Unit 4: Inference for numerical data
4. ANOVA

Sta 101 - Spring 2019

Duke University, Department of Statistical Science

Dr. Abrahamsen

Slides posted at *https://stat.duke.edu/courses/Spring19/sta101.002*

Announcements

- Team evaluations due today 11:59pm
  - Search your inbox for "TEAMWORK Admin"
  - Check your spam folder
- No office hours today, they were switched to Wednesday
- Come to Monday's lab prepared to work on your project

Why the name ANOVA?

Hypothesis:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. It may seem odd that the technique is called "Analysis of Variance" rather than "Analysis of Means". As you will see, the name is appropriate because inferences about means are made by analyzing variance.

NEWS FLASH!

Jelly beans rumored to cause acne!!!

How would you check this rumor? Imagine that doctors can assign an "acne score" to patients on a 0-100 scale.

- What would your research question be?
- How would you conduct your study?
- What statistical test would you use?

http://imgs.xkcd.com/comics/significant.png

Clicker question

Suppose $\alpha = 0.05$. What is the probability of making a Type 1 error and rejecting a null hypothesis like

$$H_0 : \mu_{\text{purple jelly bean}} - \mu_{\text{placebo}} = 0$$

when it is actually true?

(a) 1%
(b) 5%
(c) 36%
(d) 64%
(e) 95%

Clicker question

Suppose we want to test 20 different colors of jelly beans versus a placebo with hypotheses like

$$H_0 : \mu_{\text{purple jelly bean}} - \mu_{\text{placebo}} = 0$$
$$H_0 : \mu_{\text{brown jelly bean}} - \mu_{\text{placebo}} = 0$$
$$H_0 : \mu_{\text{peach jelly bean}} - \mu_{\text{placebo}} = 0$$
$$\ldots$$

and we use $\alpha = 0.05$ for each of these tests. What is the probability of making at least one Type 1 error in these 20 independent tests?

(a) 1%
(b) 5%
(c) 36%
(d) 64%
(e) 95%

Conditions on ANOVA

1. *Independence*:
   (a) within group: sampled observations must be independent
   (b) between group: groups must be independent of each other
2. *Approximate normality*: distribution should be nearly normal within each group
3. *Equal variance*: groups should have roughly equal variability
   (a) Not necessary if all group sizes are the same
   (b) As a rule of thumb, the condition is not met if $s_{\max} \geq 2s_{\min}$

ANOVA tests for <u>some</u> difference in means of many different groups

Null hypothesis:

$$H_0 : \mu_{\text{placebo}} = \mu_{\text{purple}} = \mu_{\text{brown}} = \ldots = \mu_{\text{peach}} = \mu_{\text{orange}}.$$

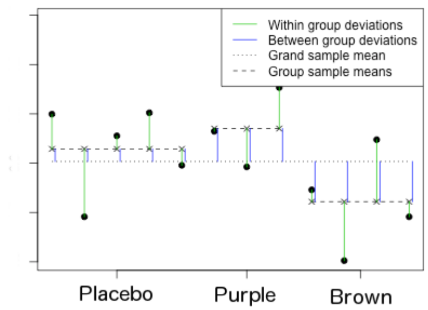
Clicker question

Which of the following is a correct statement of the alternative hypothesis?

(a) For any two groups, including the placebo group, no two group means are the same.
(b) For any two groups, not including the placebo group, no two group means are the same.
(c) Amongst the jelly bean groups, there are at least two groups that have different group means from each other.
(d) Amongst all groups, there are at least two groups that have different group means from each other.

ANOVA compares between group variation to within group variation

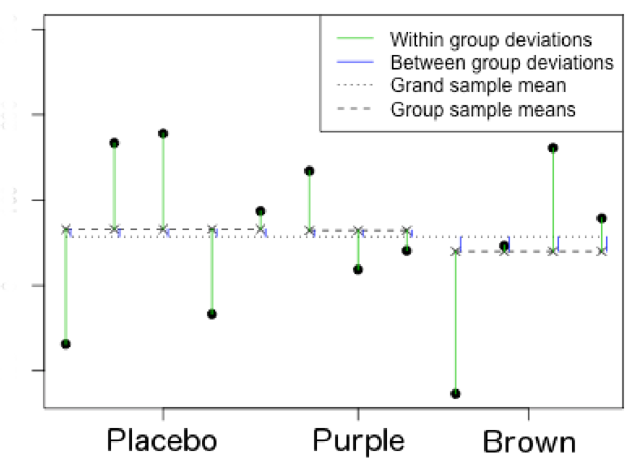
$$\sum |^2 \Big/ \sum |^2 = \text{BETWEEN} / \text{WITHIN} = SSG/SSE$$



- Relatively large WITHIN group variation: little apparent difference
- Relatively large BETWEEN group variation: there may be a difference

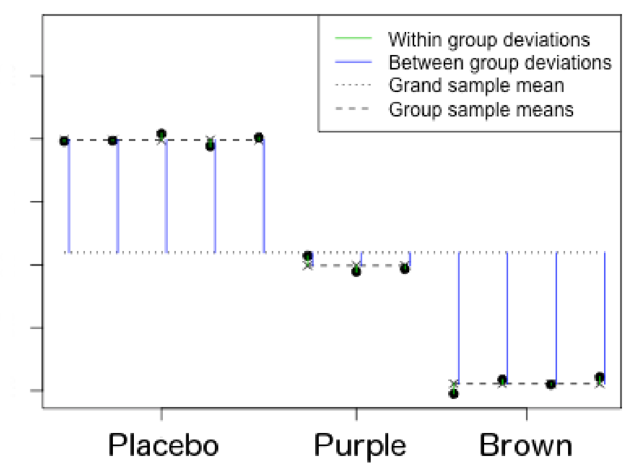
Relatively large WITHIN group variation: little apparent difference

$$\sum |^2 \Big/ \sum |^2$$



Relatively large BETWEEN group variation: there may be a difference

$$\sum |^2 \Big/ \sum |^2$$



For historical reasons, we use a modification of this ratio called the F-statistic:

$$F = \frac{SSG \;/\; (k-1)}{SSE \;/\; (n-k)} \quad = \quad \frac{MSG}{MSE}$$

k: # of groups; n: # of obs.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Between groups	$k-1$	SSG	MSG	F_{obs}	p_{obs}
Within groups	$n-k$	SSE	MSE		
Total	$n-1$	SSG+SSE			

Note: F distribution is defined by two dfs: $df_G = k-1$ and $df_E = n-k$

F-table: Reject H_0 at table's significance level if $F_{obs} \geq F^{\star}_{df_G, df_E}$

To identify which means are different, use t-tests and the Bonferroni correction

- If the ANOVA yields a significant results, next natural question is: "Which means are different?"
- Use t-tests comparing each pair of means to each other,
  - with a common variance (MSE from the ANOVA table) instead of each group's variances in the calculation of the standard error,
  - and with a common degrees of freedom (df_E from the ANOVA table)
- Compare resulting p-values to a modified significance level

$$\alpha^{\star} = \frac{\alpha}{K}$$

where $K = \frac{k(k-1)}{2}$ is the total number of pairwise tests

Application exercise: 4.4 ANOVA

See the course webpage for details.

Summary of main ideas

1. Comparing many means requires care
2. ANOVA tests for <u>some</u> difference in means of many different groups
3. ANOVA compares between group variation to within group variation
4. To identify which means are different, use t-tests and the Bonferroni correction