Unit 4: Inference for numerical data
4. ANOVA

Sta 101 - Spring 2019

Duke University, Department of Statistical Science

Dr. Ellison

Slides posted at
https://www2.stat.duke.edu/courses/Spring19/sta101.001/

Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**
Solution: Use ANOVA!

1. Why do we need ANOVA? Comparing many means requires care
2. ANOVA Step 1: ANOVA tests for some difference in means of many different groups
3. How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Announcements

Coming up…

- Project Stage 1 is due **Thursday just before your lab section time.**
- Problem Set 4 is due **Friday 3/8 11:55pm**
- Performance Assessment 4 is due **Sunday 3/17 11:55pm** (opens Wednesday)
- Readiness Assessment 5 is **Monday 3/18**
- New TA for 10:05 section Melanie Lai Wai
- New TA for 3:05 section Tess Chandler

Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**
Solution: Use ANOVA!

1. Why do we need ANOVA? Comparing many means requires care
2. ANOVA Step 1: ANOVA tests for some difference in means of many different groups
3. How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Would we use the same analysis for the following research questions?

- "Is there an association between income and gender?"
- "Is there an association between income and political affiliation?"

Sample Data

Gender *(Categorical with 2 Levels)*	Political Affiliation *(Categorical with 3 Levels)*	Voted in 2018? *(Categorical with 2 Levels)*	Age *(Numerical - Continuous)*	Income *(Numerical - Continuous)*	Number of Dependents *(Numerical - Discrete)*
Male	Democrat	Yes	22	$20,000	0
Female	Republican	Yes	39	$40,000	3
Male	Independent	Yes	47	$400,000	0
Male	Republican	Yes	18	$129,000	0
Female	Republican	No	29	$85,000	2
Female	Democrat	No	80	$72,000	1
Female	Democrat	Yes	56	$55,000	0
Male	Independent	Yes	72	$34,000	0
...	...	...	...	...	...

- "Is there an association between income and gender?"

Numerical response variable

Categorical explanatory variable ***with 2 levels***

https://www.wsj.com/articles/parsing-the-gender-pay-gap-1542917969

- "Is there an association between income and gender?"

Numerical response variable

Categorical explanatory variable ***with 2 levels***

Independent Means Hypothesis Test
***provided necessary conditions are met**

https://www.wsj.com/articles/parsing-the-gender-pay-gap-1542917969

$Ho: \mu_{male} - \mu_{female} = 0$
$Ha: \mu_{male} - \mu_{female} \neq 0$

α=0.05

- "Is there an association between income and gender?"

Numerical response variable

Categorical explanatory variable ***with 2 levels***

Independent Means Hypothesis Test
***provided necessary conditions are met**

https://www.wsj.com/articles/parsing-the-gender-pay-gap-1542917969

$Ho: \mu_{male} - \mu_{female} = 0$ → No association
$Ha: \mu_{male} - \mu_{female} \neq 0$ → Association

α=0.05

- "Is there an association between **income** and **political affiliation**?"

Numerical response variable

Categorical explanatory variable

with >2 levels

- "Is there an association between **income** and **political affiliation**?"

Numerical response variable

Categorical explanatory variable

with >2 levels

ANOVA

***provided necessary conditions are met**

α=0.05

Ho: $\mu_{Rep} = \mu_{Dem} = \mu_{Ind}$

Ha: ***at least one of the pairs of means are different***

- "Is there an association between **income** and **political affiliation**?"

Numerical response variable

Categorical explanatory variable

with >2 levels

ANOVA

***provided necessary conditions are met**

α=0.05

Ho: $\mu_{Rep} = \mu_{Dem} = \mu_{Ind}$

Ha: ***at least one of the pairs of means are different***

→ No association

→ Association

- "Is there an association between **income** and **political affiliation**?"

Numerical response variable

Categorical explanatory variable

with >2 levels

Why should we not just do the following?

Ho: $\mu_{Dem} - \mu_{Rep} = 0$
Ha: $\mu_{Dem} - \mu_{Rep} \neq 0$ α=0.05

Ho: $\mu_{Dem} - \mu_{Ind} = 0$
Ha: $\mu_{Dem} - \mu_{Ind} \neq 0$ α=0.05

Ho: $\mu_{Ind} - \mu_{Rep} = 0$
Ha: $\mu_{Ind} - \mu_{Rep} \neq 0$ α=0.05

- **"Is there an association between income and political affiliation?"**

Numerical response variable

Categorical explanatory variable ***with >2 levels***

See next slides!

$Ho: \mu_{Dem} - \mu_{Rep} = 0$
$Ha: \mu_{Dem} - \mu_{Rep} \neq 0$ α=0.05

$Ho: \mu_{Dem} - \mu_{Ind} = 0$
$Ha: \mu_{Dem} - \mu_{Ind} \neq 0$ α=0.05

$Ho: \mu_{Ind} - \mu_{Rep} = 0$
$Ha: \mu_{Ind} - \mu_{Rep} \neq 0$ α=0.05

Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**

Solution: Use ANOVA!

1. NEW ⚙ Why do we need ANOVA? Comparing many means requires care
2. Step 1: ANOVA tests for some difference in means of many different groups
3. ⚙ How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ⚙ ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Outline

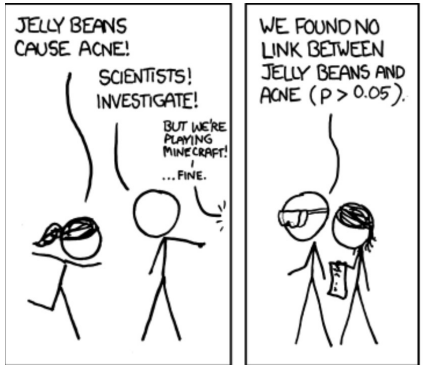
Why do we need ANOVA?

Comparing more than two means requires care!

NEW ⚙

NEWS FLASH!

Jelly beans rumored to cause acne!!!



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

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?

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?

Research Question: *Does eating jelly beans cause acne?*

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?

Ate Jelly Beans or Placebo?	Acne Score
Jelly Beans	100
Jelly Beans	1
Placebo	50
Jelly Beans	22
Placebo	74
Jelly Beans	27
Placebo	33
Placebo	13
Jelly Beans	4
Jelly Beans	89
...	...

Research Question: *Does eating jelly beans cause acne?*

Random Experiment

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?

Ate Jelly Beans or Placebo?	Acne Score
Jelly Beans	100
Jelly Beans	1
Placebo	50
Jelly Beans	22
Placebo	74
Jelly Beans	27
Placebo	33
Placebo	13
Jelly Beans	4
Jelly Beans	89
...	...

Research Question: *Does eating jelly beans cause acne?*

Random Experiment

Use an Independent Means Hypothesis Test:

$H_0: \mu_{jelly\ beans} - \mu_{placebo} = 0$

$H_A: \mu_{jelly\ beans} - \mu_{placebo} > 0$

NEWS FLASH!

Now they're saying that it's *only a certain color* of jelly bean that causes 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?

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

NEWS FLASH!

Now they're saying that it's *only a certain color* of jelly bean that causes 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?

Type of Jelly Bean or Placebo	Acne Score
Green Jelly Bean	100
Green Jelly Bean	1
Green Jelly Bean	22
...	...
Purple Jelly Bean	22
Purple Jelly Bean	7
...	...
Placebo	22
Placebo	56
...	...

Research Question: *Does eating a certain color of jelly bean cause you to get acne?*

Random Experiment

Clicker question

Suppose α = 0.05. What is the probability of making a Type 1 error (aka: 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 α = 0.05. What is the probability of making a Type 1 error (aka: 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%

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%

Hint: What type of distribution should we consider using if we are asked to find the probability that ***k*** (or at least/at most *k)* out of **n** independent trials have a certain outcome when we know the probability of any given trial having this outcome is **p?**

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%

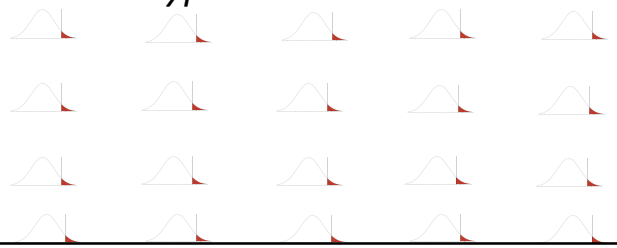
X=# of tests out of 20 that make a type 1 error
X~Bin(n=20,p=0.05)

P(X≥1)= 1-P(X=0)

$$= 1 - \binom{20}{0} 0.05^0 (1 - 0.05)^{20-0}$$

Outline

Running multiple independent means hypothesis tests to answer just one research question will result in a *higher probability of making at least one Type 1 Error*.



Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**

Solution: Use ANOVA!

1. Why do we need ANOVA? Comparing many means requires care
2. ANOVA Step 1: ANOVA tests for some difference in means of many different groups
3. How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : _____

ANOVA tests for some difference in means of many different groups

Null hypothesis:

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

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 tests for some difference in means of many different groups

Null hypothesis:

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

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.

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : Amongst all groups, there are at least two groups that have different group means from each other.

Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**

Solution: Use ANOVA!

1. Why do we need ANOVA? Comparing many means requires care
2. ANOVA Step 1: ANOVA tests for some difference in means of many different groups
3. How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : Amongst all groups, there are at least two groups that have different group means from each other.

How do we make a conclusion about these hypotheses with ANOVA?

Step 1a: Fill out the ANOVA table.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Step 1a: Fill out the ANOVA table.

Test Statistic p-value

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Step 1a: Fill out the ANOVA table.

Test Statistic p-value

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Rstudio

p-value=pf(test statistic, k-1, n-k, lower.tail=FALSE)

Order matters!

Step 1a: Fill out the ANOVA table.

Test Statistic p-value

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Other useful information…

Step 1a: Fill out the ANOVA table.

Test Statistic p-value

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Other useful information…

Step 1a: Fill out the ANOVA table.

Test Statistic p-value

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

Step 1b: Evaluate the p-value to make a conclusion.

- p-value<α: Reject Ho. There is sufficient evidence to suggest Ha.
- p-value≥α: Fail to reject Ho. There is not sufficient evidence to suggest Ha.

Outline

ANOVA *seems* like it follows a different hypothesis testing structure than what we've done so far. Let's see why ANOVA works and how it's different!

	Do we know how to create a confidence interval/hypothsis test yet? If so, how?	
Population Parameter	**Confidence Interval**	**Hypothesis Testing**
μ_1-μ_2	CLT Confidence Interval (Unit 4)	Randomization Testing (Unit 1) CLT Hypothesis Testing (Unit 4)
p_1-p_2		Randomization Testing (Unit 1)
$Median_1$-$Median_2$		Randomization Testing (Unit 1)
μ	CLT Confidence Interval (Unit 3 and 4) Bootstrap Confidence Interval (Unit 4)	CLT Hypothesis Testing (Units 3 and 4) Bootstrap Hypothesis Testing (Unit 4)
μ_{diff}	CLT Confidence Interval (Unit 4) Bootstrap Confidence Interval (Unit 4)	CLT Hypothesis Testing (Units 3 and 4) Bootstrap Hypothesis Testing (Unit 4)
Median	Bootstrap Confidence Interval (Unit 4)	Bootstrap Hypothesis Testing (Unit 4)
p	Bootstrap Confidence Interval (Unit 4)	Bootstrap Hypothesis Testing (Unit 4)

Outline

Let's compare <u>Step 1 of ANOVA</u> and an <u>Independent Means Test</u>.

Hypotheses Different

ANOVA compares between group variation to within group variation

Similarities/Differences with the independent means test:

- <u>ANOVA (step 1) Hypotheses:</u>

$$Ho: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$Ha: at\ least\ one\ pair\ of\ means\ is\ different$$

- <u>Independent Means Hypotheses:</u>

$Ho: \mu_1 = \mu_2$ $Ha: \mu_1 > \mu_2$ | $Ho: \mu_1 = \mu_2$ $Ha: \mu_1 < \mu_2$ | $Ho: \mu_1 = \mu_2$ $Ha: \mu_1 \neq \mu_2$

Outline

Let's compare Step 1 of ANOVA and an Independent Means Test.

Test Statistics Come from Different Distributions

ANOVA compares between group variation to within group variation

Test statistic from **ANOVA Step 1** *always* follows the F-distribution:

$$p - value = P(F \geq MSG/MSE)$$

Test statistic for **independent means hypothesis test** follows the Z-distribution or T-distribution:

$$p - value = P\left(Z \geq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \qquad p - value = P\left(T \geq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$$

$$p - value = P\left(Z \leq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \qquad p - value = P\left(T \leq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$$

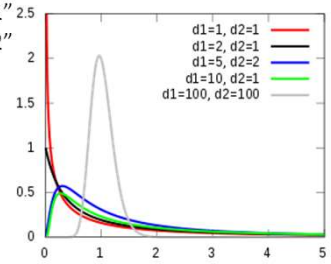
$$p - value = 2 \cdot P\left(Z \geq \left|\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right|\right) \qquad p - value = 2 \cdot P\left(T \geq \left|\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|\right)$$

ANOVA compares between group variation to within group variation

F-distribution

- F-distribution is always positive.
- Has two parameters:
  - "Degrees of freedom 1"
  - "Degrees of freedom 2"

ORDER MATTERS!



https://en.wikipedia.org/wiki/F-distribution

R-Studio

pf(f-stat, df1, df2, lower.tail=FALSE) ≠ pf(f-stat, df2, df1, lower.tail=FALSE)

Outline

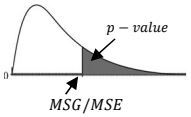
Let's compare Step 1 of ANOVA and an Independent Means Test.

Shape of p-value of only one type for ANOVA Step 1.

ANOVA compares between group variation to within group variation

P-value from **ANOVA Step 1** *always* is a RIGHT TAIL in the F-distribution!

$$p-value = P(F \geq MSG/MSE)$$



P-value for a **independent means hypothesis test** can be a right/left/two-sided tail in the Z-distribution or T-distribution:

$$p-value = P\left(Z \geq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \qquad p-value = P\left(T \geq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$$

$$p-value = P\left(Z \leq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \qquad p-value = P\left(T \leq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$$

$$p-value = 2 \cdot P\left(Z \geq \left|\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right|\right) \qquad p-value = 2 \cdot P\left(T \geq \left|\frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|\right)$$

Outline

Let's compare Step 1 of ANOVA and an Independent Means Test.

***Both** test statistics measure:*

$$\frac{\textit{between group variability}}{\textit{within group variability}}$$

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

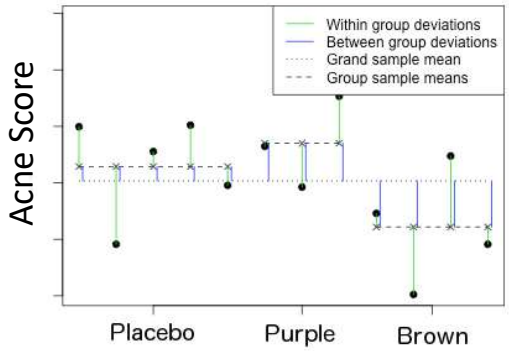
$$F = \frac{\sum|^2 \ / \ (k-1)}{\sum|^2 \ / \ (n-k)} = \frac{MSG}{MSE}$$

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$\sum\|^2 \ SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$\sum\|^2 \ SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

ANOVA compares between group variation to within group variation

Test Statistic=MSG/MSE=$(\sum|^2 / \sum|^2)\cdot$(n-k/k-1)



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

$$F = \frac{\sum|^2 \ / \ (k-1)}{\sum|^2 \ / \ (n-k)} = \frac{MSG}{MSE}$$

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Between groups	k-1	$\sum\|^2 \ SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	MSG = SSG/(k-1)	MSG/MSE	P(F > MSG/MSE)
Within Groups	n-k	$\sum\|^2 \ SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	MSE = SSE/(n-k)		
Total	n-1	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$			

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

- Variability between groups
- Variability within groups
- Total variability in the response variable

ANOVA compares between group variation to within group variation

Similarities with the independent means/samples test:

Measure of:

- F-statistic in ANOVA:

$$p - value = P(F \geq MSG/MSE)$$

$$\frac{Between\ Group\ Variability}{Within\ Group\ Variability} \quad \sum|^2 \Big/ \sum|^2$$

- T-statistic in INDEPENDENT MEANS TEST (with null value=0 and right tailed):

$$p - value = P\left(T \geq \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$$

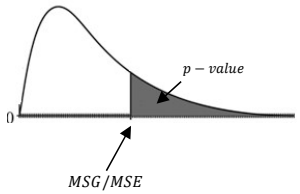
Between group variability measure

Within group variability measure

ANOVA compares between group variation to within group variation

As F-statistic MSG/MSE ↑

- P-value ↓
- More likely to reject Ho: $\mu_1 = \mu_2 = \cdots = \mu_k$



$$p - value = P(F \geq MSG/MSE)$$

Outline

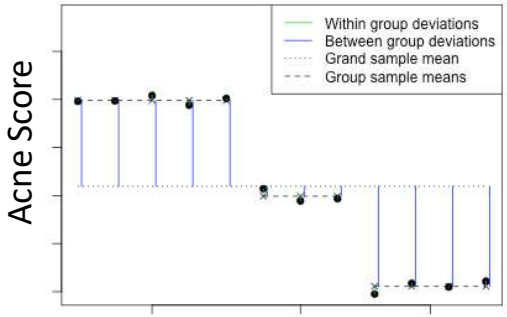
Why would a high value for

$$\frac{between\ group\ variability}{within\ group\ variability}$$

suggest that the population means aren't all the same?

Relatively large BETWEEN group variation: there may be a difference

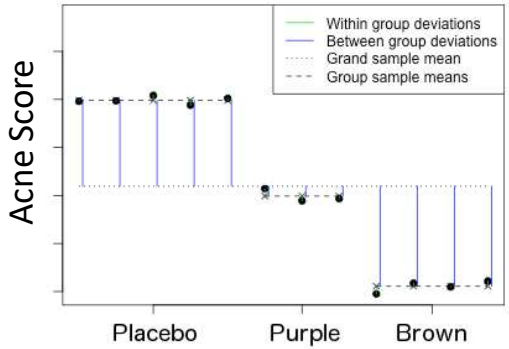
Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n\text{-}k/k\text{-}1)$



Fraction is _____, more likely to _____ in ANOVA.

Relatively large BETWEEN group variation: there may be a difference

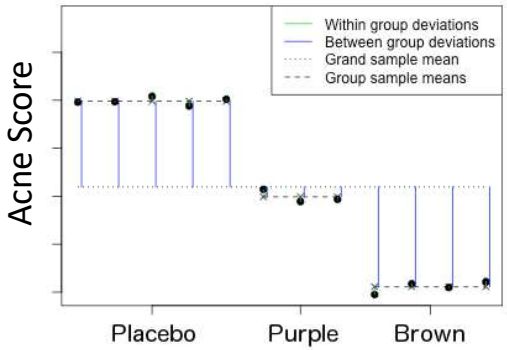
Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n\text{-}k/k\text{-}1)$



Fraction is _____, more likely to _____ in ANOVA.

Relatively large BETWEEN group variation: there may be a difference

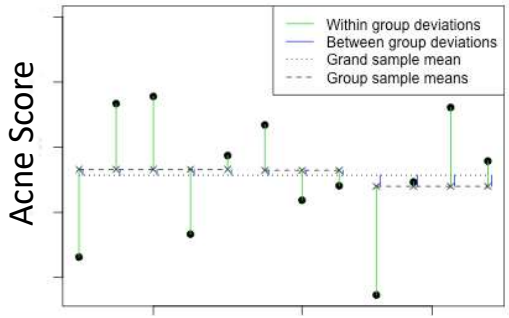
Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n\text{-}k/k\text{-}1)$



Fraction is large, more likely to reject Ho in ANOVA.

Relatively large WITHIN group variation: little apparent difference

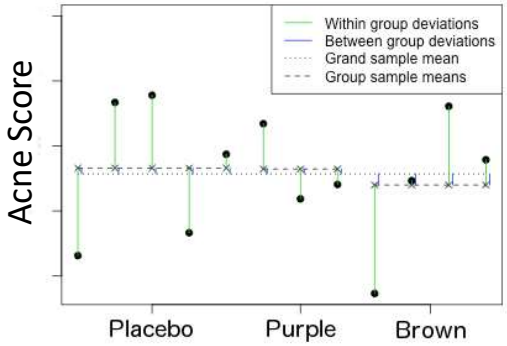
Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n\text{-}k/k\text{-}1)$



Fraction is _____, more likely to _____ in ANOVA.

Relatively large WITHIN group variation: little apparent difference

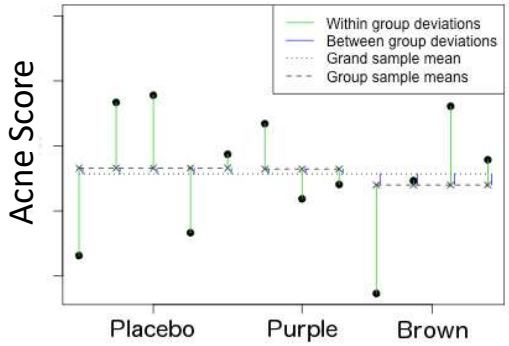
Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n-k/k-1)$



Fraction is _____, more likely to _____ in ANOVA.

Relatively large WITHIN group variation: little apparent difference

Test Statistic=MSG/MSE= $(\sum|^2 / \sum|^2)\cdot(n-k/k-1)$



Fraction is small, more likely to fail to reject Ho in ANOVA.

Outline

1. Housekeeping

2. Main ideas

Independent Means Hypothesis Testing: Test for an Association Between a Numerical Variable and a Categorical Variable **with Two Levels**

Problem: What if we want to test for an Association Between a Numerical Variable and a Categorical Variable **with MORE THAN TWO Levels?**

Solution: Use ANOVA!

1. Why do we need ANOVA? Comparing many means requires care
2. ANOVA Step 1: ANOVA tests for some difference in means of many different groups
3. How/Why does ANOVA Step 1 Work? ANOVA compares between group variation to within group variation
4. ANOVA Step 2: To identify which means are different, use t-tests and the Bonferroni correction

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : Amongst all groups, there are at least two groups that have different group means from each other.

Step 2 of ANOVA:

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : Amongst all groups, there are at least two groups that have different group means from each other.

Step 2 of ANOVA:

If we rejected Ho in Step 1…. find out **which** pairs have different group means from each other.

Outline

Step 0 of ANOVA: Check conditions.

Step 1 of ANOVA:

Assess the following hypotheses.

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

H_a : Amongst all groups, there are at least two groups that have different group means from each other.

Step 2 of ANOVA:

Post-hoc multiple pairwise comparison tests.

Conduct a 2-Tailed Independent Means Test for Each Possible Mean Pair… BUT CHANGE THE FOLLOWING 3 THINGS IN EACH TEST!

NEXT SLIDE

$H_0 : \mu_{placebo} = \mu_{purple}$ $H_a : \mu_{placebo} \neq \mu_{purple}$

$H_0 : \mu_{placebo} = \mu_{peach}$ $H_a : \mu_{placebo} \neq \mu_{peach}$

$H_0 : \mu_{purple} = \mu_{brown}$ $H_a : \mu_{purple} \neq \mu_{brown}$

…

$H_0 : \mu_{placebo} = \mu_{brown}$ $H_a : \mu_{placebo} \neq \mu_{brown}$

$H_0 : \mu_{placebo} = \mu_{orange}$ $H_a : \mu_{placebo} \neq \mu_{orange}$

$H_0 : \mu_{purple} = \mu_{peach}$ $H_a : \mu_{purple} \neq \mu_{peach}$

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

Things that are Different	Regular Independent Means Test (with null value =0)	Post-hoc (ANOVA Step 2) Multiple Pairwise Comparisons Tests (do this for each one)
Significance Level	α	$\alpha^* = \frac{\alpha}{k(k-1)/2}$

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

Things that are Different	Regular Independent Means Test (with null value =0)	Post-hoc (ANOVA Step 2) Multiple Pairwise Comparisons Tests (do this for each one)
Significance Level	α	$\alpha^* = \frac{\alpha}{k(k-1)/2}$

Bonferroni Correction

From Step 1 of ANOVA

Total number of pairings of means

k=# of groups

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

Things that are Different	Regular Independent Means Test (with null value =0)	Post-hoc (ANOVA Step 2) Multiple Pairwise Comparisons Tests (do this for each one)
Significance Level	α	$\alpha^* = \frac{\alpha}{k(k-1)/2}$
T-Statistics	$T = \frac{(\bar{x}_a - \bar{x}_b) - 0}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$	$T = \frac{(\bar{x}_a - \bar{x}_b) - 0}{\sqrt{\frac{MSE}{n_a} + \frac{MSE}{n_b}}}$

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

Things that are Different	Regular Independent Means Test (with null value =0)	Post-hoc (ANOVA Step 2) Multiple Pairwise Comparisons Tests (do this for each one)
Significance Level	α	$\alpha^* = \frac{\alpha}{k(k-1)/2}$
T-Statistics	$T = \frac{(\bar{x}_a - \bar{x}_b) - 0}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$	$T = \frac{(\bar{x}_a - \bar{x}_b) - 0}{\sqrt{\frac{MSE}{n_a} + \frac{MSE}{n_b}}}$
Degrees of Freedom to use in T-distribution	$df = \min(n_a - 1, n_b - 1)$	$df = df_E = n - k$ • n=total number of observations • k=# of groups

Outline

- random sample / assignment
- each n_j less than 10% of respective population

Conditions for ANOVA

1. ***Independence:***
   - ✓ **within groups:** sampled observations must be independent
   - ✓ **between groups:** the groups must be independent of each other (non-paired)
2. ***Approximate normality:*** distributions should be nearly normal within each group
3. ***Equal variance:*** groups should have roughly equal variability

Application exercise: 4.4 ANOVA

See the course webpage for details.