Example 3.1: MythBusters and the Yawning Experiment

MythBusters, a popular television program on the Discovery Channel, once conducted an experiment to investigate whether or not yawning is contagious. The premise of the experiment was to invite a stranger to sit in a booth for an extended period of time. Fifty subjects were said to be tested in total, of which 34 were "seeded" with a yawn by the person conducting the experiment. The other 16 were not given a yawn seed. Using a two-way mirror and a hidden camera, the experimenters observed and recorded the results which are given below.

	Yawned	Did not Yawn	Total
Seeded with a yawn	10	24	34
Not seeded with a yawn	4	12	16

Research Hypothesis: Those "seeded" with a yawn are more likely to actually yawn than those who are not seeded.

When we analyze data on two variables, our first step is to distinguish between the **response** variable and the **explanatory** (or **predictor**) variable.

Response variable: The outcome variable on which comparisons are made.

Explanatory (or Predictor) variable: This defines the groups to be compared.

Questions:

1. What variables must be measured in the MythBusters Yawning experiment in order to investigate the research hypothesis? Are these variables categorical or numerical?

2. Which is the response variable? Which is the explanatory variable?

Descriptive Methods for Two Categorical Variables:

A **contingency table** shows the joint frequencies of two categorical variables. The rows of the table denote the categories of the first variable, and the columns denote the categories of the second variable.
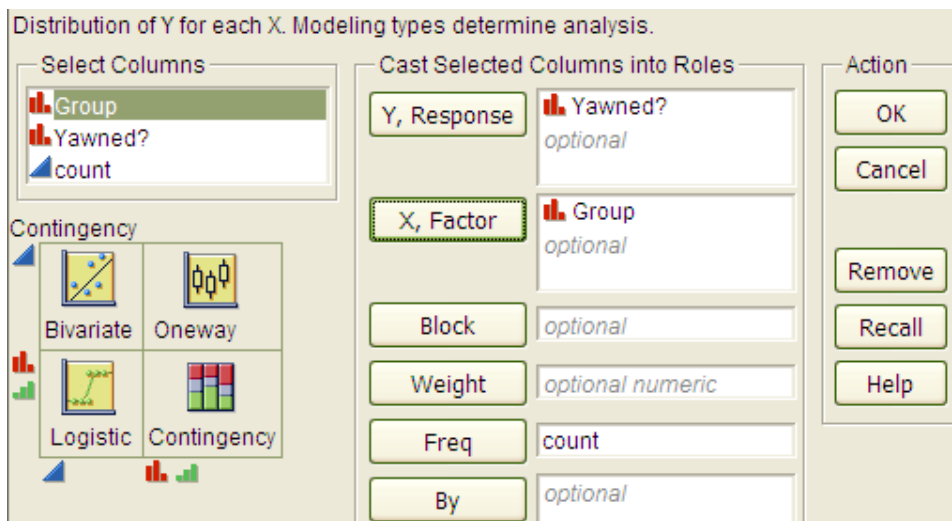
A **mosaic plot** gives a visual representation of the relationship between two categorical variables. A mosaic plot graphically presents the information given in the contingency table.

To obtain these descriptive summaries from JMP, the data should be entered as follows:

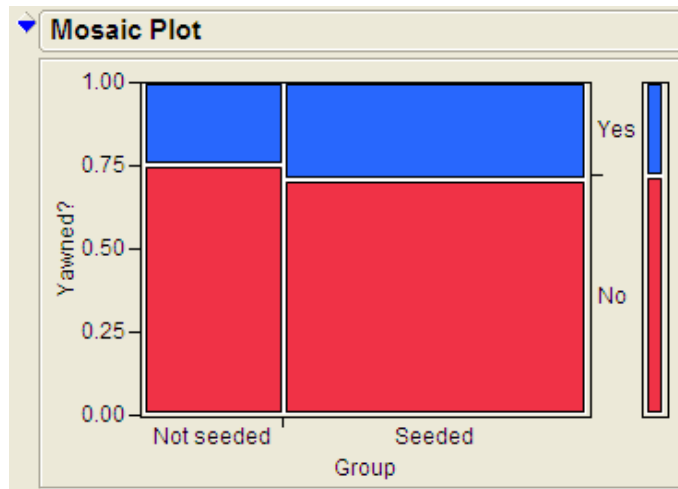Right-click on this column and choose Preselect Role > Frequency

	Group	Yawned?	count
1	Seeded	Yes	10
2	Seeded	No	24
3	Not seeded	Yes	4
4	Not seeded	No	12

Select **Analyze > Fit Y by X**. Place the response variable in the **Y, Response** box and the explanatory variable in the **X, Factor** box.

Distribution of Y for each X. Modeling types determine analysis.

Select Columns
- Group
- Yawned?
- count

Contingency
- Bivariate Oneway
- Logistic Contingency

Cast Selected Columns into Roles

Role	Value
Y, Response	Yawned? *optional*
X, Factor	Group *optional*
Block	*optional*
Weight	*optional numeric*
Freq	count
By	*optional*

Action
- OK
- Cancel
- Remove
- Recall
- Help

Click OK, and JMP returns the following:

Contingency Table

Count Total % Col % Row %	Yawned? No	Yes	
Not seeded	12 24.00 33.33 75.00	4 8.00 28.57 25.00	16 32.00
Seeded	24 48.00 66.67 70.59	10 20.00 71.43 29.41	34 68.00
	36 72.00	14 28.00	50

(Group is the row label; Yawned? is the column label)

Mosaic Plot

Questions:

1. Find the proportion that yawned in the Seeded group.

2. Find the proportion that yawned in the Not Seeded group.

3. Find the difference in the proportion that yawned between these two groups. Do these proportions differ in the direction conjectured by the researchers?

4. Even if the seeding of a yawn had absolutely no effect on whether or not a subject actually yawned, is it *possible* to have obtained a difference such as this by random chance alone? Explain.

Statistical Inference: Conducting a Simulation Study with Playing Cards

The above *descriptive analysis* tells us what we have learned about the 50 subjects in the study. Our next goal is to make *inferences* beyond what happened in the study (i.e., we want to make general statements about the population). Does the higher proportion of yawners in the seeded group provide convincing evidence that being seeded with a yawn actually makes a person more likely to yawn? Note that it is possible that random chance alone could have led to this large of a difference. That is, while it is *possible* that the yawn seeding had no effect and the MythBusters happened to observe more yawners in the seeded group just by chance, the key question is whether it is *probable*.

We will answer this question by replicating the experiment over and over again, but in a situation where we know that yawn seeding has no effect (the null model). We'll start with 14 yawners and 36 non-yawners, and we'll randomly assign 34 of these 50 subjects to the seeded group and the remaining 16 to the non-seeded group.

Note that we could use playing cards to replicate this experiment: let 14 red cards represent the yawners, and let 36 black cards represent the non-yawners. Shuffle the cards well, and randomly deal out 34 to be the seeded group. This is the first run of our simulation study. Construct the contingency table to show the number of yawners and non-yawners in each group from the first run of our simulation study.

	Yawned	Did not Yawn	Total
Seeded with a yawn			34
Not seeded with a yawn			16
Total	14	36	50

Next, note that if you know the number of yawners in the Seeded group, then you can fill in the rest of the cells in the contingency table. So, for the rest of the simulated runs, **we need only keep track of the number of yawners in the Seeded group**. Record the number of yawners in the Seeded group from your first simulated experiment in the table below. Then, repeat this randomization process nine more times, recording your results in the table.

Simulated Experiment	1	2	3	4	5	6	7	8	9	10
Number of Yawners in Seeded Group										

Next, create a dotplot of the results:

Number of Yawners Randomly Assigned to the Seeded Group
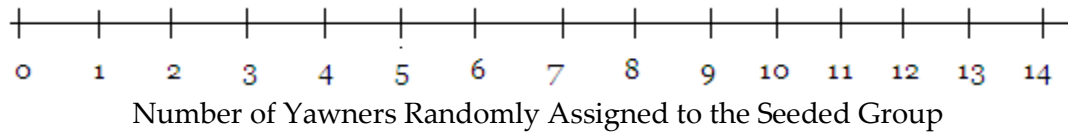
<u>Questions</u>:

1. How many randomizations were performed by the class as a whole? What proportion gave results *at least as extreme* as the actual study (10 or more yawners in the seeded group)?

2. Note that the random process used in the simulation study models the situation where the yawn seeding has no effect on whether the subject actually yawns—we simply assume there are 14 people who will yawn no matter what group they are in, and they are assigned to the two groups at random . Based on this simulation study, does it appear that random assignment of yawners to groups will result in 10 or more yawners in the seeded group just by chance? Explain.

3. Recall that the MythBusters obtained 10 yawners in the Seeded group. Given your answer to Question 2, would you say that the data provide convincing statistical evidence to support the research question? Explain.

Statistical Inference: Conducting a Simulation Study with Tinkerplots

Note that we could use technology to simulate this process much more efficiently. This will enable us to easily obtain many more simulated results so that we can be more confident in our answer to Question 3 above.

Open the file **YawnExperiment.tp** from the course website. Note that each mixer contains 50 "subjects." The first mixer contains 14 yawners and 36 non-yawners; the second mixer contains 34 subjects in the Seeded group and 16 in the Non-seeded group.

Make sure the **Repeat** value is set to 50 (the number of subjects in the study), and set BOTH mixers to sample **Without Replacement** (as shown below).
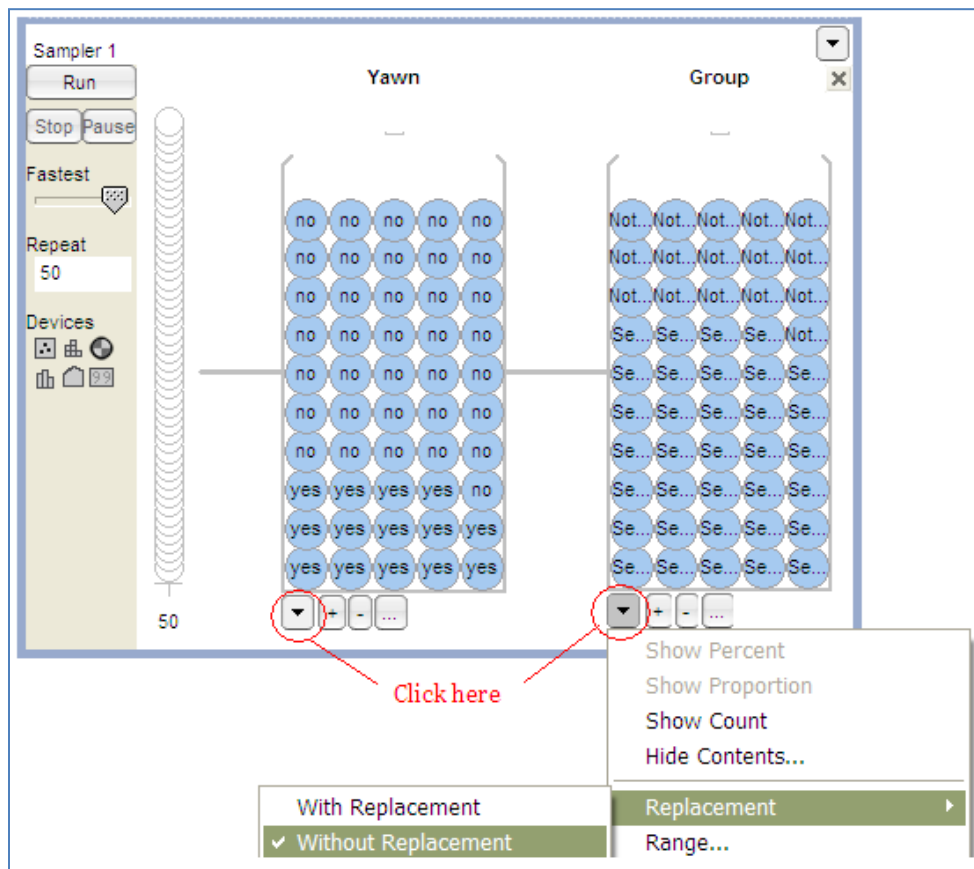
Click Run, and you will see the results of the first simulated trial:

Results of Yawning Experiment	Attributes ▼			
	Join	Yawn	Group	\<new\>
1	no,Seeded	no	Seeded	
2	yes,See...	yes	Seeded	
3	no,Not s...	no	Not seed...	
4	no,Seeded	no	Seeded	
5	yes,See...	yes	Seeded	
6	no,Seeded	no	Seeded	
7	no,Not s...	no	Not seed...	
8	no,Not s...	no	Not seed...	
9	yes,Not	yes	Not seed...	

To plot these results, drag a new **Plot** to the workspace. Drag the predictor variable (Group) to the x-axis and the response variable (Yawn) to the y-axis, as shown below.
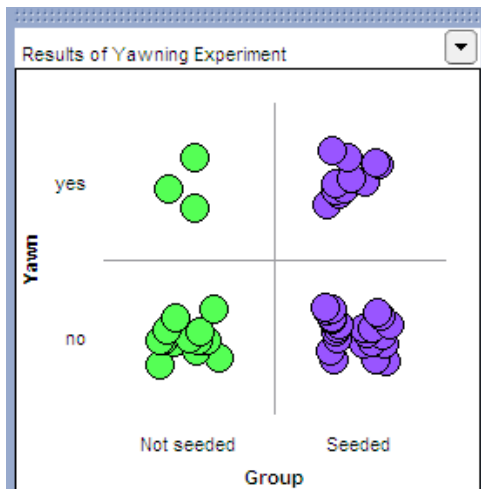
Click on **n** to count the number in each cell, and you can also vertically stack the dots using the **Stack** tool.

Note that we are interested in recording the number of yawners in the Seeded group. Place your cursor over this number (circled in red on the below graph), right-click, and choose **Collect Statistic.** Set the **Collect** value to 99 in order to obtain a total of 100 simulated results, and click **Collect**.
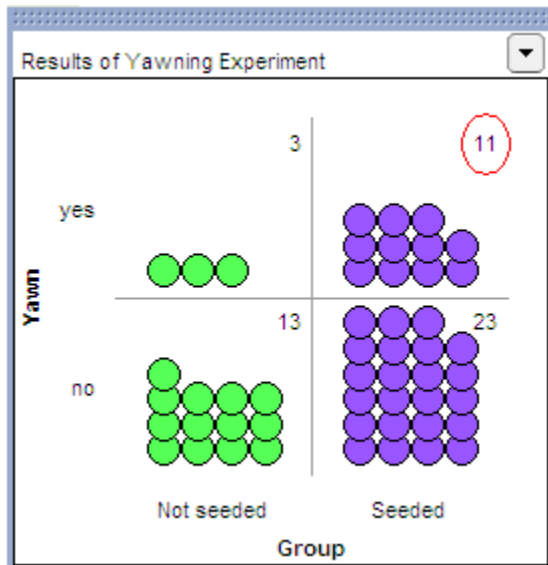
Finally, to graph these outcomes, highlight the "count_Yawn_yes_Group_Seeded" column and drag a new plot into the workspace. Drag a point all the way to the right to organize the dots, and use the **Stack** tool to vertically stack the dots. Finally, right-click on either endpoint on the x-axis and set "Bin Width" to 1, and use the count tool (**n**) to count the dots in each bin.
You should see something similar to the following:
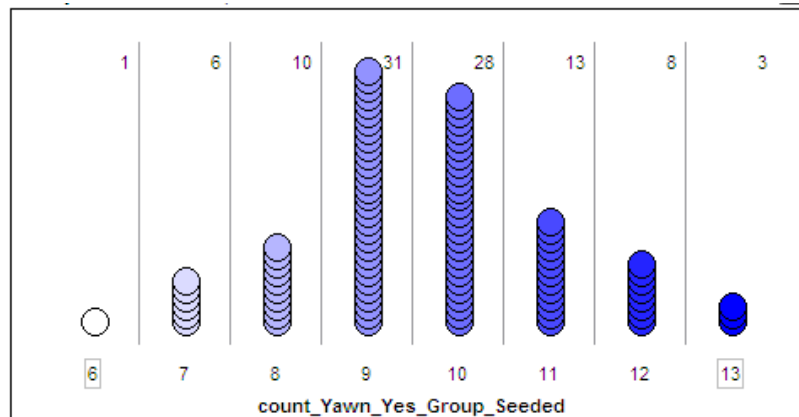
Questions:

1. What does each dot on the above plot represent?

2. What value(s) occurred most often by chance under the null model? Explain why this makes sense.

3. How often did we see results at least as extreme as the observed data (10 or more yawners in the seeded group) under the null model? Calculate the proportion of simulated results in which we observed 10 or more yawners in the seeded group. Note that this is an approximate p-value!

4. The MythBusters reported the following results: 25% yawned of those not given a yawn seed, and 29% yawned of those given a yawn seed.   Then, they cited the "large sample size" and the 4% difference in the proportion that yawned between the seeded and non-seeded group to confidently conclude that yawn seed had a significant effect on the subjects. Therefore, they concluded that the yawn is decisively contagious. Do you agree or disagree with their answer? Justify your reasoning.

Statistical Inference: Fisher's Exact Test to Obtain Exact p-values

We just used a simulation study to investigate the Mythbusters' research hypothesis. This is a valid analysis; however, it provides only an approximate p-value. We can obtain the exact p-value using probability theory and a distribution known as the hypergeometric distribution. Details about the hypergeometric distribution are beyond the scope of this course; however, we will discuss **Fisher's Exact Test** which uses the hypergeometric distribution to calculate p-values. This is easily carried out in JMP.

Note the output obtained when you used the **Analyze > Fit Y by X** option:

Tests

N	DF	-LogLike	RSquare (U)
50	1	0.05318858	0.0018

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.106	0.7443
Pearson	0.105	0.7459

Fisher's

Exact Test	Prob	Alternative Hypothesis
Left	0.7417	Prob(Yawned?=Yes) is greater for Group=Not seeded than Seeded
Right	0.5128	Prob(Yawned?=Yes) is greater for Group=Seeded than Not seeded
2-Tail	1.0000	Prob(Yawned?=Yes) is different across Group

JMP has already used the hypergeometric probability distribution to calculate the probability of observing results at least as extreme as those observed in the MythBusters actual experiment. Now, you can use this output to carry out the test.

Carrying out Fisher's Exact Test

This test is based on the probability of observing data at least as extreme as the actual observed data. The procedure is carried out as follows.

Step 1:	Convert the research hypothesis into H_o and H_a. H_0: Yawn seeding has no effect on whether or not a person yawns. H_a: Those seeded with a yawn are more likely to yawn. We can also state the hypotheses in terms of the following parameters: $\pi_{\text{yawn | seeded}}$ = the proportion that yawn (i.e., the probability of yawning) when Seeded $\pi_{\text{yawn | seeded}}$ = the proportion that yawn (i.e., the probability of yawning) when Not Seeded H_0: H_a:
Step 2:	Determine the p-value. p-value:
Step 3:	Write a conclusion addressing the original research hypothesis.

Example 3.2: Vested Interest and Task Performance

This example is from Investigating Statistical Concepts, Applications, and Methods by Beth Chance and Allan Rossman. 2006. Thomson-Brooks/Cole.

"A study published in the *Journal of Personality and Social Psychology* (Butler and Baumeister, 1998) investigated a conjecture that having an observer with a vested interest would decrease subjects' performance on a skill-based task. Subjects were given time to practice playing a video game that required them to navigate an obstacle course as quickly as possible. They were then told to play the game one final time with an observer present. Subjects were randomly assigned to one of two groups. One group (A) was told that the participant and observer would each win $3 if the participant beat a certain threshold time, and the other group (B) was told only that the participant would win the prize if the threshold were beaten. The threshold was chosen to be a time that they beat in 30% of their practice turns. The following results are very similar to those found in the experiment: 3 of the 12 subjects in group A beat the threshold, and 8 of 12 subjects in group B achieved success."

	A: Vested Interest	B: No Vested Interest	Total
Achieved success	3	8	11
Did not achieve success	9	4	13
Total	12	12	24

Research Hypothesis: **Having an observer with a vested interest decreases performance on a skill-based task.**
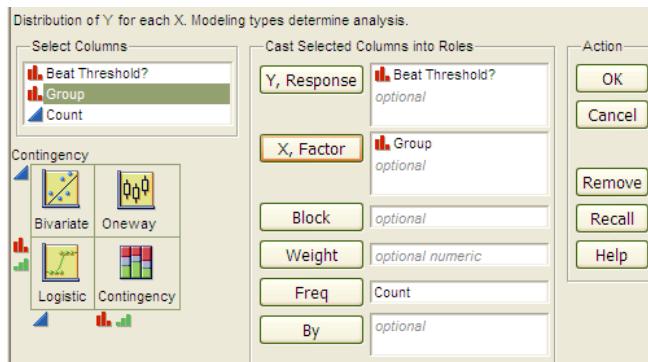
Questions:

1. What variables must be measured in order to address the research question? Are they categorical or numerical?

2. Which is the response variable? Which is the explanatory variable?
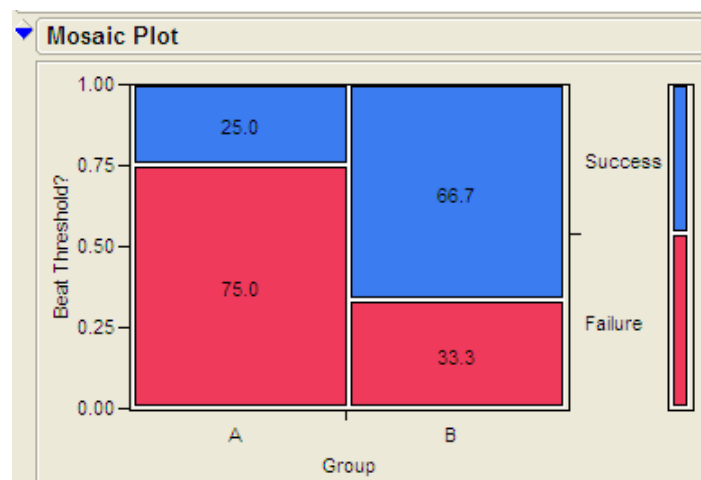
The data can be entered into JMP as follows:

To create the contingency table and the mosaic plot in JMP, select **Analyze > Fit Y by X**. As always, place the response variable in the **Y, Response** box and the explanatory variable in the **X, Factor** box.

Click OK, and JMP returns the following:

Questions:

3. What is the proportion of successes (beating the threshold) for each group?

4. What is the difference in the proportion of successes between these two groups? Do these proportions differ in the direction conjectured by the researchers?

5. Even if the observer's interest had absolutely no effect on subjects' performance, is it *possible* to have obtained a difference such as this by random chance alone? Explain.

6. Next, let's think about some outcomes that might occur by chance even if having an observer with a vested interest has no real effect on the success rate. We will assume that 11 of the 24 subjects are going to succeed, no matter what group they're in. Then, we'll randomly assign them to the two groups of size 12 and see how many of the successes end up in each group purely by chance. Note that this implies that the row and column totals are fixed. What would the counts for the "most extreme" table look like in this case? That is, what possible outcome would give us the most convincing evidence in support of the research hypothesis?

	Beat Threshold?		
Count	Failure	Success	
A			12
B			12
	13	11	24

(Group labels shown on vertical axis)

7. Give a few more examples of tables that are *more extreme* than the observed data but not *as extreme* as the "most extreme" table shown above.
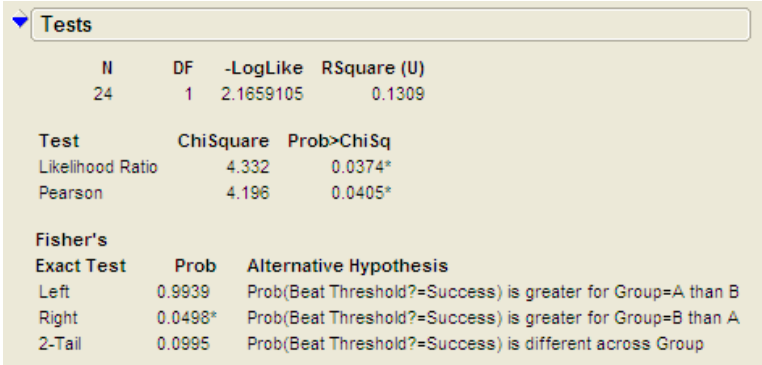
	Beat Threshold?		
Count	Failure	Success	
A			12
B			12
	13	11	24

	Beat Threshold?		
Count	Failure	Success	
A			12
B			12
	13	11	24

	Beat Threshold?		
Count	Failure	Success	
A			12
B			12
	13	11	24

Keep in mind that a p-value is always the probability of seeing results at least as extreme as the observed data. So, JMP uses what is known as the hypergeometric distribution to find the probability of seeing the observed data table and EACH table that is more extreme, assuming that there is no difference in the two groups being compared. The sum of these probabilities is the p-value from Fisher's exact test.

Fisher's Exact Test to Obtain Exact p-values

Step 1:	Convert the research hypothesis into H_0 and H_a. H_0: Whether the observer has a vested interest or not has no impact on performance on a skill-based task. H_a: Having an observer with a vested interest tends to decrease a subjects' performance on a skill-based task. State the hypotheses in terms of the following parameters: $\pi_{\text{succeed | vested}}$ = the proportion that succeed (i.e., the probability of succeeding) when Vested $\pi_{\text{succeed | not vested}}$ = the proportion that succeed (i.e., the probability of succeeding) when Not Vested H_0: H_a:
Step 2:	Determine the p-value. **Tests** 		N	DF	-LogLike	RSquare (U)
	24	1	2.1659105	0.1309 | | Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.332	0.0374*
Pearson	4.196	0.0405*	 **Fisher's** 	Exact Test	Prob	Alternative Hypothesis
Left	0.9939	Prob(Beat Threshold?=Success) is greater for Group=A than B
Right	0.0498*	Prob(Beat Threshold?=Success) is greater for Group=B than A
2-Tail	0.0995	Prob(Beat Threshold?=Success) is different across Group	 p-value:
Step 3:	Write a conclusion addressing the original research hypothesis.

Example 3.3: Claritin and Nervousness

An advertisement by the Schering Corporation in 1999 for the allergy drug Claritin mentioned that in a pediatric randomized clinical trial, symptoms of nervousness were shown by 4 of 188 patients on Claritin and 2 of 262 patients taking a placebo.

Research Hypothesis: The proportion who experience nervousness is <u>greater</u> for those who take Claritin than for those who take the placebo.

The data can be found in the file **Claritin.JMP**.

	Drug	Nervousness?	Count
1	Claritin	Yes	4
2	Claritin	No	184
3	Placebo	Yes	2
4	Placebo	No	260

Questions:

1. What variables must be measured in order to address the research hypothesis?

2. Which is the response variable? Which is the explanatory variable?

3. Fill in the following contingency table based on the data table from JMP.

	Nervousness?		
Drug	Yes	No	Row Totals (Fixed)
Claritin			
Placebo			
Total			

4. Find the proportion that experience nervousness in each group.

Next, use JMP to carry out Fisher's Exact Test for these data.

<u>Step 1</u>:	Convert the research hypothesis into H_0 and H_a. H_0: H_a:
<u>Step 2</u>:	Determine the <u>p-value</u>. p-value:
<u>Step 3</u>:	Write a conclusion addressing the original research hypothesis.

Observational Studies vs. Designed Experiments

Reconsider the "Vested Interest and Task Performance" example. Fisher's exact test provided evidence that the proportion of successes was in fact smaller for the vested interest group (p-value = .0498). Now, the question is this: can we conclude that having a vested interest really is the *cause* of the decreased performance?

The answer to this question lies in whether the experiment itself was a designed experiment or an observational study.

An **observational study** involves collecting and analyzing data *without randomly* assigning treatments to experimental units.

On the other hand, in a **designed experiment**, a treatment is *randomly* imposed on individual subjects in order to observe whether the treatment causes a change in the response.

Key statistical idea:

The random assignment of treatments used by researchers in a designed experiment should balance out between the treatment groups any other factors that might be related to the response variable. Therefore, designed experiments *can* be used to establish a cause-and-effect relationship (as long as the p-value is small).

On the other hand, observational studies establish only that an association exists between the predictor and response variable. With observational studies, it is always possible that there are other lurking variables not controlled for in the study that may be impacting the response. Since we can't be certain these other factors are balanced out between treatment groups, it is possible that these other factors could explain the difference between treatment groups.

Note that the "Vested Interest and Task Performance" study is an example of a designed experiment since participants were randomly assigned to the two groups. We were trying to show that having a vested interest *caused* a decreased task performance. The small p-value rules out observing the decreased performance in the vested interest group simply by chance, and the randomization of subjects to treatment groups should have balanced out any other factors that might explain the difference. So, the most likely explanation left is that having a vested interest really is the cause of the decreased task performance.

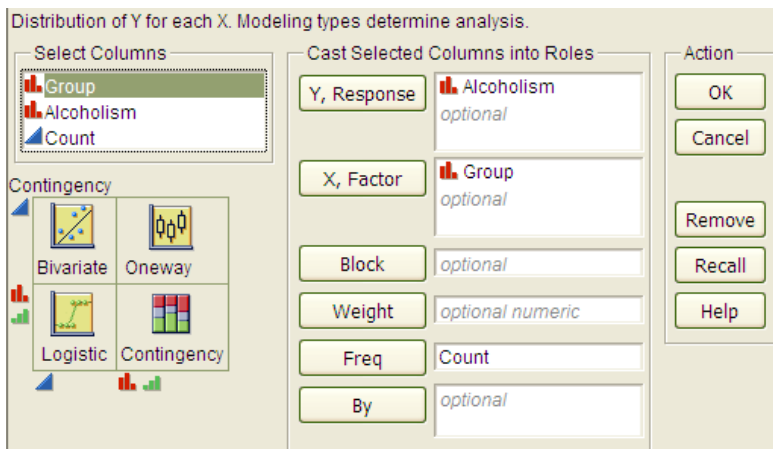### Example 3.4: Alcoholism and Depression

Past research has suggested a high rate of alcoholism in families of patients with primary unipolar depression. A study of 210 families of females with primary unipolar depression found that 89 had alcoholism present. A set of 299 control families found 94 present.

Research Hypothesis: **The proportion of families afflicted by alcoholism <u>differs</u> depending on whether or not the female in the family had primary unipolar depression.**
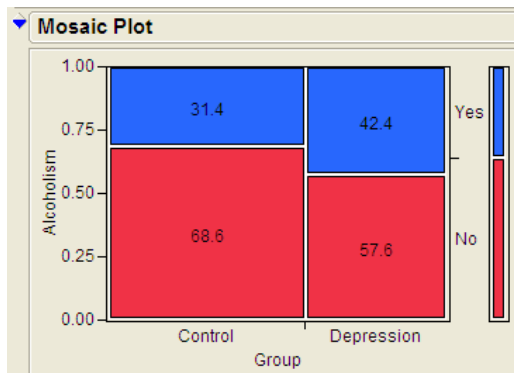
The data can be found in the file **Depression.JMP**:

	Group	Alcoholism	Count
1	Control	Yes	94
2	Depression	Yes	89
3	Control	No	205
4	Depression	No	121

To analyze these data, choose **Analyze > Fit Y by X** and enter the following.

The JMP output is shown below:

Questions:

1. Identify the proportion of families with alcoholism in both groups.

2. Is there evidence that the proportion of families afflicted by alcoholism <u>differs</u> depending on whether or not the female in the family had primary unipolar depression? Use the JMP output to answer this question.

3. Can we say that a female having unipolar depression *causes* alcoholism in her family? Explain your reasoning.

CHI-SQUARE TEST (ALTERNATIVE TO FISHER'S EXACT TEST)

Fisher's exact test from the previous section can be used for either upper-tailed, lower-tailed, or two-tailed hypothesis tests for differences in two proportions. The Chi-square test is another procedure we can use to test for differences between two proportions. However, this procedure can be used for **ONLY TWO-SIDED TESTS**! Consider the data from Example 4.4.

H_0: The proportion of families afflicted by alcoholism is the same regardless of whether or not the female in the family suffers from primary unipolar depression.

H_a: The proportion of families afflicted by alcoholism differs depending on whether or not the female in the family suffers from primary unipolar depression.

With the chi-square test, the evidence from our sample will consist of a test statistic, which for this test is given as follows:

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

Question: Where do we get the expected counts?

The Observed Counts	The Expected Counts

Contingency Table

Alcoholism

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

Group

Contingency Table

Alcoholism

Count	No	Yes	
Control			299
Depression			210
	326	183	509

Group

We can also use JMP to calculate the expected counts, deviances, and cell contributions to find the test statistic:

Contingency Table

Count Expected Deviation Cell Chi^2	No	Yes	
Control	205 191.501 13.499 0.9516	94 107.499 -13.499 1.6951	299
Depression	121 134.499 -13.499 1.3548	89 75.501 13.499 2.4135	210
	326	183	509

(Row label: Group; Column label: Alcoholism)

Test Statistic =

$$\sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(205 - 191.5)^2}{191.5} + \frac{(94 - 107.5)^2}{107.5} + \frac{(121 - 134.5)^2}{134.5} + \frac{(89 - 75.5)^2}{75.5}$$

$$= \frac{(13.5)^2}{191.5} + \frac{(-13.5)^2}{107.5} + \frac{(-13.5)^2}{134.5} + \frac{(13.5)^2}{75.5}$$

$$= 0.9516 + 1.6951 + 1.3548 + 2.4135$$

$$= 6.415$$

Questions:

1. What does it mean when the test statistic is "large"?

2. At what point does the test statistic provide evidence to support our research question?

Find the p-value:

Recall that the Chi-Square test uses a distribution known as the chi-square (χ^2) distribution. The chi-square distribution takes on only positive numbers and is a continuous distribution. In addition, this distribution is indexed by its degrees of freedom (or df). For this test, this is given by df = (r - 1)(c - 1).

When the null hypothesis is true, the test-statistic follows the chi-square distribution with df = (r - 1)(c - 1). .

The following graph shows the chi-square distribution with df = 1. 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 the p-value is very small for this study.

Observed Chi-square
Test Statistic = 6.4
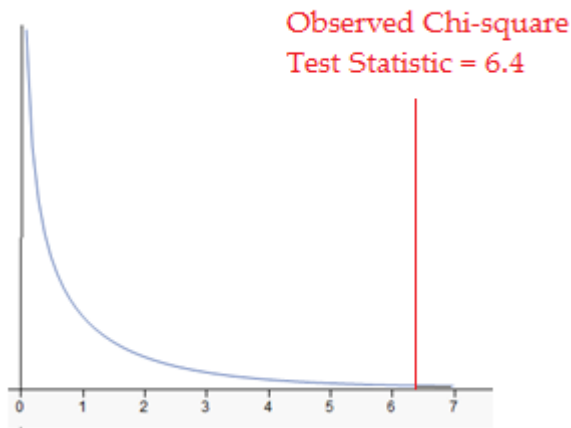
To find this p-value, we will use JMP. Note that the output for the chi-square test automatically appears when you select **Analyze > Fit Y by X.** The p-value for this study is 0.0113.

Tests

Source	DF	-LogLike	RSquare (U)
Model	1	3.19237	0.0096
Error	507	329.25916	
C. Total	508	332.45153	
N	509		

Chi-square test statistic

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.385	0.0115*
Pearson	6.415	0.0113*

p-value associated with the Chi-square test statistic

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9956	Prob(Alcoholism=Yes) is greater for Depression=No than Yes
Right	0.0075*	Prob(Alcoholism=Yes) is greater for Depression=Yes than No
2-Tail	0.0146*	Prob(Alcoholism=Yes) is different across Depression

Assumptions behind the Chi-Square Test:

The chi-square test for 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. JMP and most other statistical software packages will warn you when the results of the chi-square test are suspect.

Also, each observation in the study can be classified into only one cell of the contingency table, and the observations must be independent.

Example 3.5: Mice and Tobacco Smoke

Ten mice (6–8 weeks old) were randomly assigned to one of two groups; five were exposed to simulated environmental tobacco smoke for 6 h/day, 5 days/week for 5 months. The other 5 mice were kept in clean air during this time period. Then, all of the mice were allowed to recover for a further 4 months in filtered air before being killed for analysis of lung tumor incidence. The results are shown below.

	Tumor	No tumor
Treated	5	0
Control	2	3

Research hypothesis: The proportion of mice that develop a lung tumor differs between those exposed to tobacco smoke and the control group.

Step 1: Convert the Research Hypothesis into H_0 and H_a.

H_0: The proportion of mice that develop a lung tumor when exposed to tobacco does not differ from the proportion that develop a lung tumor when not exposed to tobacco.

H_a: The proportion of mice that develop a lung tumor differs between those exposed to tobacco smoke and those not exposed.

Equivalently, we can state the hypotheses as follows:

H_0:

H_a:

Step 2: Find the p-value for the data.

⊿ **Tests**

N	DF	-LogLike	RSquare (U)
10	1	2.7435847	0.4491

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.487	0.0192*
Pearson	4.286	0.0384*

Warning: Average cell count less than 5, LR ChiSquare suspect.

Fisher's

Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Tumor=yes) is greater for Group=Control than Treated
Right	0.0833	Prob(Tumor=yes) is greater for Group=Treated than Control
2-Tail	0.1667	Prob(Tumor=yes) is different across Group

p-value from Fisher's exact test: _____

p-value from Chi-Square test: _____

> Keep the assumptions behind the chi-square test in mind:
>
> The chi-square test for 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, each observation in the study can be classified into only one cell of the contingency table, and the observations must be independent.

Questions:

1. Do the assumptions behind the chi-square test appear to be met? Explain.

2. Which test should we use – Fisher's exact or Chi-square? Explain.

Step 3: Write a conclusion addressing the original research hypothesis.