UCLA STAT 13

Introduction to Statistical Methods for the Life and Health Sciences

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

Teaching Assistants:
Brandi Shanata & Tiffany Head

University of California, Los Angeles, Fall 2007
http://www.stat.ucla.edu/~dinov/courses_students.html

Chapter 11

Analysis of Variance - ANOVA

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)

Comparing the Means of I Independent Samples

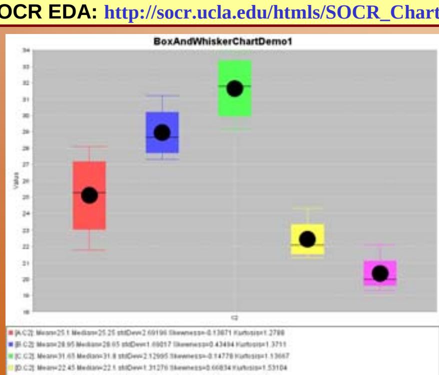
Example: 5 varieties of peas are currently being tested by a large agribusiness cooperative to determine which is best suited for production. A field was divided into 20 plots, with each variety of peas planted in four plots. The yields (in bushels of peas) produced from each plot are shown in the table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

Variety of Pea

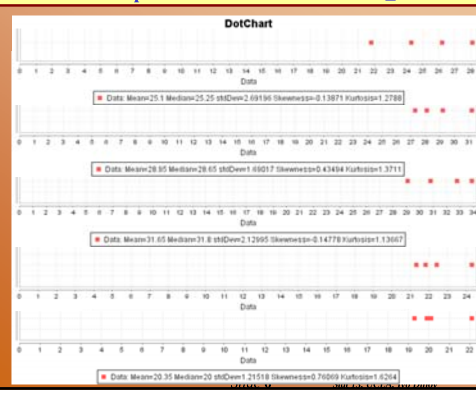
A	B	C	D	E
26.2	29.2	29.1	21.3	20.1
24.3	28.1	30.8	22.4	19.3
21.8	27.3	33.9	24.3	19.9
28.1	31.2	32.8	21.8	22.1

A	26.2,24.3,21.8,28.1
B	29.2,28.1,27.3,31.2
C	29.1,30.8,33.9,32.8
D	21.3,22.4,24.3,21.8
E	20.1,19.3,19.9,22.1

Comparing the Means of I Independent Samples
SOCR EDA: http://socr.ucla.edu/htmls/SOCR_Charts.html

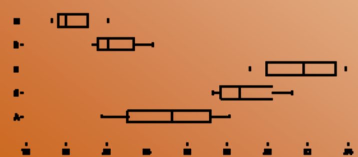


Comparing the Means of I Independent Samples
SOCR EDA: http://socr.ucla.edu/htmls/SOCR_Charts.html



Comparing the Means of I Independent Samples

- In applying ANOVA, the data are regarded as random samples from k populations
- Notation (let sub-indices1 = A, 2 = B, etc…):
  - Population means: $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$
  - Population standard deviations: $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$



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?

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).

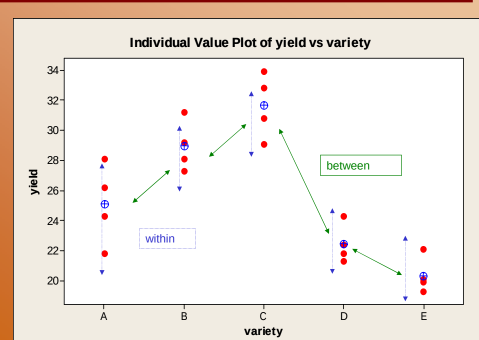
Issues in ANOVA

- 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

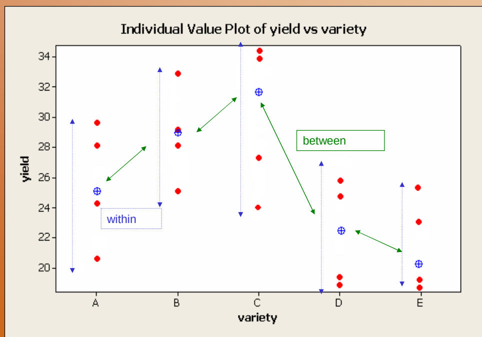
Issues in ANOVA

- 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

Issues in ANOVA



Issues in ANOVA



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

The Basic ANOVA

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

The Basic ANOVA

- Formulae:

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

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

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

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(between) = $\sum_{i=1}^{I} n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$

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

Variation Between Groups

- As our other measures of variation have used in the past is degrees of freedom, SS(between) also has degrees of freedom

  df (between) = I – 1

- Finally our measure of between group variability is mean square between:

  MS(between) = $\dfrac{SS(between)}{df\,(between)}$

  This measures variability between the sample means

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

Can be though of as the combination of variation within the I groups

Variation Within Groups

- SS(within) also has degrees of freedom
  df (within) = n* - I
- Finally our measure of within variability is mean square within:

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

This is a measure of variability within the groups

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

df(within) = n* - 1

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

This was s^2 from chapter 2!

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?

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

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

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(total) = SS(within) + SS(between)

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

df(total) = df(within) + df(between) = (n* - I) + (I – 1) = n* - 1

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

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
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  (----*----)
                         ----+---------+---------+---------+-----
                          20.0      24.0      28.0      32.0
Pooled StDev = 1.889
```


ANOVA Calculations
SOCR Analyses: socr.ucla.edu/htmls/SOCR_Analyses.html

Values	Group
26.2	A
24.3	A
21.8	A
28.1	A
29.2	B
28.1	B
27.3	B
31.2	B
29.1	C
30.8	C
33.9	C
32.8	C
21.3	D
22.4	D
24.3	D
21.8	D
20.1	E
19.3	E
19.9	E
22.1	E

ANOVA Calculations - Results
SOCR Analyses: socr.ucla.edu/htmls/SOCR_Analyses.html

Sample Size = 20
Independent Variable = Variety
Dependent Variable = Yield
Results of One-Way Analysis of Variance:
Model:
Degrees of Freedom = 4
Residual Sum of Squares = 342.040
Mean Square Error = 85.510
Error:
Degrees of Freedom = 15
Residual Sum of Squares = 53.520
Mean Square Error = 3.568
Corrected Total:
Degrees of Freedom = 19
Residual Sum of Squares = 395.560
F-Value = 23.966
P-Value = 2.2855121203368967E-6
R-Square = .865

Values	Group
26.2	A
24.3	A
21.8	A
28.1	A
29.2	B
28.1	B
27.3	B
31.2	B
29.1	C
30.8	C
33.9	C
32.8	C
21.3	D
22.4	D
24.3	D
21.8	D
20.1	E
19.3	E
19.9	E
22.1	E

The ANOVA table

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

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

The Global F Test

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

The Global F Test

- #2 The test statistic:

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

- 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_o

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)

Example

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

	A	B	C
	0	1	4
	1	0	5
		2	
N	2	3	2
S	1	3	9
μ	0.5	1	4.5

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

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

http://socr.stat.ucla.edu/test/SOCR_Analyses.html
http://socr.stat.ucla.edu/Applets.dir/Normal_T_Chi2_F_Tables.htm

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

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

$$SSE = SS(Within) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i.\right)^2 =$$

$$= \left[0.5^2 + 0.5^2\right] + \left[0^2 + 1^2 + 1^2\right] + \left[0.5^2 + 0.5^2\right] = 3$$

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_I$$

$$H_A: \text{some } \mu_i \neq \mu_j \Rightarrow \text{Reject } H_o!$$

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_o: $\mu_1 = \mu_2 = \ldots = \mu_I$

where 1 = A, 2 = B, etc…

H_a: at least two of the μ_k's are different

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

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    (----*----)
                         ----+---------+---------+---------+-----
                          20.0      24.0      28.0      32.0
Pooled StDev = 1.889
```

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

p = 0.000

The Global F Test

- Because 0.000 < 0.05 we will reject H_o.
- 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!

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_o

  Don't need to worry about doubling!

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 forth groups** received no special treatment and differed only in that the third group's progress was checked weekly and the forth was checked only at the end of the study. Observations on age (months) when the child began to walk are on the next slide

Practice

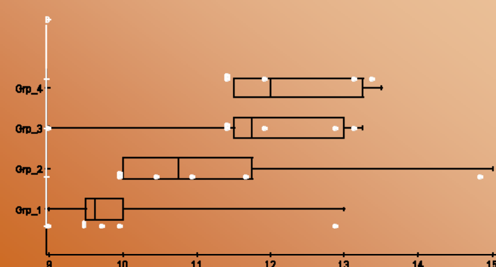
Grp_1	Grp_2	Grp_3	Grp_4
9	11	11.5	13.25
9.5	10	12	11.5
9.75	10	9	12
10	11.75	11.5	13.5
13	10.5	13.25	11.5
9.5	15	13	

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

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

Values	Group
9	1
9.5	1
9.75	1
10	1
13	1
9.5	1
11	2
10	2
10	2
11.75	2
10.5	2
15	2
11.5	3
12	3
9	3
11.5	3
13.25	3
13	3
13.25	4
11.5	4
12	4
13.5	4
11.5	4

Practice



ANOVA Practice

SOCR Analysis: socr.ucla.edu/htmls/SOCR_Analyses.html

Sample Size = 23
Independent Variable = TreatmentGroup
Dependent Variable = WalkingAge
Results of One-Way Analysis of Variance:
Model:
Degrees of Freedom = 3
Residual Sum of Squares = 14.778
Mean Square Error = 4.926
Error:
Degrees of Freedom = 19
Residual Sum of Squares = 43.690
Mean Square Error = 2.299
Corrected Total:
Degrees of Freedom = 22
Residual Sum of Squares = 58.467
F-Value = 2.142
P-Value = 0.128545596738329
R-Square = .253

Practice

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

H_a: at least two of the μ_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		

Practice

With 3 numerator df and 19 denominator df

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

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*

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  -+---------+---------+---------+--------
1      6  10.125  1.447  (--------*--------)
2      6  11.375  1.896          (--------*-------)
3      6  11.708  1.520            (--------*--------)
4      5  12.350  0.962                (--------*---------)
                         -+---------+---------+---------+--------
                        9.0      10.5      12.0      13.5
Pooled StDev = 1.516
```


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

$$\sigma_1 = \sigma_2 = \ldots = \sigma_I$$

* normality is less crucial if the sample sizes are large

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

Multiple Comparisons

- Once we reject Ho for the ANOVA, we know that at least two of the μ_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

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  ----+---------+---------+---------+-----
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  (----*----)
                         ----+---------+---------+---------+-----
                          20.0      24.0      28.0      32.0
```

Multiple Comparisons

- A better solution is to compare each group with an overall α 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

      Each CI calculated with (overall error rate)/(# of comparisons)
    - Newman-Keuls: less conservative/more powerful, but complicated
    - Tukey procedure: easy to use with MTB

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 α
    - RECALL: The zero rule
  - We will rely on the computer to calculate these intervals

Multiple Comparisons

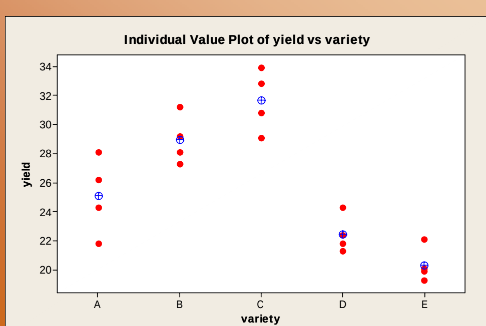
```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of variety
Individual confidence level = 99.25%

variety = A subtracted from:
variety   Lower  Center   Upper  ---------+---------+---------+---------+
B        -0.277   3.850   7.977           (----*----)
C         2.423   6.550  10.677              (----*----)
D        -6.777  -2.650   1.477      (----*----)
E        -8.877  -4.750  -0.623   (----*----)
                                 ---------+---------+---------+---------+
                                       -8.0       0.0       8.0      16.0

variety = B subtracted from:
variety    Lower  Center   Upper  ---------+---------+---------+---------+
C         -1.427   2.700   6.827           (----*-----)
D        -10.627  -6.500  -2.373   (----*----)
E        -12.727  -8.600  -4.473 (----*----)
                                  ---------+---------+---------+---------+
                                        -8.0       0.0       8.0      16.0

variety = C subtracted from:
variety    Lower   Center   Upper  ---------+---------+---------+---------+
D        -13.327   -9.200  -5.073  (-----*----)
E        -15.427  -11.300  -7.173 (----*----)
                                   ---------+---------+---------+---------+
                                         -8.0       0.0       8.0      16.0

variety = D subtracted from:
variety   Lower  Center  Upper  ---------+---------+---------+---------+
E        -6.227  -2.100  2.027         (----*-----)
                                ---------+---------+---------+---------+
                                      -8.0       0.0       8.0      16.0
```


Multiple Comparisons

What does this mean?

- The best was 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)

Multiple Comparisons



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

Multiple Comparisons

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

Individual confidence level = 98.89%

treatment = 1 subtracted from:
treatment   Lower  Center  Upper  ----+---------+---------+---------+-----
2          -1.214   1.250  3.714          (---------*---------)
3          -0.881   1.583  4.047           (---------*---------)
4          -0.359   2.225  4.809              (---------*---------)
                                  ----+---------+---------+---------+-----
                                   -2.5       0.0       2.5       5.0
treatment = 2 subtracted from:
treatment   Lower  Center  Upper  ----+---------+---------+---------+-----
3          -2.131   0.333  2.797      (---------*---------)
4          -1.609   0.975  3.559         (---------*---------)
                                  ----+---------+---------+---------+-----
                                   -2.5       0.0       2.5       5.0
treatment = 3 subtracted from:
treatment   Lower  Center  Upper  ----+---------+---------+---------+-----
4          -1.942   0.642  3.226       (----------*---------)
                                  ----+---------+---------+---------+-----
                                   -2.5       0.0       2.5       5.0
```