Announcements:

- Special office hours for final exam:
  Friday, Jason will have 1-3 (instead of 1-2)
  Monday (just before exam) I will have 10-noon.
  Exam is Monday, 1:30 – 3:30.
- Homework assigned today and Wed not due, but solutions on website.
- Final exam review sheets posted on web (in list of lectures).
- Form and instructions for disputing points will be sent by email. Watch for it if you plan to do that. Fill out and bring to final exam to hand in.

Homework: Chapter 16: 1, 7, 8, 16, 17 (not due)

Chapter 16

Analysis of Variance

ANOVA = *Analysis of variance*

- Compare means for *more than 2* groups.
- We have *k* independent samples and measure a quantitative variable on all units in all *k* samples.
- We want to know if the *population* means are all equal (null hypothesis) or if at least one is different (alternative hypothesis).
- This is called ***one-way ANOVA*** because we *analyze variability* in order to compare means.

Example: Friend or Pet to Relieve Stress?

- Randomized experiment using 45 women volunteers who said they love dogs.
- Each woman assigned to do a stressful task:
  - 15 did it alone
  - 15 did it with a good friend present
  - 15 did it with their dog present
- Response variable = heart rate

16.1 Comparing Means with an ANOVA F-Test

H$_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$

H$_a$: The means are not all equal.

Example: Parameters of interest are the *population* mean heart rates for all such women if they were to do the task under the 3 conditions:

μ_1: if doing the task alone.

μ_2: if doing the task with good friend present.

μ_3: if doing the task with dog present.

Step 1 for the example (defining hypotheses)

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

H_a: The means are not all equal.

Step 2 (In general):

The test statistic is called an *F*-statistic. In words, it is defined as:

$$F = \frac{\text{Variation among sample means}}{\text{Natural variation within groups}}$$

$$F = \frac{\text{Variation among sample means}}{\text{Natural variation within groups}}$$

Variation among sample means is 0 if all k sample means are equal and gets larger the more spread out they are.

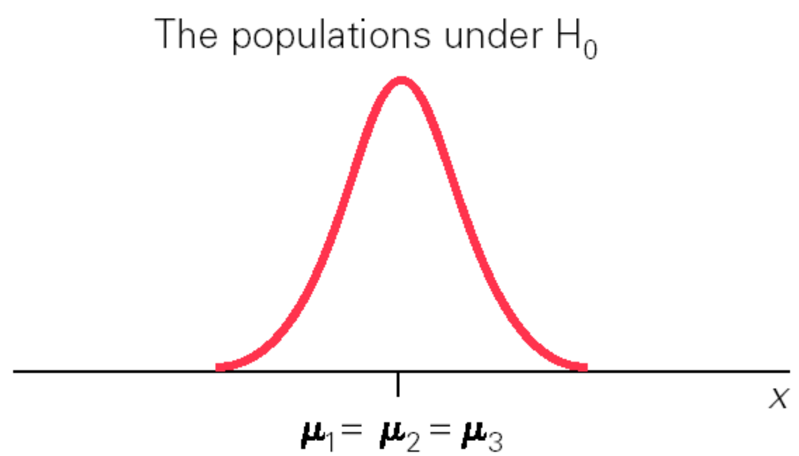
If large enough => evidence at least one population mean is different from others => **reject null hypothesis**.

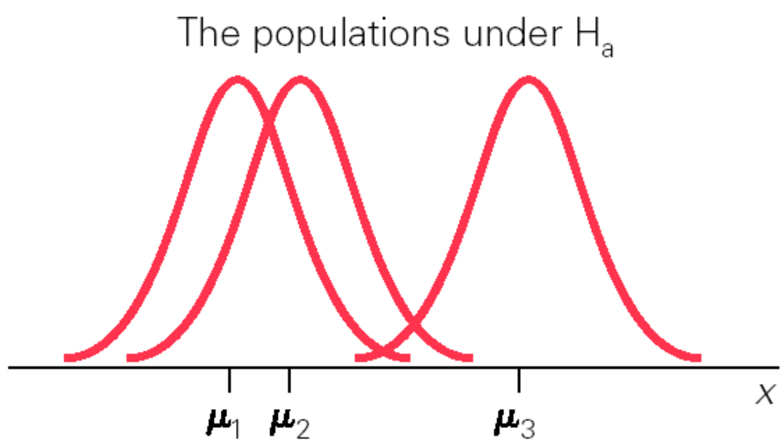
***p*-value** found using an ***F*-distribution** (more later)

Assumptions for the *F*-Test

- Samples are **independent random samples**.
- Distribution of response variable is a **normal** curve within each population (but ok as long as large *n*).
- Different populations **may have different means**.
- All populations have **same standard deviation**, σ.

e.g. How $k = 3$ populations might look …





Conditions for Using the *F*-Test

- *F*-statistic can be used if data are **not extremely skewed**, there are **no extreme outliers**, and group standard deviations are not markedly different.
- Tests based on *F*-statistic are valid for data with skewness or outliers **if sample sizes are large**.
- A rough criterion for standard deviations is that the largest of the sample standard deviations should not be more than twice as large as the smallest of the sample standard deviations.

Notation for Summary Statistics

k = number of groups

$\bar{x}_i$, s_i, and n_i are the mean, standard deviation, and sample size for the i^{th} sample group

N = total sample size ($N = n_1 + n_2 + \ldots + n_k$)

Example: *Friends, Pets and Stress*

Three different conditions => $k = 3$

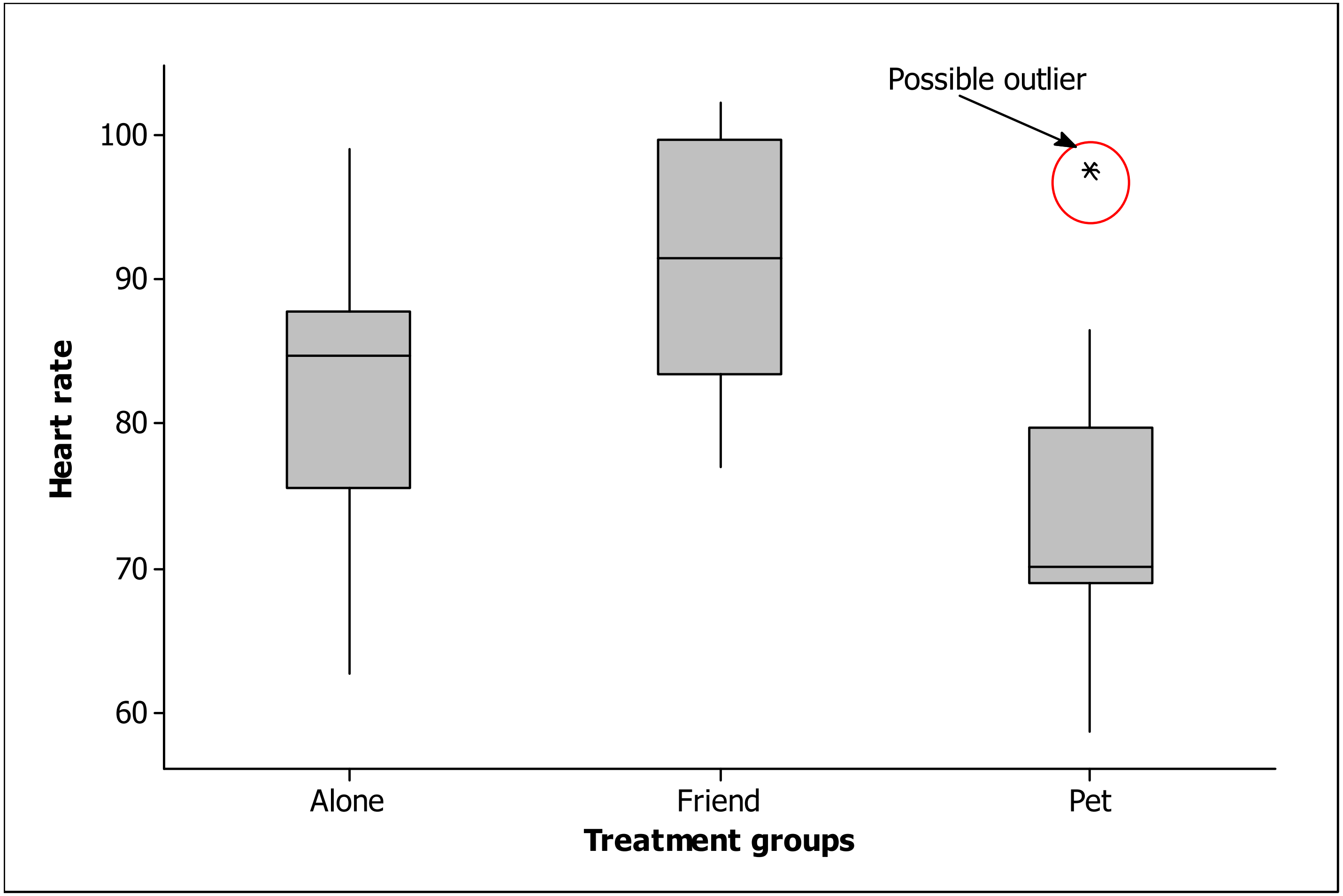
$n_1 = n_2 = n_3 = 15$; so $N = 45$

$$\bar{x}_1 = 82.52,\ \bar{x}_2 = 91.33,\ \bar{x}_3 = 73.48$$

$$s_1 = 9.24,\quad s_2 = 8.34,\quad s_3 = 9.97$$

Example, continued

Do friends, pets, or neither reduce stress more than when alone? Boxplots of *sample* data:



Conditions for Using the *F*-Test: Does the example qualify?

- *F*-statistic can be used if data are **not extremely skewed**, there are **no extreme outliers**.
  - ✓ In the example there is one possible outlier but with a sample of only 15 it is hard to tell.
- A rough criterion for standard deviations is that the largest of the sample standard deviations should not be more than twice as large as the smallest of the sample standard deviations.
  - ✓ That condition is clearly met. The sample standard deviations *s* are very similar.

16.2 Details of the F Statistic for Analysis of Variance

Fundamental concept: the variation among the data values in the overall sample can be separated into:

(1) differences **between group** means

(2) natural variation among observations **within a group**

Total variation =
Variation **between** groups + Variation **within** groups

"ANOVA Table" displays this information in summary form, and gives F statistic and *p*-value.

Measuring variation *between* groups: How far apart are the means?

Sum of squares for groups = SS Groups

$$\text{SS Groups} = \sum_{groups} n_i \left(\bar{x}_i - \bar{x}\right)^2$$

Numerator of *F*-statistic = **mean square for groups**

$$\text{MS Groups} = \frac{\text{SS Groups}}{k-1}$$

Measuring variation *within* groups: How variable are the individuals?

Sum of squared errors = SS Error

$$\text{SS Errors} = \sum_{groups} (n_i - 1)(s_i)^2$$

Denominator of F-statistic = **mean square error**

$$\text{MSE} = \frac{\text{SS Error}}{N - k}$$

Pooled standard deviation: $s_p = \sqrt{\text{MSE}}$

Measures internal variability *within* each group.

Measuring Total Variation

Total sum of squares = SS Total = SSTO

$$\text{SS Total} = \sum_{values} \left(x_{ij} - \bar{x}\right)^2$$

SS Total = SS Groups + SS Error

General Format of a One-Way ANOVA Table

Source	df	SS	MS	F
Between groups (due to factor)	$k - 1$	$\text{SS Groups} = \sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$	$\dfrac{\text{SS Groups}}{k-1}$	$F = \dfrac{\text{MS Groups}}{\text{MSE}}$
Error (within groups)	$N - k$	$\text{SSE} = \sum_{\text{groups}} (n_i - 1)s_i^2$	$\dfrac{\text{SSE}}{N-k}$	
Total	$N - 1$	$\text{SSTO} = \sum_{\text{values}} (x_{ij} - \bar{x})^2$		

Example: *Stress, Friends and Pets*

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

H_a: The means are not all equal.

One-way ANOVA: Heart rate versus treatment group

Source	DF	SS	MS	F	P
Group	2	2387.7	1193.8	14.08	0.000
Error	42	3561.3	84.8		
Total	44	5949.0			

The *F*-statistic is 14.08 and *p*-value is 0.000...

p-value so small => reject H_0 and accept Ha. Conclude there are differences among *population* means. *MORE LATER ABOUT FINDING P-VALUE.*

Conclusion in Context:

- The *population* mean heart rates would differ if we subjected all people similar to our volunteers to the 3 conditions (alone, good friend, pet dog).

- Now we want to know which one(s) differ! Individual 95% confidence intervals (next slide for formula):

```
                        Individual 95% CIs For Mean Based on
                        Pooled StDev
Group   Mean           --+---------+---------+---------+-------
Alone   82.524                       (------*------)
Friend  91.325                                    (-----*------)
Pet     73.483          (------*-------)
                       --+---------+---------+---------+-------
                       70.0      77.0      84.0      91.0
```

95% Confidence Intervals for the Population Means

In one-way analysis of variance, a **confidence interval for a population mean** μ_i is

$$\bar{x}_i \pm t^* \left(\frac{s_p}{\sqrt{n_i}} \right)$$

where $s_p = \sqrt{\text{MSE}}$ and ***t**** is from Table A.2:

t* is such that the confidence level is the probability between -t* and t* in a t-distribution with df $= N - k$.

Multiple Comparisons

Multiple comparisons: Problem is that *each* C.I. has 95% confidence, but we want *overall* 95% confidence. Can do multiple C.I.s and/or tests at once.

Most common: ***all pairwise comparisons of means***.

Ways to make inferences about each pair of means:

- **Significance test** to assess if two means significantly differ.

- **Confidence interval** for difference computed and if 0 is *not* in the interval, there is a statistically significant difference.

Multiple Comparisons, continued

Many statistical tests or C.I.s => increased risk of making at least one type I error (erroneously rejecting a null hypothesis). Several procedures to control the overall **family type I error rate** or overall **family confidence level.**

- **Family error rate** for set of significance tests is probability of making one or more type I errors when more than one significance test is done.
- **Family confidence level** for procedure used to create a set of confidence intervals is the proportion of times all intervals in set capture their true parameter values.

Tukey method: Family confidence level of 0.95

Confidence intervals for $\mu_i - \mu_j$ (differences in means)

```
Difference is Alone subtracted from:
            Lower   Center    Upper
Friend      2.015    8.801   15.587
Pet       -15.827   -9.041   -2.255
```

(Friend – Alone)

(Pet – Alone)

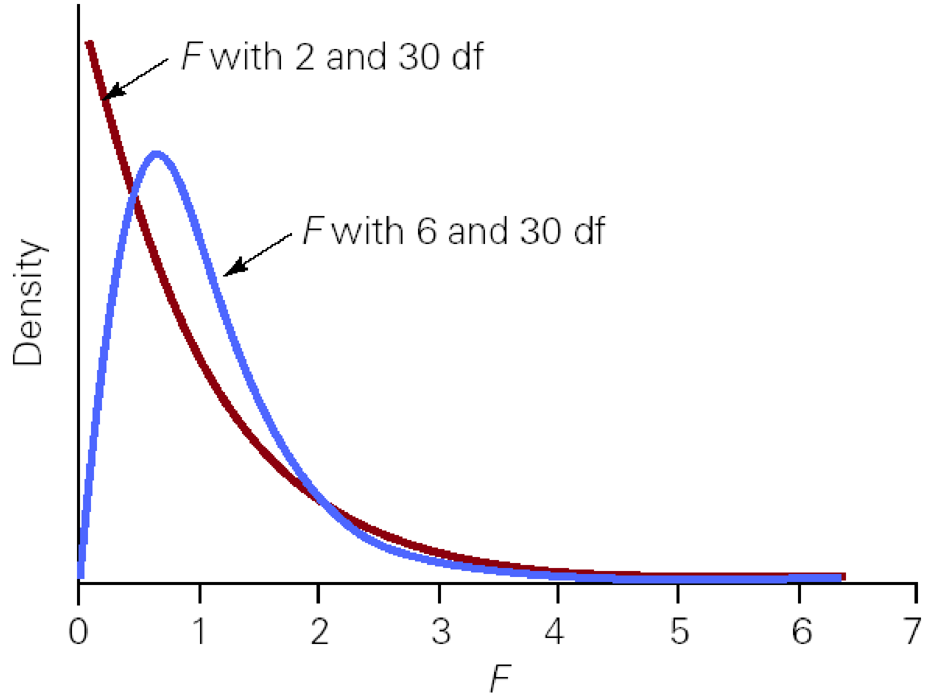
```
Difference is Friend subtracted from:
            Lower    Center     Upper
Pet       -24.628   -17.842   -11.056
```

(Pet – Friend)

None of the confidence intervals cover 0! That indicates that all 3 population means differ. Ordering is *Pet < Alone < Friend*

The Family of *F*-Distributions

- Skewed distributions with minimum value of 0.
- Specific *F*-distribution indicated by two parameters called *degrees of freedom:* **numerator degrees of freedom** and **denominator degrees of freedom.**
- In one-way ANOVA, numerator df = $k - 1$, and denominator df = $N - k$
- *Looks similar to chi-square distributions*



Determining the *p*-Value

Statistical Software reports the *p*-value in output.

Table A.4 provides **critical values** (to find rejection region) for 1% and 5% significance levels.

- If the *F*-statistic is > than the 5% critical value, the *p*-value < 0.05.
- If the *F*-statistic is > than the 1% critical value, the *p*-value < 0.01 .
- If the *F*-statistic is between the 1% and 5% critical values, the *p*-value is between 0.01 and 0.05.

Example: ***Stress, Friends and Pets***

Reported F-statistic was $F = 14.08$ and p-value < 0.000

$N = 15$ women:
num df = $k - 1 = 3 - 1 = 2$
den df = $N - k = 45 - 3 = 42$

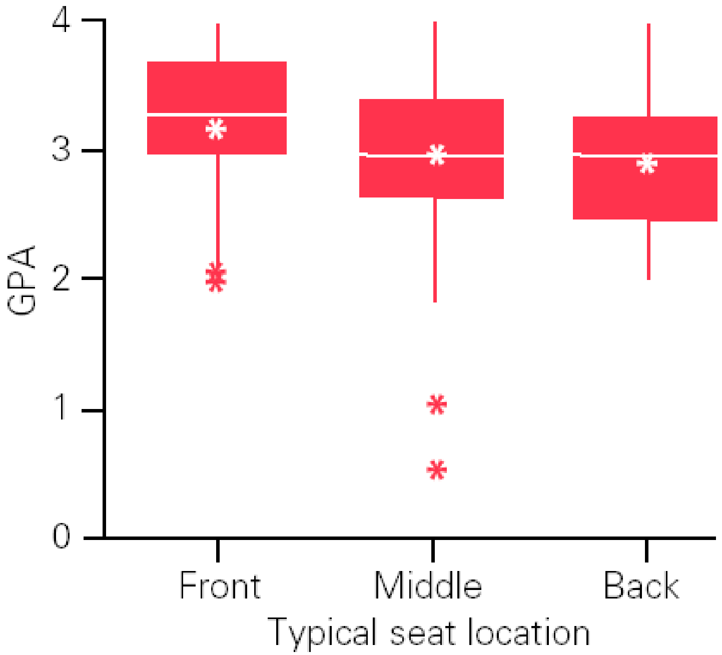
Table A.4 with df of (2, 42), closest available is (2, 40):
The 5% critical value is 3.23.
The 1% critical value is 5.18 and
The F-statistic was much larger, so p-value < 0.01.

Example 16.1 *Seat Location and GPA*

Q: Do best students sit in the front of a classroom?

Data on seat location and GPA for n = 384 students;
88 sit in front, 218 in middle, 78 in back

Comparative boxplot (white asterisk within box indicates sample mean)



Level	N	Mean	StDev
Front	88	3.2029	0.5491
Middle	218	2.9853	0.5577
Back	78	2.9194	0.5105

Students sitting in the front generally have slightly higher GPAs than others.

Example 16.1 *Seat Location and GPA (cont)*

- The boxplot showed **two outliers** in the group of students who typically sit in the middle of a classroom, but there are 218 students in that group so these outliers don’t have much influence on the results.
- The **standard deviations** for the three groups are **nearly the same**.
- Data **do not appear to be skewed**.

Necessary conditions for F-test seem satisfied.

Notation for Summary Statistics

k = number of groups

$\bar{x}$, s_i, and n_i are the mean, standard deviation, and sample size for the i^{th} sample group

N = total sample size ($N = n_1 + n_2 + \ldots + n_k$)

Example 16.1 *Seat Location and GPA (cont)*

Three seat locations => $k = 3$

$n_1 = 88$, $n_2 = 218$, $n_3 = 78$; $N = 88+218+78 = 384$

$$\bar{x}_1 = 3.2029, \bar{x}_2 = 2.9853, \bar{x}_3 = 2.9194$$

$$s_1 = 0.5491, s_2 = 0.5577, s_3 = 0.5105$$

Example 16.1 *Seat Location and GPA (cont)*

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

H_a: The means are not all equal.

Analysis of Variance for GPA

Source	DF	SS	MS	F	**P**
Location	2	3.994	1.997	6.69	**0.001**
Error	381	113.775	0.299		
Total	383	117.769			

The *F*-statistic is 6.69 and the *p*-value is 0.001.

p-value so small => reject H_0 and conclude there are differences among the means.

Example 16.1 *Seat Location and GPA (cont)*

95% Confidence Intervals for 3 population means:

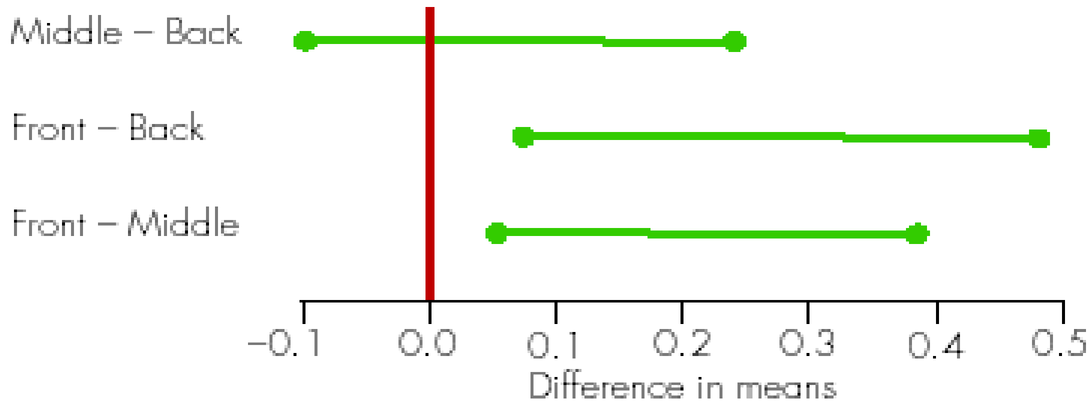
Interval for "front" ***does not overlap*** with the other two intervals => ***significant difference*** between mean GPA for front-row sitters and mean GPA for other students

```
                                     Individual 95% CIs For Mean
                                     Based on Pooled StDev
Level      N      Mean     StDev   ----+---------+---------+---------+--
Front     88    3.2029    0.5491                     (-------*------)
Middle   218    2.9853    0.5577          (----*----)
Back      78    2.9194    0.5105    (-------*-------)
                                   ----+---------+---------+---------+--
Pooled StDev = 0.5465                 2.85      3.00      3.15      3.30
```

Example 16.1 *Seat Location and GPA (cont)*

Pairwise Comparison Output:

Tukey: Family confidence level of 0.95



Only one interval covers 0, $\mu_{\text{Middle}} - \mu_{\text{Back}}$

Appears *population* mean GPAs differ for front and middle students and for front and back students.

Example 16.4 ***Testosterone and Occupation***

To illustrate how to find p-value.

Study: Compare mean testosterone levels for $k = 7$ occupational groups:

Ministers, salesmen, firemen, professors, physicians, professional football players, and actors.

Reported F-statistic was $F = 2.5$ and p-value < 0.05

$N = 66$ men:

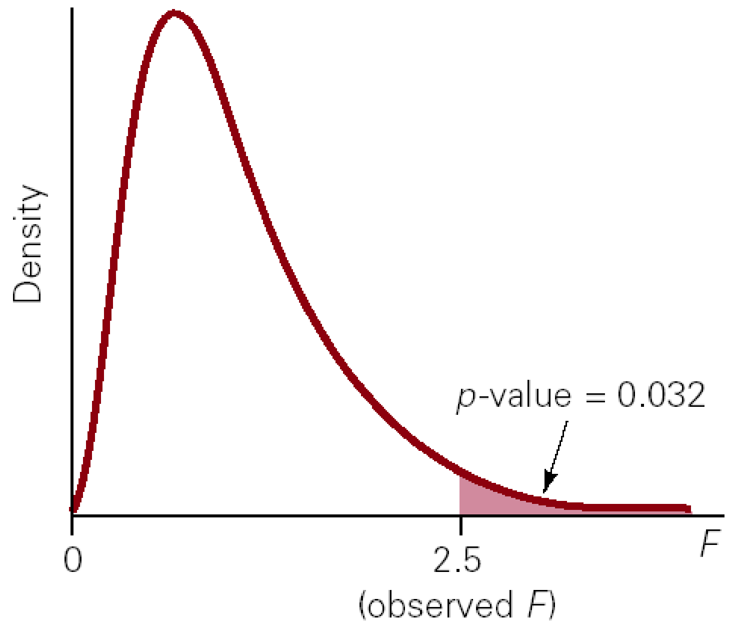
num df $= k - 1 = 7 - 1 = 6$

den df $= N - k = 66 - 7 = 59$

From Table A.4, rejection region is $F \geq 2.25$. Since the calculated $F = 2.5 > 2.25$, reject null hypothesis.

Example 16.2 *Testosterone and Occupation*

P-value picture shows exact p-value is 0.032:



There are 21 possible comparisons!
Significant differences were found for only these occupations:

Actors > Ministers
Football players > Ministers

USING R COMMANDER
Statistics – Means – One-way ANOVA
Then click "pairwise comparisons"

```
Df Sum Sq Mean Sq F value    Pr(>F)
group          2 2387.7 1193.84  14.079 2.092e-05 ***
Residuals   42 3561.3   84.79
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
           mean        sd  n
Alone  82.52407 9.241575 15
Friend 91.32513 8.341134 15
Pet    73.48307 9.969820 15
```

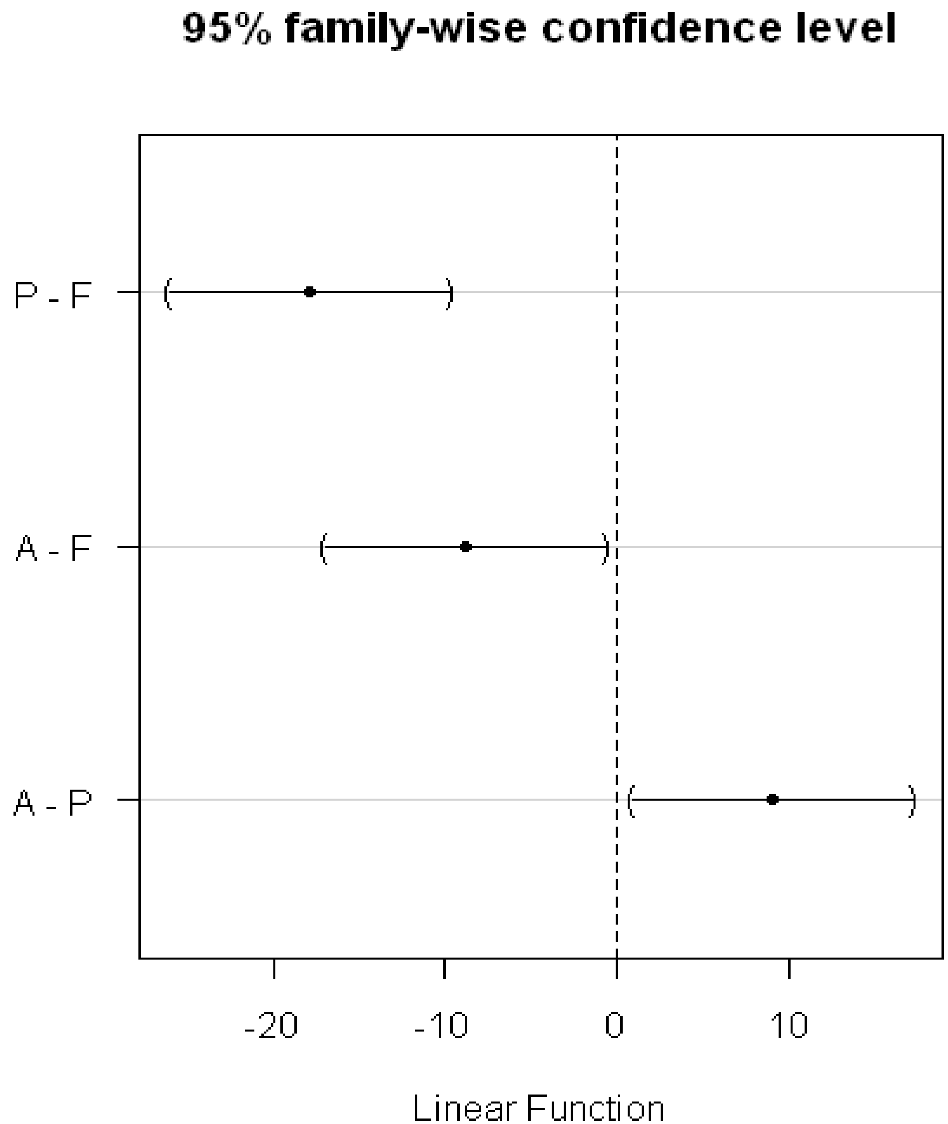
```
         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = rate ~ group, data = PetStress)

Quantile = 2.4298
95% family-wise confidence level


Linear Hypotheses:
                    Estimate lwr      upr
Friend - Alone == 0   8.8011   0.6313  16.9709
Pet - Alone == 0     -9.0410 -17.2108  -0.8712
Pet - Friend == 0   -17.8421 -26.0119  -9.6723
```



95% family-wise confidence level
P - F
A - F
A - P
-20
-10
0
10
Linear Function