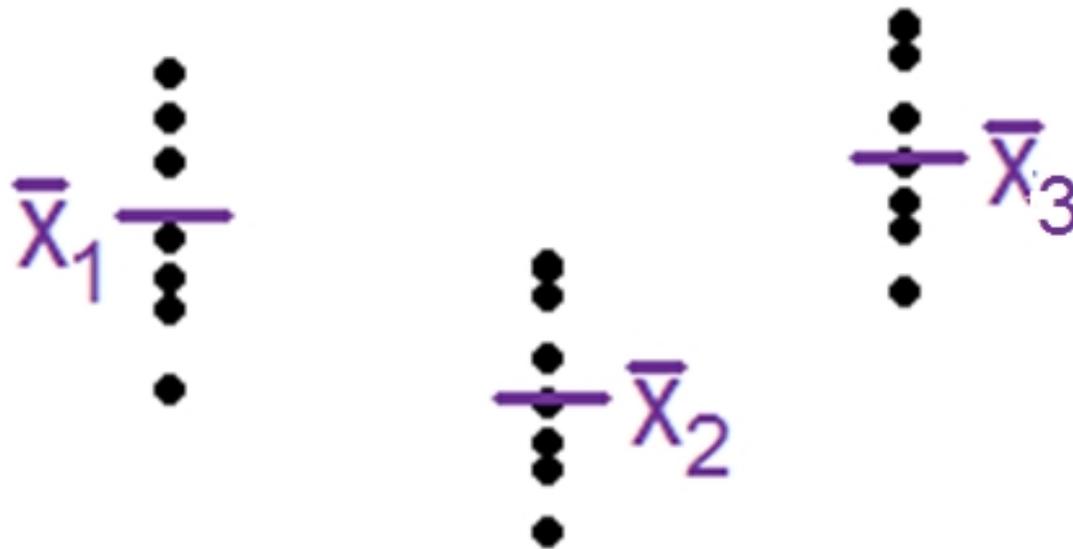
Standard error also gets smaller when there are more data points.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

In every case that the t-test is used, you're ultimately just answering one question over and over again:

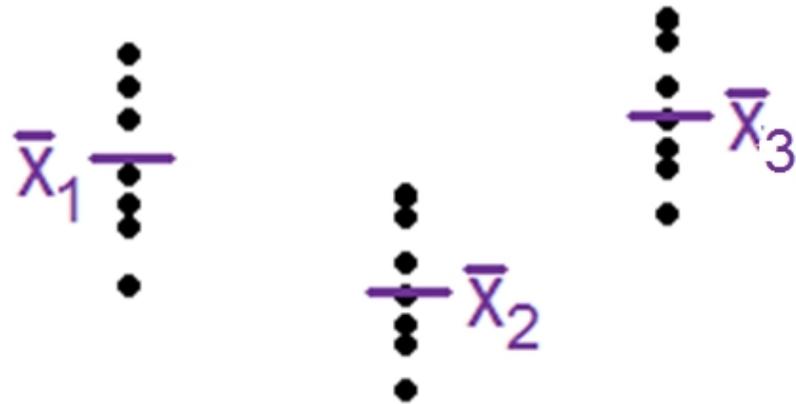
Are the differences ***between*** the two groups large compared to the differences ***within*** each group?

Can we use the t-test to determine if there are differences between any of the three means from three samples?



We can't do this all as a single t-test, because the t-test is only a comparison between **two** sample means. We have three

We could test each pair of groups and look for differences.



If we found a significant difference between two means, that would imply that not all the means are the same.

We'd need to test:

Mean of group 1 vs group 2

Mean of group 2 vs group 3

Mean of group 1 vs group 3

Doing multiple t-tests takes time, and what's worse: It opens up the issue of multiple testing (the more tests you do, more likely you are to commit an error like falsely rejecting the null)

A much cleaner solution is the *F-test* of ANOVA.

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$



Please axolotl questions if you are confused.

$$\bar{X}_1 \begin{matrix} \bullet \\ \bullet \end{matrix}$$

MS stands for ***mean square***, and MS_{within} is the average squared difference from a data point to the average for the that group. It's the mean squared WITHIN a group

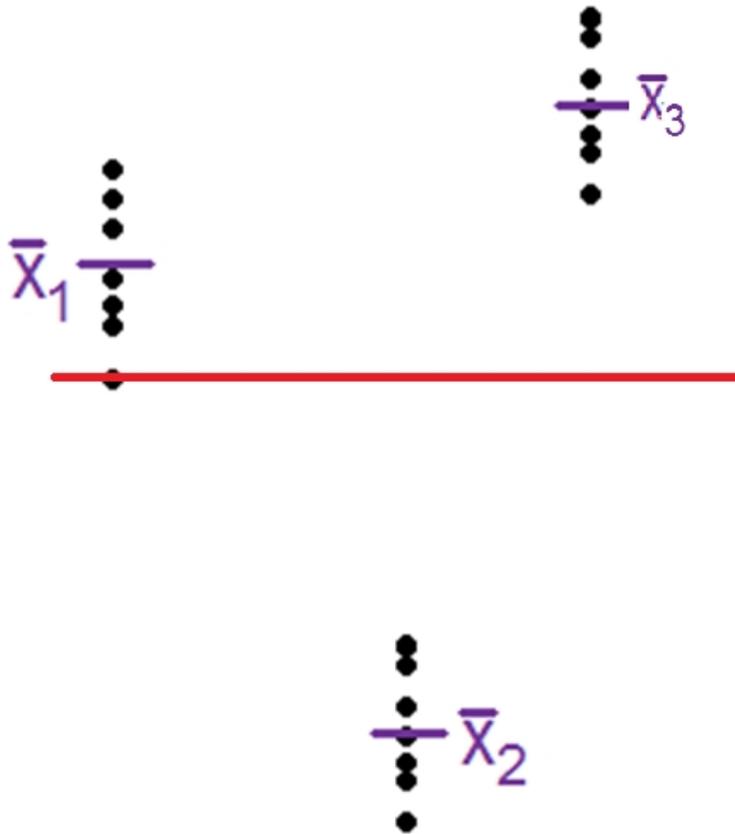
If we were just looking at a single group, this average squared distance would be the standard deviation squared, or the ***variance***.

MS_{within} is large when the spread within the samples is large.

$$F = \frac{MS_{\text{between}}^{\text{small}}}{MS_{\text{within}}^{\text{large}}}$$

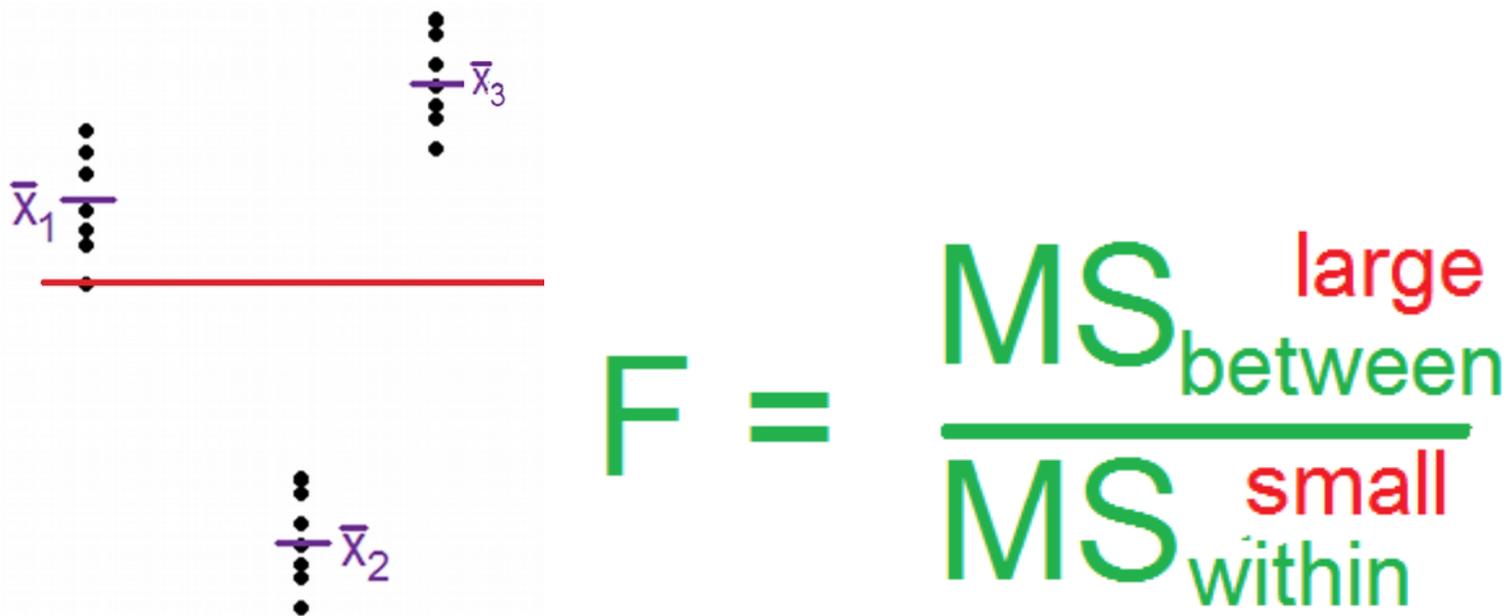
Spread/variance within a sample makes it hard to detect differences between the samples, and so the F-statistics gets smaller, just like the t-statistic.

MS_{between} is large when there are large differences between the sample means. MS_{between} stands for the differences between means, instead of within them.



$$F = \frac{MS_{\text{between}}^{\text{large}}}{MS_{\text{within}}^{\text{small}}}$$

Here, the average (squared) difference from a group mean to the **grand mean**, the average of data points from all the groups put together, is much larger than the differences between each point and its group mean.



F will be large and there is strong evidence that there is some difference in the true means between the groups.

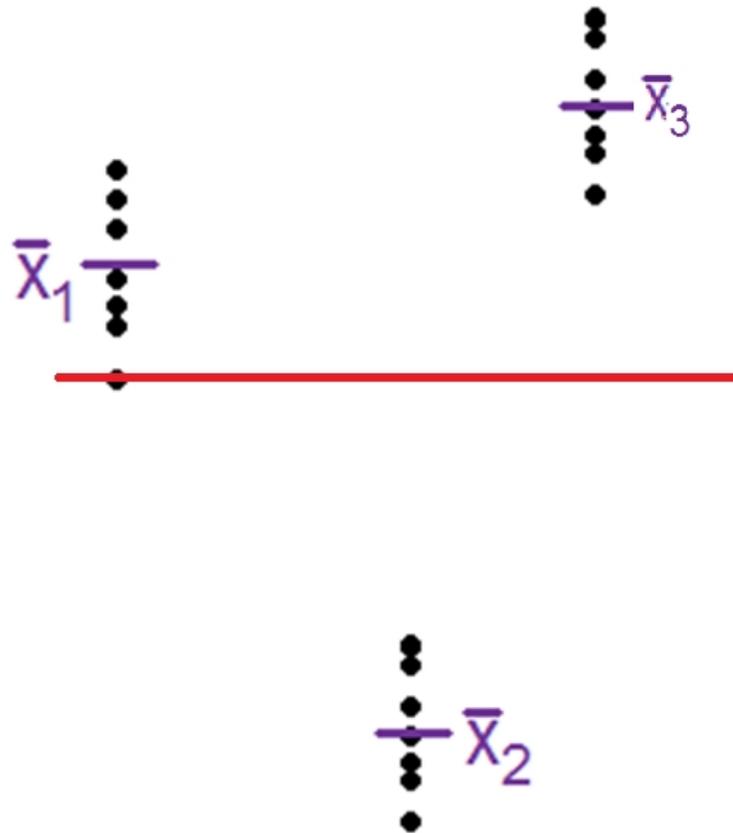


ANOVA is a big-picture tool.

AnOVA is a set of statistical methods designed to answer one question “Where is the variance coming from?”

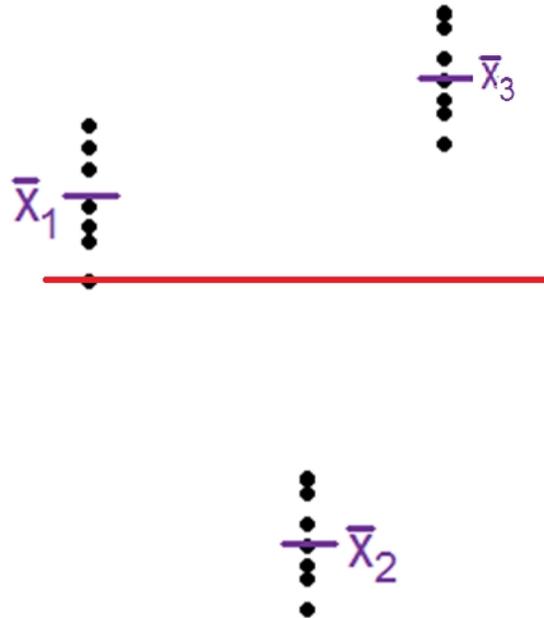
A less formal way to ask this question is:

“Why are the data values from my sample different? How can I explain these differences?”



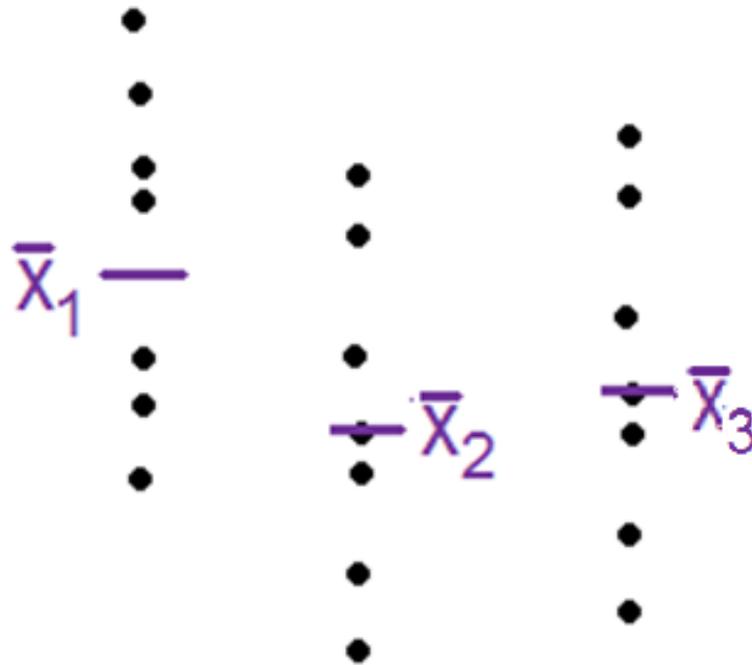
Sometimes the values are different because they come from groups that have different true means.

Doing an ANOVA will tell us that the variation is due to the different group means in this case.



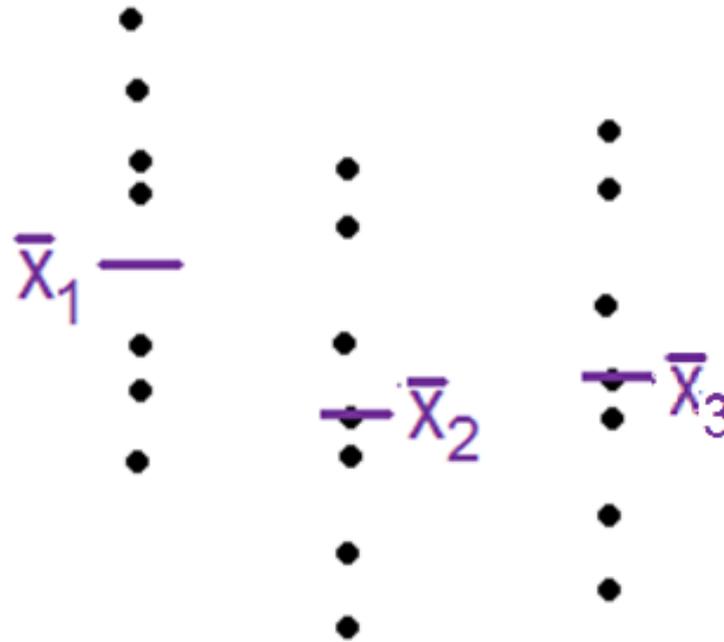
ANOVA can tell us how much evidence there is against there being no group differences. (This is the ***null hypothesis***)

Here, we would reject the null hypothesis because most of the variation can be explained by the differences ***between*** groups.



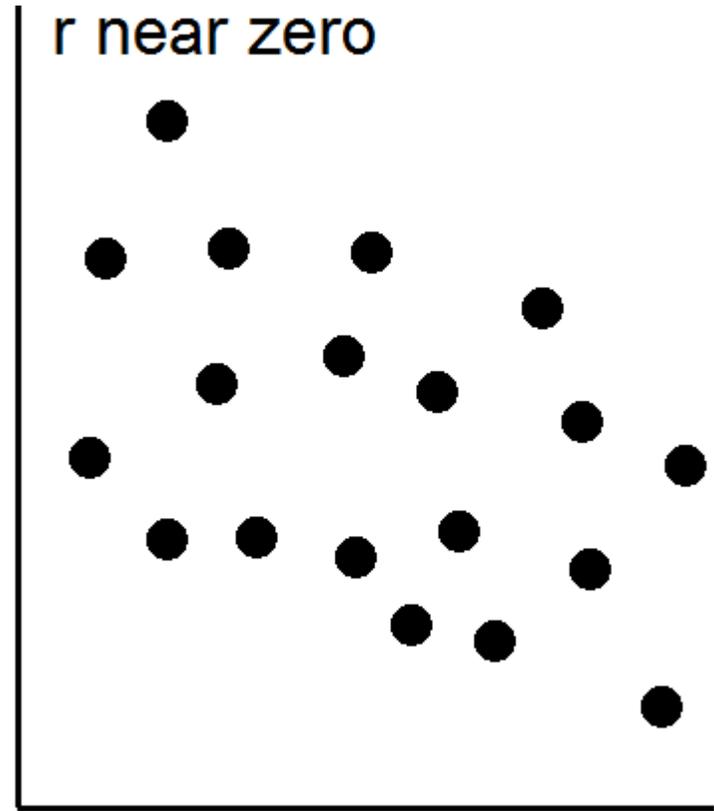
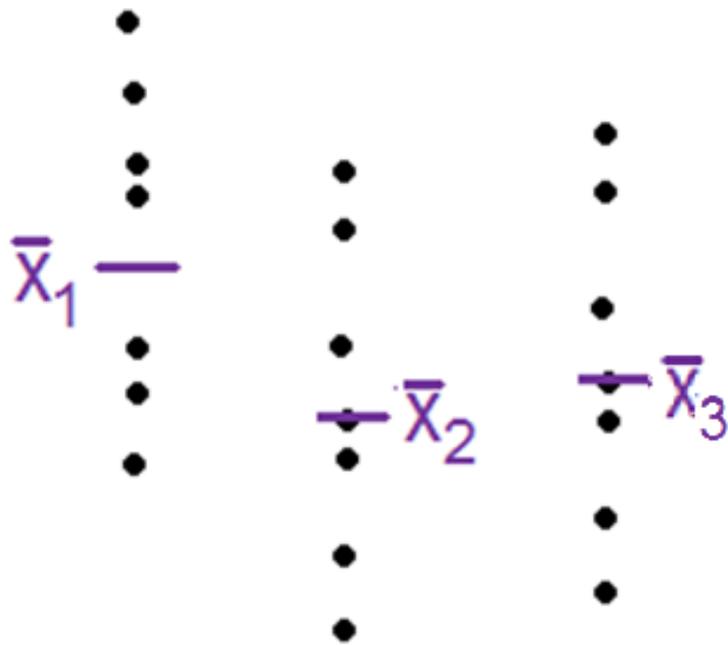
Sometimes the group means are not very different compared to the differences between values a group.

Here, doing an ANOVA will tell us that the variation is from random scatter.



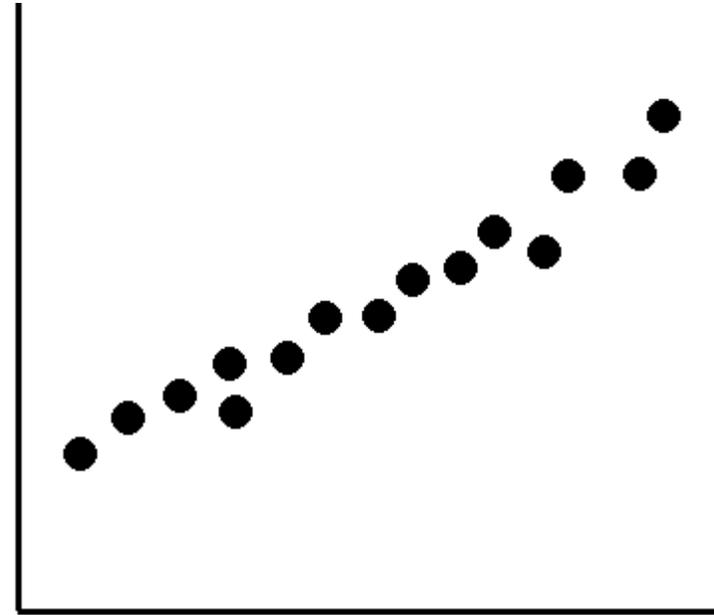
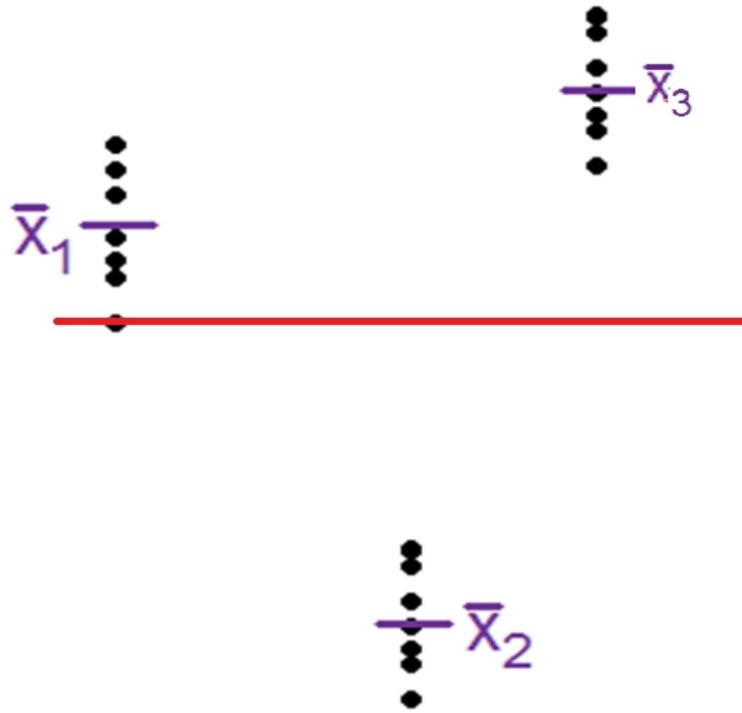
In other words, the groups won't explain very much of the variation in the response.

The group means are close enough we would the hypothesis that the true means were different.



Small differences between the group means is a lot like a weak correlation in its use:

The independent variable (nominal in ANOVA, and interval in correlation) doesn't explain much of the variation in the dependent variable (interval in both cases)



Large differences between group means are akin to a strong correlation.

Knowing the group will tell you a lot about the values to expect, just as knowing the independent X value tells you a lot about the Y values to expect.

If a correlation is significant, that means that our sample showed it to be far enough from zero to reject the hypothesis that the true correlation was zero.

It also means that at least some of the variance in Y is explained by X . (Because r -squared isn't zero)

The same is true for the ANOVA F-Test. If it yields a small p -value, that means the sample means are far enough to reject the hypothesis that the difference between true means is zero.

It also means that some of the variance is explained by groups.

In correlation, the closer values get to a straight line, the more variance is explained (r^2 gets closer to 1)

In ANOVA, the closer values get to their group means, the more variance is explained (again, proportion explained gets closer to 1)

Just as when X has nothing to do with Y in correlation/regression $r^2=0$, if the group has nothing to do with the measured values, none of the variance is explained.



Let's sandwich in an example before break.

Yarn Breakage Example

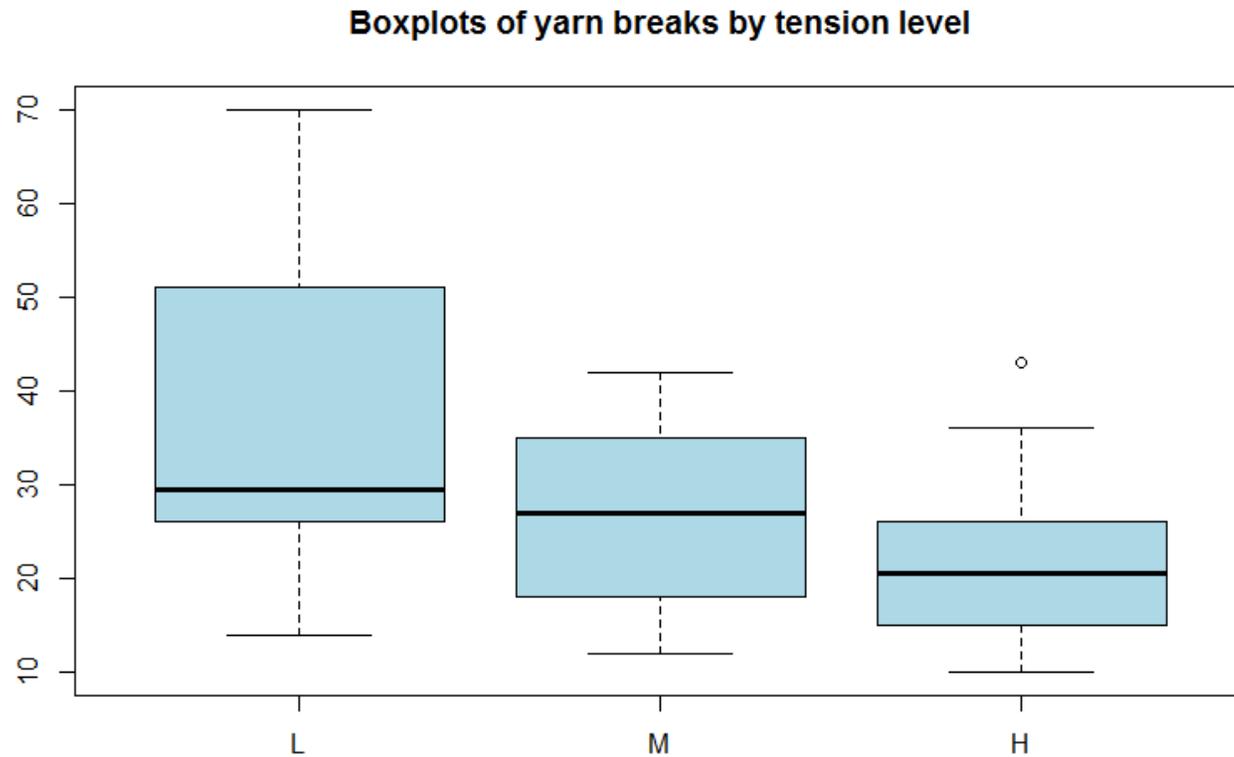
In R, there is a dataset called “warpbreaks”. It's also available as 'warpbreaks.csv'

The responses are counts of number of breaks in yarn (breaks) of two different kinds of wool (wool) under three different levels in tension (tension).

The number of breaks is the response (Y), and the levels of tension are the explanatory variable (X).

In ANOVA, the variance within each group is assumed to be the same. We can check this assumption informally with a side-by-side boxplot.

If a box is much larger than the others, ANOVA won't work as well.



More formally, we can do *Levene's* a **Bartlett** test on the data.

Levene's test is a hypothesis test of equal variance.

For three groups, this is:

$$H_0: \sigma_1 = \sigma_2 = \sigma_3$$

H_A : At least one pair of variances is unequal.

```
bartlett.test( response, grouping_variable)
```

```
> bartlett.test(warpbreaks$breaks, warpbreaks$tension)
```

```
      Bartlett test of homogeneity of variances
```

```
data:  warpbreaks$breaks and warpbreaks$tension
```

```
Bartlett's K-squared = 9.6571, df = 2, p-value = 0.007998
```

$p < 0.01$, so there is evidence of unequal variance.

Differences in variance between groups generally have to be VERY large to matter. A good rule of thumb is that the standard deviation of one group has to be at least 3 times as big as another.

That's why, even though the first group has a bigger standard deviation visually, **and the Bartlett test detected unequal variance, that we can proceed with the ANOVA as planned.**

With unequal variance, groups with larger variance 'count' for more in the ANOVA calculation.

After we have created a linear model, the `anova()` function gives a clean ***ANOVA table*** using the response and explanatory variances from the model.

```
> mod = lm(breaks ~ tension, data=warpbreaks)
> anova(mod)
Analysis of Variance Table

Response: breaks
          Df Sum Sq Mean Sq F value    Pr(>F)
tension    2  2034.3  1017.13   7.2061 0.001753 **
Residuals 51  7198.6   141.15
```

What does the ANOVA table mean? 1/6

tension					
Residuals					

“Tension” in this case is the explanatory variable, or ***grouping variable***.

“Residuals” represents the 'error' or 'within group' variation.

What does the ANOVA table mean? 2/6

	Df				
tension	2				
Residuals	51				

“Df” is *degrees of freedom*.

There are 3 groups, so $3 - 1 = 2$ degrees of freedom are used.

There are 54 observations, and three *parameters* (one for each group) so there are $54 - 3 = 51$ degree remaining for residuals.

What does the ANOVA table mean? 3/6

	Df	Sum Sq			
tension	2	2034			
Residuals	51	7199			

“Sum sq” stands for Sum of Squares.

It's akin to the $\sum (x_i - \bar{x})^2$ term in the variance and standard deviation formula.

“Sum Sq” is NOT a priority to know.

What does the ANOVA table mean? 4/6

	Df	Sum Sq	Mean Sq		
tension	2	2034	1017		
Residuals	51	7199	141		

“Mean sq” stands for Mean Squares.

It is computed by $(\text{Sum Sq}) / \text{df}$.

The residual Mean Squares is the amount of variance we would expect from randomness. (i.e. the variance ***within groups***)

The “tension” mean squares is the variance ***between groups***.

What does the ANOVA table mean? 5/6

	Df	Sum Sq	Mean Sq	F-value	
tension	2	2034	1017	7.20	
Residuals	51	7199	141		

We are comparing variances, so an ***F-test*** is appropriate.

F-value is the value from the f-test. This value means the variance of the group means is 7.20 times as large as the values within them.

It is computed by ***Mean sq. (group) / Mean sq. (Residuals)***

What does the ANOVA table mean? 6/6

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
tension	2	2034	1017	7.20	.001753
Residuals	51	7199	141		

Pr(>F) is the p-value of the F test.

A p-value of .001753 is strong evidence that the means are not all equal.

The most important parts are in yellow.

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
tension	2	2034	1017	7.20	.001753
Residuals	51	7199	141		

The p-value is especially important because it shows what you really want to know:

Are the group means different or not?

In this case, yes.



Don't rush ahead! Take a break first.