STAT 110: Practice Problem 3 - Solutions
Fall 2017

---

1. A recent U.S. Census report indicated that 71% of the nation's college undergraduates work while in school. Suppose that in a random sample of 185 undergraduates at Winona State, 124 reported working while in school. The research hypothesis is as follows: **The proportion of Winona State University undergraduates who work while in school is less than the national average.**

   a. Identify both the population of interest and the sample in this study.

      Population: **all Winona State University undergraduate students**
      Sample: **the 185 WSU undergraduate students who were surveyed**

   b. Let π = the proportion of all WSU undergraduates who work while in school. Calculate the statistic of interest in this study.

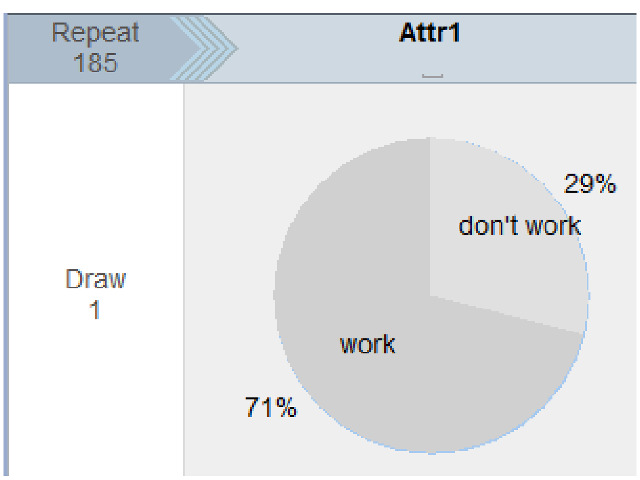
      $\hat{\pi}$ = 1**24/185 = 67%**

   c. Set up the null and alternative hypotheses to test the research hypothesis.

      H_o: **π = 71% (the proportion of WSU undergraduates who work while in school is the same as the national average)**

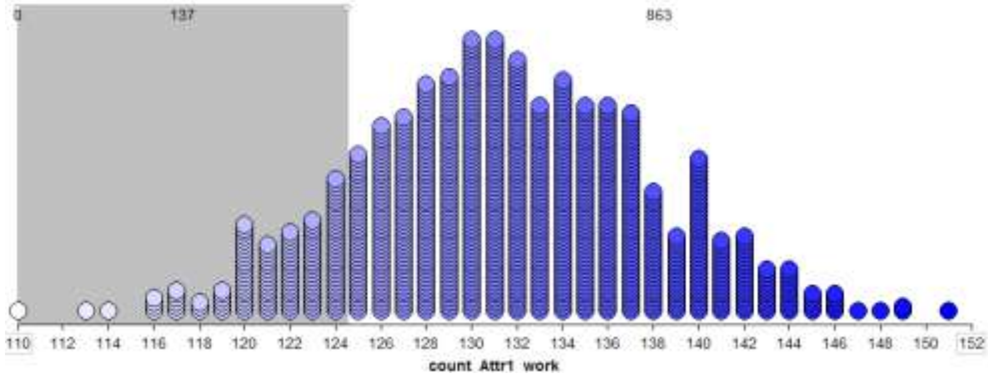
      H_a: **π < 71% (the proportion of WSU undergraduates who work while in school is less than the national average)**

   d. Next, set up a simulation in Tinkerplots that will allow you estimate a p-value for investigating this research hypothesis. Sketch your spinner, including (1) the labels you used, (2) the probabilities attached to these labels, and (3) the Repeat value.



e. Carry out 1,000 runs of your simulation study. Sketch your results below.

Your results should be similar to the following:



f. Recall that in this study, 124 of the 185 WSU undergraduates surveyed reported working while in school. Use the results of your simulation to estimate the p-value.

Estimated p-value: **To compute this, find the proportion of simulated outcomes that were at 124 or less. From my simulation study (shown above), the estimated p-value is 137/1000 = 0.137.**

g. Write a conclusion in the context of the problem to address whether or not the evidence supports the research hypothesis.

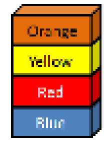
Though fewer than 71% of the sample reported working while in school, this survey study overall did not provide enough statistical evidence to conclude that the proportion of *all* Winona State University undergraduates who work while in school is less than the national average.

2. Consider the following study in which researchers were interested in trying to determine how much information a child gathers from conversations held between others.

A child was told by the experimenter that they were going to soon be setting up a game to play and that the child might win a prize at the end. The experimenter asked the child to sit quietly off to the side and play with a toy for a few moments while she talked with her friend, who was a bystander in the room. The experimenter then started a conversation (that was purposely scripted to be uninteresting to the child) with the bystander in the room, and the following statement was made in the middle of this conversation.

Experimenter: "Oh, wait. I don't remember which drawer has the prize in it. Is it the _____ drawer?"
Bystander: "Yes, it is in the _______ drawer." The blanks were filled in at random with one of the four colors of the actual drawers: orange, yellow, red, or blue. These colors were counterbalanced amongst the participants.

The conversation continued for a few more moments, and then the experimenter returned to the child and continued setting up their game. The experimenter then made the following statement, "Oh, I can't remember which drawer the prize is in!" They asked the child, "Which drawer should I open?" The researchers recorded whether the child chose the correct drawer or not (the correct drawer was the one that had been mentioned in the conversation). **This was repeated with 51 children, and 33 of them identified the correct drawer.**

Conversation with child playing in same room	One drawer contains the prize
	

Source: "Little Pitchers Use Their Big Ears: Preschoolers Solve Problems by Listening to Others Ask Questions" by Mills, C.M, Danovitch, J.H, Grant, M.G, and Elashi, F.B. (2012). Child Development, March/April 2012, Volume 83, Number 2, Pages 568-580.

The researchers in this study want to show that children are able to apply information overheard from a third-party conversation that took place while their attention was engaged in another task and with no prior instruction to listen to that conversation. In other words, they want to show that **more children are choosing the correct drawer than we would expect by guessing.**

a. Identify both the population of interest and the sample in this experiment.

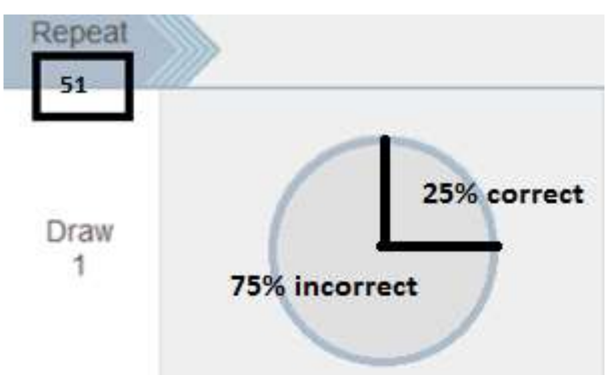
Population: **All preschoolers (ages 3-5)**
Sample: **The 51 preschoolers used in this study**

b. Set up the null and alternative hypotheses to address the research hypothesis.

Let π = the proportion of <u>all</u> preschoolers that would identify the correct drawer (or, you could think of this as the true probability of a preschooler choosing the correct drawer).

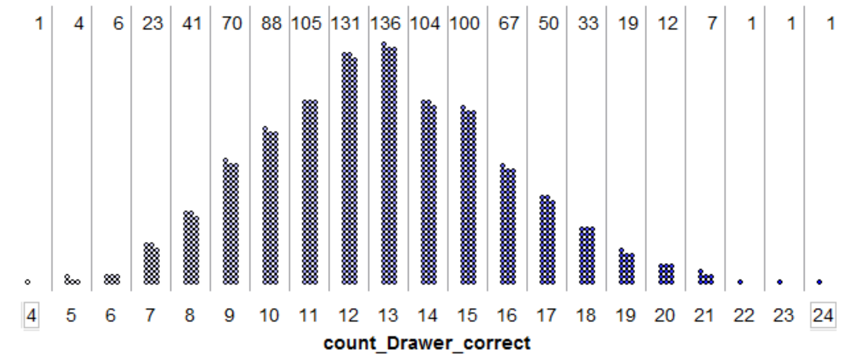
H_0: $\pi = .25$ (i.e., the children are simply guessing which drawer the prize is in)
H_a: $\pi > .25$ (i.e., more children are choosing the correct drawer than we would expect by guessing, which may indicate they are applying information they overheard from a third-party conversation)

c. Set up a simulation in Tinkerplots that will allow you estimate a p-value for investigating the research hypothesis. Sketch your spinner, including the labels you used, the probabilities attached to these labels, and the Repeat value.



d. Carry out the simulation study by tracking the number of children who <u>correctly</u> identify the correct drawer just by guessing, repeating this process a total of 1000 times. Sketch your results below.

Your results should be similar to the following:



e. Recall that in this study, 33 of the 51 children identified the correct drawer. Use the results of your simulation to estimate the p-value.

Estimated p-value: **Count the proportion of simulated outcomes that were at 33 or higher. In my simulation study, this was 0/1000 = 0.**

f. Write a conclusion in the context of the problem to address whether the evidence supports the research hypothesis.

The probability of obtaining such a study result if the children are guessing is *extremely* small. Since this p-value is well below 0.05, we have very strong evidence that the children are choosing the correct drawer more often than we would expect by guessing and thus believe they are applying information overheard from third-party conversations.

g. The researchers have asked you to complete the following statement to include in their paper: "Above-chance performance as measured by a binomial test required that at least ___ of the 51 children identify the correct drawer." Basically, they are asking you how many of the 51 children have to identify the correct drawer in order for them to have statistical evidence supporting their research question. Give this cutoff and explain your reasoning.

Cutoff = 19

Reasoning: Based on the binomial distribution, the probability of 19 or more preschoolers choosing the correct drawer when guessing is 0.035, which means that the p-value for this observed outcome is 0.035 (which is less than .05). However, if only 18 preschoolers were to choose the correct drawer, the p-value would be 0.066 (which is above .05) and we would not have evidence to support the research hypothesis using the .05 rule.

3. In the 1980s, it was generally believed that autism affected about 5% of the nation's children. Some people believe that an increase in the number of chemicals in the environment has led to an increase in the incidence of autism today. A recent study examined 384 children and found that 46 of them showed signs of some form of autism. Do these results provide evidence that the incidence of autism has increased since 1980? *Comment: Note that even if there is evidence that the incidence of autism has increased, we can't be sure that it's due to chemicals – this is just a theory. It's also reasonable to assume that the incidence of autism has increased simply because more people are diagnosed today than in 1980.*

   a. Identify both the population and the sample in this experiment.

      Population: **the nation's children**
      Sample: **the 384 children studied**

   b. Define the parameter of interest, π.

      π = the proportion of all children in our nation today who have autism.

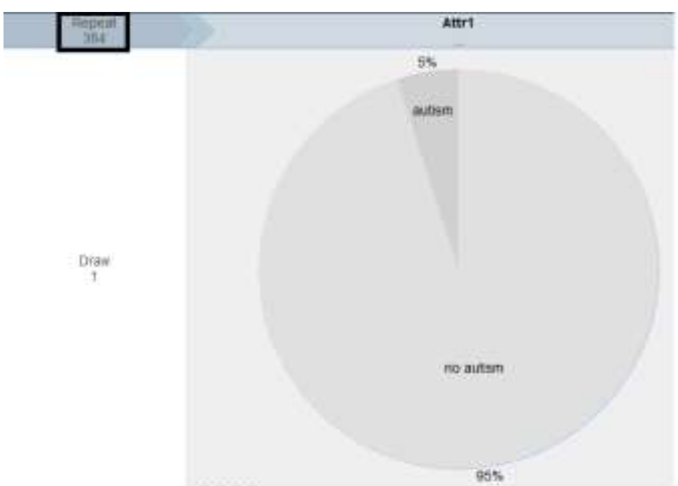
   c. Calculate the statistic and **explain how this differs from the parameter**.

      $\hat{\pi}$ = 46/384 = 12%. This is the proportion of the <u>sample</u> that has autism; the parameter represents the proportion of the <u>population</u> with autism.

   d. Set up the null and alternative hypotheses to address the research hypothesis.

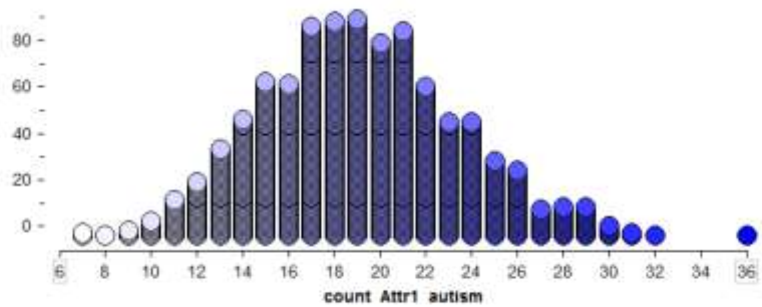
      H_o: **π = .05 (or π = 5%; the incidence of autism is the same today as in 1980)**
      H_a: **π > .05 (or π > 5%; the incidence of autism has increased since 1980)**

   e. Set up a simulation in Tinkerplots that will allow you estimate a p-value for investigating the research hypothesis. Sketch your spinner, including the labels you used, the probabilities attached to these labels, and the Repeat value.



f. Carry out the simulation study by tracking the number of children who have autism, repeating this process a total of 1000 times. Sketch your results below.

Your results should be similar to the following:



g. Recall that in this study, 46 of the 384 children studied had autism. Use the results of your simulation to estimate the p-value.

Estimated p-value: **Count the proportion of simulated outcomes that were at 46 or higher. In my simulation study, this was 0/1000 = 0.**

h. Write a conclusion in the context of the problem to address whether the evidence supports the research hypothesis.

The probability of obtaining such a study result if the incidence of autism is still only 5% is *extremely* small. Since this p-value is well below .05, we have very strong evidence that the incidence of autism today is higher than 5%; that is, we have evidence that it has increased since the 1980s (recall, however, that it could simply be that more are diagnosed today).