Hôpital général juif
Jewish General Hospital

McGill

INSTITUT LADY DAVIS DE RECHERCHES MÉDICALES / LADY DAVIS INSTITUTE FOR MEDICAL RESEARCH

LDI Seminar Series in Biostatistics: Lecture 4

Kevin McGregor

October 16, 2017

Last time

- Defined Pearson's correlation coefficient and how to interpret.
- Simple linear regression: how to fit
- Basic residual diagnostics
- Multiple regression interpretation

Today

- Today we'll focus on **An**alysis **o**f **Va**riance (ANOVA)
- Recall the independent-samples *t*-test:
  - Comparing mean of some continuous variable within two groups
  - Assume outcome variable follows a normal distribution within each group
- One-way ANOVA is the generalization of the *t*-test to two or more groups
  - One-way ANOVA with two groups is *exactly* equivalent to independent-samples *t*-test.

ANOVA as a hypothesis test

- A one-way ANOVA looks at the mean of a variable y within k different groups.
- Let the true population mean of y in each of the k groups be $\mu_1, \mu_2, \ldots, \mu_k$.
- ANOVA does not look at pairwise differences between means. Only gives an overall test of differences between means:

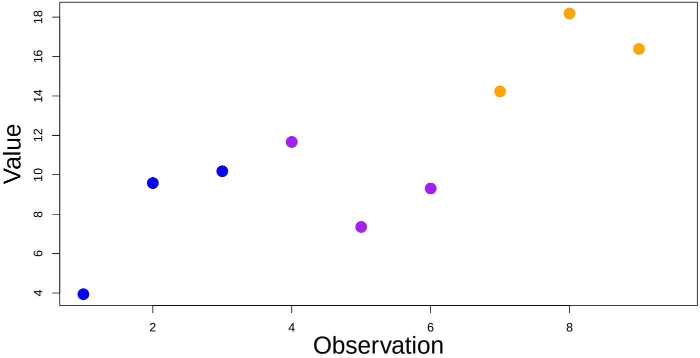
$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \text{At least one of the means is not equal to the rest} \end{aligned}$$

- If the null is rejected, we don't necessarily know which group(s) are significantly different from the rest.

ANOVA notation

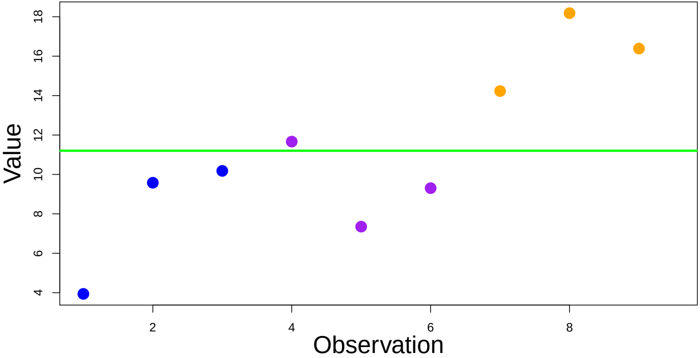
- Assume that the observed value for individual i in group j is y_{ij}.
- The number of subjects in group j is n_j.
  - i.e. The individuals in group 1 are $y_{11}, y_{21}, \ldots, y_{n_1 1}$
- The total number of observations (over all groups) is $n = \sum_{j=1}^{k} n_j$

Points from 3 groups



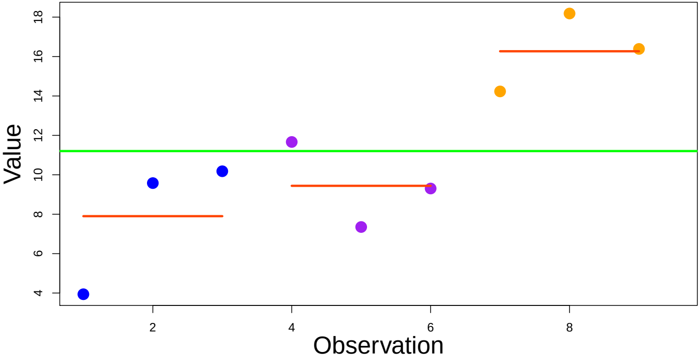
Assume we have n=9 observations from $k = 3$ different groups: blue, purple, and orange.

Grand mean



The mean of all the observations is called the grand mean. Call this $\overline{y}_{grand} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij}$.

Group means

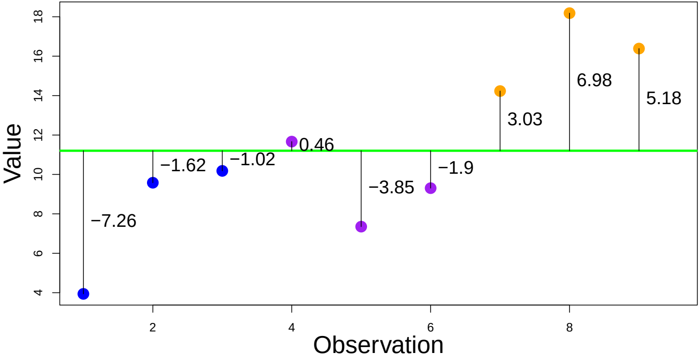


The mean for group j is calculated as $\overline{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$.

ANOVA notation

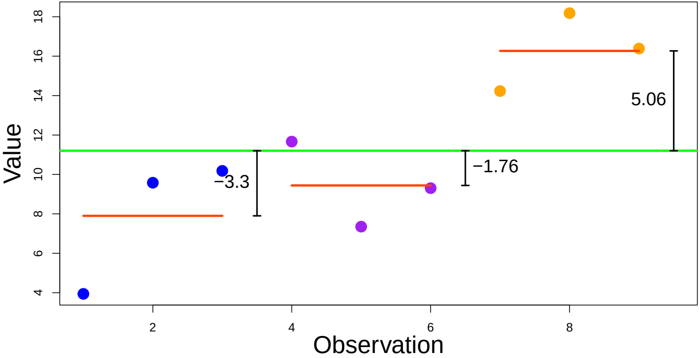
- We're testing if the group means are all equal.
- Like before, we need to consider the variability in the data to see if the differences are significant.
- In ANOVA there are different ways to characterize variability.
- As usual we look at the sum of the squares of differences of quantities. Each measure of variablility is called a certain type of "sum of squares".

Total sum of squares



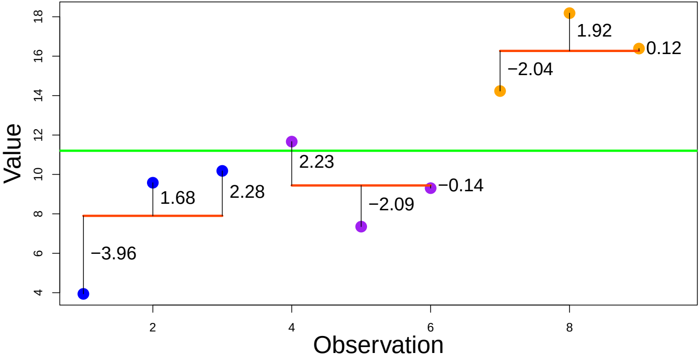
Take the difference between the points and the grand mean. Called the total sum of squares: $SST = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_{grand})^2$

Sum of squares between groups



Take the difference between the group means and the grand mean. Called the sum of squares between groups: $SSB = \sum_{j=1}^{k}(\overline{y}_j - \overline{y}_{grand})^2$

Sum of squares within groups



Take the difference between the individual values and the group means. Called the sum of squares within groups: $SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_j)^2$

ANOVA notation

- The total sum of squares can be written as $SST = SSB + SSW$
- Therefore, the total variation in the data is split up into two parts:
  - The sum of squares between groups (SSB) measures how different each group mean is from the grand mean. The larger this value is, the larger the differences between the group means.
  - The sum of squares within groups (SSW) measures the amount of variation from each data point to its corresponding group mean. The smaller this value, the more confident we are that the differences between the group means are significant.
- So we want to test whether the SSB is large relative to the SSW.

Mean square

- Before we can compare SSB and SSW, we need to make a correction based on the number of groups we're comparing.
- Have to consider the *degrees of freedom* for each quantity.
  - The degrees of freedom for the between group comparison is $df_B = k - 1$ (i.e. the number of groups minus one)
  - The degrees of freedom for the within group comparison is $df_W = n - k$ (i.e. the total number of subjects minus the number of groups)
- Can calculate the *mean* square terms:
  - $MSB = \frac{SSB}{df_B} = \frac{SSB}{k-1}$
  - $MSW = \frac{SSW}{df_W} = \frac{SSW}{n-k}$

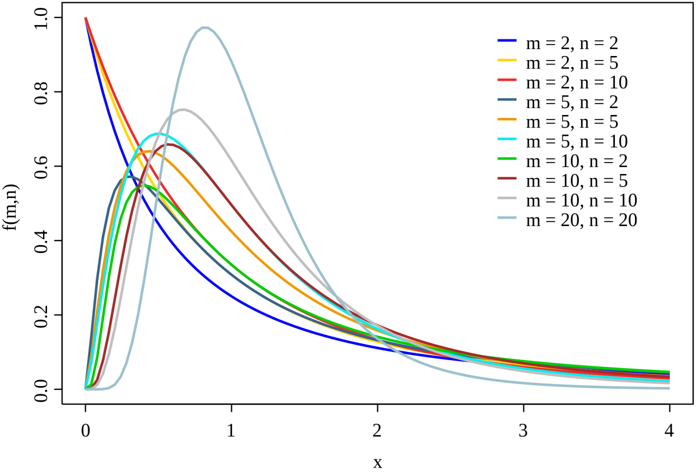
ANOVA F-test

- The test statistic for the one-way anova is then the ratio of the two mean square terms:

$$F = \frac{MSB}{MSW} = \frac{SSB/(k-1)}{SSW/(n-k)}$$

- This statistic is compared to the critical value from the F-distribution. This distribution depends on the degrees of freedom parameters df_B and df_W (sometimes called the *numerator* and *denominator* degrees of freedom).

F-distribution



https://commons.wikimedia.org/wiki/File:Dichte_F-Verteilung.svg

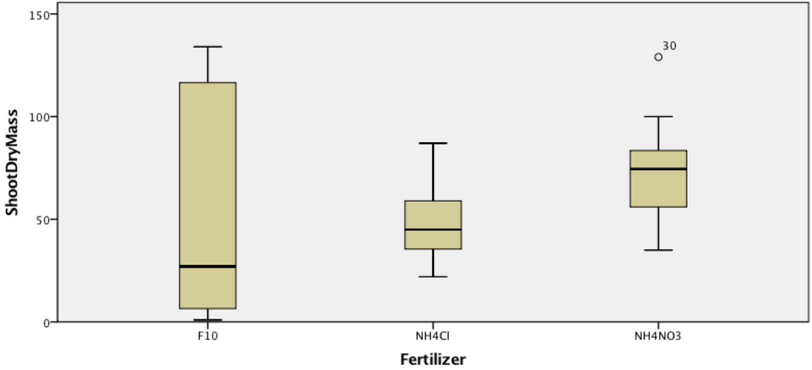
Significance in F-test

- Like the t-test, we compare the test statistic F to the upper tail of the appropriate F-distribution.
- For significance level $0 < \alpha < 1$ the critical value is: $F_{\alpha,df_B,df_W} = F_{\alpha,k-1,n-k}$. This can be obtained from a table or through software.
- Comparing the test statistic:
  - If $F > F_{\alpha,k-1,n-k}$, then we reject H_0... evidence that the group means are not all equal
  - Otherwise, do not reject H_0.

Assumptions of the ANOVA

- Values of the outcome variable are normally-distributed within each group.
- Equal variance between the groups.
- All observations are independent of one another.

ANOVA example



Data: a 2001 experiment (Perrine et al.) looking at the dry mass of rice shoots using three types of fertilizer treatments: (F10, NH4Cl, and NH4NO3). n=72, with 24 plants in each treatment group

ANOVA example

- Shot dry mass means in each group: F10=57.83, NH4Cl=48.42, NH4NO3=72.42
- $df_B = 3 - 1 = 2$, $df_W = 72 - 3 = 69$
- Common to summarize results in ANOVA table:

ANOVA

ShootDryMass

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7018.778	2	3509.389	2.820	.066
Within Groups	85869.000	69	1244.478		
Total	92887.778	71			

- In this case, $F_{0.05,2,69} = 3.13$. Since $F = 2.82 < 3.13$ we do **not** reject H_0 at level $\alpha = 0.05$. Not enough evidence to declare differences in means.

Post-hoc tests

- Remember that this is an *omnibus* test. It does not tell us which particular groups differ.
- If an ANOVA test is significant, we can then do *post-hoc* tests. These allow us to directly compare the individual means of the different groups.
- We have to be careful! This is called *multiple testing*.
  - Every time we add a new statistical test to our analysis, we slightly increase the probability of getting at least one false positive.
- Post-hoc tests are set up to minimize the chances of getting a false positive.
- Many, many, many types of post-hoc test are available. They have different advantages and disadvantages.

Post-hoc test examples

- Bonferroni correction: just divide the confidence level α by the total number of tests done. Then calculate *t*-statistics based on new significance threshold.
  - Very simple procedure... but also tends to be very conservative. Doing many tests makes it very hard to find any significant pairs.
- Tukey's test: Less stringent than Bonferroni
  - A very good choice when sample sizes are not equal in groups, and when you need confidence intervals for pairwise differences.
- Stepwise procedures (various options available). Involves ordering the *p*-values and using a more stringent testing threshold on the smallest ones.
  - Better when equal sample size among groups and don't need confidence intervals.
- Scheffé's method can be used to test more complicated comparisons (called *contrasts*)
  - e.g. Could test the mean of two groups vs. the mean of a third group.

Choosing a post-hoc test

- Despite the large number of post-hoc test available, several are standard.
  - Make sure the one you choose is best for your study and what kind of comparison you want to make.
- Can run multiple post-hoc tests, but have to be careful... this is reintroducing multiple testing!
- If you do run multiple post-hoc tests, then the results from all of them should be reported.
- Don't just run multiple tests and then just choose the one that gives significance!

Factorial ANOVA

Factorial ANOVA

- Up to now, we were considering only one grouping variable.
- The factorial ANOVA can look at the means of a continuous variable with respect to multiple categorial variables.
- e.g. The case where we consider two grouping variables is called 2-way ANOVA.
- This kind of analysis goes nicely with the classic factorial experiment (i.e. multiple treatments each with multiple possible levels)
- Won't go into the mathematical details. But still involves decomposing the variation into different terms that explain different parts of the model, and unexplained variance. Basically the same assumptions as in the one-way ANOVA.

Types of effects

Consider a 2-way ANOVA comparing effects of drug A in men/women on systolic blood pressure. Assume there are 3 dose groups for the drug: A_1, A_2, and A_3.

The *main* effects test the effect of one variable while the other variable is fixed:

- The main effect for the drug looks at the overall effectiveness of the drug in the different doses.
- The main effect for sex investigates whether mean blood pressure differs overall between males/females.

The *interaction* effect tests the combined effect of the two grouping variables:

- The interaction effect between the drug and sex investigates whether the effect of the drug on blood pressure is different for men and women.

Higher order ANOVA

- Can do ANOVA with 3 or more categorical variables
- Interpretation of main effects is the same, but would usually only consider interactions between pairs of variables. Higher order interactions too difficult to interpret.
- Each effect is tested through a separate F-test.
- Remember that, for each effect, a significant result doesn't tell which particular group(s) differ from the rest.

2-way ANOVA example

- Recall the example looking at shoot dry mass in rice plants
- Now consider both the fertilizer type (F10, NH4Cl, and NH4NO3) and rice variety (wild type vs. ANU843).
- This is a balanced experiment... still have the same number of units in each experimental group (12 in each fertilizer/variety group)

Dependent Variable: ShootDryMass

Fertilizer type	Rice variety	Mean	Std. Deviation	N
F10	ANU843	7.33	4.774	12
	wt	108.33	26.722	12
	Total	57.83	54.896	24
NH4Cl	ANU843	46.58	14.712	12
	wt	50.25	18.261	12
	Total	48.42	16.325	24
NH4NO3	ANU843	71.50	20.336	12
	wt	73.33	23.078	12
	Total	72.42	21.293	24
Total	ANU843	41.81	30.377	36
	wt	77.31	32.910	36
	Total	59.56	36.170	72

2-way ANOVA example

Tests of Between-Subjects Effects

Dependent Variable: ShootDryMass

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	68325.611[a]	5	13665.122	36.719	.000
Intercept	255374.222	1	255374.222	686.206	.000
fert_num	7018.778	2	3509.389	9.430	.000
variety_num	22684.500	1	22684.500	60.955	.000
fert_num * variety_num	38622.333	2	19311.167	51.890	.000
Error	24562.167	66	372.154		
Total	348262.000	72			
Corrected Total	92887.778	71			

a. R Squared = .736 (Adjusted R Squared = .716)

Separate rows for each main/interaction effect test in the 2-way ANOVA.

2-way ANOVA example result

- The fertilizer main effect is significant (it wasn't before!). We have evidence that the fertilizer level is associated with shoot dry mass.
- The variety main effect is significant. We have evidence that the rice variety is associated with shoot dry mass.
- The interaction term is significant. We have evidence that the effect of the fertilizer on shoot dry mass depends on which variety of rice is being considered.

Take-home message

- Doing an ANOVA usually leads to better power (ability to detect true association) than a bunch of pairwise t-tests.
- Works nicely with experiments you might be using in the lab.
- Plethora of post-hoc tests available to test differences between individual groups. Just don't do a bunch and report the only significant one.
- Though normality is assumed, ANOVA is robust to violation of this assumption (in particular when the sample size is large).

References

Rice experiment dataset:

1. Perrine, Francine M., et al. "Rhizobium plasmids are involved in the inhibition or stimulation of rice growth and development." *Functional Plant Biology* 28.9 (2001): 923-937.

Thank you! - Merci!

Questions?