

# BIOS 600: Principles of Statistical Inference

## Analysis of Variance (ANOVA)

Fall 2012

- ▶ Pagano and Gauvreau, Chapter 12
- ▶ Gelman, “Of Beauty, Sex, and Power”

# Methods for Testing Association

When we wish to assess the association between variables, one factor that determines which method should be used is the distribution of each variable. Examples include the following.

- ▶ Categorical predictor, categorical outcome: chi-squared test, Fisher's exact test, logistic regression (multiple predictors)

# Methods for Testing Association

When we wish to assess the association between variables, one factor that determines which method should be used is the distribution of each variable. Examples include the following.

- ▶ Categorical predictor, categorical outcome: chi-squared test, Fisher's exact test, logistic regression (multiple predictors)
- ▶ Categorical predictor, continuous outcome: t-test, ANOVA, sign test, signed-rank test, or Kruskal-Wallis test

# Methods for Testing Association

When we wish to assess the association between variables, one factor that determines which method should be used is the distribution of each variable. Examples include the following.

- ▶ Categorical predictor, categorical outcome: chi-squared test, Fisher's exact test, logistic regression (multiple predictors)
- ▶ Categorical predictor, continuous outcome: t-test, ANOVA, sign test, signed-rank test, or Kruskal-Wallis test
- ▶ Continuous predictor, categorical outcome: logistic regression

# Methods for Testing Association

When we wish to assess the association between variables, one factor that determines which method should be used is the distribution of each variable. Examples include the following.

- ▶ Categorical predictor, categorical outcome: chi-squared test, Fisher's exact test, logistic regression (multiple predictors)
- ▶ Categorical predictor, continuous outcome: t-test, ANOVA, sign test, signed-rank test, or Kruskal-Wallis test
- ▶ Continuous predictor, categorical outcome: logistic regression
- ▶ Continuous predictor, continuous outcome: linear regression

# Motivating Example: Pets and Stress

An investigator was interested in how response to a psychological stressor may be affected by the presence of a pet. A small study was conducted in which 45 people were randomized to three groups and then exposed to the stressor, with the pulse rate measured after exposure. In the first group, a pet was present when the pulse rate was measured, in the second group a human friend was present, and in the third group neither pet nor friend were present. The investigator would like to know whether there are differences in the post-exposure pulse rate among the three groups.

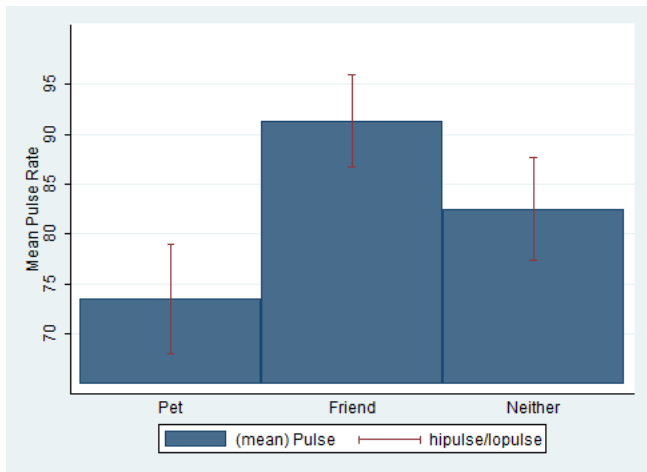
# Pets and Stress Pulse Rates



Group 1 (Pet present)	Group 2 (Friend present )	Group 3 (Neither pet nor friend present)
69.17	99.69	84.74
68.86	91.35	87.23
70.17	83.40	84.88
64.17	100.88	80.37
58.69	102.15	91.75
79.66	89.82	87.45
69.23	80.28	87.78
75.98	98.20	73.28
86.45	101.06	84.52
97.54	76.91	77.80
85.00	97.05	70.88
69.54	88.02	90.02
70.08	81.60	99.05
72.26	86.98	75.48
65.45	92.49	62.65



# CI's for Group Means



Hmm, what kind of friends are those? Regardless, how do we make comparisons across three groups? Which groups are different?

# CI Code...because it took me FOREVER to do this!

```
collapse (mean) meanpulse= Pulse (sd) sdpulse=Pulse  
      (count) n=Pulse, by(Group)  
  
generate hipulse = meanpulse +  
      invttail(n-1,0.025)*(sdpulse / sqrt(n))  
generate lopulse = meanpulse -  
      invttail(n-1,0.025)*(sdpulse / sqrt(n))  
  
sort(Group)  
  
twoway (bar meanpulse Group,yscale(range(65 100)))  
      (rcap hipulse lopulse Group), xlabel(1 "Pet"  
      2 "Friend" 3 "Neither", noticks) xtitle("Group")  
      ytitle("Mean Pulse Rate")
```

## Example: Pets and Stress

We are interested in testing

$$H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER}$$

against the alternative that at least one mean is different.

One way to do this would be to use t-tests on all possible pairs of tests (here there are just three). However, if we have more groups, this becomes quite complicated. For example, with 10 groups we need to do  $\binom{10}{2} = 45$  tests! In addition to being time-consuming, carrying out multiple tests can lead to an inflated type I error rate.

# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests

# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests
- ▶ Suppose also the tests are independent and done at a 0.05 significance level

# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests
- ▶ Suppose also the tests are independent and done at a 0.05 significance level
- ▶ Then the probability we fail to reject all three tests (the right decision) is  $(1 - 0.05)^3 = 0.95^3 = 0.857$ , and the probability of rejecting at least one of the three null hypotheses, called the *family-wise error rate*, is  $1 - 0.857 = 0.143 \gg 0.05$

# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests
- ▶ Suppose also the tests are independent and done at a 0.05 significance level
- ▶ Then the probability we fail to reject all three tests (the right decision) is  $(1 - 0.05)^3 = 0.95^3 = 0.857$ , and the probability of rejecting at least one of the three null hypotheses, called the *family-wise error rate*, is  $1 - 0.857 = 0.143 \gg 0.05$
- ▶ With 45 tests, the probability of rejecting at least 1 of them is almost 1

# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests
- ▶ Suppose also the tests are independent and done at a 0.05 significance level
- ▶ Then the probability we fail to reject all three tests (the right decision) is  $(1 - 0.05)^3 = 0.95^3 = 0.857$ , and the probability of rejecting at least one of the three null hypotheses, called the *family-wise error rate*, is  $1 - 0.857 = 0.143 \gg 0.05$
- ▶ With 45 tests, the probability of rejecting at least 1 of them is almost 1
- ▶ In reality, this is a little more complicated (tests are not independent), but we still have the problem of an inflated type I error rate.



# Multiple Comparisons

- ▶ Suppose all means are truly equal, and we conduct all three pairwise tests
- ▶ Suppose also the tests are independent and done at a 0.05 significance level
- ▶ Then the probability we fail to reject all three tests (the right decision) is  $(1 - 0.05)^3 = 0.95^3 = 0.857$ , and the probability of rejecting at least one of the three null hypotheses, called the *family-wise error rate*, is  $1 - 0.857 = 0.143 \gg 0.05$
- ▶ With 45 tests, the probability of rejecting at least 1 of them is almost 1
- ▶ In reality, this is a little more complicated (tests are not independent), but we still have the problem of an inflated type I error rate.
- ▶ ANOVA is one way to control the overall type I error rate at a fixed level  $\alpha$ , if we only test pairwise differences when the “chunk” or overall test is rejected

# ANOVA

ANOVA stands for *analysis of variance*. We use ANOVA when we want to compare more than two groups.

For the two-sample t-test comparing the means of two groups, we could consider

$$H_0 : \mu_{PET} = \mu_{FRIEND}$$

and

$$H_A : \mu_{PET} \neq \mu_{FRIEND}.$$

What if we have three groups? Our null might be that the groups are all the same, e.g.

$$H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER},$$

and our alternative would be the complement of  $H_0$ , or that there is some type of difference between the groups.

# ANOVA Null Hypothesis

In ANOVA, we typically follow this testing procedure.

1. Conduct an *overall* or *chunk* test of the null hypothesis that the means of all of the groups are equal.

# ANOVA Null Hypothesis

In ANOVA, we typically follow this testing procedure.

1. Conduct an *overall* or *chunk* test of the null hypothesis that the means of all of the groups are equal.
  - 1.1 If you reject this null hypothesis, then we usually *step down* to see which means are different from each other. A *multiple comparisons* correction is often used for these pairwise comparisons of means.

# ANOVA Null Hypothesis

In ANOVA, we typically follow this testing procedure.

1. Conduct an *overall* or *chunk* test of the null hypothesis that the means of all of the groups are equal.
  - 1.1 If you reject this null hypothesis, then we usually *step down* to see which means are different from each other. A *multiple comparisons* correction is often used for these pairwise comparisons of means.
  - 1.2 If you do not reject the null hypothesis, this means our data are consistent with a population in which the means are all equal. Generally, no further testing should be done.

# ANOVA Alternative Hypothesis

For ANOVA with three groups, our null hypothesis is

$H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER}$ . What could happen under the alternative?

►  $\mu_{PET} \neq \mu_{FRIEND} \neq \mu_{NEITHER}$

If we reject the null hypothesis, any of these situations could be true, and we may wish to conduct further tests to discover what setting we are in. Conducting further tests without rejecting the overall test of  $H_0$  will lead to an inflated type I error rate unless we use another method to adjust for multiple comparisons.

# ANOVA Alternative Hypothesis

For ANOVA with three groups, our null hypothesis is  $H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER}$ . What could happen under the alternative?

- ▶  $\mu_{PET} \neq \mu_{FRIEND} \neq \mu_{NEITHER}$
- ▶  $\mu_{PET} = \mu_{FRIEND}$  but  $\mu_{NEITHER}$  is different

If we reject the null hypothesis, any of these situations could be true, and we may wish to conduct further tests to discover what setting we are in. Conducting further tests without rejecting the overall test of  $H_0$  will lead to an inflated type I error rate unless we use another method to adjust for multiple comparisons.

# ANOVA Alternative Hypothesis

For ANOVA with three groups, our null hypothesis is  $H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER}$ . What could happen under the alternative?

- ▶  $\mu_{PET} \neq \mu_{FRIEND} \neq \mu_{NEITHER}$
- ▶  $\mu_{PET} = \mu_{FRIEND}$  but  $\mu_{NEITHER}$  is different
- ▶  $\mu_{PET} = \mu_{NEITHER}$  but  $\mu_{FRIEND}$  is different

If we reject the null hypothesis, any of these situations could be true, and we may wish to conduct further tests to discover what setting we are in. Conducting further tests without rejecting the overall test of  $H_0$  will lead to an inflated type I error rate unless we use another method to adjust for multiple comparisons.



# ANOVA Alternative Hypothesis

For ANOVA with three groups, our null hypothesis is  $H_0 : \mu_{PET} = \mu_{FRIEND} = \mu_{NEITHER}$ . What could happen under the alternative?

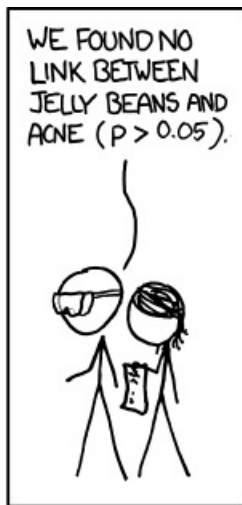
- ▶  $\mu_{PET} \neq \mu_{FRIEND} \neq \mu_{NEITHER}$
- ▶  $\mu_{PET} = \mu_{FRIEND}$  but  $\mu_{NEITHER}$  is different
- ▶  $\mu_{PET} = \mu_{NEITHER}$  but  $\mu_{FRIEND}$  is different
- ▶  $\mu_{FRIEND} = \mu_{NEITHER}$  but  $\mu_{PET}$  is different

If we reject the null hypothesis, any of these situations could be true, and we may wish to conduct further tests to discover what setting we are in. Conducting further tests without rejecting the overall test of  $H_0$  will lead to an inflated type I error rate unless we use another method to adjust for multiple comparisons.

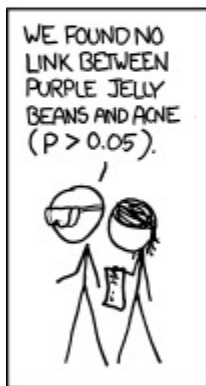
# Conducting Further Tests

So what if we go ahead and conduct further tests if the overall test is not rejected?

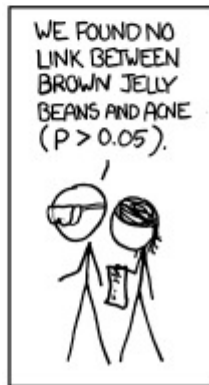
# Multiple Comparisons in Action: Overall Test



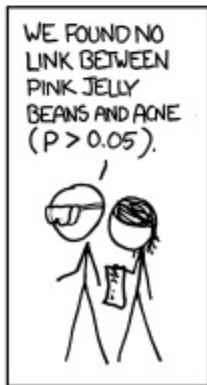
## Extra Test 1: Wild Blackberry



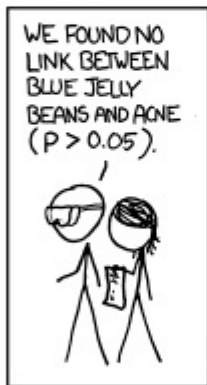
## Extra Test 2: Root Beer



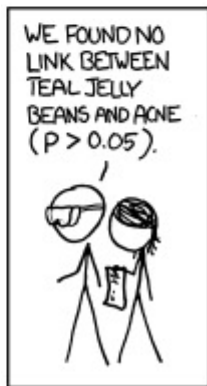
## Extra Test 3: Cotton Candy



## Extra Test 4: Plum

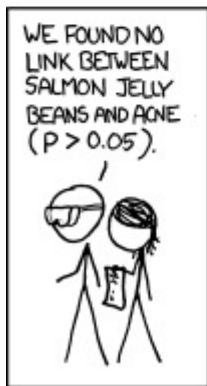


## Extra Test 5: Blueberry

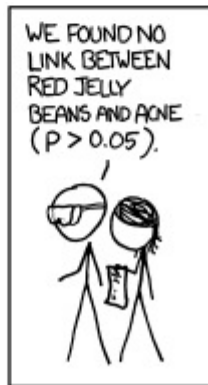




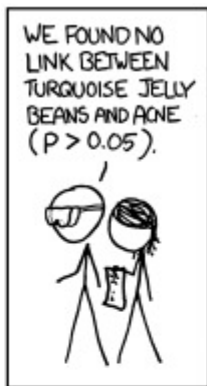
## Extra Test 6: Pink Grapefruit



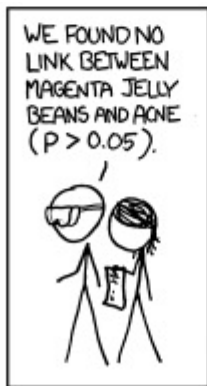
## Extra Test 7: Strawberry Jam



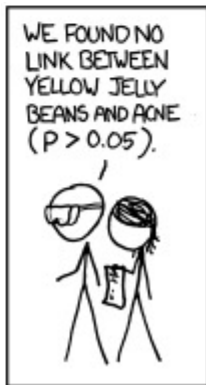
## Extra Test 8: Berry Blue



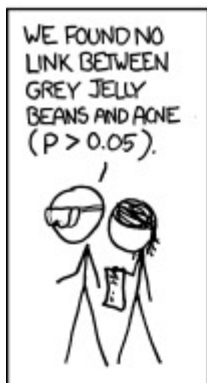
## Extra Test 9: Pomegranate



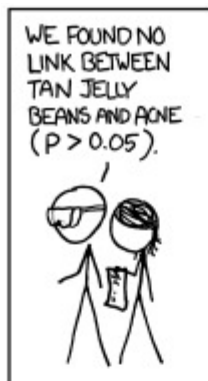
## Extra Test 10: Buttered Popcorn



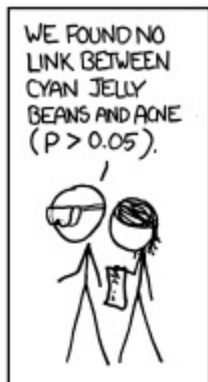
## Extra Test 11: Black Pepper (A Bertie Bott Specialty!)



## Extra Test 12: Ear Wax (Another Bertie Bott Bean!)



## Extra Test 13: Soap (Our last Bertie Bott Bean!)

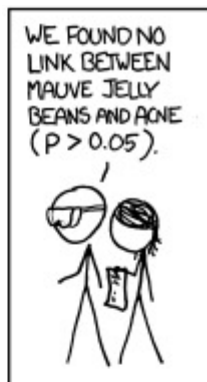


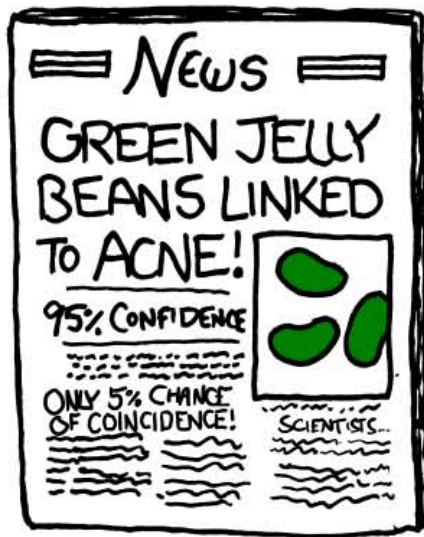


## Extra Test 14: Green Apple



## Extra Test 15: Strawberry Cheesecake





Disclaimer: So, uh, we did the green study again and got no link. It was probably a – “RESEARCH CONFLICTED ON GREEN JELLY BEAN-ACNE LINK; MORE STUDY RECOMMENDED!” headline.

# Why Analyze Variance?

ANOVA stands for *analysis of variance*.

What does variance have to do with our null hypothesis, which is about equality of means, say  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$  for  $K$  means?

Stay tuned!

# General ANOVA Setup

►  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$

# General ANOVA Setup

- ▶  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$
- ▶  $H_A : \text{at least one of the means is different}$

# General ANOVA Setup

- ▶  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$
- ▶  $H_A$  : at least one of the means is different
  - ▶ NOT ONLY  $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_K$  (this would force all means to be different)



# General ANOVA Setup

- ▶  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$
- ▶  $H_A$  : at least one of the means is different
  - ▶ NOT ONLY  $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_K$  (this would force all means to be different)
- ▶ Let the response of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  
 $i = 1, 2, \dots, K, j = 1, \dots, n_i$

# General ANOVA Setup

- ▶  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$
- ▶  $H_A$  : at least one of the means is different
  - ▶ NOT ONLY  $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_K$  (this would force all means to be different)
- ▶ Let the response of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  
 $i = 1, 2, \dots, K, j = 1, \dots, n_i$
- ▶  $K$  is the number of groups

# General ANOVA Setup

- ▶  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$
- ▶  $H_A$  : at least one of the means is different
  - ▶ NOT ONLY  $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_K$  (this would force all means to be different)
- ▶ Let the response of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  
 $i = 1, 2, \dots, K, j = 1, \dots, n_i$
- ▶  $K$  is the number of groups
- ▶  $n_i$  is the number of subjects in group  $i$

# General ANOVA Setup

- ▶ ANOVA model is

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

where we assume the  $\varepsilon_{ij}$  are independent and normally distributed with mean 0 and variance  $\sigma^2$ .

# General ANOVA Setup

- ▶ ANOVA model is

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

where we assume the  $\varepsilon_{ij}$  are independent and normally distributed with mean 0 and variance  $\sigma^2$ .

- ▶  $\varepsilon_{ij}$  represent random error about each group mean,  $\mu_i$

# General ANOVA Setup

- ▶ ANOVA model is

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

where we assume the  $\varepsilon_{ij}$  are independent and normally distributed with mean 0 and variance  $\sigma^2$ .

- ▶  $\varepsilon_{ij}$  represent random error about each group mean,  $\mu_i$
- ▶ Model assumes the  $y_{ij}$  are independent and normally distributed with mean  $\mu_i$  and variance  $\sigma^2$

# General ANOVA Setup

- ▶ ANOVA model is

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

where we assume the  $\varepsilon_{ij}$  are independent and normally distributed with mean 0 and variance  $\sigma^2$ .

- ▶  $\varepsilon_{ij}$  represent random error about each group mean,  $\mu_i$
- ▶ Model assumes the  $y_{ij}$  are independent and normally distributed with mean  $\mu_i$  and variance  $\sigma^2$
- ▶ Sometimes you see the model broken out with  $\mu_i = \mu + \alpha_i$ , where  $\mu$  represents the overall or *grand mean* and  $\alpha_i$  represents each group's deviation from the overall mean

# Assumptions of ANOVA

- ▶ Populations are normal
- ▶ Homoscedastic variance (same variance of individual observations in each group)
- ▶ Samples are independent

If these assumptions are violated, then results from ANOVA may not be valid. We will discuss some alternatives later in the course.



# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  
 $i = 1, 2, 3, j = 1, \dots, 15$

# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  $i = 1, 2, 3$ ,  $j = 1, \dots, 15$
- ▶ We have three groups (pets, friends, neither) and  $n_1 = n_2 = n_3 = 15$  subjects in each group for a total of 45 people with measurements.

# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  $i = 1, 2, 3$ ,  $j = 1, \dots, 15$
- ▶ We have three groups (pets, friends, neither) and  $n_1 = n_2 = n_3 = 15$  subjects in each group for a total of 45 people with measurements.
- ▶ For each group, calculate the group mean and variance, so now we have  $\bar{y}_1$  and  $s_1^2$  (pet group mean and variance),  $\bar{y}_2$  and  $s_2^2$  (friend group mean and variance), and  $\bar{y}_3$  and  $s_3^2$  (mean and variance for neither).

# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  $i = 1, 2, 3$ ,  $j = 1, \dots, 15$
- ▶ We have three groups (pets, friends, neither) and  $n_1 = n_2 = n_3 = 15$  subjects in each group for a total of 45 people with measurements.
- ▶ For each group, calculate the group mean and variance, so now we have  $\bar{y}_1$  and  $s_1^2$  (pet group mean and variance),  $\bar{y}_2$  and  $s_2^2$  (friend group mean and variance), and  $\bar{y}_3$  and  $s_3^2$  (mean and variance for neither).
- ▶ ANOVA is concerned with two different variances:

# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  $i = 1, 2, 3$ ,  $j = 1, \dots, 15$
- ▶ We have three groups (pets, friends, neither) and  $n_1 = n_2 = n_3 = 15$  subjects in each group for a total of 45 people with measurements.
- ▶ For each group, calculate the group mean and variance, so now we have  $\bar{y}_1$  and  $s_1^2$  (pet group mean and variance),  $\bar{y}_2$  and  $s_2^2$  (friend group mean and variance), and  $\bar{y}_3$  and  $s_3^2$  (mean and variance for neither).
- ▶ ANOVA is concerned with two different variances:
  - ▶ *Within-groups variance*: variance of the individual observations around their respective group means

# ANOVA for Pet Data

- ▶ Let the pulse rate of subject  $j$  in group  $i$  be given by  $y_{ij}$ ,  $i = 1, 2, 3$ ,  $j = 1, \dots, 15$
- ▶ We have three groups (pets, friends, neither) and  $n_1 = n_2 = n_3 = 15$  subjects in each group for a total of 45 people with measurements.
- ▶ For each group, calculate the group mean and variance, so now we have  $\bar{y}_1$  and  $s_1^2$  (pet group mean and variance),  $\bar{y}_2$  and  $s_2^2$  (friend group mean and variance), and  $\bar{y}_3$  and  $s_3^2$  (mean and variance for neither).
- ▶ ANOVA is concerned with two different variances:
  - ▶ *Within-groups variance*: variance of the individual observations around their respective group means
  - ▶ *Between-groups variance*: variance of the group means around the overall mean of all observations,  $\bar{y}$ .

## Reminder: Variance

Variance of  $x$  is estimated as

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}.$$

# Within-groups variance

Each group's variance  $s^2$  is a measure of the variance of the individuals around that population group mean. To get a pooled estimate of the common variance of individuals around their group means, we can calculate

$$s_W^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_K - 1)s_K^2}{n - K},$$

where  $K$  is the number of groups and  $n = n_1 + n_2 + \cdots + n_K$ . We can think of the within-groups variance as the inherent variability in the population.



# Between-groups variance

The between groups variance is estimated by

$$s_B^2 = \frac{n_1(\bar{y}_1 - \bar{y}.)^2 + n_2(\bar{y}_2 - \bar{y}.)^2 + \cdots + n_K(\bar{y}_K - \bar{y}.)^2}{K - 1}.$$

We can think of the between-groups variance as the sum of inherent variability *and* any kind of systematic variability due to the group effect.

Both  $s_W^2$  and  $s_B^2$  are estimates of the population variance  $\sigma^2$  if  $H_0$  is true. Now, if the sample means vary around the overall mean (this variance measured by  $s_B^2$ ) more than the individual observations vary around the sample means (measured by  $s_W^2$ ), we have evidence that the corresponding population means are in fact different (so that all  $K$  means are not the same).

How do we compare the variances? Consider

$$F = \frac{\text{inherent variability} + \text{treatment effect}}{\text{inherent variability}}.$$

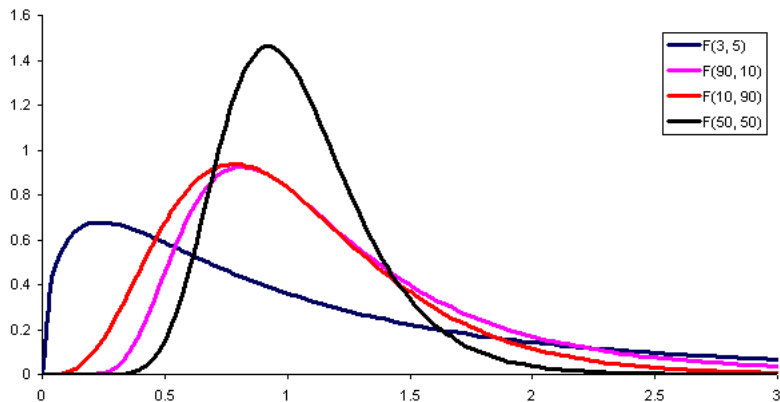
If there is little to no treatment effect (no treatment effect implies  $H_0$  is true), this ratio will be very close to 1.

Thus the F statistic is given by

$$F = \frac{s_B^2}{s_W^2},$$

which if  $H_0$  is true has an  $F$  distribution with  $K - 1$  and  $n - K$  df. The df associated with  $s_B^2$  are called the *numerator degrees of freedom* and correspond to the total number of groups minus 1. The df associated with  $s_W^2$  are called the *denominator degrees of freedom* and equal the total sample size minus the number of groups. For ANOVA the test is inherently one-tailed, rejecting  $H_0$  only if  $F$  is considerably larger than one. (This does not mean we have a one-sided alternative; we just look at one tail of the F distribution to get the p-value.)

# F distribution



The numerator and denominator degrees of freedom (ndf and ddf) determine the cutpoint for a given  $\alpha$  level. The ndf is one less than the number of groups, and the ddf is the total number of observations minus the number of groups.

$$F = \frac{s_B^2}{s_W^2}$$

If there are only two groups, then the F test gives the same result as the t-test.

This nice visual ANOVA illustrator shows how variances can give us information about the equality of means.

# Stata Code: One Way ANOVA for Pet Data

```
. oneway Pulse Group, tabulate means standard bonferroni
```

Group	Summary of Pulse	
	Mean	Std. Dev.
Pet prese	<b>73.483334</b>	<b>9.9702037</b>
Friend pr	<b>91.325333</b>	<b>8.3403223</b>
Neither p	<b>82.525333</b>	<b>9.2407657</b>
Total	<b>82.444667</b>	<b>11.627437</b>

# F Test for Pet Data

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	2387.67352	2	1193.83676	14.08	0.0000
Within groups	3561.00762	42	84.7858958		
Total	5948.68115	44	135.197299		

Bartlett's test for equal variances:  $\chi^2(2) = 0.4306$  Prob> $\chi^2 = 0.806$

$F = \frac{s_B^2}{s_W^2} = \frac{1193.8}{84.8} = 14.08$ , with  $\text{ndf} = 3 - 1 = 2$  and  $\text{ddf} = 45 - 3 = 42$ , has a p-value of  $< 0.001$ . We conclude that at least one of the three groups comes from a population with a different mean from the others.

Next: which groups are different?

# Bonferroni Correction

As we showed earlier, conducting multiple tests on a data set increases the *family-wise error rate*. One very conservative way to ensure this is not the case is to divide  $\alpha$  by the number of tests to be done and to use that as the significance level. This procedure is called the Bonferroni correction. For example for two tests, to preserve an overall 0.05 type I error rate, the Bonferroni correction would use  $\frac{\alpha}{2} = 0.025$  as the significance level for each individual test instead of 0.05. Bonferroni is a conservative multiple comparisons correction, making it harder to reject the null hypothesis, but it is a safe bet in controlling the type I error rate.



# Pets and Stress: Group Differences

We can compare the groups using a Bonferroni correction (here we have three tests, so the significance level for each test is  $\frac{\alpha}{3}$ ). Stata handles this by multiplying each p-value by 3 before showing it to you, so you can still use  $\alpha$  to assess significance of these pairwise comparisons.

Comparison of Pulse by Group (Bonferroni)			
Row Mean- Col Mean	Pet pres	Friend p	
Friend p	<b>17.842</b> <b>0.000</b>		
Neither	<b>9.042</b> <b>0.031</b>	<b>-8.8</b> <b>0.037</b>	

What do we conclude?

# Two-way ANOVA

ANOVA can be used with more than one group. For example, one group in our study might be the companion (friend, pet, or neither), and another group could be the underlying trait anxiety level (e.g., high anxiety or not).

You can learn more about ANOVA in BIOS 545 or BIOS 663 (the latter is for students comfortable with linear algebra as vector and matrix operations are important). A follow-up course will teach you more about models with more than one explanatory variable (of major practical importance!) and also how to check whether the ANOVA assumptions are valid.

## In practice: ANOVA

In 2009, *Women's Health* reported results of a study by Ro and Choi examining social status correlates of reports of lifetime racial discrimination (measured as a continuous score obtained from a well-validated scale) in a sample of 754 women.

<b>Social Status Characteristic</b>	<b>Mean Discrimination Score</b>
<b>Race</b>	
African American	0.65**
Asian	0.32
Latina/Hispanic	0.32
Caucasian	0.17
<b>Financial Difficulty</b>	
Yes	0.30**
No	0.16

\*\*  $p < 0.05$  How do you interpret the results based on their table?

# In practice: ANOVA

In 2012, de Goede et al. published results of their study of the degree to which Dutch local health officials utilized epidemiologic research in their public health practice in *Health Research Policy and Systems*. They surveyed  $N = 155$  local health officials and constructed a (non-validated, pretty non-normal, but we'll close our eyes for now) scale to measure direct use of research in decision making. We'll examine the roles that prior experience with research and media attention on an issue play in the decision to use epidemiologic research in public health practice.

# In practice: ANOVA

Factor	Mean research use score	p-value
<b>Personal experience</b>		
<b>with research</b>		$< 0.01^\dagger$
None (referent)	2.2	NA
Mainly experience with qualitative research	2.6	0.13
Mainly experience with quantitative research	1.7	0.09
Experience with both	2.6	0.09

$^\dagger$  3 df F test

How do you explain this table?

# In practice: ANOVA

Factor	Mean research use score	p-value
<b>Type of media attention</b>		$< 0.05^\dagger$
None (referent)	2.2	NA
Mainly positive media	1.6	0.11
Mainly negative media	3.2	0.29
Mainly neutral media	1.7	0.16
Variable media	1.6	0.20
No familiarity with media	1.4	0.02

<sup>†</sup> 5 df F test

How do you explain this table?

# Reading for After Fall Break!

- ▶ Pagano and Gauvreau, Chapter 13