

**UCLA STAT 13**  
**Introduction to Statistical Methods for the  
 Life and Health Sciences**

**Instructor: Ivo Dinov,**  
 Asst. Prof. of Statistics and Neurology

**Teaching Assistants:**  
 Jacqueline Dacosta & Chris Barr

University of California, Los Angeles, Fall 2006

[http://www.stat.ucla.edu/~dinov/courses\\_students.html](http://www.stat.ucla.edu/~dinov/courses_students.html)

Slide 1 Stat 13, UCLA, Ivo Dinov

Chapter 11  
 Analysis of Variance - ANOVA

Slide 2 Stat 13, UCLA, Ivo Dinov

**Comparing the Means of I Independent Samples**

- In Chapter 7 we considered the comparisons of two independent group means using the independent t test
- We need to expand our thinking to compare I independent samples
- The procedure we will use is called Analysis of Variance (ANOVA)

Slide 3 Stat 13, UCLA, Ivo Dinov

**Comparing the Means of I Independent Samples**

**Example:** 5 varieties of peas are currently being tested by a large agribusiness cooperative to determine which is best suited for production. A field was divided into 20 plots, with each variety of peas planted in four plots. The yields (in bushels of peas) produced from each plot are shown in the table below:

Variety of Pea		A	B	C	D	E
	1	26.2	29.2	29.1	21.3	20.1
	2	24.3	28.1	30.8	22.4	19.3
	3	21.8	27.3	33.9	24.3	19.9
	4	28.1	31.2	32.8	21.8	22.1

Slide 4 Stat 13, UCLA, Ivo Dinov

**Comparing the Means of I Independent Samples**

- In applying ANOVA, the data are regarded as random samples from k populations
- Notation (let sub-indices 1 = A, 2 = B, etc...):
  - Population means:  $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$
  - Population standard deviations:  $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$

Slide 5 Stat 13, UCLA, Ivo Dinov

**Issues in ANOVA**

- We have five group means to compare
- Why not just carry out a bunch of t tests?
  - Repeated t tests would mean:
    - Ho:  $\mu_1 = \mu_2$       Ho:  $\mu_2 = \mu_3$       Ho:  $\mu_3 = \mu_4$
    - Ho:  $\mu_4 = \mu_5$       Ho:  $\mu_2 = \mu_4$       Ho:  $\mu_2 = \mu_5$
    - Etc...
- We would have to make  $\binom{5}{2} = 10$  comparisons
- What is so bad about that?

Slide 6 Stat 13, UCLA, Ivo Dinov

## Issues in ANOVA

- Each test is carried out at  $\alpha = 0.05$ , so a type I error is 5% for each
- The overall risk of a type I error is larger than 0.05 and gets larger as the number of groups (I) gets larger
- SOLUTION: Need to make multiple comparisons with an overall error of  $\alpha = 0.05$  (or whichever level is specified).

Slide 7 Stat 13, UCLA, Jon Dinger

## Issues in ANOVA

- There are other positive aspects of using ANOVA:
  - Can see if there is a trend within the I groups; low to high
  - Estimation of the standard deviation
    - Global sharing of information of all data yields precision in the analysis
- The main idea behind ANOVA is that we need to know how much inherent variability there is in the data before we can judge whether there is a difference in the sample means

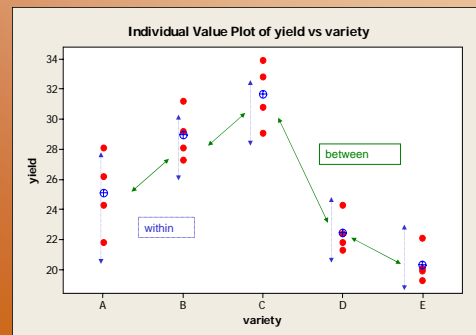
Slide 8 Stat 13, UCLA, Jon Dinger

## Issues in ANOVA

- To make an inference about means we compare two types of variability:
  - variability between sample means
  - variability within each group
- It is very important that we keep these two types of variability in mind as we work through the following formulas
- It is our goal to come up with a numeric quantity that describes each of these variability's

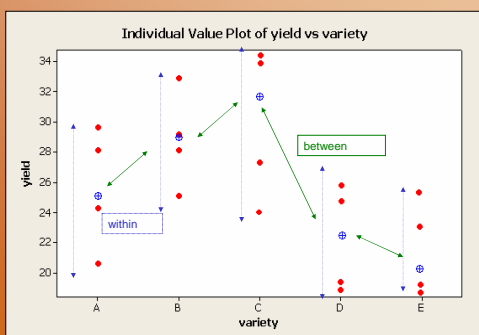
Slide 9 Stat 13, UCLA, Jon Dinger

## Issues in ANOVA



Slide 10 Stat 13, UCLA, Jon Dinger

## Issues in ANOVA



Slide 11 Stat 13, UCLA, Jon Dinger

## The Basic ANOVA

- Because we now have I groups each with it's own observations, we need to modify our notation
  - Notation:  $y_{ij}$  = group i observation j
- For the pea example:
  - $y_{11} = 26.2$
  - $y_{12} = 24.3$
  - ...
  - $y_{21} = 29.2$
  - ...
  - $y_{54} = 22.1$

Slide 12 Stat 13, UCLA, Jon Dinger

## The Basic ANOVA

- More notation:
  - $I$  = number of groups
  - $n_i$  = number of observations in group  $i$
  - $n^*$  = total number of observations =  $n_1 + n_2 + \dots + n_i$

Slide 13 Stat 13, UCLA, Jon Dineen

## The Basic ANOVA

- Formulae:

The group mean for group  $i$  is: 
$$\bar{y}_{i\cdot} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

The grand mean is: 
$$\bar{y}_{\cdot\cdot} = \frac{\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}}{n^*}$$

To compute the difference between the means we will compare each group mean to the grand mean

Slide 14 Stat 13, UCLA, Jon Dineen

## Variation Between Groups

- Goal #1 is to describe the variation between the groups means
- RECALL: For the independent  $t$  test we described the difference between two group means as  $\bar{y}_1 - \bar{y}_2$
- In ANOVA we describe the difference between  $I$  means as sums of squares between:
 
$$SS(\text{between}) = \sum_{i=1}^I n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

Can be thought of as the difference between each group mean and the grand mean – look at the formula

Slide 15 Stat 13, UCLA, Jon Dineen

## Variation Between Groups

- As our other measures of variation have used in the past is degrees of freedom,  $SS(\text{between})$  also has degrees of freedom
 
$$df(\text{between}) = I - 1$$
- Finally our measure of between group variability is mean square between:

$$MS(\text{between}) = \frac{SS(\text{between})}{df(\text{between})}$$

This measures variability between the sample means

Slide 16 Stat 13, UCLA, Jon Dineen

## Variation Within Groups

- Goal #2 is to describe the variation within the groups
- RECALL: To measure the variability within a single sample we used:
 
$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}}$$
- In ANOVA to describe the combined variation within  $I$  groups we use sums of squares within:

$$SS(\text{within}) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

Can be thought of as the combination of variation within the  $I$  groups

Slide 17 Stat 13, UCLA, Jon Dineen

## Variation Within Groups

- $SS(\text{within})$  also has degrees of freedom
 
$$df(\text{within}) = n^* - I$$
- Finally our measure of within variability is mean square within:

$$MS(\text{within}) = \frac{SS(\text{within})}{df(\text{within})}$$

This is a measure of variability within the groups

Slide 18 Stat 13, UCLA, Jon Dineen

### More on MS (within)

- The quantity for MS(within) is a measure of variability within the groups
- If there were only one group with n observations, then

$$SS(\text{within}) = \sum_{j=1}^n (y_j - \bar{y})^2$$

$$df(\text{within}) = n - 1$$

$$MS(\text{within}) = \frac{\sum_{j=1}^n (y_j - \bar{y})^2}{n - 1}$$

← This was  $s^2$  from chapter 2!

Slide 19 Stat 13, UCLA, Ivo Dinno

### More on MS (within)

- ANOVA deals with several groups simultaneously. MS(within) is a combination of the variances of the groups
- It is pooling together measurements of variability from the different groups
- With similar logic MS(within) for two groups can be transformed into the pooled standard deviation

- remember our talk in chapter 7 about the pooled and unpooled methods?

$$S_{\text{pooled}} = \sqrt{MS(\text{within})}$$

Slide 20 Stat 13, UCLA, Ivo Dinno

### A Fundamental Relationship of ANOVA

- The last formula based discussion we need to have is regarding the total variability in the data

$$(y_{ij} - \bar{y}_{..}) = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})$$

↑  
Deviation of an observation from the grand mean  
= Total variability

↑  
within

↑  
between

Slide 21 Stat 13, UCLA, Ivo Dinno

### A Fundamental Relationship of ANOVA

- This also corresponds to the sums of squares:

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^I n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

This means  $SS(\text{total}) = SS(\text{within}) + SS(\text{between})$

SS(total) measures the variability among all  $n^*$  observations in the  $I$  groups

$$df(\text{total}) = df(\text{within}) + df(\text{between}) = (n^* - I) + (I - 1) = n^* - 1$$

Slide 22 Stat 13, UCLA, Ivo Dinno

### ANOVA Calculations

- You've probably noticed that we haven't crunched any of these numbers yet
- Calculations are fairly intense
- Computers are going to rescue us: SOCR

Slide 23 Stat 13, UCLA, Ivo Dinno

### ANOVA Calculations

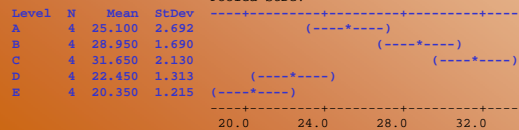
#### Example: peas (cont')

NOTE: between and within variances may be referred to as: SST (treatment=variety) and SSE (Error, within)

One-way ANOVA: yield versus variety

Source	DF	SS	MS	F	P
variety	4	342.04	85.51	23.97	0.000
Error	15	53.52	3.57		
Total	19	395.56			

S = 1.889 R-Sq = 86.47% R-Sq(adj) = 82.86%  
Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 1.889 Slide 24 Stat 13, UCLA, Ivo Dinno

## The ANOVA table

- Standard for all ANOVA's
- also helps keep you organized

Source	df	SS	MS
Between	$l - 1$	$\sum_{i=1}^l n_i (\bar{y}_i - \bar{y}_{..})^2$	$\frac{SS(\text{between})}{df(\text{between})}$
Within	$n - l$	$\sum_{i=1}^l \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$\frac{SS(\text{within})}{df(\text{within})}$
Total	$n - 1$	$\sum_{i=1}^l \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$	

Slide 25 Stat 13, UCLA, Ivo Dinno

## The Global F Test

- This is our hypothesis test for ANOVA
- #1 General form of the hypotheses:
  - $H_0: \mu_1 = \mu_2 = \dots = \mu_l$
  - $H_a$ : at least two of the  $\mu_k$ 's are different
- $H_0$  is compound when  $l > 2$ , so rejecting  $H_0$  doesn't tell us which  $\mu_k$ 's are not equal, only that two are different

Slide 26 Stat 13, UCLA, Ivo Dinno

## The Global F Test

- #2 The test statistic:

$$F_s = \frac{MS(\text{between})}{MS(\text{within})}$$

- $F_s$  will be large if there is a lot of between variation when compared to the within variation
  - discrepancies in the means are large relative to the variability within the groups
- Large values of  $F_s$  provide evidence against  $H_0$

Slide 27 Stat 13, UCLA, Ivo Dinno

## The Global F Test

- #3 The p-value
  - based on the F distribution
  - named after Fisher
  - depends on numerator df and denominator df
  - Table 10 pgs 687 – 696 (or SOCR resource)
- #4 The conclusion (TBD)

Slide 28 Stat 13, UCLA, Ivo Dinno

## Example

- Is there a significant difference between these 3 groups at  $\alpha = 0.05$ ?

$$F = \frac{MS(\text{between})}{MS(\text{within})} = \frac{MSST}{MSSE} = \frac{SST / df}{MSSE / df} \sim F(2, 4)$$

$$\frac{SST / 2}{MSSE / 4} = \frac{19.86 / 2}{3 / 4} = 13.24; p\text{-value} = 0.017 \text{ (SOCR)}$$

$$\bar{y} = \frac{13}{7} = 1.86; SST = SS(\text{Between}) =$$

$$\sum_{i=1}^3 n_i (\bar{y}_i - \bar{y}_{..})^2 = 2(1.36)^2 + 3(0.86)^2 + 2(2.64)^2 = 19.86$$

$$SSE = SS(\text{Within}) = \sum_{i=1}^3 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 =$$

$$= [0.5^2 + 0.5^2] + [0^2 + 1^2 + 1^2] + [0.5^2 + 0.5^2] = 3$$

	A	B	C
0	1	4	
1	0	5	
2		2	
N	2	3	2
S	1	3	9
$\mu$	0.5	1	4.5

$$H_0: \mu_1 = \mu_2 = \dots = \mu_l$$

$$H_a: \text{some } \mu_i \neq \mu_j \Rightarrow \text{Reject } H_0!$$

Slide 29 Stat 13, UCLA, Ivo Dinno

## The Global F Test

### Example: Peas (cont')

Do the data provide evidence to suggest that there is a difference in yield among the five varieties of peas? Test using  $\alpha = 0.05$ .

$$H_0: \mu_1 = \mu_2 = \dots = \mu_l$$

where 1 = A, 2 = B, etc...

$H_a$ : at least two of the  $\mu_k$ 's are different

Slide 30 Stat 13, UCLA, Ivo Dinno

## The Global F Test

### One-way ANOVA: yield versus variety

Source	DF	SS	MS	F	P
variety	4	342.04	85.51	23.97	0.000
Error	15	53.52	3.57		
Total	19	395.56			

$F = \frac{85.51}{3.57} = 23.97$   
 $p = 0.000$

S = 1.889 R-Sq = 86.47% R-Sq(adj) = 82.86%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
A	4	25.100	2.692
B	4	28.950	1.690
C	4	31.650	2.130
D	4	22.450	1.313
E	4	20.350	1.215

Pooled StDev = 1.889

Slide 31 Stat 13, UCLA, Ivo Dinno

## The Global F Test

- Because  $0.000 < 0.05$  we will reject  $H_0$ .
- CONCLUSION: The data show that at least two of the true mean yields of the five varieties of peas, are statistically significantly different ( $p = 0.000$ ).
- Notice we can only say that at least two of the means are different
  - not which two are different!
  - not all means are different!

Slide 32 Stat 13, UCLA, Ivo Dinno

## The Global F Test

### Example: Peas (cont')

- Suppose we need to get the p-value using the table:
- Back to bracketing!
  - numerator df = 4
  - denominator df = 15

$p < 0.0001$ , so we will again reject  $H_0$   
 Don't need to worry about doubling!

Slide 33 Stat 13, UCLA, Ivo Dinno

## Practice

**Example:** Parents are frequently concerned when their child seems slow to begin walking. In 1972 *Science* reported on an experiment in which the effects of several different treatments on the age at which a child's first walks were compared. Children in the first group were given special walking exercises for 12 minutes daily beginning at the age 1 week and lasting 7 weeks. The second group of children received daily exercises, but not the walking exercises administered to the first group. The third and fourth groups received no special treatment and differed only in that the third group's progress was checked weekly and the fourth was checked only at the end of the study. Observations on age (months) when the child began to walk are on the next slide

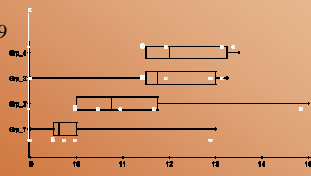
Slide 34 Stat 13, UCLA, Ivo Dinno

## Practice

Grp_1	Grp_2	Grp_3	Grp_4
9	11	11.5	13.25
9.5	10	12	11.5
9.75	10	9	12
10	11.75	11.5	13.5
13	10.5	13.25	11.5
9.5	15	13	

Suppose  $\sum_{i=1}^4 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = 43.69$

and  $\sum_{i=1}^4 n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = 14.78$



Slide 35 Stat 13, UCLA, Ivo Dinno

## Practice

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_a$ : at least two of the  $\mu_k$ 's are different

Source	df	SS	MS	F
Between	$4 - 1 = 3$	14.78	$\frac{14.78}{3} = 4.93$	$\frac{4.93}{2.30} = 2.14$
Within	$23 - 4 = 19$	43.69	$\frac{43.69}{19} = 2.30$	
Total	$23 - 1 = 22$	58.47		

Slide 36 Stat 13, UCLA, Ivo Dinno

## Practice

With 3 numerator df and 19 denominator df

$0.1 < p < 0.2$ , so we fail to reject  $H_0$

CONCLUSION: These data show that [a child's true mean walking age is not statistically significantly different among any of the four treatment groups](#) ( $0.1 < p < 0.2$ ).

Now you try to replicate these results using the computer and the file [walking.mtw](#)

Slide 37 Stat 13, UCLA, Ivo Dinno

## Practice

One-way ANOVA: age versus treatment

Source	DF	SS	MS	F	P
treatment	3	14.78	4.93	2.14	0.129
Error	19	43.69	2.30		
Total	22	58.47			

S = 1.516 R-Sq = 25.28% R-Sq(adj) = 13.48%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
1	6	10.125	1.447	8.25	11.99
2	6	11.375	1.896	8.50	14.25
3	6	11.708	1.520	9.66	13.75
4	5	12.350	0.962	11.42	13.28

Pooled StDev = 1.516

Slide 38 Stat 13, UCLA, Ivo Dinno

## Applicability of Methods

- *Standard Conditions*
  - ANOVA is valid if:
    1. Design conditions:
      - a. Reasonable that groups of observations are random samples from their respective populations. Observations within each group must be independent of one another.
      - b. The I samples must be independent
    2. Population conditions:
      - The I population distributions must be approximately **normal** with equal standard deviations
$$\sigma_1 = \sigma_2 = \dots = \sigma_I$$
- \* normality is less crucial if the sample sizes are large

Slide 39 Stat 13, UCLA, Ivo Dinno

## Applicability of Methods

- *Verification of Conditions*
  - look for bias, hierarchy, and dependence.
  - normality and normal probability plot of each group.
  - standard deviations are approximately equal if (RULE OF THUMB):

$$\frac{\text{largest sd}}{\text{smallest sd}} < 2$$

If not, we cannot be confident in our p-value from the F distribution

Slide 40 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons

- Once we reject  $H_0$  for the ANOVA, we know that at least two of the  $\mu_k$ 's are different
- We need to find which group means are different, but we shouldn't use a bunch of independent t tests
  - We discussed in section 11.1 that each independent t test for each two group combination can inflate the overall risk of a type I error

Slide 41 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons

- A naïve approach would be to calculate one sample CI's for the mean using the pooled standard deviation
  - assumption that the sd's were approx. equal
  - look for overlap in the CI's, but the problem is that these are still 95% CI's with each  $\alpha = 0.05$

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
A	4	25.100	2.692	21.71	28.49
B	4	28.950	1.690	27.56	30.34
C	4	31.650	2.130	27.39	35.91
D	4	22.450	1.313	20.83	24.07
E	4	20.350	1.215	18.92	21.78

Slide 42 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons

- A better solution is to compare each group with an overall  $\alpha$  of 0.05.
  - for this we use a technique called a multiple comparison (MC) procedure
  - The idea is to compare means two at a time at a reduced significance level, to ensure an "overall"
  - There are many different MC
    - Bonferroni: simple and conservative
    - Newman-Keuls: less conservative/more powerful, but complicated
    - Tukey procedure: easy to use with MTB

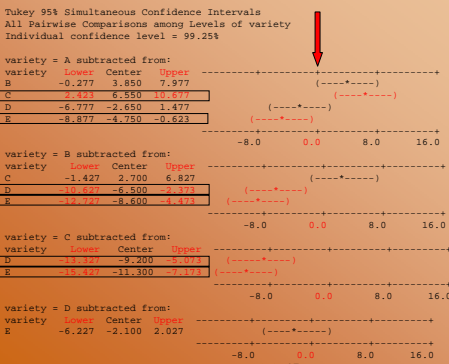
Slide 43 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons

- We will focus on the **Tukey method**
  - Uses confidence intervals for the difference in means
  - Confidence intervals similar to those in Chapter 7, for the difference of two means using an adjusted  $\alpha$ 
    - RECALL: The zero rule
  - We will rely on the computer to calculate these intervals

Slide 44 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons



Slide 45 Stat 13, UCLA, Ivo Dinno

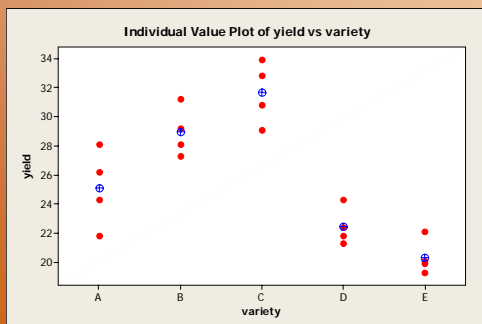
## Multiple Comparisons

What does this mean?

- The best way to summarize would be to think of the means in order from large to small and sight the differences (not necessary to repeat):
  - The true mean yield for variety A is statistically significantly different than the true means of varieties C, and E
  - The true mean yield for variety B is statistically significantly different than the true means of varieties D, and E
  - The true mean yield for variety C is statistically significantly different than the true means of variety A, D, and E (however C vs. A was previously mentioned)

Slide 46 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons



Slide 47 Stat 13, UCLA, Ivo Dinno

## Multiple Comparisons

**Example:** Walking age (cont')

We do not need to carry out Tukey's test for this data – why?

$$F = 2.14, p = 0.129$$

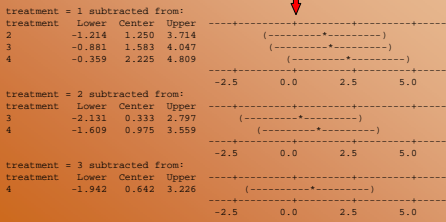
Slide 48 Stat 13, UCLA, Ivo Dinno



## Multiple Comparisons

Tukey 95% Simultaneous Confidence Intervals  
All Pairwise Comparisons among Levels of treatment

Individual confidence level = 98.89%



Slide 49

Stat 13, UCLA, Jon Dineen