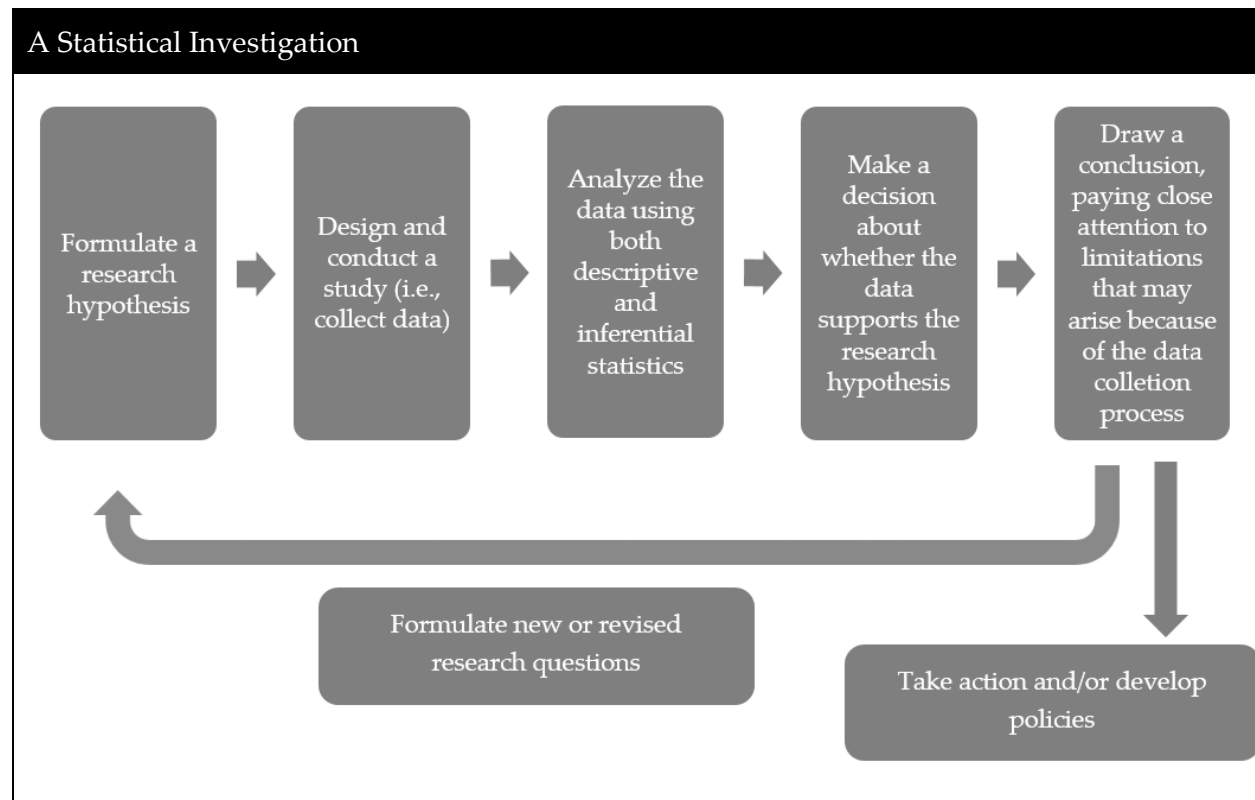


The general approach used in a statistical investigation is shown in the following diagram. Note that every investigation starts with a question that gets converted to a research hypothesis. Though we will rarely design our own studies in this course, we will discuss good strategies and best practices for collecting data throughout the semester. Most of our attention will be devoted to descriptive and inferential methods for analyzing the data, how we decide whether the data supports the research question, and how to write conclusions that appropriately summarize the results of the study.



In Chapter 2, we will discuss methods for making decisions concerning research hypotheses involving only a single categorical variable. Even though we'll be adding lots of terminology and formalities along the way, we will use the same logical approach to investigating research questions that was introduced in Chapter 1.

### DESCRIPTIVE VS. INFERENCE STATISTICS

**Definitions**

- **Descriptive statistics** are methods that describe, show, or summarize the data from our sample in a meaningful way.
- **Inferential statistics** are methods that allow us to draw conclusions about the larger population that the sample represents.

For example, recall the Helper vs. Hinder study from Example 1.2. In the sample of 16 infants studied, 14 of the 16 picked the helper. What descriptive methods might we use to *describe* the results of this study?

**FORMAL HYPOTHESIS TESTING**

You are most likely already familiar with several methods for descriptive statistics (e.g., calculating percentages, constructing bar charts or pie charts, etc.). Inferential methods, however, may be new to you. Again, inferential statistics involves drawing conclusions about the larger population that our sample was supposed to represent. In several examples introduced in Chapter 1, we used a logical process to make statistical decisions concerning a population of interest – we were actually using inferential statistics already!

Now, we will add more structure to these statistical investigations by introducing an inferential statistics procedure known as *hypothesis testing*. Note that in this chapter, we look at hypothesis testing for only research questions involving a single categorical variable. Before we discuss this procedure, however, we will discuss a few more definitions.

**Population Parameters vs. Sample Statistics**

In each of the previous examples, we tested a claim about a population **parameter** of interest.

Definitions	
<ul style="list-style-type: none"> <li>▪ A <b>parameter</b> is a numerical descriptive measure of a population. This value is almost always unknown, and our goal in a statistical investigation is typically to either <u>estimate</u> this parameter or <u>test a claim</u> regarding it.</li> <li>▪ A <b>statistic</b> is a numerical descriptive measure of a sample. This value is calculated from the observed data.</li> </ul>	

Example	Statistic	Parameter
Example 1.2: Helper vs. Hinderer		

Example	Statistic	Parameter
Example 1.3: Are women passed over for managerial training?		
Example 1.4: Font Preference		

*Hypothesis testing* is a procedure, based on sample evidence and probability, used to test a claim regarding a population parameter. The test will measure how well our observed sample statistic agrees with some assumption about this population parameter.

Before you begin a hypothesis test, you should clearly state your research hypothesis. For instance, let's reconsider the research hypotheses from three of our previous examples.

Example	Research Hypothesis
Example 1.2: Helper vs. Hinderer	Ten-month-old infants prefer the helper toy over the hinderer toy.
Example 1.3: Are women passed over for managerial training?	This particular company is discriminating against females in the management selection process.
Example 1.4: Font Preference	Consumers prefer one font over the other.

Once the research hypothesis has been developed, we typically formulate what are known as the null and alternative hypotheses. The null and alternative hypotheses are both statements about the parameter of interest in the study.

### **Setting Up the Null and Alternative Hypothesis**

- The null hypothesis,  $H_0$ , is what we will assume to be true (i.e., we will assume for the time being that whatever effect we want to detect doesn't exist in reality). We will then evaluate the observed outcome from our study against what outcomes we expected to see under the null hypothesis. This will always contain a statement saying that the population parameter is **equal** to some value.

- The alternative hypothesis,  $H_a$ , is what we are trying to show. Therefore, the research hypothesis is simply restated here as if it were true in the alternative hypothesis. This will always contain statements of inequality, saying that the population parameter is **less than, greater than, or different from** the value in the null hypothesis.

For our three examples, the null and alternative hypotheses are shown below.

Research Hypothesis	Null and Alternative Hypotheses
Ten-month-old infants tend to prefer the helper toy over the hinderer toy.	<p><math>H_0</math>: The proportion of all 10-month-old infants that prefer the helper toy is 50%.</p> <p><math>H_a</math>: The proportion of all 10-month-old infants that prefer the helper toy is greater than 50%.</p>
This particular company is discriminating against females in the management selection process.	<p><math>H_0</math>: The probability of a woman being selected for management is 60%.</p> <p><math>H_a</math>: The probability of a woman being selected for management is less than 60%.</p>
Consumers prefer one font over the other.	<p><math>H_0</math>: The proportion of all consumers that pick the Signet font is equal to 50%.</p> <p><math>H_a</math>: The proportion of all consumers that pick the Salem font is different from 50%.</p>

Note that we can also state these hypotheses in terms of the population parameter of interest using formal notation:

Research Hypothesis	Null and Alternative Hypotheses
Ten-month-old infants tend to prefer the helper toy over the hinderer toy.	<p><math>H_0</math>:</p> <p><math>H_a</math>:</p>
This particular company is discriminating against females in the management selection process.	<p><math>H_0</math>:</p> <p><math>H_a</math>:</p>
Consumers prefer one font over the other.	<p><math>H_0</math>:</p> <p><math>H_a</math>:</p>

**Evaluating Evidence Using P-Values**

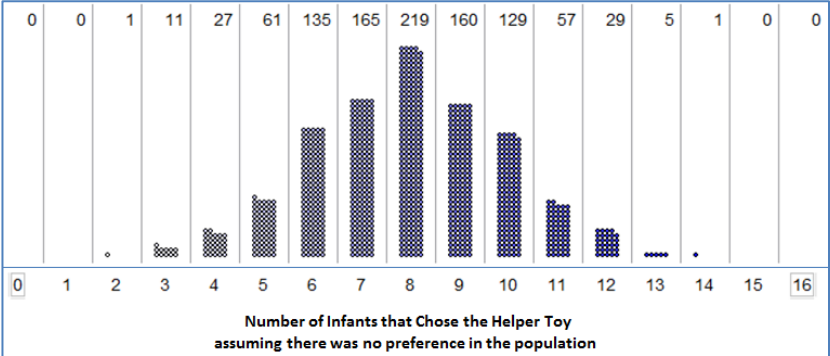
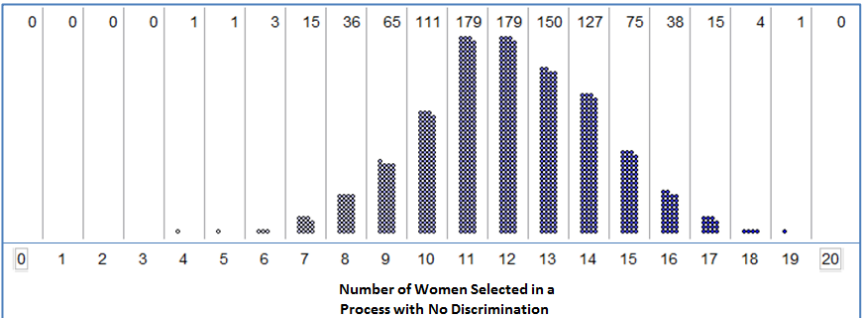
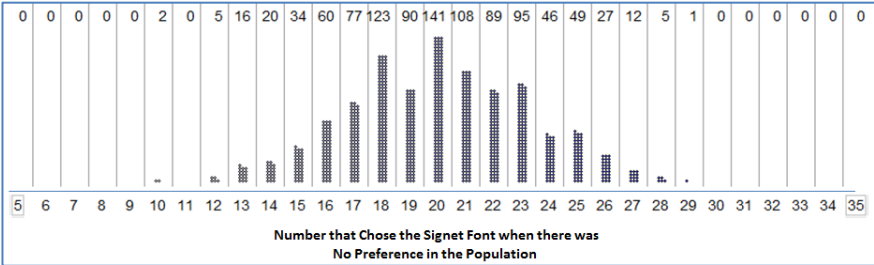
In each of our three examples, we essentially assumed the null hypothesis was true when setting up our spinner for the Tinkerplots investigation. Then, we used the results simulated under this scenario to help us decide whether observing results such as our sample data would be an unusual event if the null hypothesis were true.

Up to this point, whether an observed study result was considered unusual (or extreme) has been a rather subjective decision. Now, we will discuss the guidelines used by statisticians to determine whether an observed study result is extreme enough under the null hypothesis for us to conclude that the evidence supports the research hypothesis.

First, note that in our three examples, we examined different parts of the distribution of simulated outcomes when deciding whether the observed study data was extreme. Each of these cases is an example of a specific type of hypothesis test.

Research Hypothesis	Hypotheses	Type of Test
Ten-month-old infants tend to prefer the helper toy over the hinderer toy.	$H_0: \pi = .50$ $H_a: \pi > .50$  where $\pi$ = the true proportion of all 10-month-olds that choose the helper	Upper-tailed Test
This particular company is discriminating against females in the management selection process.	$H_0: \pi = .60$ $H_a: \pi < .60$  where $\pi$ = the true proportion of those selected for management who are female	Lower-tailed Test
Consumers prefer one font over the other.	$H_0: \pi = .50$ $H_a: \pi \neq .50$  where $\pi$ = the true proportion of all consumers that choose Signet font	Two-tailed Test

Statisticians typically use the following guidelines to determine whether the observed data supports the research question:

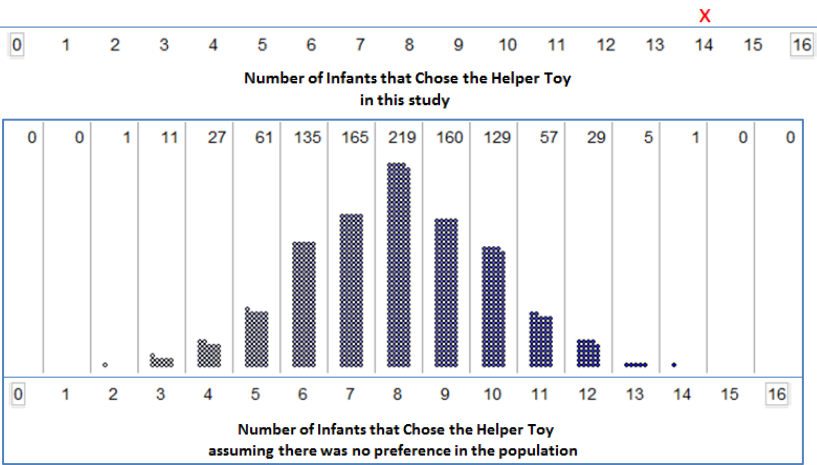
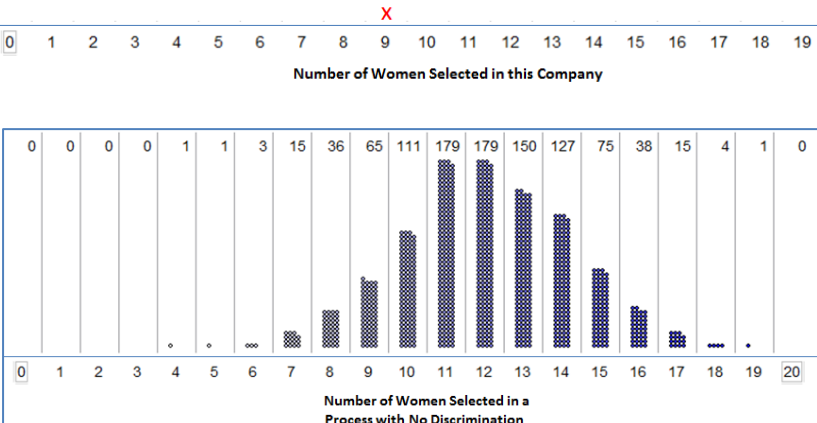
Research Hypothesis	Statistician’s Guideline for When Observed Outcome Supports the Research Question
<p>Ten-month-old infants tend to prefer the helper toy over the hinderer toy.</p>	 <p><b>Upper-tailed test:</b> The observed outcome must fall in the <u>upper</u> 5% of the distribution obtained under the null hypothesis.</p>
<p>This particular company is discriminating against females in the management selection process.</p>	 <p><b>Lower-tailed test:</b> The observed outcome must fall in the <u>lower</u> 5% of the distribution obtained under the null hypothesis.</p>
<p>Consumers prefer one font over the other.</p>	 <p><b>Two-tailed test:</b> The observed outcome must fall in either the upper 2.5% or the lower 2.5% of the distribution obtained under the null hypothesis.</p>

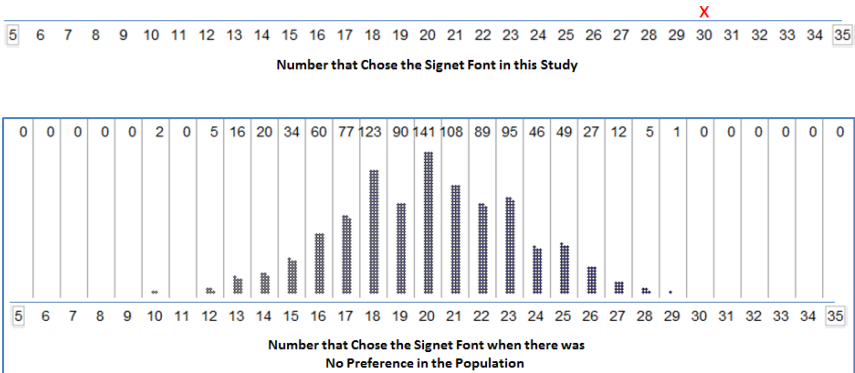
Statisticians use what is called a **p-value** to quantify the amount of evidence that an observed outcome from a set of data provides for a research question.

**Definition**

**p-value:** The probability of observing an outcome as extreme (or even more extreme in favor of the research hypothesis) than the observed study result, assuming the null hypothesis is true.

Note that in each of the above examples, we obtained the simulation results assuming the null hypothesis was true. Therefore, to *estimate* the p-value, we simply determine how often outcomes as extreme (or even more extreme) than the observed study results appeared in our simulation study.

Example	Estimate of p-value																					
<p><b>Helper vs. Hinderer?</b></p>  <p>Number of Infants that Chose the Helper Toy in this study</p> <table border="1"> <tr> <td>0</td><td>0</td><td>1</td><td>11</td><td>27</td><td>61</td><td>135</td><td>165</td><td>219</td><td>160</td><td>129</td><td>57</td><td>29</td><td>5</td><td>1</td><td>0</td><td>0</td> </tr> </table> <p>Number of Infants that Chose the Helper Toy assuming there was no preference in the population</p>	0	0	1	11	27	61	135	165	219	160	129	57	29	5	1	0	0					
0	0	1	11	27	61	135	165	219	160	129	57	29	5	1	0	0						
<p><b>Are Women Passed Over for Managerial Training?</b></p>  <p>Number of Women Selected in this Company</p> <table border="1"> <tr> <td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>3</td><td>15</td><td>36</td><td>65</td><td>111</td><td>179</td><td>179</td><td>150</td><td>127</td><td>75</td><td>38</td><td>15</td><td>4</td><td>1</td><td>0</td> </tr> </table> <p>Number of Women Selected in a Process with No Discrimination</p>	0	0	0	0	1	1	3	15	36	65	111	179	179	150	127	75	38	15	4	1	0	
0	0	0	0	1	1	3	15	36	65	111	179	179	150	127	75	38	15	4	1	0		

Example	Estimate of p-value
<p>Font Preference?</p> 	

### Making a Decision with p-values

- If the p-value is less than .05 (5%), then the data provide enough statistical evidence to support the research question.
- If the p-value is not less than .05 (5%), then the data do not provide enough statistical evidence to support the research question.

Why does this decision rule work? Consider the “Helper vs. Hinderer” example. Because the p-value falls below 5%, the observed result *must* have been in the upper 5% of the reference distribution. As stated earlier, this implies that the observed study result is very unlikely to happen by chance under the null hypothesis, which supports the research question.

On the other hand, consider the “Are Women Passed Over for Managerial Training” example. Because the p-value was larger than 5%, the observed result *can’t* have been in the lower 5% of the reference distribution. This implies that the observed study result is not all that unusual and could have easily happened by chance under the null hypothesis. Therefore, the null hypothesis *could* be true, and we have no evidence to support the research question.

This decision rule is widely accepted for determining whether study results are statistically significant; however, some researchers do advocate using a more flexible rule similar to the following:

### Making a Decision with p-values, Revised

- If the p-value falls below .05, we have strong statistical evidence to support the alternative hypothesis (i.e., the research question).
- If the p-value falls below .10 but above .05, we have “marginal” statistical evidence to support the alternative hypothesis (i.e., the research question).
- If the p-value is above .10, we have no evidence to support the research question.



Next, we will review the steps involved in a formal hypothesis test for each of our three examples. Note that our conclusions are written in the context of the problem. Moreover, even a person with no statistical background should be able to understand these conclusions (i.e., a conclusion should NOT say something like “We reject the null hypothesis.”)

<b>Helper vs. Hinderer</b>	
Research Hypothesis	Ten-month-old infants tend to prefer the helper toy over the hinderer toy.
Null and Alternative Hypotheses	$H_0$ : The proportion of all 10-month-olds that select the helper toy is 50%. $H_a$ : The proportion of all 10-month-olds that select the helper toy is greater than 50%.
p-value estimated from simulation	
Conclusion	

<b>Are Women Passed Over for Managerial Training?</b>	
Research Hypothesis	This particular company is discriminating against females in the management selection process.
Null and Alternative Hypotheses	$H_0$ : The probability a woman is selected is 60%. $H_a$ : The probability a woman is selected is less than 60%.
p-value estimated from simulation	
Conclusion	

Font Preference	
Research Hypothesis	Consumers prefer one font over the other.
Hypotheses	H <sub>0</sub> : The proportion of all consumers that select the Signet font is 50%. H <sub>a</sub> : The proportion of all consumers that select the Signet font is different from 50%.
p-value estimated from simulation	
Conclusion	

Next, let's carry out a formal hypothesis test for a few new examples.

**Example 2.1: Claims of Numbness after an Automobile Accident**

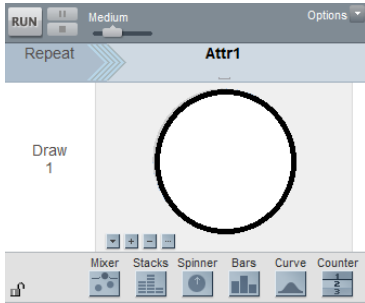
A 28-year-old white woman developed pain involving the spine and the left side of her body after an automobile collision. She was actively involved in a personal litigation against the company that owned the other vehicle, and she reported constant pain and numbness in the left arm. To test her claims, researchers touched her left arm with either 1 finger or 2 fingers simultaneously while her eyes were closed. The word “touch” was said simultaneously with the presentation of the tactile stimulus so that the subject knew when to respond. She then had to indicate whether she felt 1 single touch or 2 simultaneous touches (with the double-touch stimulus, the fingertips were always spaced 2 inches apart). The subject received 100 stimuli overall; she was correct on 30 of them. Is there statistical evidence that she is intentionally answering incorrectly?

Questions:

1. Identify both the population and sample of interest.
2. Identify the single categorical variable of interest.

3. Identify both the parameter and statistic of interest.

4. Carry out the formal hypothesis test to address the research question.

Claims of Numbness After Automobile Accident																																																																																		
Research Hypothesis	She is intentionally answering incorrectly.																																																																																	
Null and Alternative Hypotheses	<p><math>H_0</math>:</p> <p><math>H_a</math>:</p>																																																																																	
p-value estimated from simulation	<p>Carry out the simulation study to investigate this p-value. Sketch in the spinner that you used:</p>  <p>Sketch in the results of your simulation (keep track of the number of CORRECT responses on each trial):</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">30</td><td style="width: 20px; text-align: center;">31</td><td style="width: 20px; text-align: center;">32</td><td style="width: 20px; text-align: center;">33</td><td style="width: 20px; text-align: center;">34</td><td style="width: 20px; text-align: center;">35</td><td style="width: 20px; text-align: center;">36</td><td style="width: 20px; text-align: center;">37</td><td style="width: 20px; text-align: center;">38</td><td style="width: 20px; text-align: center;">39</td><td style="width: 20px; text-align: center;">40</td><td style="width: 20px; text-align: center;">41</td><td style="width: 20px; text-align: center;">42</td><td style="width: 20px; text-align: center;">43</td><td style="width: 20px; text-align: center;">44</td><td style="width: 20px; text-align: center;">45</td><td style="width: 20px; text-align: center;">46</td><td style="width: 20px; text-align: center;">47</td><td style="width: 20px; text-align: center;">48</td><td style="width: 20px; text-align: center;">49</td><td style="width: 20px; text-align: center;">50</td><td style="width: 20px; text-align: center;">51</td><td style="width: 20px; text-align: center;">52</td><td style="width: 20px; text-align: center;">53</td><td style="width: 20px; text-align: center;">54</td><td style="width: 20px; text-align: center;">55</td><td style="width: 20px; text-align: center;">56</td><td style="width: 20px; text-align: center;">57</td><td style="width: 20px; text-align: center;">58</td><td style="width: 20px; text-align: center;">59</td><td style="width: 20px; text-align: center;">60</td><td style="width: 20px; text-align: center;">61</td><td style="width: 20px; text-align: center;">62</td><td style="width: 20px; text-align: center;">63</td><td style="width: 20px; text-align: center;">64</td><td style="width: 20px; text-align: center;">65</td><td style="width: 20px; text-align: center;">66</td><td style="width: 20px; text-align: center;">67</td><td style="width: 20px; text-align: center;">68</td><td style="width: 20px; text-align: center;">69</td><td style="width: 20px; text-align: center;">70</td> </tr> <tr> <td colspan="40" style="text-align: center;"><small>count_Attr1_correct</small></td> </tr> </table> <p>Use the simulation results to estimate the p-value:</p>	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	<small>count_Attr1_correct</small>																																							
30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70																																										
<small>count_Attr1_correct</small>																																																																																		

<b>Claims of Numbness After Automobile Accident</b>	
Conclusion	

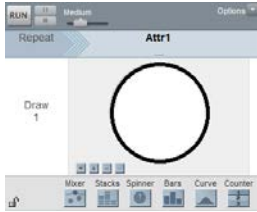
<b>A Statistical Investigation for Example 2.1 from Start to Finish</b>
<p><b><u>Research Question</u></b></p> <p>Is there statistical evidence that she can in reality feel the touch and is intentionally answering incorrectly?</p>
<p><b><u>Design Study/Collect Data</u></b></p> <p>The researchers presented the subject with 100 trials of the experiment. They kept track of the number of trials in which she gave a correct answer.</p>
<p><b><u>Descriptive Statistics - Explore/Summarize Data</u></b></p> <p>Of the 100 trials, she answered only 30 correct (30%).</p>
<p><b><u>Inferential Statistics - Draw Appropriate Inferences Beyond the Sample</u></b></p> <p>The subject's outcome (only 30 correct out of 100 trials) was much lower than what was expected if she were truly experiencing numbness, and it was shown to be very unlikely to occur by chance if she were truly numb. Even though these 100 trials provide just a sample of the subject's overall behavior, the statistical evidence is strong enough to indicate that the actual long-run probability of the subject answering correctly is much lower than we would expect if she were truly experiencing numbness (i.e., there is evidence she is faking her symptoms).</p>

**Example 2.2: Effectiveness of an Experimental Drug**

Suppose a commonly prescribed drug for relieving nervous tension is believed to be only 70% effective. Experimental results with a new drug administered to a random sample of 20 adults who were suffering from nervous tension show that 18 received relief. Is there statistical evidence that the new experimental drug is more than 70% effective?

Questions:

1. Identify both the population and sample of interest.
  
2. Identify the single categorical variable of interest.
  
3. Identify both the parameter and statistic of interest.
  
4. Carry out the formal hypothesis test to address the research question.

<b>Effectiveness of an Experimental Drug</b>	
Research Hypothesis	The new drug is more than 70% effective.
Null and Alternative Hypotheses	<p><math>H_0</math>:</p> <p><math>H_a</math>:</p>
p-value estimated from the simulation	<p>Carry out the simulation study to estimate this p-value. Sketch in the spinner that you used:</p>  <p>Sketch in the results of your simulation (keep track of the number that experience RELIEF on each trial):</p>

<b>Effectiveness of an Experimental Drug</b>	
	<p style="text-align: center;">Use the simulation results to estimate the p-value:</p>
Conclusion	

<b>A Statistical Investigation for Example 2.2 from Start to Finish</b>	
<b><u>Research Question</u></b>	
Is there statistical evidence that the new experimental drug is more than 70% effective?	
<b><u>Design Study/Collect Data</u></b>	
The new drug was administered to a random sample of 20 adults who were suffering from nervous tension. Researchers found that 18 received relief.	
<b><u>Descriptive Statistics - Explore/Summarize Data</u></b>	
Of the 20 adults in the study, 18 found relief (90%).	
<b><u>Inferential Statistics - Draw Appropriate Inferences Beyond the Sample</u></b>	
The study's outcome (90% of the adults in the sample found relief) was much higher than what we would have expected if the new drug were only as good as the old (with which 70% found relief). Furthermore, the study's outcome was shown to be very unlikely to have occurred by chance if in reality the new drug were only as good as the old. This provides strong statistical evidence that the new drug is more than 70% effective.	
<b><u>Big idea:</u></b> Even though the 20 subjects studied are just a sample of the population of all adults suffering from nervous tension, we can draw a conclusion about the effectiveness of this drug for the population of <i>all</i> adults suffering from nervous tension.	

**USING THE BINOMIAL DISTRIBUTION TO FIND EXACT P-VALUES**

There is one caveat regarding our current approach to obtaining a p-value. Certainly, different simulations will produce slightly different simulated distributions. The general pattern will be the same, but variations do exist. For example, consider the Helper vs. Hinderer study.

<b>Helper vs. Hinderer</b>	
Research Hypothesis	Ten-month-olds show a preference for the helper toy over the hinderer toy.
Null and Alternative Hypotheses	Let $\pi$ = the proportion of all ten-month-olds that choose the helper toy; equivalently, $\pi$ = the probability a randomly selected 10-month-old selects the helper toy,  $H_0: \pi = 50\%$ . $H_a: \pi > 50\%$ .

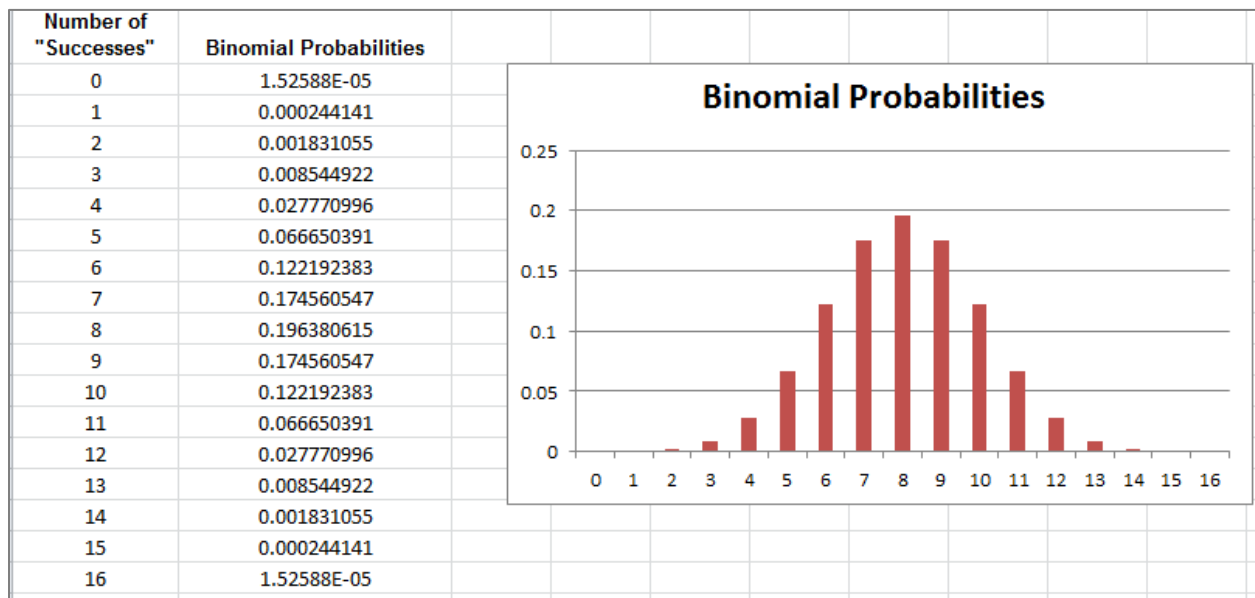
The study’s observed result was as follows: 14 out of 16 infants chose the helper toy. What if two different researchers each carried out their own simulation study to estimate the p-value?

p-value	<p>Simulation #1: p-value: _____</p> <p>Simulation #2: p-value: _____</p>
Conclusion	We do have evidence that 10-month-old infants prefer the helper toy.

Fortunately, regardless of which simulation study we use in the previous example, the final conclusion is the same, and the discrepancy between the two estimated p-values is minimal; still, it's not ideal that two different researchers get different results.

Note that as the number of trials in our simulation study increases, we expect less discrepancy between these two estimates of the p-value. So, instead of using a simulation study with only 1,000 trials to *estimate* the p-value, we would ideally like to simulate this experiment over and over again, say an infinite number of times. This would provide us with the theoretical probabilities of interest so that can get exact p-values instead of an estimate of the p-value.

The following graphic shows what the distribution would look like if we kept repeating the simulation study over and over again, each time counting and plotting the number of infants that chose the helper toy (assuming there was no real preference in the population of all infants). This theoretical probability distribution is known as the **binomial distribution**.



We can calculate these probabilities using the Excel file **BinomialProbabilities.xls**, which can be found on the course website.

Questions:

1. Does the general pattern in the above graph agree with the simulated distributions obtained from the simulation study in Tinkerplots?
2. The binomial probability value for 14 is 0.0018, or .18%. What does this value mean? Explain.



3. When we estimated the p-value using the results of the simulation study, we calculated the proportion of dots at 14 or above. How would we obtain the p-value using binomial probabilities? Explain.
  
4. What is the p-value using the binomial probabilities?

Statisticians often use the binomial distribution to calculate p-values when testing claims about a population proportion. However, before using this distribution, we should check to make sure the following conditions are met (note that these same conditions have to be met in order to estimate the p-value via a simulation study, as well).

**The Binomial Distribution - When can we use it?**

This distribution can be used whenever the following assumptions are met:

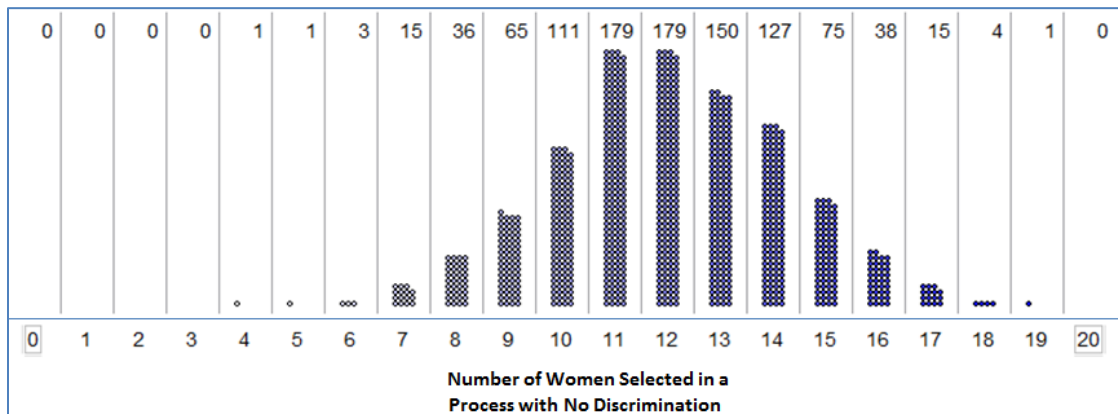
- The study involves a fixed number of trials,  $n$ .
- There are only two possible outcomes on each trial (we call these a “success” or “failure”).
- The probability of “success” ( $\pi$ ) remains constant from trial to trial.
- The  $n$  trials are independent.

Check whether these assumptions seem reasonable for the Helper vs. Hinderer study.

- There exist a fixed number of trials,  $n$ .
  
- There are only two possible outcomes on each trial (“success” or “failure”).
  
- The probability of success ( $\pi$ ) remains constant from trial to trial.
  
  
- The  $n$  trials are independent.

**Example: Gender Discrimination**

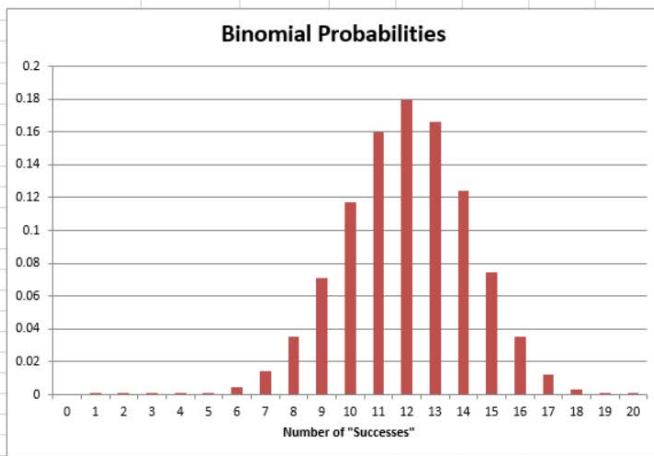
Recall that we already estimated the probability of observing 9 or fewer women selected out of 20 if there was no discrimination (i.e., we estimated the p-value with a simulation study).



Estimated p-value = \_\_\_\_\_

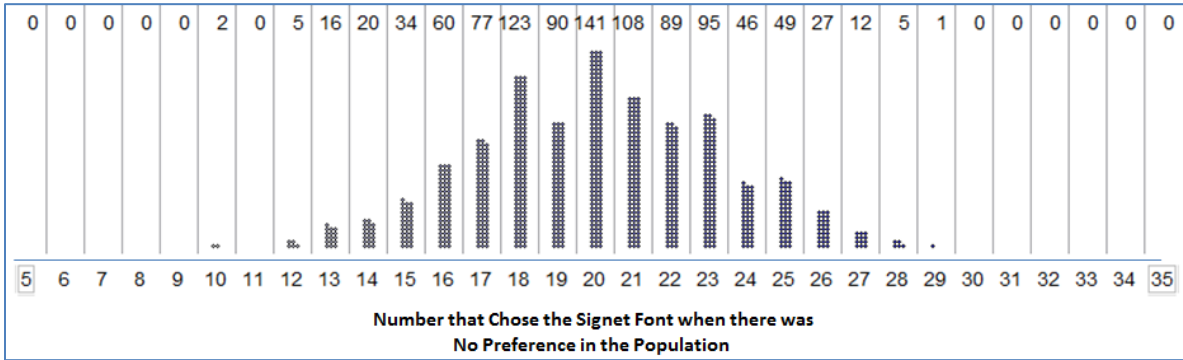
Next, we will use the binomial distribution to find the exact p-value for the Gender Discrimination Study. First, check whether the assumptions behind the binomial distribution seem reasonable in this case.

- There exist a fixed number of trials,  $n$ .
  
- There are only two possible outcomes on each trial (“success” or “failure”).
  
- The probability of success ( $\pi$ ) remains constant from trial to trial.
  
- The  $n$  trials are independent.

Are Women Passed Over for Managerial Training?																																													
Research Hypothesis	The company is discriminating against women when they are selecting employees for management training.																																												
Null and Alternative Hypotheses	<p>Let <math>\pi</math> = the probability the company selects a woman.</p> <p><math>H_0: \pi = 60\%</math>.</p> <p><math>H_a: \pi &lt; 60\%</math>.</p>																																												
p-value	<p>To find the exact p-value, we will use the binomial distribution with...</p> <p><math>n =</math> _____</p> <p><math>\pi =</math> _____</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;">Number of "Successes"</th> <th style="text-align: center;">Binomial Probabilities</th> </tr> </thead> <tbody> <tr><td>0</td><td>1.09951E-08</td></tr> <tr><td>1</td><td>3.29853E-07</td></tr> <tr><td>2</td><td>4.70041E-06</td></tr> <tr><td>3</td><td>4.23037E-05</td></tr> <tr><td>4</td><td>0.000269686</td></tr> <tr><td>5</td><td>0.001294494</td></tr> <tr><td>6</td><td>0.004854351</td></tr> <tr><td>7</td><td>0.014563052</td></tr> <tr><td>8</td><td>0.03549744</td></tr> <tr><td>9</td><td>0.070994879</td></tr> <tr><td>10</td><td>0.117141551</td></tr> <tr><td>11</td><td>0.159738478</td></tr> <tr><td>12</td><td>0.179705788</td></tr> <tr><td>13</td><td>0.165882266</td></tr> <tr><td>14</td><td>0.124411699</td></tr> <tr><td>15</td><td>0.07464702</td></tr> <tr><td>16</td><td>0.03499079</td></tr> <tr><td>17</td><td>0.012349691</td></tr> <tr><td>18</td><td>0.003087423</td></tr> <tr><td>19</td><td>0.000487488</td></tr> <tr><td>20</td><td>3.65616E-05</td></tr> </tbody> </table>  </div> <p>p-value: _____</p>	Number of "Successes"	Binomial Probabilities	0	1.09951E-08	1	3.29853E-07	2	4.70041E-06	3	4.23037E-05	4	0.000269686	5	0.001294494	6	0.004854351	7	0.014563052	8	0.03549744	9	0.070994879	10	0.117141551	11	0.159738478	12	0.179705788	13	0.165882266	14	0.124411699	15	0.07464702	16	0.03499079	17	0.012349691	18	0.003087423	19	0.000487488	20	3.65616E-05
Number of "Successes"	Binomial Probabilities																																												
0	1.09951E-08																																												
1	3.29853E-07																																												
2	4.70041E-06																																												
3	4.23037E-05																																												
4	0.000269686																																												
5	0.001294494																																												
6	0.004854351																																												
7	0.014563052																																												
8	0.03549744																																												
9	0.070994879																																												
10	0.117141551																																												
11	0.159738478																																												
12	0.179705788																																												
13	0.165882266																																												
14	0.124411699																																												
15	0.07464702																																												
16	0.03499079																																												
17	0.012349691																																												
18	0.003087423																																												
19	0.000487488																																												
20	3.65616E-05																																												
Conclusion	We do not have enough evidence that the company is discriminating against females when selecting employees for management training.																																												

**Example: Font Preference**

Finally, we will use the binomial distribution to find the exact p-value for the Font Preference Study. Recall that we have already estimated this p-value using a simulation study:



Estimated p-value = \_\_\_\_\_

Before we use the binomial distribution to find the exact p-value, we will first check whether the assumptions behind the binomial distribution seem reasonable in this case.

- There exist a fixed number of trials,  $n$ .
  
- There are only two possible outcomes on each trial (“success” or “failure”).
  
- The probability of success ( $\pi$ ) remains constant from trial to trial.
  
- The  $n$  trials are independent.

<b>Font Preference Study</b>																																																																																					
<b>Research Hypothesis</b>	Consumers prefer one font over the other.																																																																																				
<b>Null and Alternative Hypotheses</b>	<p>Let <math>\pi</math> = the proportion of all consumers that would choose the Signet font; equivalently, let <math>\pi</math> = the probability of a randomly selected consumer choosing the Signet font.</p> <p><math>H_0: \pi = 50\%</math>.  <math>H_a: \pi \neq 50\%</math>.</p>																																																																																				
<b>p-value</b>	<p>To find the exact p-value, we will use the binomial distribution with...</p> <p><math>n = \underline{\hspace{2cm}}</math> and <math>\pi = \underline{\hspace{2cm}}</math></p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 50%;">Number of "Successes"</th> <th style="width: 50%;">Binomial Probabilities</th> </tr> </thead> <tbody> <tr><td>0</td><td>9.09495E-13</td></tr> <tr><td>1</td><td>3.63798E-11</td></tr> <tr><td>2</td><td>7.09406E-10</td></tr> <tr><td>3</td><td>8.98581E-09</td></tr> <tr><td>4</td><td>8.31187E-08</td></tr> <tr><td>5</td><td>5.98455E-07</td></tr> <tr><td>6</td><td>3.49099E-06</td></tr> <tr><td>7</td><td>1.69562E-05</td></tr> <tr><td>8</td><td>6.99444E-05</td></tr> <tr><td>9</td><td>0.000248691</td></tr> <tr><td>10</td><td>0.000770943</td></tr> <tr><td>11</td><td>0.002102571</td></tr> <tr><td>12</td><td>0.005081214</td></tr> <tr><td>13</td><td>0.010944152</td></tr> <tr><td>14</td><td>0.02110658</td></tr> <tr><td>15</td><td>0.036584738</td></tr> <tr><td>16</td><td>0.057163653</td></tr> <tr><td>17</td><td>0.080701628</td></tr> <tr><td>18</td><td>0.103118747</td></tr> <tr><td>19</td><td>0.119400655</td></tr> <tr><td>20</td><td>0.125370688</td></tr> <tr><td>21</td><td>0.119400655</td></tr> <tr><td>22</td><td>0.103118747</td></tr> <tr><td>23</td><td>0.080701628</td></tr> <tr><td>24</td><td>0.057163653</td></tr> <tr><td>25</td><td>0.036584738</td></tr> <tr><td>26</td><td>0.02110658</td></tr> <tr><td>27</td><td>0.010944152</td></tr> <tr><td>28</td><td>0.005081214</td></tr> <tr><td>29</td><td>0.002102571</td></tr> <tr><td>30</td><td>0.000770943</td></tr> <tr><td>31</td><td>0.000248691</td></tr> <tr><td>32</td><td>6.99444E-05</td></tr> <tr><td>33</td><td>1.69562E-05</td></tr> <tr><td>34</td><td>3.49099E-06</td></tr> <tr><td>35</td><td>5.98455E-07</td></tr> <tr><td>36</td><td>8.31187E-08</td></tr> <tr><td>37</td><td>8.98581E-09</td></tr> <tr><td>38</td><td>7.09406E-10</td></tr> <tr><td>39</td><td>3.63798E-11</td></tr> <tr><td>40</td><td>9.09495E-13</td></tr> </tbody> </table>	Number of "Successes"	Binomial Probabilities	0	9.09495E-13	1	3.63798E-11	2	7.09406E-10	3	8.98581E-09	4	8.31187E-08	5	5.98455E-07	6	3.49099E-06	7	1.69562E-05	8	6.99444E-05	9	0.000248691	10	0.000770943	11	0.002102571	12	0.005081214	13	0.010944152	14	0.02110658	15	0.036584738	16	0.057163653	17	0.080701628	18	0.103118747	19	0.119400655	20	0.125370688	21	0.119400655	22	0.103118747	23	0.080701628	24	0.057163653	25	0.036584738	26	0.02110658	27	0.010944152	28	0.005081214	29	0.002102571	30	0.000770943	31	0.000248691	32	6.99444E-05	33	1.69562E-05	34	3.49099E-06	35	5.98455E-07	36	8.31187E-08	37	8.98581E-09	38	7.09406E-10	39	3.63798E-11	40	9.09495E-13
Number of "Successes"	Binomial Probabilities																																																																																				
0	9.09495E-13																																																																																				
1	3.63798E-11																																																																																				
2	7.09406E-10																																																																																				
3	8.98581E-09																																																																																				
4	8.31187E-08																																																																																				
5	5.98455E-07																																																																																				
6	3.49099E-06																																																																																				
7	1.69562E-05																																																																																				
8	6.99444E-05																																																																																				
9	0.000248691																																																																																				
10	0.000770943																																																																																				
11	0.002102571																																																																																				
12	0.005081214																																																																																				
13	0.010944152																																																																																				
14	0.02110658																																																																																				
15	0.036584738																																																																																				
16	0.057163653																																																																																				
17	0.080701628																																																																																				
18	0.103118747																																																																																				
19	0.119400655																																																																																				
20	0.125370688																																																																																				
21	0.119400655																																																																																				
22	0.103118747																																																																																				
23	0.080701628																																																																																				
24	0.057163653																																																																																				
25	0.036584738																																																																																				
26	0.02110658																																																																																				
27	0.010944152																																																																																				
28	0.005081214																																																																																				
29	0.002102571																																																																																				
30	0.000770943																																																																																				
31	0.000248691																																																																																				
32	6.99444E-05																																																																																				
33	1.69562E-05																																																																																				
34	3.49099E-06																																																																																				
35	5.98455E-07																																																																																				
36	8.31187E-08																																																																																				
37	8.98581E-09																																																																																				
38	7.09406E-10																																																																																				
39	3.63798E-11																																																																																				
40	9.09495E-13																																																																																				

<p>Conclusion</p>	<p>There is evidence that consumers have a preference towards one font over the other (more chose the Signet font).</p>

**Practice Problems**

1. Consider Example 2.1 (Claims of Numbness after an Automobile Accident).
  - a. Check the conditions for the binomial distribution in the context of this example.
  - b. Find the exact p-value using the binomial distribution.
  
2. Consider Example 2.2 (Effectiveness of an Experimental Drug).
  - a. Check the conditions for the binomial distribution in the context of this example.
  - b. Find the exact p-value using the binomial distribution.

### MORE ON USING P-VALUES TO MAKE DECISIONS

As mentioned earlier, some researchers advocate using the following guidelines:

- If the p-value falls below .05, there is strong evidence to support the research hypothesis.
- If the p-value falls below .10 but above .05, there is “marginal” evidence to support the research hypothesis.
- If the p-value is above .10, there is not enough evidence to support the research hypothesis.

In general, the smaller the p-value, the less likely the results of the study are due to random chance; thus, the more evidence we have that to support the research hypothesis. In some disciplines, the p-value must be much smaller than .05 in order to support a research hypothesis. For example, physics journals often like to see  $p < .001$ .

Though the above guidelines allow for claims of “marginal” evidence when p-values fall between .05 and .10, some statisticians caution against this. For example, Irwin Bross argues that such modifications would be detrimental in evaluating evidence.

*Anyone familiar with certain areas of the scientific literature will be well aware of the need for curtailing language-games. Thus if there were no 5% level firmly established, then some persons would stretch the level to 6% or 7% to prove their point. Soon others would be stretching to 10% and 15% and the jargon would become meaningless. Whereas nowadays a phrase such as statistically significant difference provides some assurance that the results are not merely a manifestation of sampling variation, the phrase would mean very little if everyone played language-games. To be sure, there are always a few folks who fiddle with significance levels--who will switch from two-tailed to one-tailed tests or from one significance test to another in an effort to get positive results. However such gamesmanship is severely frowned upon.*

*Source: Bross IDJ (1971), "Critical Levels, Statistical Language and Scientific Inference," in Foundations of Statistical Inference.*

The “.05 rule” is usually attributed to R.A. Fisher. His published thoughts on the matter are given below.

*In the investigation of living beings by biological methods statistical tests of significance are essential. Their function is to prevent us being deceived by accidental occurrences, due not to the causes we wish to study, or are trying to detect, but to a combination of the many other circumstances which we cannot control. An observation is judged significant, if it would rarely have been produced, in the absence of a real cause of the kind we are seeking. It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once*

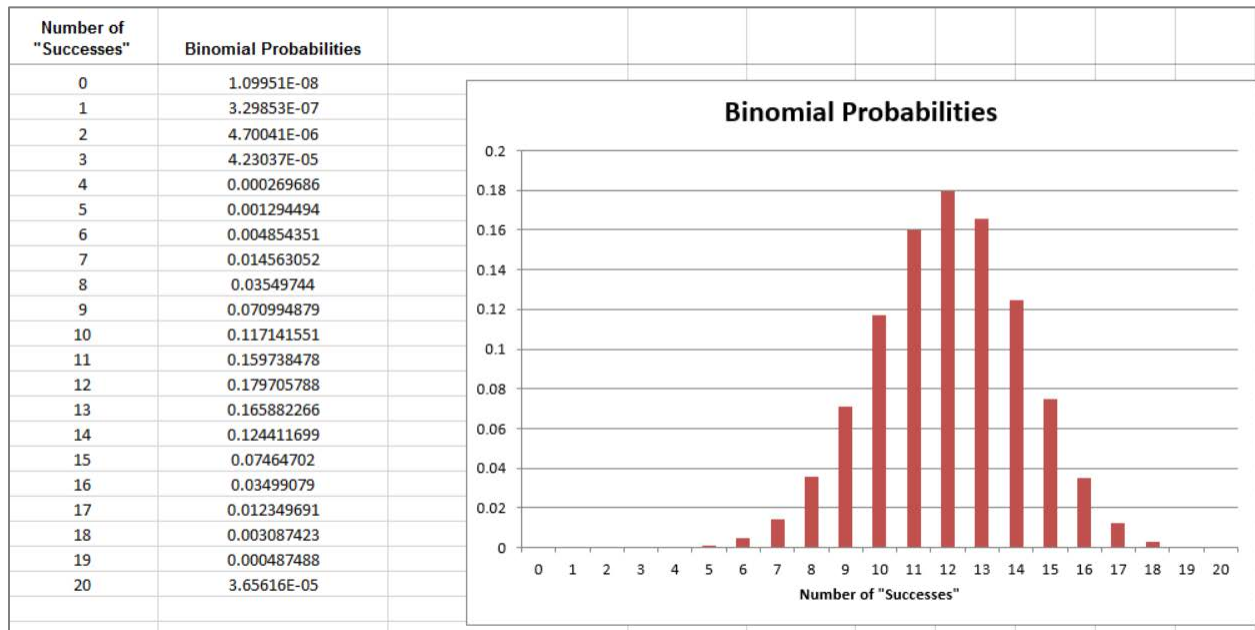
*in twenty trials. This is an arbitrary, but convenient, level of significance for the practical investigator, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation.*

Source: R.A. Fisher (1929), "The Statistical Method in Psychological Research," from the Proceedings of the Society for Psychical Research, 39, 189-191.

**EXAMINING THE EFFECT OF SAMPLE SIZE**

**Example 2.3: Gender Discrimination, Revisited**

Once again, consider the Gender discrimination example. Recall that of the 20 persons selected for management so far, only 9 (or  $9/20 = 45\%$ ) were female. In an unbiased selection process, we expected to see a woman selected 60% of the time. We found that the probability of observing 9 or fewer women selected by chance if the company was in fact using a fair selection process was .1275.



Now, suppose that the results had actually been as follows: Of the last 40 persons selected for management so far, only 18 (or  $18/40 = 45\%$ ) were female. How does this change the calculation of the p-value?



The binomial probabilities for this scenario are shown below.

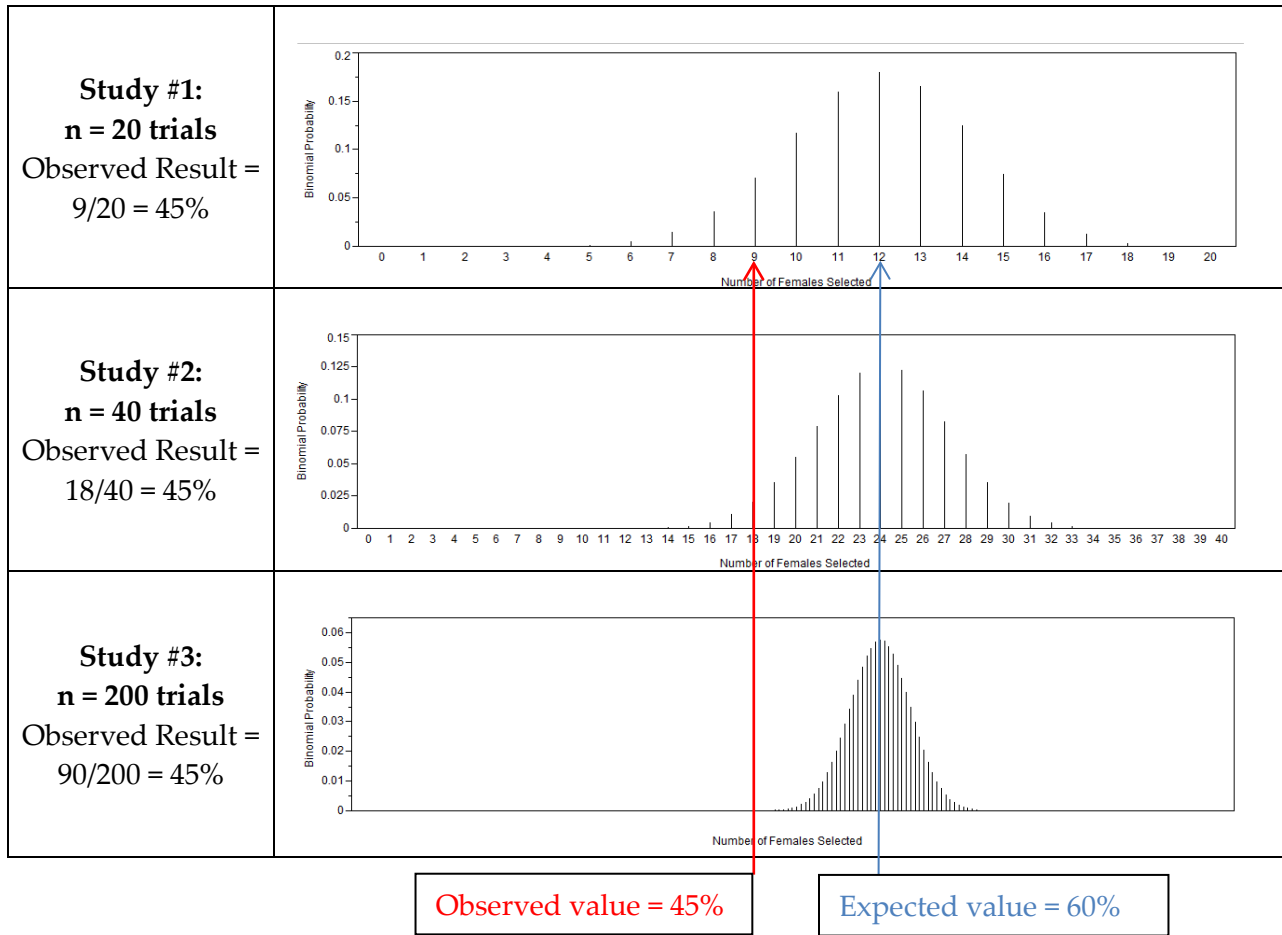
Number of "Successes"	Binomial Probabilities
0	1.20893E-16
1	7.25355E-15
2	2.12166E-13
3	4.03116E-12
4	5.59324E-11
5	6.0407E-10
6	5.28561E-09
7	3.85094E-08
8	2.38277E-07
9	1.27081E-06
10	5.90928E-06
11	2.41743E-05
12	8.76319E-05
13	0.000283118
14	0.000819021
15	0.002129454
16	0.004990908
17	0.010568983
18	0.020257217
19	0.035183587
20	0.055414149
21	0.07916307
22	0.102552159
23	0.120387317
24	0.127911524
25	0.122795063
26	0.106264959
27	0.082650523
28	0.057560186
29	0.035727012
30	0.019649857
31	0.009507995
32	0.004011185
33	0.001458613
34	0.000450454
35	0.000115831
36	2.41315E-05
37	3.91321E-06
38	4.63406E-07
39	3.56467E-08
40	1.33675E-09

For this scenario, the p-value is given by \_\_\_\_\_.

Finally, we can also consider a third study in which of the last 200 persons selected for management so far, only 90 (or  $90/200 = 45\%$ ) were female. How would you find the p-value for this scenario?

For this scenario, the p-value is given by \_\_\_\_\_.

Recall that these binomial distributions give us an idea of what outcomes occur by chance when the selection process does *not* discriminate based on gender (because we used  $\pi = 60\%$ ).



Question:

Which of the following statements is most correct?

- i. The three studies provide equally convincing statistical evidence that the selection process discriminates against women.
- ii. Study #1 provides the most convincing statistical evidence that the selection process discriminates against women.
- iii. Study #2 provides the most convincing statistical evidence that the selection process discriminates against women.
- iv. Study #3 provides the most convincing statistical evidence that the selection process discriminates against women.

Explain your reasoning.

**PRACTICAL SIGNIFICANCE VERSUS STATISTICAL SIGNIFICANCE**

Consider the previous example. As discussed earlier, the result in Study #1 is not “statistically significant” because the observed outcome does not fall in the bottom 5% of simulated outcomes (i.e., the p-value is not below .05). On the contrary, the results in Study #2 and Study #3 are “statistically significant.” The previous example illustrates that statistical significance depends on the sample size. All three studies resulted in an outcome of 45% of those selected for management being female, but this result was only statistically significant in the studies with the larger sample size.

This presents a conundrum: If a study’s results are not statistically significant, it could be that the effect under study is real, but the sample size wasn’t large enough to detect that effect (this relates to a concept known as the *power* of a hypothesis test which is discussed in upper-level statistics courses). On the other hand, if the sample size is large enough, very small differences between the observed results and the expected value in the null hypothesis can lead to statistically significant differences.

To counter this, researchers often consider “practical significance” in addition to “statistical significance.” A result is known as “practically significant” if the difference between the observed and expected result is large enough to be of value in the practical sense.

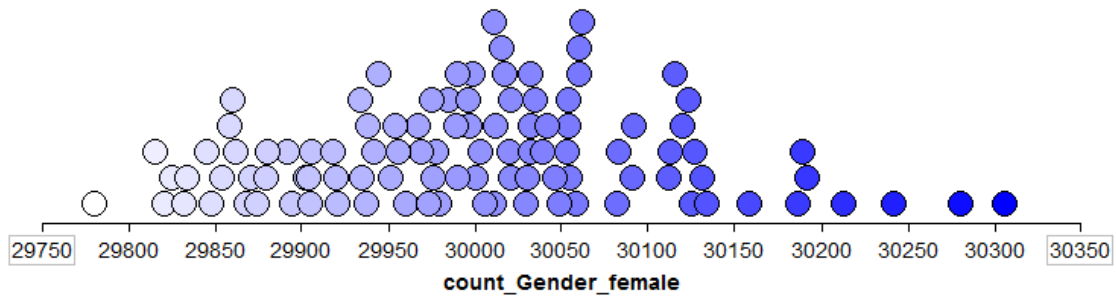
**Example 2.4: Dukes vs. Wal-mart Stores, Inc.**

The lead plaintiff in this case, Betty Dukes, was a Wal-Mart employee. She and others alleged gender discrimination in promotion policies and practices in Wal-Mart stores.

As the nation’s largest private employer, Wal-Mart makes tens of thousands of promotion decisions each year. The following data was provided during this trial: Wal-Mart promoted roughly 50,000 individuals to management between 1997 and 2002. Female employees constituted about 60% of the group eligible for these promotions.

**Questions:**

1. How many of the 50,000 individuals promoted do you expect to be female if Wal-Mart is not discriminating based on gender?
2. Suppose that 29,780 of the individuals promoted were women (note that this was not the actual outcome in the court case). What percentage is this?
3. A simulation study with 100 trials was conducted to see what outcomes occur by chance when the selection process is *not* discriminatory. The results are shown below.



Recall the hypothetical observed value of 29,780 females being selected. The research hypothesis is that Wal-Mart is discriminating against women in their promotion policies and practices. Based on the results of the simulation study, is this result “statistically significant”? Explain.

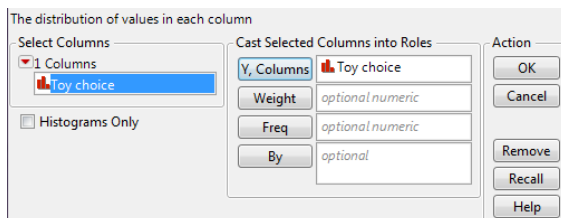
4. In the previous question, you estimated the p-value for this research hypothesis. Use the binomial distribution to find the exact p-value. Again, is the result “statistically significant”?
  
5. Is this result “practically significant”? Explain.
  
6. Do you foresee any problems if we rely on only statistical significance when making decisions? Explain.

### USING JMP TO FIND P-VALUES FOR THE BINOMIAL EXACT TEST

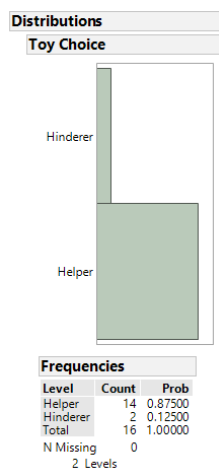
Earlier, we used Excel to calculate binomial probabilities and to find the p-value associated with the binomial exact test. This test is also easily implemented in JMP when given the raw data. For example, the data from the Helper/Hinderer example could be entered into JMP as follows:

	Toy choice
1	Helper
2	Helper
3	Helper
4	Helper
5	Helper
6	Helper
7	Helper
8	Helper
9	Helper
10	Helper
11	Helper
12	Helper
13	Helper
14	Helper
15	Hinderer
16	Hinderer

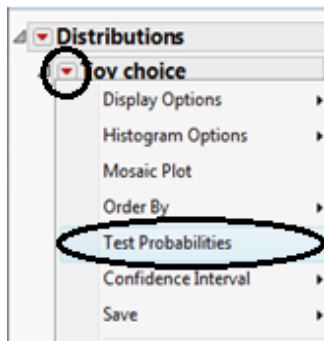
To get the test statistic and p-value in JMP, select **Analyze > Distribution**. Enter the following:



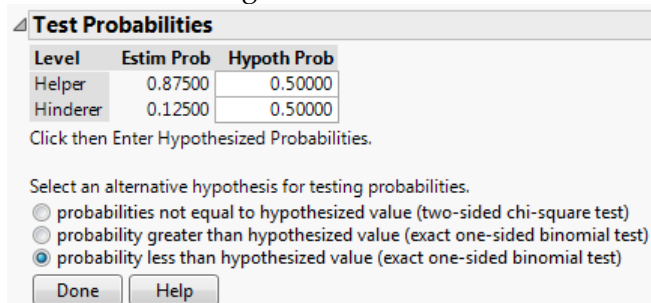
When you click OK, JMP displays descriptive statistical summaries:



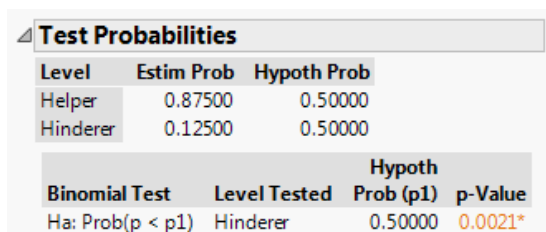
To carry out the inferential statistics, on the output that appears, select “Test Probabilities” from the red drop-down arrow next to the variable name:



Enter the following:



Click “Done” and JMP should return this output:

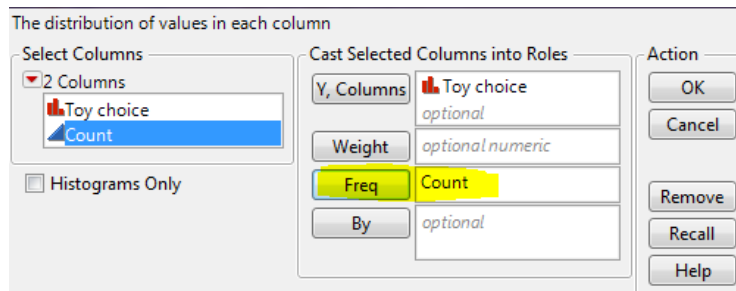


### An Alternative Method of Setting up the Data Sheet

You can also set up the data as shown below:

	Toy choice	Count
1	Helper	14
2	Hinderer	2

If you provide the summarized data to JMP as shown above, then enter the following after selecting **Analyze > Distribution**:



From here on out, the steps are the same as shown on the previous page.