Class 19: Two Way Tables, Conditional Distributions, Chi-Square (Text: Sections 2.5; 9.1)

Big Picture: More than Two Samples
In Chapter 7: We looked at quantitative variables and compared the means of *two* samples.
In Chapter 8: We looked at categorical variables and compared the proportions in *two* samples.

What if we have more than two samples?
For example, suppose we want to compare proportions of men, women, and children? Or we want to compare the mean effect of a drug on Whites, African-Americans, Hispanics, and Asians? The tests we know so far don't work because there are more than two samples.

The two tests we now work with more than two samples:

- **Chi-Square Test**: For categorical variables, more than two samples (Chapter 9)
- **ANOVA** (Analysis of Variance): For quantitative variables, more than two samples (Chapter 12)

First we'll do categorical variables and we'll start with a two sample example that can be done both by the z-test and by the new method—-Chi-square.

Example: ABC Poll of 1500 adults, 750 men, 750 women on Presidential approval rating

Ex**: On March 10, 2002, 623 men, 607 women approved of Bush's handling of his job. Was there a gender gap in the approval ratings significant at the 5% level?**
This is a two-sample test of proportions. Population 1 is all men; Population 2 is all women.
Step 1: Hypotheses: $H_0: p_1 = p_2 \; H_a: p_1 \neq p_2$
Step 2: Test Statistics: We find the sample proportions:
Men: $\hat{p}_1 = 623/750 = 0.83067$, Women: $\hat{p}_2 = 607/750 = 0.80933$, Pooled:

$$\hat{p} = \frac{623+607}{750+750} = 0.82.$$

Then, using the pooled proportion for the standard error:

$$z = \frac{\frac{623}{750} - \frac{607}{750}}{\sqrt{0.82(1-0.82)\left(\frac{1}{750}+\frac{1}{750}\right)}} = \frac{0.83067 - 0.80933}{0.019839} = 1.075.$$

Step 3: P value $= 2P(z > 1.075) = 2(0.14) = 28\%$.
Step 4: We fail to reject the null hypothesis. There was no significant gender gap.

Ex**: On Sept 9, 2001, 443 men out of 750 and 390 women out of 750 approved of Bush's handling of job. Is there a gender gap in the approval ratings significant at the 5% level?**
Step 1: Hypotheses: $H_0: p_1 = p_2 \; H_a: p_1 \neq p_2$
Step 2: Test Statistics: We find the sample proportions
Men: $\hat{p}_1 = 0.5907$, Women: $\hat{p}_2 = 0.5200$ Pooled:

$$\hat{p} = \frac{443+390}{750+750} = 0.556$$

$$z = \frac{\frac{443}{750} - \frac{390}{750}}{\sqrt{0.556(1-0.556)\left(\frac{1}{750}+\frac{1}{750}\right)}} = \frac{0.5907 - 0.520}{0.025657} = 2.754$$

Step 3: P value $= 2P(z > 2.754) = 2(0.00294) = 0.0059 = 0.59\%$
Step 4: We reject the null hypothesis. There **was** a significant gender gap.

Aside: What might have caused the difference between the two results? 9/11. This dramatically changed the way people regarded the president, and in particular the way women reacted.

RELATIONSHIPS BETWEEN CATGORICAL VARIABLES: TWO WAY TABLES

In Chapter 2, we looked at relationships between quantitative variables (using regression and correlation). Now we look at relationships between categorical variables, using counts. In this example of presidential approval ratings, gender is **explanatory variable**; approval is **response variable**.

We give data in a two-way table, which records counts of all the possible counts (including, for example, the total number of approvals, which we did not have before).

March 10, 2002

	Men	Women	
Approve	623	607	1230
Don't	127	143	270
	750	750	1500

Sept 9, 2001

	Men	Women	
Approve	443	390	833
Don't	307	360	667
	750	750	1500

Look at proportions of whole population: this is called the **joint distribution**. Each proportion is out of the whole 1500 (so, for example, 0.415 = 623/1500 and 0.82 = 1230/1500). Here all the proportions in the interior four cells of the table should add to 1.

Joint Distribution: March 10, 2002

	Men	Women	
Approve	0.415	0.405	0.820
Don't	0.085	0.095	0.180
	0.500	0.500	1.000

Joint Distribution: Sept 9, 2001

	Men	Women	
Approve	0.295	0.260	0.555
Don't	0.205	0.240	0.445
	0.500	0.500	1.000

To compare men and women, we need a **conditional distribution,** conditioned on gender

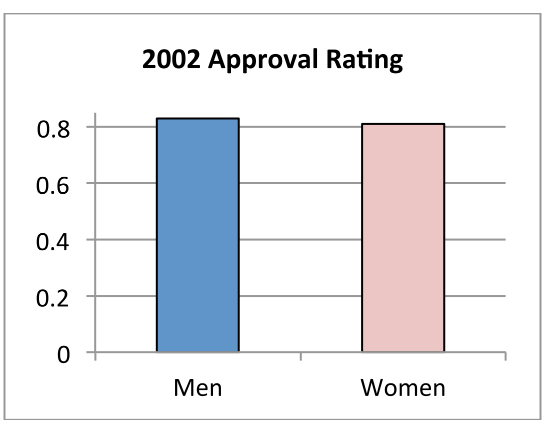
Conditional Distribution: March 10, 2002

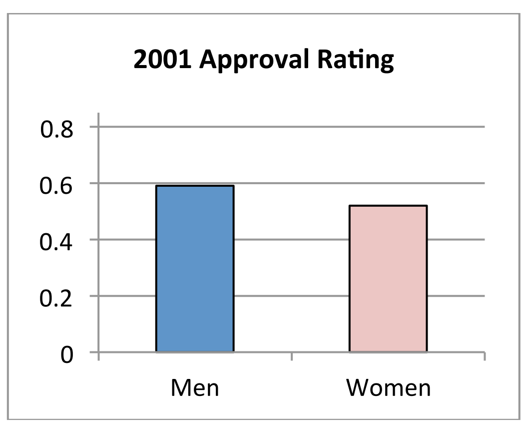
	Men	Women
Approve	0.83	0.81
Don't	0.17	0.19
	1.00	1.00

Conditional Distribution: Sept 9, 2001

	Men	Women
Approve	0.59	0.52
Don't	0.41	0.48
	1.00	1.00

Bar graphs illustrate the difference between the conditional distributions. How can we tell if the difference between man and woman is large enough to be significant? We need a hypothesis test.





HYPOTHESIS TEST: CHI-SQUARED

To decide if the difference between the genders is significant, we can use a two-sample difference in proportions (as earlier in this class) or a chi-squared test. For chi-squared, we compare the observed two-way table of counts with the table of **expected counts**, which is obtained by assuming the row and column totals are the same, given no interaction.

Expected Counts: 2002

	Men	Women	
Approve	=(0.82)750 615	=(0.82)750 615	**1230** (=0.82)
Don't	=(0.18)750 135	=(0.18)750 135	**270** (=0.18)
	750	**750**	**1500**

Expected Counts: 2001

	Men	Women	
Approve	416.5 =(0.555)750	416.5 =(0.555)750	833 (0.555)
Don't	333.5 =(0.445)750	333.5 =(0.445)750	667 (0.445)
	750	750	1500

We use expected counts to compare whether each column has same percent breakdown, that is, whether men and women approve in the same proportions.
[Aside: Alternatively, which turns out to be same thing, we can look at whether those who approve/disapprove have same sex breakdown.]

For reference, here is the original table of observed counts:

Observed Counts: 2002

	Men	Women	
Approve	623	607	1230
Don't	127	143	270
	750	750	1500

Observed Counts: 2001

	Men	Women	
Approve	443	390	833
Don't	307	360	667
	750	750	1500

To calculate expected counts

↙total men ↙total approve

$$\frac{750}{1500}\cdot 1230 = \frac{750\cdot 1230}{1500} = 0.5\cdot 1230 = 615$$

↑total overall ↑men, approve

$$\text{Expected cell count} = \frac{\text{Row total}\times\text{Column total}}{n}$$

Now we make a table of differences, with Observed – Expected:

Difference: 2002	Men	Women
Approve	8 =623-615	−8 =607-615
Don't	−8 =127-135	8 =143-135

Difference: 2001	Men	Women
Approve	26.5	−26.5
Don't	−26.5	26.5

Looking ahead:

Ex: What differences do you see between the tables of differences that may indicate significance?
Recall: 2002 is not significant; 2001 is significant. You can see the difference appearing in the size of the differences.
Ex: What do you notice about the sums of the differences in one table? How many values are needed to determine all the others?
Sums in any one row or any one column are 0, so knowing one value gives us all the others.
We say the table has **one degree of freedom**.

INFERENCE FOR TWO-WAY TABLES: CHI SQUARE TEST FOR INDEPENDENCE

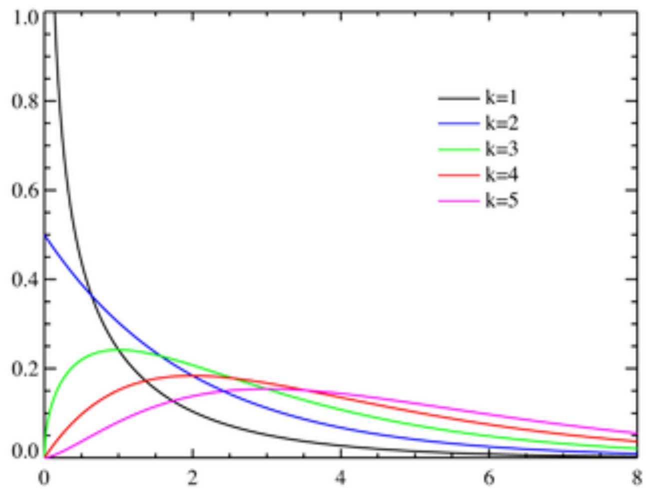
Null hypothesis: There is no association between variables.
This means the columns have same % breakdown. In addition, the rows have same % break down.
Alternative hypothesis: There **is** an association between the variables, but we do not specify what.

Chi-Square Test Statistic

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

With $df = (\#\text{ rows} - 1)(\#\text{ cols} - 1)$

We see that $\chi^2 \geq 0$. For various degrees of freedom, the chi-squared distribution is shown below:[1]



Note that when there is an association between variables, the observed and expected counts are further apart, so we are going to get large values of χ^2. Thus

- Large values of χ^2 give significance.

Conversely, if there is no association, the expected and observed counts are similar (they may not be identical because of random variation), so we get a small value of χ^2. Thus

- Small values of χ^2 give no significance.

Ex: Perform the Chi-Square hypothesis test for 2002

Step1:
H_0: There is no association between gender and approval rating
H_a: There is an association between gender and approval rating

Step 2:
The test statistic is

$$\chi^2 = \frac{(623-615)^2}{615} + \frac{(607-615)^2}{615} + \frac{(127-135)^2}{135} + \frac{(143-135)^2}{135} = 1.156 \quad \text{with} \quad \text{df} = 1$$

Step3:
From the table
P value $= \text{P}(\chi^2 > 1.156) > 0.25 = 25\%$
From calculator
$P(\chi^2 > 1.156) = \chi^2\text{cdf}(1.156,20,1) = 0.282 = 28.2\%$ [

Step 4:
We do not reject the null hypothesis. There is no evidence for association—there is no evidence for a gender gap.

[1] http://en.wikipedia.org/wiki/Chi-square_distribution

<u>Ex</u>: Perform the Chi-Square hypothesis test for 2001
Step1:
H_0: There is no association between gender and approval rating
H_a: There is an association between gender and approval rating
Step 2:
The test statistic is

$$\chi^2 = \frac{(443-416.5)^2}{416.5} + \frac{(390-416.5)^2}{416.5} + \frac{(307-333.5)^2}{333.5} + \frac{(360-333.5)^2}{333.5} = 7.584 \quad \text{with} \quad \text{df} = 1$$

Step3:
From table
$P(\chi^2 > 7.584)$ is between 0.01 and 0.005, so 0.5% to 1%.
From calculator
$\text{p}(\chi^2 > 7.584) = \chi^2\text{cdf}(7.584, 100, 1) = 0.0059 = 0.59\%$
Step 4:
We reject the null hypothesis. There is evidence for association—that is, for a gender gap.

<u>Ex</u>: Interpret the P-value in the previous example.
Assuming there is no interaction between gender and approval, there's 0.59% chance that we see observed values as far, or further, from the expected values by chance.

<u>Comparison between the Chi-Square and the previous z-test</u>

2002

$\chi^2 = 1.156$	P-value = 28%	$1.1560 = (1.075)^2$
$z = 1.075$	P-value = 28%	

2001

$\chi^2 = 7.584$	P-value = 0.59%	$7.584 = (2.754)^2$
$z = 2.754$	P-value = 0.59%	

<u>Ex</u>: What do you notice about the P-values of the two ways of doing each test? What does it tell us?
Provided the z-test is two-sided, the P-values of the two tests are the same, so we always come to the same conclusion.

<u>Notes on Using the Chi-Square Test</u>

- Rejecting null hypothesis does not tell us which way the interaction happens.
- Since χ^2 is always positive; there is only an upper tail. We never multiply the probabilities by 2.
- **Conditions**: We can use the Chi-Square test when
  - For tables larger than 2×2: Average of expected counts ≥ 5; all expected counts ≥ 1
  - For 2×2 tables: All expected counts ≥ 5