

CONFIDENCE INTERVALS FOR THE PROPORTIONS

Recall our conclusion from Example 3.4 after conducting a hypothesis test: “We have evidence the proportion of families afflicted by alcoholism differs depending on whether or not the female in the family had primary unipolar depression.” The next natural question is as follows: *how different* are these proportions?

One could consider using the methods we discussed in Chapter 2 to construct a confidence interval for the true proportion of families afflicted by alcoholism for each group separately:

		Alcoholism		
Count		No	Yes	
Group	Control	205	94	299
	Depression	121	89	210
		326	183	509

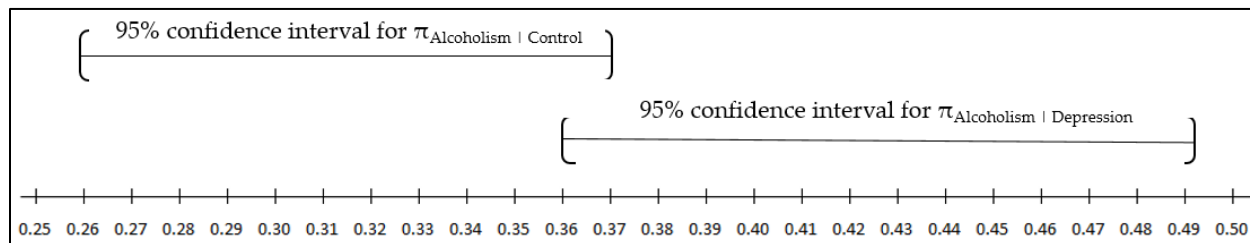
Depression Group:

- $\hat{\pi}_{\text{Alcoholism}|\text{Depression}} = 0.4238$
- margin of error = $1.96\sqrt{\frac{(0.4238)(1 - 0.4238)}{210}} = 0.0668$
- 95% confidence interval: $0.36 \leq \pi_{\text{Alcoholism}|\text{Depression}} \leq 0.49$

Control Group:

- $\hat{\pi}_{\text{Alcoholism}|\text{Control}} = 0.3144$
- margin of error = $1.96\sqrt{\frac{(0.3144)(1 - 0.3144)}{299}} = 0.0526$
- 95% confidence interval: $0.26 \leq \pi_{\text{Alcoholism}|\text{Control}} \leq 0.37$

These intervals can be visualized as shown below:



If these confidence intervals did *not* overlap at all, then we could be confident that a significant difference exists between the two proportions of interest. The opposite, however, is not necessarily true.

For example, note that the confidence intervals slightly overlap in this case. This may tempt you to conclude that there is not enough evidence to conclude the proportions differ between groups. Recall, however, that both Fisher's exact test (p-value = 0.0146) and the chi-square test (p-value = 0.0113) indicated that there *was* a significant difference in the alcoholism rate between the two groups.

The fact that we *can* compute separate confidence intervals for the two population proportions and then see if they overlap doesn't mean that we *should* compute separate confidence intervals. A better approach to quantifying the difference in the two population proportions of interest is discussed in the next section.

CONFIDENCE INTERVALS FOR THE DIFFERENCE IN PROPORTIONS

If the question really is, "How large is the difference in these proportions?", then why not focus on what we really care about – the difference!

Recall the general form of a confidence interval:

General form of a confidence interval
sample statistic \pm margin of error

In this context, our sample statistic of interest is now the *difference* in the two sample proportions:

$$\hat{\pi}_{\text{Alcoholism|Depression}} - \hat{\pi}_{\text{Alcoholism|Control}} = 0.4238 - 0.1344 = 0.1094$$

This difference in the sample proportions can be interpreted as follows: "In our *sample* of families, the proportion with alcoholism was about 11 percentage points higher for the Depression group than for the Control group."

The difference we just calculated describes the size of the difference in the proportions obtained in our SAMPLES. To generalize these results to the POPULATIONS of interest and to make statements about how large the difference is between the true population proportions, we must calculate a confidence interval. In other words, we will consider not only our point estimate for the difference in proportions but also the margin of error associated with this difference.

The confidence interval for the difference in proportions is calculated as follows.

Let $\hat{\pi}_1 = \hat{\pi}_{\text{Alcoholism}|\text{Depression}}$ and $\hat{\pi}_2 = \hat{\pi}_{\text{Alcoholism}|\text{Control}}$.

1. Start with the sample statistic (i.e., **point estimate**) of interest: $\hat{\pi}_1 - \hat{\pi}_2 =$
2. Calculate the **standard error** associated with this point estimate. The standard error associated with the difference between two proportions is calculated as follows.

$$\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} =$$

3. Calculate the **margin of error**. This is defined as 1.96 standard errors for a 95% confidence interval (it's good enough to think of this as "about two" standard errors).

Note that if you desired to construct a 90% or a 99% confidence interval, instead, the formula for the margin of error would change slightly.

Level of Confidence	90%	95%	99%
Margin of Error	1.645 standard errors	1.96 standard errors	2.58 standard errors

For this problem, the margin of error is:

4. Find the endpoints of the confidence interval for the difference in proportions:

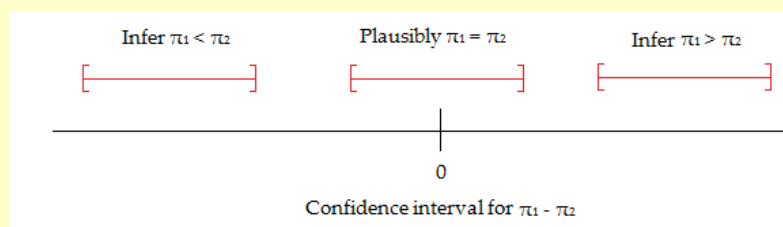
Lower endpoint = $(\hat{\pi}_1 - \hat{\pi}_2) - \text{margin of error} =$ _____

Upper endpoint = $(\hat{\pi}_1 - \hat{\pi}_2) + \text{margin of error} =$ _____

We can also write this interval as follows:

Guidelines for Interpreting a Confidence Interval for a Difference in Proportions:

1. Check whether zero falls in the interval or not. If it does, it is *plausible* (but not proof!) that the two proportions are equal. In this case, we do not have enough evidence that the proportions differ.
2. If all values in a 95% confidence interval for $\pi_1 - \pi_2$ are positive, then you can infer that $\pi_1 > \pi_2$. The interval tells you how much bigger (with 95% certainty) π_1 is. Similarly, if all values in a confidence interval for $\pi_1 - \pi_2$ are negative, then you can infer that $\pi_1 < \pi_2$. The interval tells you how much bigger π_2 is.

Questions:

1. Interpret the 95% confidence interval for the difference in proportions in this problem.
2. Does this confidence interval agree with the results of the hypothesis test? Explain.
3. Note that we could have also reversed the order of the proportions when finding the difference. How would the result have changed had we found a 95% confidence interval for $\pi_{\text{Alcoholism} \mid \text{Conrol}} - \pi_{\text{Alcoholism} \mid \text{Depression}}$, instead?

Example 3.6: Adoption and Suicide Attempts

In September of 2013, researchers from the University of Minnesota published a study in the journal *Pediatrics*. This study was described in a *Fox News* article titled “Adopted teens may be at higher risk of suicide.” This article described the results as follows:

[The researchers] examined data from an existing University of Minnesota study of 692 adopted children and 540 non-adopted siblings in Minnesota... All of the adopted kids, who were between 11 and 21 years old during the study period, had been taken in by their families before age two. Almost three quarters of the adopted children were born abroad, most of the foreign-born children were from South Korea and 60 percent of those were girls. At the beginning of the study, and again about three years later, the researchers asked participating families if the children had made a suicide attempt. Over the three years of the study, 56 children attempted suicide at least once, according to the family members’ reports. Of those kids, 47 were adopted and nine were not adopted.

Questions:

1. Create a table of counts based on the data obtained in this study:

	Suicide Attempt	No Suicide Attempt	Totals
Adopted			
Not adopted			
Totals			

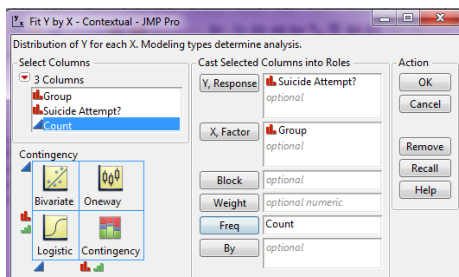
2. Find the proportion of adopted children that attempted suicide.
3. Find the proportion of non-adopted children that attempted suicide.
4. Find and interpret a 95% confidence interval for the difference in proportions.

Confidence Interval for a Difference in Proportions in JMP

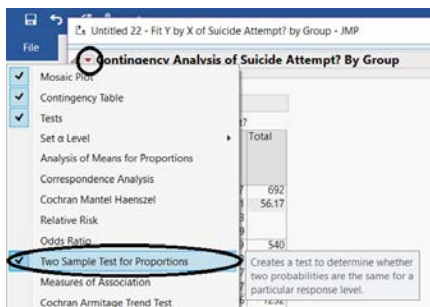
We can use JMP to find a confidence interval for a difference in proportions. Let's once again consider the Adoption and Suicide Attempt example. The data should be entered in JMP as follows.

Group	Suicide Attempt?	Count
Adopted	Yes	47
Adopted	No	645
Not adopted	Yes	9
Not adopted	No	531

In JMP, select **Analyze > Fit Y by X** and enter the following:



Click OK. On the output that appears, click on the red drop-down arrow next to **Contingency Analysis...** and select **Two Sample Test for Proportions**.



JMP returns the following:

Two Sample Test for Proportions			
Description	Proportion Difference	Lower 95%	Upper 95%
P(Yes Adopted)-P(Yes Not adopted)	0.051252	0.028698	0.07273
Adjusted Wald Test		Prob	
P(Yes Adopted)-P(Yes Not adopted) ≥ 0		<.0001*	
P(Yes Adopted)-P(Yes Not adopted) ≤ 0		1.0000	
P(Yes Adopted)-P(Yes Not adopted) = 0		<.0001*	

Response Suicide Attempt? category of interest

No

Yes

CALCULATING RISKS AND THE RELATIVE RISK

The confidence interval calculated above helps us to quantify the discrepancy in the proportion of adopted versus non-adopted children that had attempted suicide. We could also measure the amount of discrepancy between these two proportions based on something called **relative risk**. This involves taking the *ratio* of the two proportions instead of the *difference*.

Note that in the previous example, our estimate of the *proportion* of adopted children that had attempted suicide could be viewed as an estimate of the *probability* of an adopted child attempting suicide (the same could be said for the non-adopted group). When we calculate the probability of an adverse event, we sometimes refer to this probability as a **risk**.

For example, we can calculate the **risk** of attempting suicide for both groups in this study as follows:

	Suicide Attempt	No Suicide Attempt	Total
Adopted	47	645	692
Not adopted	9	531	540

- Risk of attempting suicide for the Adopted group = $\hat{\pi}_{\text{Suicide} | \text{Adopted}}$
- Risk of attempting suicide for the Not Adopted group = $\hat{\pi}_{\text{Suicide} | \text{Not Adopted}}$

The **relative risk** (also called the risk ratio) is then computed as follows:

$$\text{Relative Risk} = \frac{\hat{\pi}_{\text{Suicide} | \text{Adopted}}}{\hat{\pi}_{\text{Suicide} | \text{Not Adopted}}} =$$

Comments:

1. We interpret this number by saying that in this study, the risk of suicide attempt for adopted children was _____ times as large as the risk of suicide attempt for children who were not adopted. In other words, adopted children were _____ times as likely to attempt suicide as were non-adopted children.
2. A relative risk value of 1.0 is the reference value for making comparisons. That is, a relative risk of 1.0 says that there is no difference in the two proportions of interest.

3. When you are interpreting a relative risk, you **MUST** consider which value you have in the numerator. For example, we could have also calculated the relative risk as follows:

$$\text{Relative Risk} = \frac{\hat{\pi}_{\text{Suicide}|\text{Not Adopted}}}{\hat{\pi}_{\text{Suicide}|\text{Adopted}}} =$$

How would we interpret this value?

CALCULATING ODDS AND THE ODDS RATIO

Another quantity that is often used to describe differences in categorical outcomes between groups is the odds ratio. This ratio is used more commonly in practice than the relative risk ratio; however, it is more difficult to interpret and is sometimes harder to understand.

Odds

Before computing an odds ratio, we first need to compute the odds for both groups. The odds of an event occurring are defined as follows:

$$\text{Odds of Event} = \frac{\text{Probability Event Occurs}}{\text{Probability Event Does Not Occur}}$$

Now, consider our Adoption and Suicide Attempt example. For each group (Adopted and Not Adopted), we will compute the odds of attempting suicide. Recall the contingency table for this example.

	Suicide Attempt	No Suicide Attempt	Total
Adopted	47	645	692
Not adopted	9	531	540

Use the definition from above to find the odds of attempting suicide for both groups:

- Odds of Attempting Suicide in the Adopted Group:

- Odds of Attempting Suicide in the Not Adopted Group:

Note that we can also compute the odds as the number of subjects in each group that attempted suicide divided by the number in that group that did not attempt suicide.

	Suicide Attempt	No Suicide Attempt	Total
Adopted	47	645	692
Not adopted	9	531	540

- Odds of Attempting Suicide in the Adopted Group:

$$\frac{\text{Number that Attempted Suicide in the Adopted Group}}{\text{Number that did NOT Attempt Suicide in the Adopted Group}} =$$

- Odds of Attempting Suicide in the Non-adopted Group:

$$\frac{\text{Number that Attempted Suicide in the Non - adopted Group}}{\text{Number that did NOT Attempt Suicide in the Non - adopted Group}} =$$

Odds Ratio

The **odds ratio** is then simply the ratio of the odds for the two groups:

$$\text{Odds Ratio} = \frac{\text{Odds of Attempting Suicide in the Adopted Group}}{\text{Odds of Attempting Suicide in the Non - adopted Group}} =$$

We interpret this quantity by saying that in this study, the odds of attempting suicide were about _____ times as high for adopted children as for non-adopted children.

We could also have calculated the odds ratio as follows:

$$\text{Odds Ratio} = \frac{\text{Odds of Attempting Suicide in the Non - Adopted Group}}{\text{Odds of Attempting Suicide in the Adopted Group}} =$$

How would we interpret this value?

Comments:

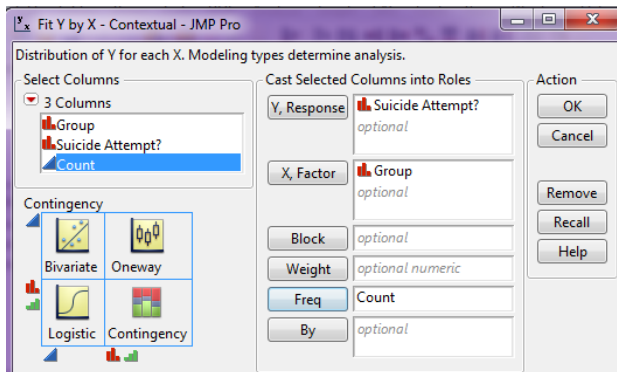
1. An odds ratio of 1.0 implies that there is no observable difference between the two odds. This is always the reference value!
2. Odds ratios are often used in health care experiments when we want to determine how much more likely a subject is to develop a disease when they have some risk factor as opposed to not having that risk factor. In a *retrospective study* (also called a *case-control study*), samples are drawn from those who have the disease and those who do not. It is then determined whether each subject has the risk factor. In this type of study, the relative risk cannot be estimated because the study design does not allow us to estimate the probability of the disease for those with or for those without the risk. So, we must use the odds ratio instead of the relative risk for these studies. This is why statisticians put so much emphasis on the odds ratio even though the relative risk is easier to understand and interpret.
3. If the probability of the outcome under study in the general population is very low, then the odds ratio is typically close to the relative risk and can thus be used as an *estimate* of the relative risk.

Relative Risk and Odds Ratios in JMP

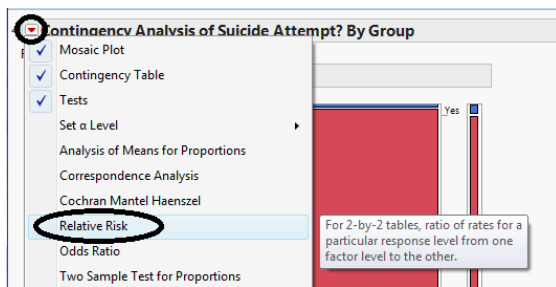
To discuss the calculation of the relative risk and the odds ratio in JMP, let's once again consider the Adoption and Suicide Attempt example. The data should be entered in JMP as follows.

Group	Suicide Attempt?	Count
Adopted	Yes	47
Adopted	No	645
Not adopted	Yes	9
Not adopted	No	531

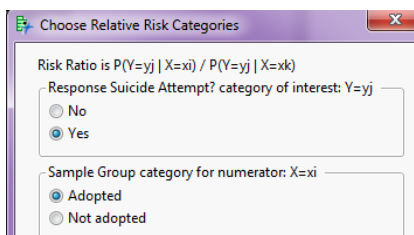
In JMP, select **Analyze > Fit Y by X** and enter the following:



Click OK. On the output that appears, click on the red drop-down arrow next to **Contingency Analysis...** and select **Relative Risk**.



Enter the following and click OK.



JMP returns the following:

Relative Risk			
Description	Relative Risk	Lower 95%	Upper 95%
P(Yes Adopted)/P(Yes Not adopted)	4.075145	2.015176	8.240868

If you also ask for the **Odds Ratio** from this red drop-down arrow, JMP returns this quantity:

Odds Ratio		
Odds Ratio	Lower 95%	Upper 95%
0.2326	0.112956	0.478973

Note that JMP is by default computing the odds of NOT attempting suicide since this column appears first in the contingency table. It is then computing the odds ratio with the Adopted Group in the numerator, since this row appears first in the table.

Contingency Table				
		Suicide Attempt?		
Count		No	Yes	
Group	Adopted	645	47	692
	Not adopted	531	9	540
		1176	56	1232

After re-ordering the columns for the *Suicide Attempt?* variable we obtain the following:

Contingency Table				
		Suicide Attempt?		
Count		Yes	No	
Group	Adopted	47	645	692
	Not adopted	9	531	540
		56	1176	1232

Odds Ratio		
Odds Ratio	Lower 95%	Upper 95%
4.299225	2.0878	8.853021

Questions:

1. Verify that the relative risk and odds ratio shown match our calculations.
2. Note that JMP gives a confidence interval for both the relative risk and the odds ratio. What would it imply if either of these confidence intervals was to include the value one? Explain.

Example 3.7: Importance of Considering Absolute Risk

In 1995, the UK Committee on Safety of Medicines issued a warning that birth control pills increased the risk of life-threatening blood clots. The data on which these conclusions were drawn were very similar to the following outcomes.

Birth Control Pill?	Blood Clot	No Blood	
		Clot	Total
Yes	2	6998	7000
No	1	6999	7000
Total	3	13,997	14,000

Questions:

1. Find the risk of blood clots for those using birth control pills.
2. Find the risk of blood clots for those not using birth control pills.
3. Find and interpret the relative risk.
4. All of the headlines related to this study mentioned the “two-fold risk increase” or “100% risk increase” in blood clots because of the birth control pill. What do you suppose happened to the number of birth control users after these results were publicized? Are there any potential negative consequences to this? Explain.
5. What do you suppose would have happened had patients also considered the absolute risk of blood clots for those using birth control pills?

INFERENCES FOR COMPARING TWO OR MORE PROPORTIONS

Chi-Square Test of Independence

In this section, we will once again discuss a hypothesis testing procedure that uses the chi-square distribution. Specifically, this is known as the *Chi-Square Test of Independence*.

Example 3.8: Support for Iraq War and Political Affiliation

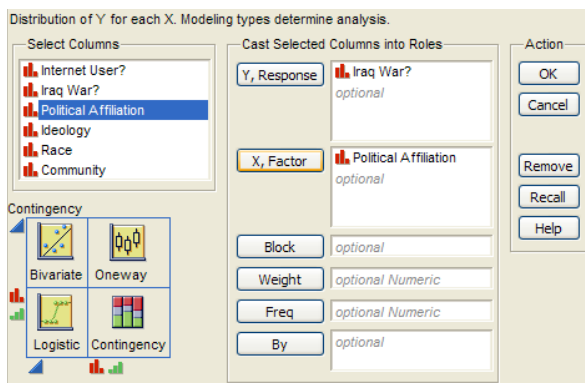
In March of 2003, the Pew Internet & American Life Project commissioned Princeton Survey Research Associates to develop and carry out a survey of what Americans thought about the recent war in Iraq. Some of the results of this survey of over 1,400 American adults are given in the JMP data file **IraqWar.JMP**. (Source: McClave & Sincich, Problem 13.33)

Responses to the following questions were recorded:

1. Do you support or oppose the Iraq War?
2. Do you ever go online to access the Internet or World Wide Web?
3. Do you consider yourself a Republican, Democrat, or Independent?
4. In general, would you describe your political views as very conservative, conservative, moderate, liberal, or very liberal?
5. What is your race?
6. Do you live in a suburban, rural, or urban community?

Research Question: Is there a significant association between Support for the Iraq War and Political Affiliation?

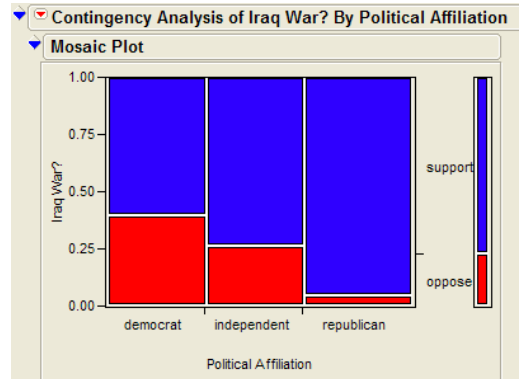
Note that this investigation requires us to focus on only two of the measured variables: *Support for the Iraq War* and *Political Affiliation*. First, let's summarize the data using JMP. Select **Analyze > Fit Y by X** and enter the following:



JMP returns the contingency table and mosaic plot:

Contingency Table

		Iraq War?		
		oppose	support	
Political Affiliation	Count			
	Row %			
	democrat	173	262	435
		39.77	60.23	
independent	112	315	427	
	26.23	73.77		
republican	21	451	472	
	4.45	95.55		
		306	1028	1334



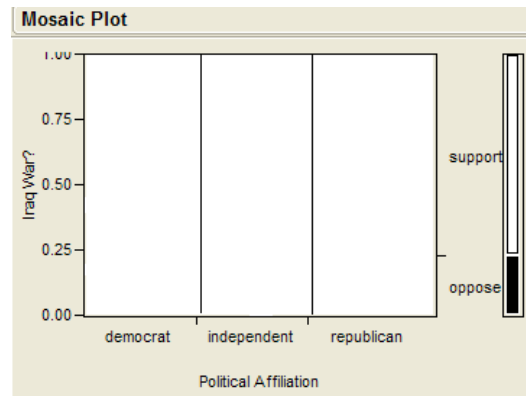
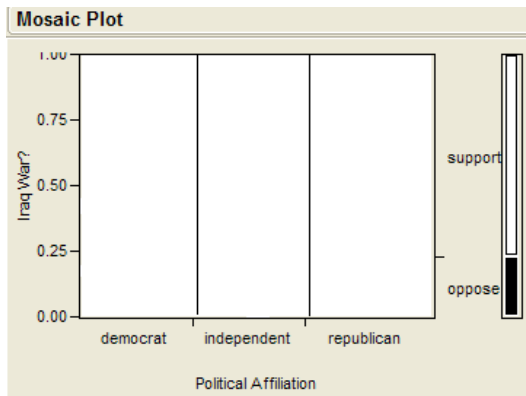
Questions:

1. What can you say about the association between Support for the Iraq War and Political Affiliation based on the data obtained in the sample?

2. On the graphs below, sketch hypothetical examples of each of the following situations.

Political Affiliation IS Associated with Support for the Iraq War

Political Affiliation IS NOT Associated with Support for the Iraq War



Carrying out the chi-square test of independence:

Assumptions behind the Chi-square Test:

The chi-square test of independence may be inappropriate for tables with very small expected cell frequencies. One rule of thumb suggests that most of the expected cell frequencies in the table should be 5 or more; otherwise, the chi-square approximation may not be reliable. Also, all observations that are counted in the contingency table should be independent of each other.

Step 1: Convert the research question into H₀ and H_a:

H₀:

H_a:

Step 2: Calculate a test statistic and p-value from the data.

$$\text{Test Statistic} = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Observed counts:

		Iraq War?		
		oppose	support	
Political Affiliation	Count			
	democrat	173	262	435
	independent	112	315	427
	republican	21	451	472
	306	1028	1334	

Finding the expected counts:

		Iraq War?		
		oppose	support	
Political Affiliation	Count			
	democrat			435
	independent			427
	republican			472
	306	1028	1334	

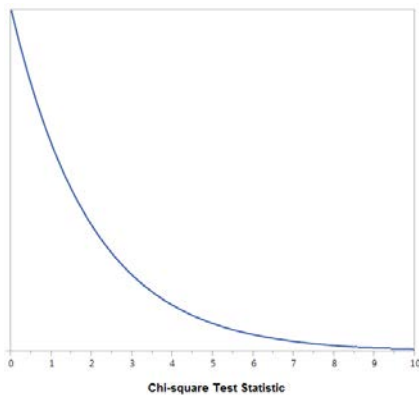
Once we have the expected counts, we can calculate the test statistic as follows:

$$\sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(173 - 99.78)^2}{99.78} + \frac{(262 - 335.22)^2}{335.22} + \frac{(112 - 97.95)^2}{97.95} + \frac{(315 - 329.05)^2}{329.05} + \frac{(21 - 108.27)^2}{108.27} + \frac{(451 - 363.73)^2}{363.73} = 53.217 + 15.992 + 2.016 + 0.600 + 70.343 + 20.939 = 163.6$$

Finding the p-value:

Recall that if the null hypothesis is true, the chi-square test statistic follows a chi-square distribution with $df = (r - 1)(c - 1)$.

The following graph shows the chi-square distribution with $df = 2$ (i.e., this shows us what test statistics we expect to see by chance when the null hypothesis is true). The p-value is found by plotting the chi-square test statistic on the x-axis and calculating the area under the curve above the test statistic (note that this represents how often we see test statistics at least as extreme as the observed when the null hypothesis is true).



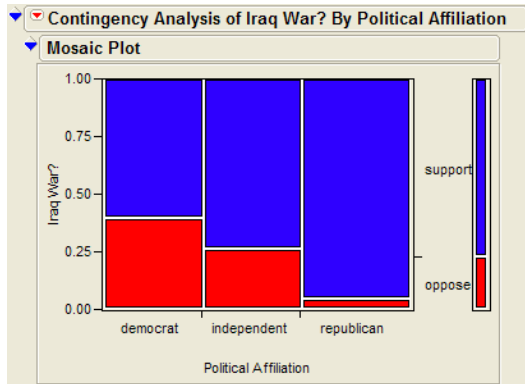
The output for the chi-square test is also given in the JMP output:

Tests				
	N	DF	-LogLike	RSquare (U)
	1334	2	94.450724	0.1315
Test	ChiSquare	Prob>ChiSq		
Likelihood Ratio	188.901	<.0001*		
Pearson	163.615	<.0001*		

Step 3: Write a conclusion in the context of the problem

“There is a significant association between support for the Iraq war and political affiliation.”

Note that we can describe this relationship based on the patterns displayed in the mosaic plot.



Chi-square Test of Independence versus Chi-square Test of Homogeneity

In the context of our Iraq War example, some statisticians might differentiate between these two versions of the chi-square test as follows.

Name of Test	Research Question	When is it Appropriate?
Chi-Square Test of Homogeneity	Does the proportion that supports the Iraq War <u>differ</u> across Political Affiliation?	When the data are collected from separate random samples (e.g., the researchers take separate random samples of Democrats, Republicans, and Independents and then measure the response variable on each subject). In general, this tests whether the response variable differs across two (or more) populations.
Chi-Square Test of Independence	Is there a significant <u>association</u> between Support for the Iraq War and Political Affiliation?	When the data are collected on a single random sample and the groupings are then retrospectively determined (e.g., the researchers take a single random sample and then classify each subject according to both political affiliation and support for the war). In general, this tests whether there is an association between two variables in a single population.

Note that the choice of homogeneity versus independence depends on only the study design and that the phrasing of the research question is slightly different between the two versions.

The actual study described here in Example 3.8 involved a single random sample, so this is technically a chi-square test of independence. If the researchers had taken separate random samples of U.S. citizens from each political affiliation group, how would our results have changed? The answer is... not much. We would have calculated the chi-square test statistic in the same way, so we would have obtained the same test statistic and the same p-value. Our conclusion might have focused on there being a significant difference in the proportion that supports the war across political affiliation instead of focusing on there being a significant relationship between the two variables, but the overall story would have remained the same.

Example 3.9: Opinions on Disciplining Children across Regions of U.S.

On September 16-17, 2014, an NBC News/Marist Poll surveyed a random sample of 606 adults nationwide. Respondents were asked the following question: “Do you think it is right or wrong for parents to discipline their children by striking them - either with a paddle, switch, or belt?” The results were separated according to the Region of the U.S. in which the respondent lived (Northeast, Midwest, South, or West). The data are summarized in the following table.

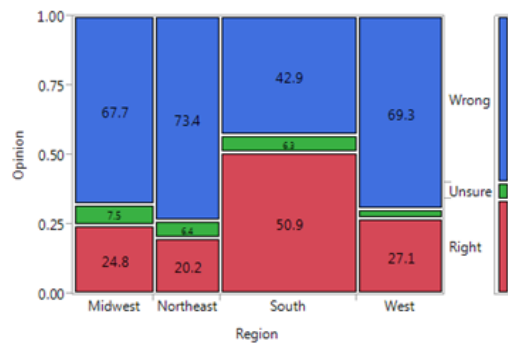
	Right	Wrong	Unsure	Totals
Northeast	22	80	7	109
Midwest	33	90	10	133
South	114	96	14	224
West	38	97	5	140
Totals	207	363	36	606

The research question is as follows: **Is there a statistically significant association between Region and Opinion on this Discipline Issue?**

First, since we don’t have access to the raw data, we can enter the data from the contingency table into JMP as follows (see the file **Discipline.jmp** posted on the course website).

Region	Opinion	count
Northeast	Right	22
Northeast	Wrong	80
Northeast	Unsure	7
Midwest	Right	33
Midwest	Wrong	90
Midwest	Unsure	10
South	Right	114
South	Wrong	96
South	Unsure	14
West	Right	38
West	Wrong	97
West	Unsure	5

We can then use the **Analyze > Fit Y by X** menu to obtain the desired output:



Step 1: Convert the research question into H_0 and H_a .

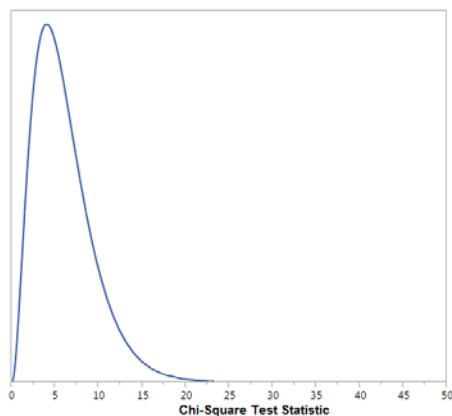
H_0 :

H_a :

Step 2: Find the test statistic and p-value from the data.

Tests			
N	DF	-LogLike	RSquare (U)
606	6	24.862750	0.0487
Test	ChiSquare	Prob>ChiSq	
Likelihood Ratio	49.725	<.0001*	
Pearson	49.548	<.0001*	

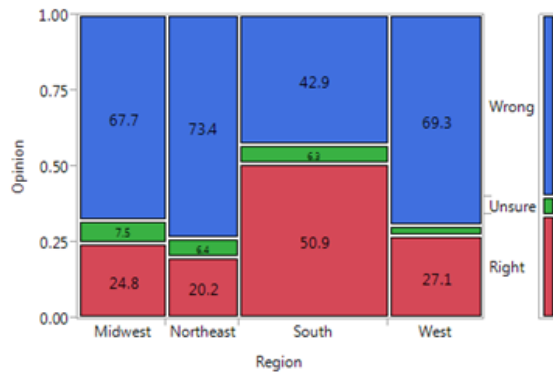
Note: The chi-square distribution with $df = 6$ is shown below. Do you see why the p-value is so small?



Step 3: Write a conclusion in the context of the problem.

“There is evidence of a significant association between region and opinion on this discipline issue.”

Finally, note that if we determine a significant relationship exists between the two variables of interest, we should take the time to describe this relationship.



Also, we can look at each cell’s contribution to the test statistic to informally determine what is driving the significance of this relationship:

Contingency Table

		Opinion			
		Right	Unsure	Wrong	
Region	Count				
	Expected				
	Cell Chi^2				
	Midwest	33 45.4307 3.4013	10 7.90099 0.5576	90 79.6683 1.3399	133
	Northeast	22 37.2327 6.2320	7 6.47525 0.0425	80 65.2921 3.3132	109
South	114 76.5149 18.3642	14 13.3069 0.0361	96 134.178 10.8630	224	
West	38 47.8218 2.0172	5 8.31683 1.3228	97 83.8614 2.0584	140	
		207	36	363	606