

# STA 570

# Spring 2011

Lecture 22

*Thursday, April 14*

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

# 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

$$z_{obs} = \frac{\text{estimate} - \text{null hypothesis value}}{\text{standard error of estimator}}$$
$$= \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$\hat{p}$  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.*

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



# Confidence Interval and Hypothesis Test for the Difference of Two Proportions: Example

- A serious side effect of many allergy medicines is that they cause drowsiness, which makes them dangerous for industrial workers. In recent years, different nondrowsy allergy medicines have been developed. One of them was Hismanal. The manufacturer claimed that this was the first once-a-day nondrowsy allergy medicine.
- The nondrowsiness claim is based on a clinical experiment in which 1,604 patients were given Hismanal, and 1,109 patients were given a placebo.
- Of the first group, 7.1% reported drowsiness; of the second group, 6.4% reported drowsiness.
- Do these results allow us to infer at the 5% significance level that Hismanal's claim is false?
- Report a 95% confidence interval for the difference of the proportions.

# Contingency Table

- The proportions are usually listed in a table called ***contingency table***
- How are the outcomes of the response variable *contingent* on the category of the explanatory variable
- Each row represents a category of one variable, and each column represents a category of the other variable
- The cells of the table contain frequency counts for the four possible combinations of outcomes

	Drowsy	Not Drowsy	
Hismanal			
Placebo			

# 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	$z_{obs} = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z >  z_{obs} )$

# 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](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>

# Wilcoxon-Mann-Whitney Test: a Nonparametric Test

*How does it work?*

- $R_{ij}$  = rank of observation  $Y_{ij}$  among *all* observations from *both* samples
- Example:

Original Observations		Ranks	
Sample 1	Sample 2	Sample 1	Sample 2
8.5	9.6	1	2
10.3	10.2	4	3
	11.7		5

# Wilcoxon-Mann-Whitney Test

- If there are ties, use *midranks*
- Example:

Original Observations		Ranks	
Sample 1	Sample 2	Sample 1	Sample 2
8.5	9.6	1	2
10.3	10.3	3.5	3.5
	11.7		5

## Two Independent Samples: ***Exact*** WMW Test (for small samples)

- Calculate the rank sum for the first sample
- Look at all possible configurations that could occur with the same sample sizes
- p-value = sum of the probabilities of all theoretically possible configurations that are “at least as extreme” as the one observed



# Two Independent Samples: *Exact* WMW Test

- All possible 10 configurations in the example

1	3	1	2	1	2	2	1	1	2	2	1	2	1	3	1	3	1	4	1
2	4	3	4	4	3	3	4	5	3	4	3	5	3	4	2	5	2	5	2
	5		5		5		5		4		5		4		4		4		3
<b>3</b>		<b>4</b>		<b>5</b>		<b>5</b>		<b>6</b>		<b>6</b>		<b>7</b>			5	<b>8</b>		<b>9</b>	

■ One-sided p-value=0.4, two-sided p-value=0.8

# 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

# Summary: Investigating Association Between Two Variables

		Response Variable		
		Unordered Categorical (Nominal)	Ordinal	Quantitative
Explanatory Variable	Nominal with 2 Levels	Comparing Proportions of 2 Independent Samples	Wilcoxon-Mann-Whitney Rank Sum Test	Comparing Means of 2 Independent Samples
	Unordered Categorical (Nominal) More than 2 Levels	?	?	Analysis of Variance
	Ordinal		?	
	Quantitative	?		Regression

# 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:  $\mu_1, \mu_2, \dots, \mu_g$
- Null hypothesis: All the means are equal, that is,  $H_0 : \mu_1 = \mu_2 = \dots = \mu_g$
- Group sample sizes:  $n_1, n_2, \dots, n_g$   
*ANOVA works best when the sample sizes are equal (balanced design)*
- Total sample size:  $N = n_1 + n_2 + \dots + n_g$
- Sample Means:  $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_g$
- Sample standard deviations:  $s_1, s_2, \dots, s_g$



# 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:

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

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

# F Test Statistic

The test statistic for the null hypothesis

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

is the ratio of the two variance estimates:

$$F = \frac{\text{Between-Groups Estimate}}{\text{Within-Groups Estimate}}$$
$$= \frac{BSS / (g - 1)}{WSS / (N - g)} = \frac{MSH}{MSE}$$



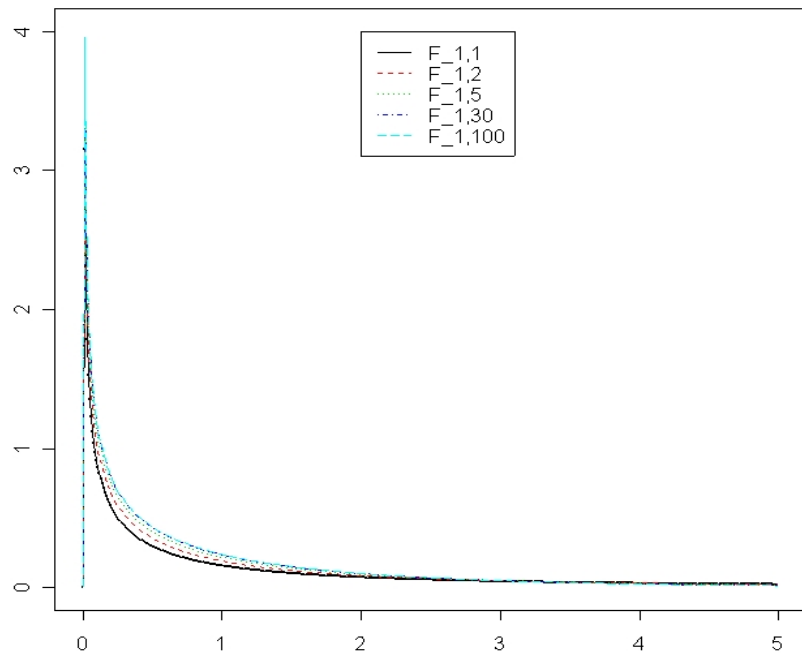
# *F* Test Statistic

- 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

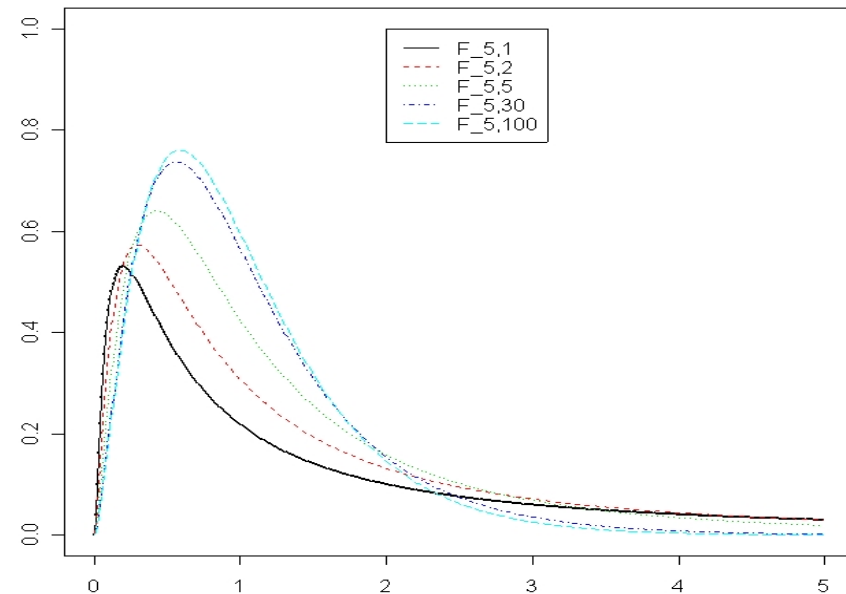
# $F$ Test Statistic

- 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)

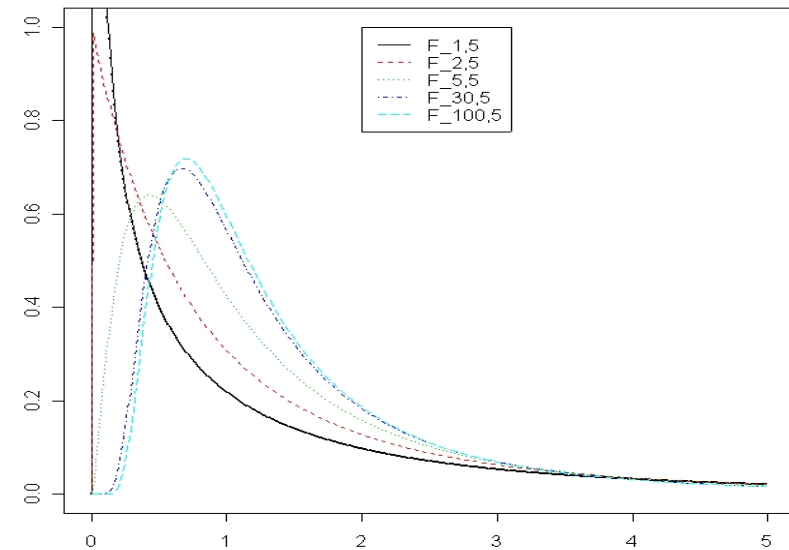
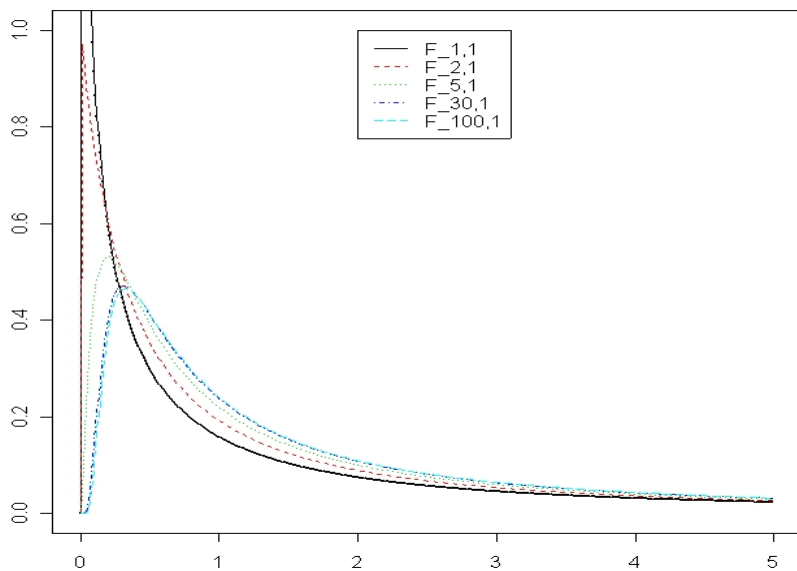


# F Distributions (5,x)



# 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 group	Levels	Values
	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	Mean	Std Dev
A	3	6.00000000	2.00000000
B	2	3.00000000	2.82842712
C	3	8.00000000	2.64575131