STA 321 Spring 2014

Lecture 21 *Tuesday, April 15*

- **Comparing Proportions of Two Independent Samples**
- **Comparing Two Independent Samples of Ordinal Data**
- **Analysis of Variance**

QUIZ I

- What are the assumptions for t-test for a small sample? List all of them.

- Normality of the population distribution
- Quantitative Data set
- Random sample

Quiz II

- In order to determine the *p*-value, which of the following is not needed?

a. The level of significance.

b. Whether the test is one-tail or two-tail.

c. The value of the test statistic.

d. All of these choices are true.

Comparing Proportions in Two (Large) Independent Samples

- Response variable: Qualitative
- Inference about the population proportions that are classified in a particular category of the response variable
- Explanatory variable: The variable that defines the group membership.
- Are the proportions different for the two groups?

$$p_2 - p_1 = 0\,?$$

- Confidence interval for the difference
- Significance test about whether the difference equals zero

Comparing Two Proportions: Examples

1. Gender Gap in Party Identification

   Explanatory variable: Male/Female

   Response variable: Party Identification

   Is the proportion of Republicans different between male and female voters?

2. Explanatory variable: Treatment (Drug / Placebo)

   Response variable: Pain (Yes/No)

   Is the proportion who suffers from pain different for the two treatment groups?

Confidence Interval for the Difference of Two Proportions

- Here, large sample means at least five observations in each category of interest in each of the samples
- The large sample confidence interval for $p_2 - p_1$ is

$$(\hat{p}_2 - \hat{p}_1) \pm z \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence Interval for the Difference of Two Proportions: Example

- Famous five-year study on the effect of Aspirin to reduce heart disease
- Study subjects: 22,071 male physicians
- Every other day, participants took either an aspirin tablet or a placebo
- 11,034 who took placebo: 189 had a heart attack
- 11,037 who took aspirin: 104 had heart attacks
- *Estimate the heart attack rates for the two groups.*
- *Construct a 95% confidence interval to compare them.*
- *Interpret.*

$$(\hat{p}_2 - \hat{p}_1) \pm z \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$$

Significance Test for the Difference of Two Proportions

- The large sample (see above) significance test for the null hypothesis that both population proportions are equal,

$H_0 : p_1 = p_2$ which is equivalent to $H_0 : p_2 - p_1 = 0$, is

$$z_{obs} = \frac{\text{estimate } - \text{null hypothesis value}}{\text{standard error of estimator}}$$

$$= \frac{\widehat{p}_2 - \widehat{p}_1}{\sqrt{\widehat{p} \cdot (1 - \widehat{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$\widehat{p}$ is the "pooled" proportion of the total sample (both samples together) in the category of interest

Significance Test for the Difference of Two Proportions

- As above, most commonly, the alternative hypothesis is two-sided
- Then, the P-value is the two-tail probability of “anything at least as extreme as observed”
- The probability is taken from a z-score applet (normal distribution)

Significance Test for the Difference of Two Proportions: Example

- Effect of Aspirin to reduce heart disease
- Study subjects: 22,071 male physicians
- Every other day, participants took either an aspirin tablet or a placebo
- 11,034 who took placebo: 189 had a heart attack
- 11,037 who took aspirin: 104 had heart attacks
- *Test whether the rates are significantly different. Report the P-value and interpret.*

$$z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} =$$

Summary

Large Sample Significance Test for the Difference of Two Proportions

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : p_1 = p_2$		
Research Hypothesis	$H_1 : p_2 < p_1$	$H_1 : p_2 > p_1$	$H_1 : p_1 \neq p_2$
Test Statistic	$z_{obs} = \dfrac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \lvert z_{obs} \rvert)$

Comparing Two Independent Samples with Ordinal Response Variable

- Response variable: Qualitative, Ordinal
- Inference about whether one of the samples tend to have the larger values of the ordinal variable
- Explanatory variable: The variable that defines the group membership.
- Confidence interval for the difference
- Significance test about whether the difference equals zero

Comparing Two Independent Samples with Ordinal Response Variable

Wilcoxon-Mann-Whitney test for independent samples

- http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html
- The test is also available in SAS, of course.
- The confidence interval is not yet included in many software packages. You can download a SAS macro that calculates the confidence interval for location difference at
- http://www.ams.med.uni-goettingen.de/de/sof/ld/TSP.SAS

Nonparametric Statistics

- The only appropriate method for analyzing ordinal data
- Provides alternative ways to analyze quantitative data
- Nonparametric methods ***have to*** be used when the assumptions for the standard methods (normal distribution) are not met (e.g., very skewed data)
- However, even when these assumptions *are* met, then nonparametric methods ***can*** be used. They will only be slightly worse than the standard methods
- So, they are always a useful and safe alternative
- STA 673

Summary: Comparing Two Samples

- Independent Samples
  - Quantitative Response Variable
    - Large samples
    - Small samples from normal population
    - Small samples, not normal
  - Ordinal Response Variable
  - Binary Variable *(Proportions)*
    - Large Samples
    - Small Samples

Summary: Comparing Two Samples

- Dependent Samples (Matched Pairs)
  - Quantitative Response Variable
    - Large number of pairs
    - Small number of pairs, normal population
    - Small number of pairs, not normal
  - Ordinal Response Variable
  - Binary Variable *(Proportions)*
    - Large number of pairs
    - Small number of pairs

Analysis of Variance

- Earlier
  - Comparing Two Samples (Binary, Ordinal, Quantitative)
  - Example: Compare mean annual income of men and women
- Now
  - Several samples, Quantitative variable
  - Comparing Several Means
  - Example: Compare mean statistics exam score for students from several different departments/colleges

Analysis of Variance

- Short: ANOVA
- Developed by Sir R.A. Fisher in the 1920s for agricultural data
- Uses the *F*-distribution (named after Fisher)
- Goal: Detect evidence of differences between population means

ANOVA Notation

- Number of groups: *g*
- Means of the response variables for the *g* populations: $\mu_1, \mu_2, \ldots, \mu_g$
- Null hypothesis: All the means are equal, that is, $H_0 : \mu_1 = \mu_2 = \ldots = \mu_g$
- Group sample sizes: $n_1, n_2, \ldots, n_g$

  ANOVA works best when the sample sizes are equal (balanced design)
- Total sample size: $N = n_1 + n_2 + \ldots + n_g$
- Sample Means: $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_g$
- Sample standard deviations: $s_1, s_2, \ldots, s_g$

ANOVA Assumptions

- Independent random samples are selected from the *g* populations ***(very important)***
- Response is truly quantitative ***(very important)***
- The standard deviations of the population distributions for the *g* groups are equal, denoted by *sigma* *(if this is not the case, there is a statistical solution: variance stabilizing transformations)*
- The population distributions of the response variable are normal for each of the *g* groups ***(important for small sample sizes and small number of samples)***

Example

- Scores on the first quiz (maximum 10 points) in a beginning French course for ninth-grade students
- Three groups of students:
  - Group A: Never studied foreign language before, but have good English skills
  - Group B: Never studied foreign language before; have poor English skills
  - Group C: Studied other foreign language

Variability Between and Within Groups

- ANOVA compares two types of variability
  - Variability of the sample observations about their separate means (*within-group variability*)
  - Variability of the sample means from the different groups about the overall mean (*between-group variability*)

Variability Between and Within Groups

- The greater the variability between sample means and

  the smaller the variability within each group of observations,

  the stronger the evidence that the null hypothesis of equal means is false
- The test statistic is the ratio of two variance estimates:
  - Between-groups estimate and
  - Within-groups estimate

Within-Groups Estimate of Variance Technical Details

- For each of the *g* groups, a variance estimate can be calculated
- Construct a ***pooled variance estimate*** by adding up weighted group variance estimates
- The weights are the degrees of freedom for each group: n_i-1
- Divide by the total degrees of freedom: $N-g$

Within-Groups Estimate of Variance Technical Details

- This estimate is a weighted average of the separate sample variances, with greater weight given to larger samples
- It is unbiased and efficient
- Mathematical Formula:

$$\hat{\sigma}^2 = \frac{WSS}{N-g} = \frac{SSE}{N-g} = MSE$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_g - 1)s_g^2}{N-g}$$

Between-Groups Estimate of Variance

- This estimate is based on the variability between each sample mean and the overall mean (from all samples together)
- Under the null hypothesis, it is unbiased
- Mathematical Formula:

$$\frac{BSS}{g-1} = MSH$$

$$= \frac{n_1(\bar{Y}_{1.} - \bar{Y}_{..})^2 + n_2(\bar{Y}_{2.} - \bar{Y}_{..})^2 + \cdots + n_g(\bar{Y}_{g.} - \bar{Y}_{..})^2}{g-1}$$

where $\bar{Y}_{..} = \frac{1}{N}\sum Y_{ij}$

F Test Statistic

The test statistic for the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_g$$

is the ratio of the two variance estimates:

$$F = \frac{\text{Between-Groups Estimate}}{\text{Within-Groups Estimate}}$$

$$= \frac{BSS/(g-1)}{WSS/(N-g)} = \frac{MSH}{MSE}$$

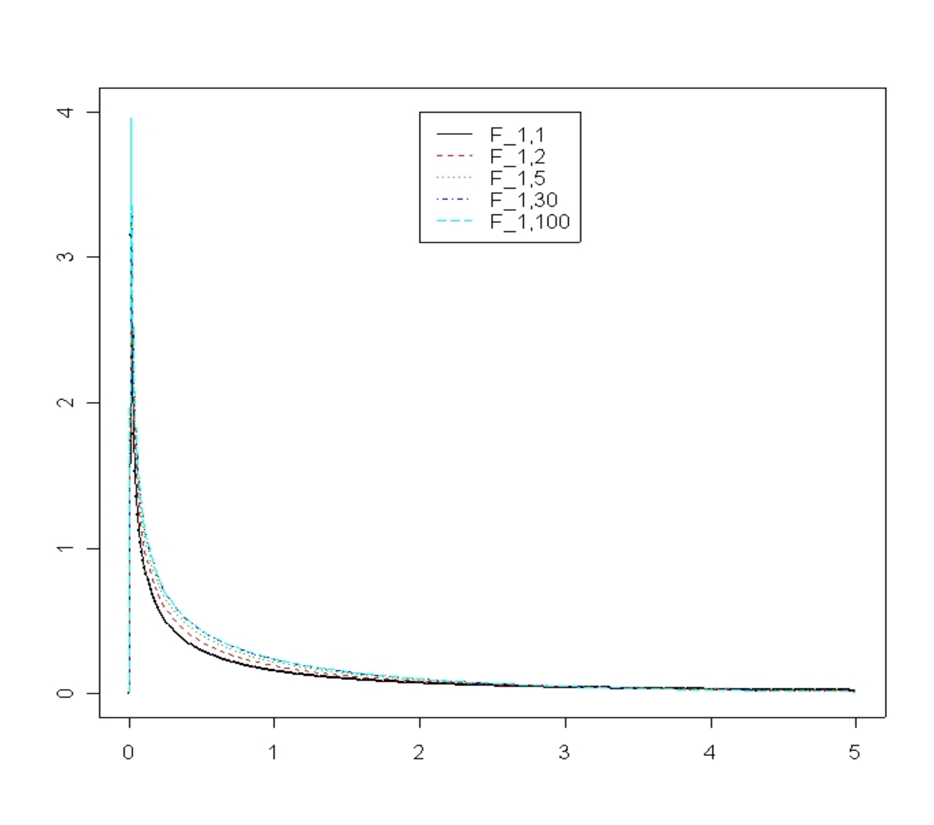
F Test Statistic

- It is called the ***analysis of variance F statistic*** or ***ANOVA F statistic***
- If the null hypothesis is true, its sampling distribution is the *F* distribution with degrees of freedom $df_1 = g-1$ and $df_2 = N-g$
- The P-value is the right-tail probability that the *F* test statistic takes a value at least as large as the observed *F* value
- The larger the *F* test statistic, the smaller the P-value

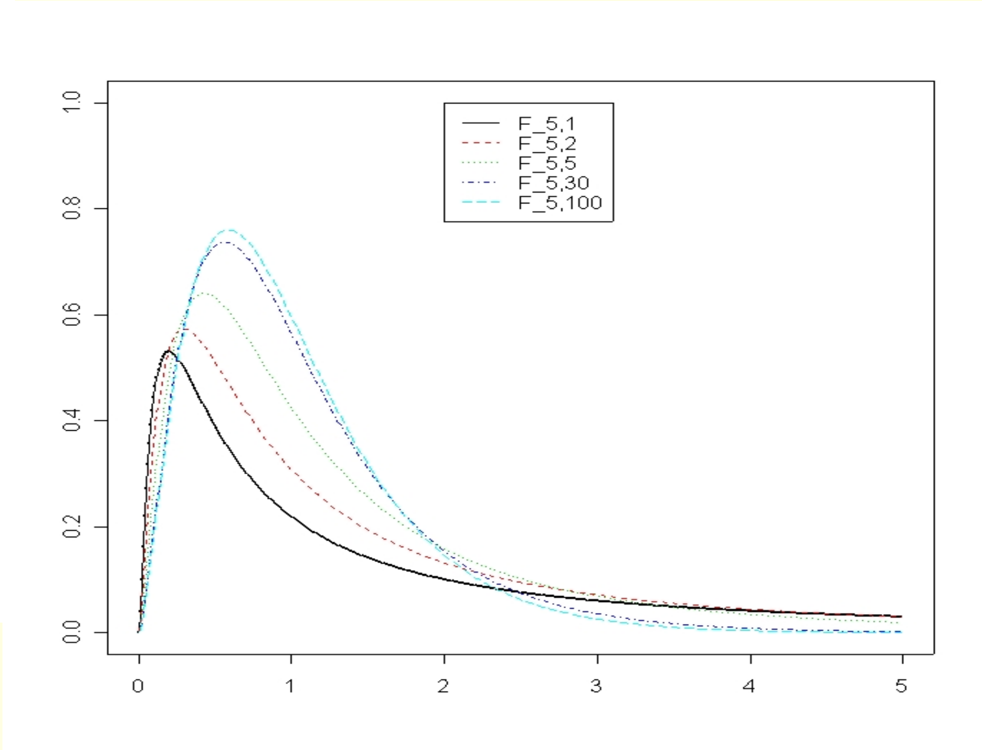
F Test Statistic

- If the null hypothesis is true, numerator and denominator are both unbiased estimates of the same quantity
- We expect the ratio to be around 1
- If the null hypothesis is not true, the numerator will be larger, and the test statistic takes larger values
- We reject the null hypothesis when the values are too large
- What is too large? Sampling distribution

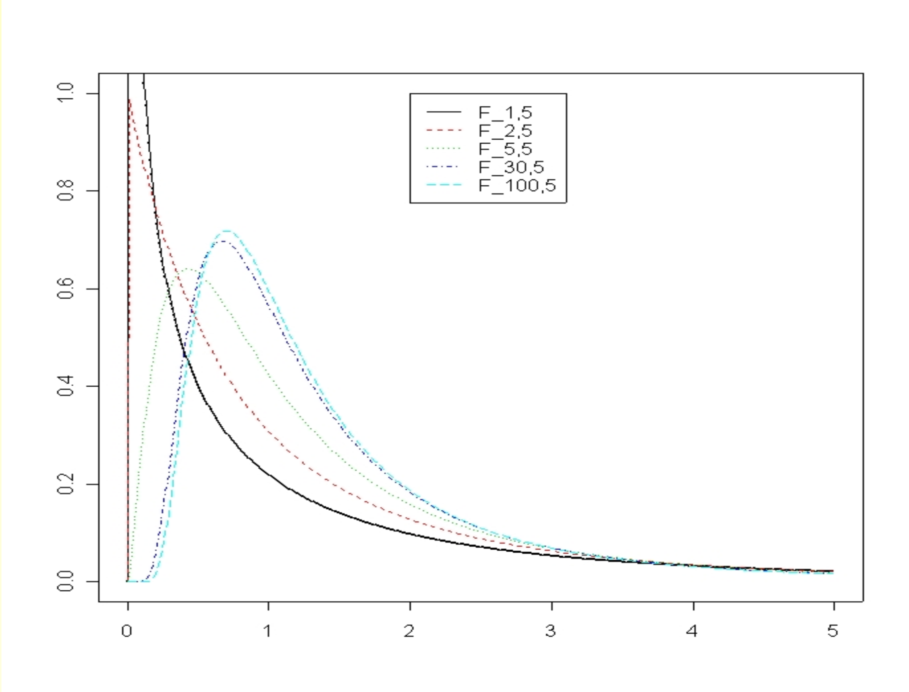
F Distributions (1,x)



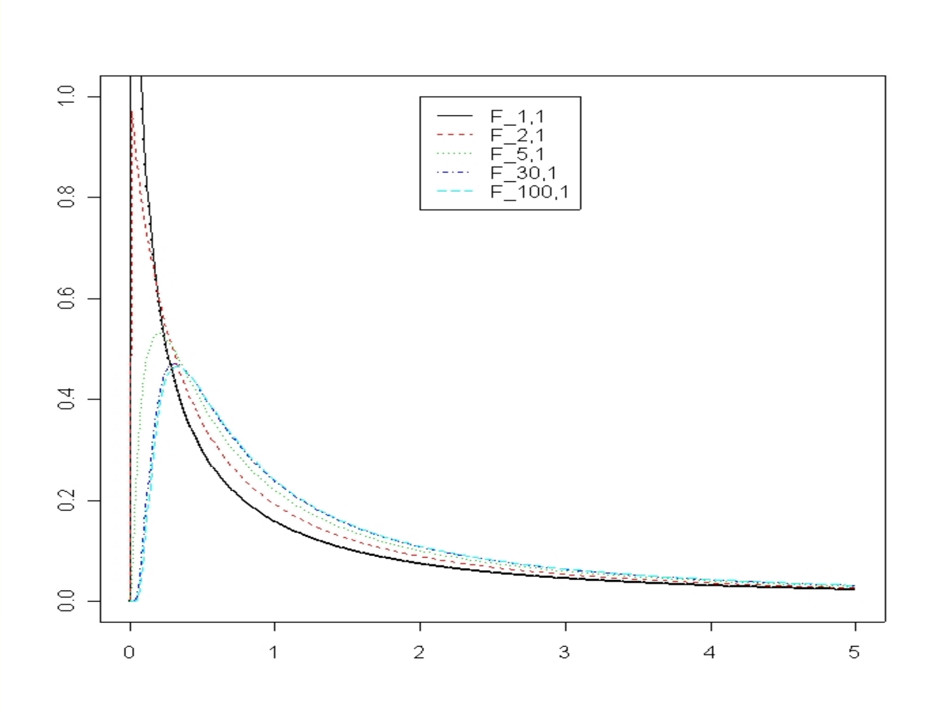
F Distributions (5,x)



F Distributions (x,5)



F Distributions (x,1)



Example (contd.)

- The quiz scores in the beginning French course are given in the table
- Calculate the F statistic for the quiz score example
- What is the P-value?
- F distribution online tool

Group A	Group B	Group C
4	1	9
6	5	10
8		5

ANOVA Table (SAS Output)

- Statistical software displays the results of ANOVA *F* tests in a table called ***ANOVA table***

The GLM Procedure
Class Level Information

Class	Levels	Values
group	3	A B C

Number of observations 8

Dependent Variable: score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	30.00000000	15.00000000	2.50	0.1768
Error	5	30.00000000	6.00000000		
Corrected Total	7	60.00000000			

Level of group	N	------------score------------ Mean	Std Dev
A	3	6.00000000	2.00000000
B	2	3.00000000	2.82842712
C	3	8.00000000	2.64575131

ANOVA Online Tool 1

ANOVA Online Tool 2

Another Example

- Three materials for making artificial teeth are compared with regard to hardness.
- The materials are Endura, Duradent, and Duracross.
- Six pairs of teeth are tested for each material.
- The response variable is the Vickers microhardness of the occlusal surfaces, measured with a load of 50 g and a loading time of 30 sec.

Example: Hardness of Artificial Teeth (contd.)

- Data table, with sample means and standard deviations

	Endura	Duradent	Duracross
Hardness	27.1 27.6 28 28.5 27.3 26.7	23.9 24.5 23.9 24.4 22.9 24.5	44.9 37.9 40.4 38.5 40.4 35.7
Sample Mean	*27.53*	*24.02*	*39.63*
Sample Standard Deviation	*0.65*	*0.61*	*3.12*

ANOVA Table (from SAS)

The SAS System 19:52 Monday, March 31, 2008
The GLM Procedure

Class Level Information

Class	Levels	Values
MATERIAL	3	Duracros Duradent Endura

Number of Observations Read 18
Number of Observations Used 18

Dependent Variable: HARDNESS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**2**	**805.3144444**	**402.6572222**	**114.71**	**<.0001**
Error	**15**	**52.6550000**	**3.5103333**		
Corrected Total	**17**	**857.9694444**			

Interpretation

- The null hypothesis for the ANOVA is

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_g$$

- The P-value from the ANOVA table is <0.0001
- At 5% level, there is sufficient evidence against the null hypothesis
- So we can conclude that not all the population means are equal

Interpretation, contd.

- However, the conclusion of the test does not specify which means are different or how different they are
- More detailed inference is necessary to determine the nature of the differences

Multiple Comparisons of Means

- Confidence intervals are usually more informative than test results
- In practice, we would be interested in estimates of the population means and confidence intervals for their differences
- Compare groups A vs. B, A vs. C, B vs. C
- We can also perform pairwise t-tests with some correction like Bonferroni procedure
- "post-hoc (*after this*) analysis"

Multiple Comparisons Using SAS

```
data teeth;
input hardness material$;
cards;
27.1 Endura
27.6 Endura
28      Endura
28.5 Endura
27.3 Endura
26.7 Endura
23.9 Duradent
24.5 Duradent
23.9 Duradent
24.4 Duradent
22.9 Duradent
24.5 Duradent
44.9 Duracross
37.9 Duracross
40.4 Duracross
38.5 Duracross
40.4 Duracross
35.7 Duracross
;
```

```
proc glm data=teeth;
class material;
model hardness=material;
means material/bon
alpha=0.05;
run;
```

SAS Output

Bonferroni (Dunn) t Tests for ***hardness***

NOTE: This test controls the Type I ***experimentwise*** error rate

Alpha	***0.05***
Error Degrees of Freedom	15
Error Mean Square	3.510333
Critical Value of t	2.69374
Minimum Significant Difference	***2.9139***

Means with the same letter are not significantly different.

Bon Grouping	**Mean**	**N**	**type**
A	***39.633***	6	Duracros
B	***27.533***	6	Endura
C	***24.017***	6	Duradent

Interpretation

- ANOVA *F* test:
  - The population means are not all the same
- Pairwise comparisons:
  - Duracross is significantly harder than Endura and than Duradent
  - Endura is significantly harder than Duradent

Summary

- Use ANOVA to check whether population means for *g* groups are identical
- Quantitative response,

  qualitative explanatory variable (group)
- If (and ONLY IF) there is enough evidence that the population means are not all identical, perform pairwise comparisons to find out which pairs are significantly different

Quiz I

When testing the null hypothesis using the confidence interval estimate of the difference between two means, one would reject the null hypothesis (two sided test) when

(a) the upper confidence limit is greater than zero
(b) the confidence interval does not include zero
(c) the confidence interval includes zero
(d) the lower confidence limit is less than zero

Quiz II

- We have created a 90% confidence interval for *mu* with the result (25, 32). What conclusion will we make if we test H_0: *mu* = 28 vs. H_1: *mu* not = 28 at <u>*alpha* = 0.10</u>?

a. Reject H_0 in favor of H_1.

b. Accept H_0 in favor of H_1.

c. Fail to reject H_0 in favor of H_1.

d. We cannot tell from the information given.