Worked Example Handouts and Interpretation Questions

Where appropriate, the correct answers of interpretation questions are indicated in italics or bold.

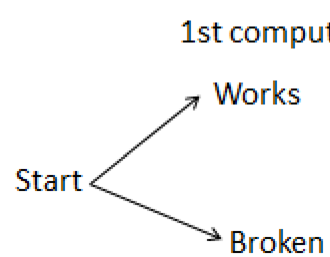
Chapter 5: Conditional Probability

1a. A company has just received a shipment of 20 computers. 5 of them are secretly broken. The head of IT randomly chooses 2 different computers to test. What is the probability that both of the computers work?

Solution: The problem says that the head of IT tests 2 *different* computers, which means that they're sampling without replacement. (After they sample one computer, they don't put it back to potentially choose again.) This means that the events "the first computer works" and "the second computer works" are dependent. We need to use conditional probability to solve this problem.

The problem doesn't explicitly tell us P(F|E), so I recommend drawing a tree. Start by drawing the first set of branches to represent the possible outcomes of **what we're conditioning on**. Usually this is the thing that happens first or the thing you know about first. In this case, it's whether the first computer works (Figure A.4).

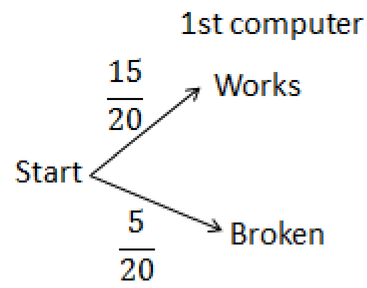
Figure A.4. First Branches of a Conditional Probability Tree.



- **Hint:** If you find yourself drawing more than 3 branches from a single point, stop and think about whether there's a more efficient way. For example, we could have drawn 20 branches, one for each computer that could have been chosen, but that would take a long time!

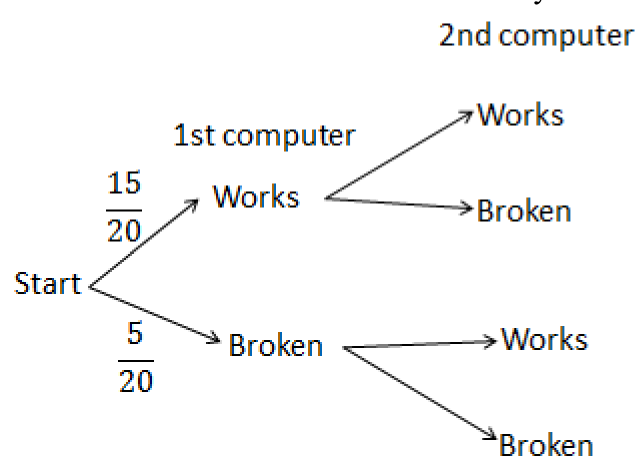
Label each arrow with the probability of that event. In this case, we start with 20 computers, and 5 of them are broken. So P(Broken) = $\frac{5}{20}$ and P(Works) = $\frac{15}{20}$. See Figure A.5.

Figure A.5. First Branches of a Conditional Probability Tree With Probability Labels.



Then, add another set of branches for the thing that **depends on the first event**. Usually this is the thing that happens second, or that we know about second. In this case, it's whether the second computer works or is broken. See Figure A.6.

Figure A.6. Second Set of Branches of a Conditional Probability Tree.

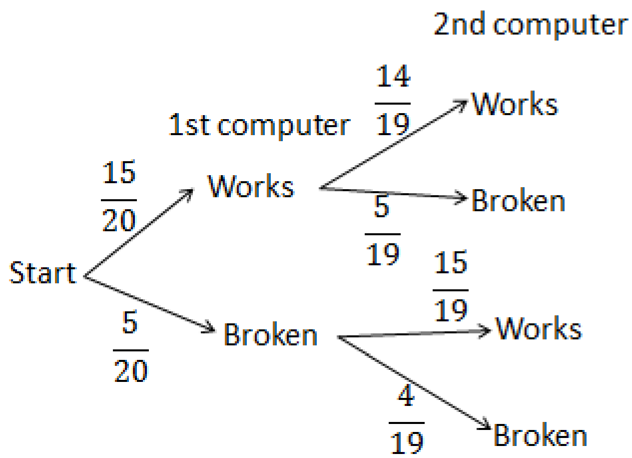


Every outcome from the first event should have its own set of branches for the second event.

Label those branches with their probabilities. These are the *conditional* probabilities of the 2nd computer working or being broken, *given* that the first computer either worked or was broken. If the first computer worked, then we've found one of the 15 working computers, so there are only 14 working computers left that could be randomly selected. There are a total of 19 possible computers to choose from (again, because we've already tested one), so P(2nd computer works | 1st computer worked) = $\frac{14}{19}$.

If the first computer worked, there are still 5 broken computers left, so P(2^{nd} computer is broken | 1^{st} computer worked) = $\frac{5}{19}$. See Figure A.7.

Figure A.7. Second Set of Branches of a Conditional Probability Tree With Probability Labels.



Hint to help you check your work: For any set of branches that starts at the same point, the sum of the probabilities should be 1. (For example, $\frac{14}{19} + \frac{5}{19} = 1$.)

Then identify which path or paths through the tree represent events that we were originally asked about. We're trying to find the probability that both of the computers work, so that means we want the top path: the 1^{st} computer works AND the 2^{nd} computer works. Multiply the probabilities along that path: $\frac{15}{20} * \frac{14}{19} = 0.5526$.

1b. Find the probability that exactly 1 of the computers works.

Solution: We can use the same probability tree to answer this question, because it involves the same scenario (same number of computers, same number that are broken), but asks about a different probability.

In this case, there are 2 paths that represent the event we want to know about. If exactly 1 computer works, that means that **either** the 1^{st} computer works and the 2^{nd} is broken, **or** the 1^{st} computer is broken and the 2^{nd} works. Multiply the probabilities along each of these paths:

P(1^{st} computer works and 2^{nd} is broken) = $\frac{15}{20} * \frac{5}{19} = 0.1973684$

P(1^{st} computer is broken and 2^{nd} works) = $\frac{5}{20} * \frac{15}{19} = 0.1973684$

Then **add** the probabilities from the different paths: $0.1973684 + 0.1973684 = 0.3947$. This is the probability that exactly 1 computer works.

(Why do we add the probabilities from different paths? Each path represents a disjoint event, so we can use the Addition Rule for Disjoint Events to find the probability of path A **or** path B.)

Interpretation Question 1
How can you check your work when writing probability labels?

A. Not more than 3 branches start at the same point.
B. All branches that start at the same point add up to 1.
C. All branches in the tree add up to 1.

Interpretation Question 2
In your opinion, what is the hardest part of this problem?

A. Deciding what events should go on which branches
B. Choosing the probability labels for branches
C. Computing the probability based on the labels
D. Some other part

(Affirm students' experience that these problems can be challenging. Give students the opportunity to ask questions about the aspects that they find hardest.)

Worked examples, continued
2. The General Social Survey asked a random sample of 2000 people to rate their happiness and their health. Table A.2 shows the results (Adapted from Sullivan 2013, p. 610).

Table A.2. Example Data on Happiness and Health.

Happiness	Excellent Health	Good Health	Fair Health	Poor Health	**Total**
Very Happy	271	304	48	17	**640**
Pretty Happy	265	533	247	53	**1098**
Not Too Happy	27	113	92	30	**262**
Total	**563**	**950**	**387**	**100**	**2000**

Suppose we choose 1 person at random from this sample. Are the events "Good Health" and "Very Happy" independent?

Solution: Start by finding the marginal probabilities:

P(E) = P(Good Health) = (# of people in good health) / (total # of people) = 950/2000 = 0.475
P(F) = P(Very Happy) = (# of very happy people) / (total # of people) = 640/2000 = 0.32

Find the product of the probabilities:

$$P(E) * P(F) = 0.475 * 0.32 = 0.152$$

Then find the probability of both events together:

P(E and F) = P(Good Health **and** Very Happy) = (# of people in good health **and** very happy) / (total # of people) = 304/2000 = 0.152

We found that P(E and F) = P(E) * P(F), so these events are independent.

Interpretation Question 3
Which method did this example use to check for independence?

A. P(E|F) = P(E)
B. P(F|E) = P(F)
*C. P(E and F) = P(E)*P(F)*

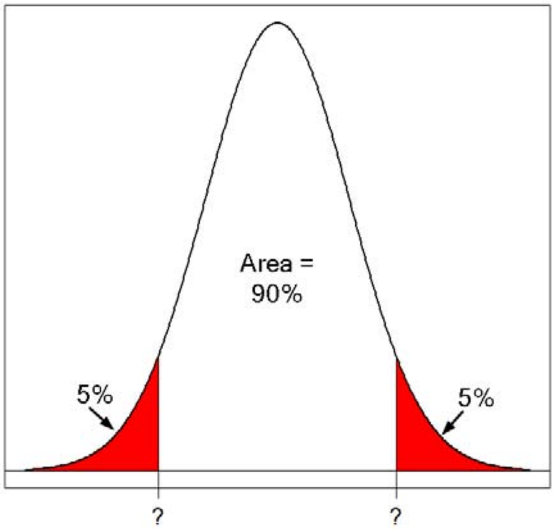
Chapter 7: Normal Distribution Quantiles

1. The scores earned on the math portion of the SAT are approximately normally distributed, with mean 516 and standard deviation 116. What scores separate the middle 90% of test takers from the bottom 5% and the top 5%?

(Note: Another way to think about this is, "What scores are the 5^{th} percentile and the 95^{th} percentile?")

Solution: Start by sketching a normal probability curve and labeling what we know, as shown in Figure A.8. We know we want 5% of the area below the lower cutoff, 5% above the upper cutoff, and 90% of the area between the two cutoffs.

Figure A.8. Illustration of Normal Probability Curve.



We want to find the values of x (the scores) that are labeled on Figure A.8 with "?". Because we want to find values of x, not probabilities, we want invNorm.

The lower cutoff (x_1) has an area of 0.05 below it, so we can find it using

2nd -> Vars -> invNorm

invNorm(.05, 516, 116)

So $x_1 = 325.1970$.

The upper cutoff (x_2) has an area of 0.9 + 0.05 = 0.95 below it, so we can find it using

2nd -> Vars -> invNorm

invNorm(.95, 516, 116)

So $x_2 = 706.8030$.

Interpretations:

- The middle 90% of SAT takers have math scores between 325 and 707.
- The probability that a randomly selected SAT taker has a math score between 325 and 707 is 0.9.

- **Note:** If the question asks for the first quartile (Q_1), then it's asking for what ***x*-value** has 25% of the data (or area) below it. So you would also want to use **invNorm**.
- Similarly, if the question asks for something like the "90th percentile", then it's asking for what ***x*-value** has 90% of the data (or area) below it.

Interpretation Question 1
How do you decide when to use **invNorm** and when to use **normalcdf**?

A. *Use **invNorm** when asked to find an x-value. Use **normalcdf** when asked to find a probability or percentage.*
B. Use **invNorm** when the problem only has 3 numbers. Use **normalcdf** when the problem has 4 numbers.
C. Use **invNorm** to find the height of the normal density curve. Use **normalcdf** to find the area under the normal density curve.

Mini-lecture
If we need to decide whether a population's distribution can be modeled by a normal distribution, we can use a *normal probability plot*. (Show an example.) This graph has the observed data (or its z-scores) on the x-axis. The y-axis has the z-scores (or percentiles) that you would expect if the data were perfectly normally distributed.

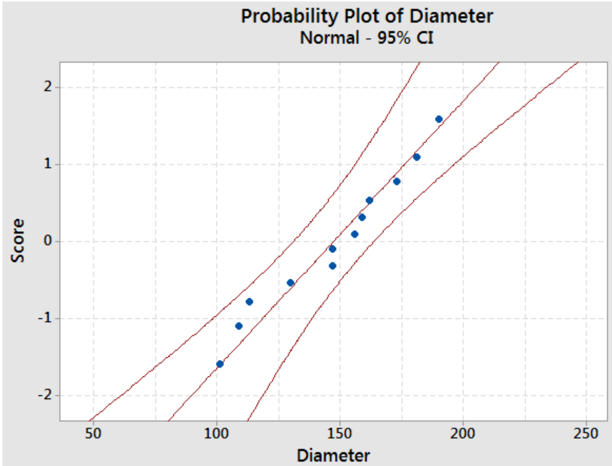
If the data are approximately normal, then the z-scores in data will be close to the expected values, so the points in the probability plot will be close to the line $y = x$. The curved boundaries provide an indication of what we consider "close."

Worked examples, continued
2. The data in Table A.3 and Figure A.9 represent the diameter (in centimeters) of a random sample of 12 Douglas fir trees (Sullivan 2013, p. 472; Minitab 2010).

Table A.3. Diameters of Douglas Fir Trees.

156	190	147	173	159	181
162	130	101	147	113	109

Figure A.9. Normal Probability Plot of Diameters of Douglas Fir Trees.

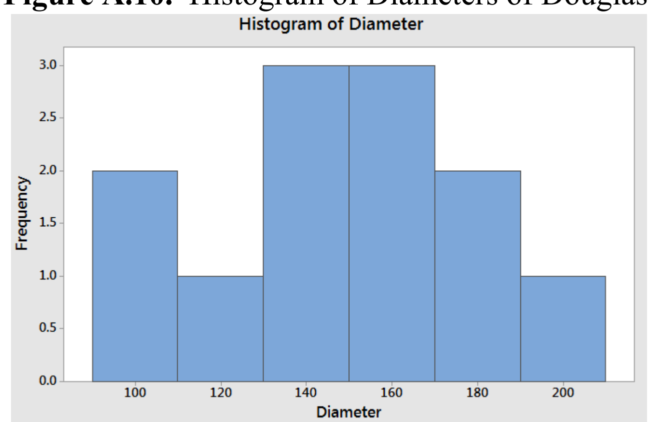


Is a normal distribution an appropriate model for the diameters of Douglas fir trees? If so, estimate the diameter that separates the largest 25% of trees from the rest.

Solution: We want to ignore minor "wiggles" and look at the general trend of the points in the probability plot. It looks like they follow the diagonal line fairly well. Also, all the points stay inside the curved boundaries, which adds to our belief that a normal distribution is an appropriate model.

In this case, we were given the data, so we can use Minitab to make a histogram of the data to help check our assessment, as shown in Figure A.10.

Figure A.10. Histogram of Diameters of Douglas Fir Trees.



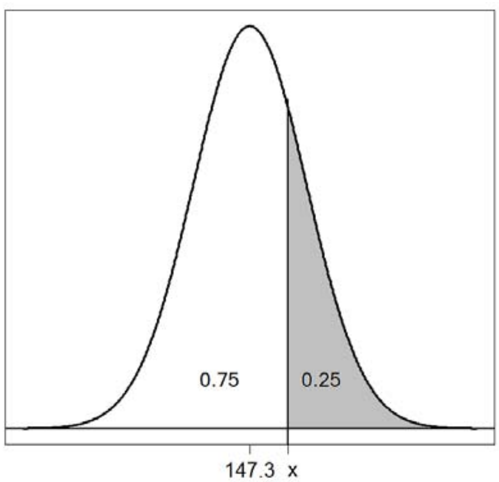
The histogram is not perfectly bell-shaped, but we don't expect that with a random sample (especially one as small as 12 observations). The histogram does not show strong evidence of skewness or outliers, which confirms our idea that a normal distribution is a reasonable model.

Next, we want to estimate the diameter that separates the largest 25% of trees from the rest. To do this, we need to know the mean and standard deviation of the normal distribution. These will be the same as the mean and standard deviation of the data.

Use Stat -> Edit -> Enter the data into L1. Then use Stat -> Calc -> 1-var stats -> 1-var stats L1. $\bar{x}$, the mean, is 147.3333. This is a sample, so look at *Sx*, the sample standard deviation. This is 28.8171.

Then, we want to know what diameter (or x-value) separates the largest 25% of trees from the rest, in a normal distribution with mean 147.3333 and standard deviation 28.8171. If the area under the normal distribution to the right of x is 0.25, then the area to the left is $1 - 0.25 = 0.75$. See Figure A.11.

Figure A.11. Example of Normal Distribution.



So, we can find the cutoff x-value using 2nd -> Vars -> invNorm.

invNorm(.75, 147.3333, 28.8171) = 166.7701

So a tree that is 166.8 cm tall is taller than 75% of all Douglas fir trees. Another interpretation is that 166.7701 is the **third quartile** or **75th percentile** of the population.

Interpretation Question 2

Think-Pair-Share: What are the “big steps” of solving problem 2? What features of the problem told you to use these steps?

Interpretation Question 3
How did we find the standard deviation for the normal distribution in this model?

A. The standard deviation was given in the problem.
B. Find the population standard deviation of the data.
C. *Find the sample standard deviation of the data.*
D. Take the square root of the variance.

Chapter 9: Confidence Intervals for a Proportion

1. Is it appropriate to use your calculator to find a confidence interval for the following? If so, find a 95% confidence interval and interpret it.

a) p is the true proportion of students in this class who are spending more than \$5 to celebrate Halloween (costume, candy, etc.). You survey 40 people and find that 20 of them are spending more than \$5 on Halloween.

Solution: The population (our class) has 60 students. Our sample size is $n = 40$.

$$20 * 40 = 800 > 60$$

so n is larger than 5% of the population size. This means that if we sample without replacement, we can't treat the results from different people as being independent. The binomial distribution is **not** a good model for X, the number of people in our sample who are spending more than \$5 on Halloween. That means that a normal distribution is **not** a good model for $\hat{p}$, so we shouldn't use the built-in function on a calculator to find a confidence interval.

(This doesn't mean that a large sample size is bad! Large samples give us lots of information about the population, which is good. It just means that if your sample is more than 5% of your population, you should consult a statistician about other ways to find a confidence interval.)

b) p is the true proportion of all students at our school who are spending more than \$5 to celebrate Halloween (costume, candy, etc.). You survey 40 people and find that 10 of them are spending more than \$5 on Halloween.

Solution: Our population is all students at our school. n, our sample size, is 40.

$$20 * 40 = 800$$

This is less than the population of our school, so n is less than 5% of the population size. This means that if we sample without replacement, we **can** treat the results from different people as being independent. The binomial distribution **is** a good model for X, the number of people in our sample who are spending more than \$5 on Halloween.

The sample proportion is

$$\hat{p} = \frac{x}{n} = \frac{10}{40} = .25$$

We want to check

$$n\hat{p}(1 - \hat{p}) = 40 * .25 * .75 = 7.5$$

$7.5 < 10$, so the binomial distribution will not be very bell-shaped. This means that a normal distribution is **not** a good model for $\hat{p}$, so we shouldn't use the built-in function on a calculator to find a confidence interval. It would be a good idea to increase the sample size, so we can compute a confidence interval (and learn more about the population.)

c) *p* is the true proportion of all students at a large university (10,000 students) who are spending more than $5 to celebrate Halloween (costume, candy, etc.). You survey 400 people and find that 100 of them are spending more than $5 on Halloween.

Solution: Our population (all students at the large university) is approximately 10,000 people. *n*, our sample size, is 400.

$$20 * 400 = 8{,}000 < 10{,}000$$

So *n* is less than 5% of the population size. This means that if we sample without replacement, we **can** treat the results from different people as being independent. The binomial distribution **is** a good model for *X*, the number of people in our sample who are spending more than $5 on Halloween.

The sample proportion is

$$\hat{p} = \frac{x}{n} = \frac{100}{400} = .25$$

We want to check

$$n\hat{p}(1 - \hat{p}) = 400 * .25 * .75 = 75$$

$75 \geq 10$, so the binomial distribution for *X* **is** approximately bell-shaped. This means that a normal distribution **is** a good model for $\hat{p}$, so we can use the built-in function on a calculator to find a confidence interval.

Find a 95% confidence interval for *p*: We want a confidence interval for a proportion, so we can use Stat -> Tests -> 1-PropZInt. *x* is the number of people in our sample who said "yes," so *x* = 100. Our sample size is *n* = 400. Finally, we want a 95% level of confidence, so choose C-level: .95.

This gives the interval

$$(.20757, .29243)$$

Interpret the 95% confidence interval for *p*:

- We are 95% confident that the true proportion of **all** students at the large university who are spending more than $5 to celebrate Halloween is between .20757 and .29243.

Note: If everyone in our class went out and gathered their own random sample of 400 people and computed the 95% confidence interval for p, we would expect 95% of the intervals to contain the true value of p. (But we wouldn't know which intervals, because the only way to know the true value of p is to survey *all* students at the university.)

Interpretation Question 1
Is it bad to have a sample size > 5% of your population?

A. Yes, because you can't use the functions on your calculator.
B. No, because it gives you more information.

Interpretation Question 2
The interpretation of a confidence interval deals with…

A. p, the proportion in the whole population
B. $\hat{p}$, the proportion in the sample
C. x, the number of people in the sample who said "yes"
D. n, the sample size

Mini-lecture
We found that when $n = 400$, $\hat{p} = .25$, and the level of confidence (LOC) is 95%, our confidence interval is (.20757, .29243).

What if we change the sample size? Try computing a confidence interval with $n = 200$, $\hat{p} = .25$, and LOC = 95%. (Give students time to do this.)

The new confidence interval is (.18999, .31001), which is wider than before. A smaller sample gives us less information to estimate p. That means our estimate is less precise, so the confidence interval is wider.

What if we keep n the same but change the LOC? Try computing a confidence interval with $n = 400$, $\hat{p} = .25$, and LOC = 99%. (Give students time to do this.)

The new confidence interval is (.19423, .30577), which is wider than before. We need a wider interval, to be more sure that it contains the true value: 99% confident instead of only 95% confident.

Think of this like those Velcro mitts that kids use to play catch. (Show the image from http://www.ltksports.com.tw/product_data.php?id=508.) The ball is the population proportion, and we're trying to "catch" it in the confidence interval (the mitt). Imagine that you're a little kid who's not very coordinated, so you're just going to hold your mitt up and hope that the ball flies into it. If you have a giant mitt (zoom in on the image, so it looks larger), you're going to be more confident of catching the ball. It's the same way with confidence intervals: If we want to be more confident of catching the parameter, we need a larger interval.

Worked examples, continued

2. Two studies tried to estimate the proportion of women who have ever had breast cancer, by examining the medical records of a sample of women who died recently. Both studies found a point estimate of $\hat{p} = 0.12$. Study A found a confidence interval of (0.04, 0.20). Study B found a confidence interval of (0.11, 0.13). Which of these are possible explanations? Select all of the possibilities.

i. Study A used a larger sample size.
ii. Study B used a larger sample size.
iii. Study A used a higher level of confidence.
iv. Study B used a higher level of confidence.

Solution: Study A had a margin of error of

$$\frac{.20 - .04}{2} = .08$$

and Study B had a margin of error of

$$\frac{.13 - .11}{2} = .01$$

so Study B had a narrower confidence interval. (In other words, Study B had a more *precise* interval.)

Increasing the sample size (while keeping the level of confidence the same) results in a **narrower** confidence interval. So option ii is a possibility. (Option i is wrong.)

Increasing the level of confidence (while keeping the sample size the same) results in a **wider** confidence interval. So option iii is a possibility. (Option iv is wrong.)

What if we increase *both* the sample size and the level of confidence at the same time? In this case, we can't predict what will happen to the margin of error (or to the precision of the confidence interval). It depends on the specific values of the sample size and the level of confidence.

Chapter 10: 1-Proportion Z-Test

1a. A carnival worker claims to be able to tell people's birth months by psychic powers, more accurately than if he were just guessing. Write the null and alternative hypotheses to test this claim.

There are 12 months in the year, so if the carnival worker guesses at random, we expect that he will be right about $\frac{1}{12}$ of the time. So the null hypothesis is

$$H_0: p = \frac{1}{12}$$

where *p* is the proportion of birth months that the carnival worker would get right, if he attempted to tell the birth months of every person in the population.

(Note that this scenario is asking about a *proportion* of birth months that he gets right, so the parameter should be *p*, not μ.)

We want to test the carnival worker's claim that he gets *more* right than if he were guessing. So, we should use the alternative hypothesis

$$H_a: p > \frac{1}{12}$$

making this a 1-sided test.

b. Out of 80 people at the fair, the carnival worker successfully guesses 9 people's birth months (Sullivan 2013, p. 493)**. Is it appropriate to use 1-prop ZTest to test the claim?**

80 * 20 = 1600, which is less than the population size of the US. So, the sample size is less than 5% of the population size. That means we can treat the observations as being independent.
In this hypothesis test, the sample size is *n* = 80. The hypothesized proportion is $p_0 = \frac{1}{12}$. So

$$np_0(1 - p_0) = 80 * \frac{1}{12} * \left(\frac{11}{12}\right) = 6.1111$$

6.1111 is less than 10, so a normal distribution will not be a good approximation for $\hat{p}$. This means that it is not appropriate to use 1-prop ZTest to test the claim.

When $np_0(1 - p_0) < 10$, you have two good options:

- Gather a larger sample, or
- Do the hypothesis test in Minitab, which can use a binomial distribution to compute the exact p-value.

Interpretation Question 1
What is the parameter in this example?

A. The proportion of birth months the carnival worker got right, out of the 80 people at the fair.
B. *The proportion of birth months the carnival worker would get right, if he attempted to tell the birth months of every person in the population.*
C. The average number of birth months the carnival worker would get right, if he attempted to tell the birth months of every person in the population.

Interpretation Question 2
What should you do when

$$np_0(1-p_0) < 10?$$

A. Gather a larger sample
B. Do the hypothesis test in Minitab
C. Panic!
D. Either A or B

Worked examples, continued
2. According to OSHA, 75% of restaurant employees say that work stress has a negative impact on their personal lives. You want to know if the proportion is different for people who work in university cafeterias, so you survey 100 people. 68% of the people you survey say that work stress has a negative impact on their personal lives.

a. Write the null and alternative hypotheses for this test.
In this problem, we're interested in a percentage or a proportion, not a mean, so the parameter should be *p*. The null hypothesis should be that *p* equals some established value—*not* the value from the sample. So, the null hypothesis is

$$H_0\colon p = .75$$

where *p* is the proportion of people in the population of all *university cafeteria workers* who say that work stress has a negative impact on their personal lives.

(Notice that *p* isn't the proportion of all restaurant workers. That's because we already know that 75% of the population of restaurant workers have too much work stress. The goal of the hypothesis test is to find out whether cafeteria workers are different from this.)

Always write your alternative hypothesis before looking at the data. In this case, you should ignore the fact that 68% of the sample had too much work stress, and just look at the words, "You want to know if the proportion is **different for** people who work in university cafeterias." That tells us that we would be equally interested to find that university cafeterias are *more* stressful than restaurants as if they're *less* stressful. In other words, we have no particular reason to expect the difference to be in one direction versus the other.

The words "different for" tell us that we want the alternative hypothesis

$$H_0\colon p \neq .75$$

which gives a 2-sided test.

Interpretation Question 3
Why is the null hypothesis H_0: $p = .75$ instead of H_0: $p = .68$?

A. .75 is larger than .68.
B. .75 is the value that appears first in the problem.
C. .75 is the established value; .68 is the value from the sample.

Worked examples, continued
b. Find the p-value for this test.

In this sample, it is appropriate to use 1-prop ZTest to find the p-value. (You can check this yourself.) Before we do so, we need to find *x*, the number of "successes" (people who said yes) in our sample. We know that our sample size was *n* = 100 people, and the sample proportion of people who said yes was $\hat{p} = .68$. So to find *x*, we use

$$x = n\hat{p} = 100 * .68 = 68$$

Now we can use Stat -> Tests -> 1-prop ZTest with $p_0 = .75$ (the value in the null hypothesis), *x* = 68, *n* = 100, prop $\neq p_0$ (the alternative hypothesis).

This gives a p-value of .10597.

Interpretation Question 4
How can you find *x* when it's not given in the problem?

A. $\boldsymbol{x = n\hat{p}}$
B. $x = np$
C. $x = \frac{\hat{p}}{n}$
D. $x = \frac{p}{n}$

Worked examples, continued
c. Using $\alpha = .05$, write the conclusion in the context of the problem.

$$.10597 > .05$$

Our p-value is larger than the level of significance, so we should retain H_0. In the context of the problem, our conclusion is that there is not enough evidence to claim that the proportion of all cafeteria workers who are negatively affected by work stress is different from .75.

Interpretation Question 5
In your opinion, what is the most challenging part of hypothesis testing?

(Allow students to choose from Table A.4; then display options for extra practice on different topics.)

Table A.4. Optional Problems to Practice Aspects of Hypothesis Testing.

Challenging aspect	Optional extra practice
A. Writing the hypotheses	Section 10.1 # 15-21
B. Deciding whether a particular test (like 1-PropZTest) is appropriate	Section 10.2 Ex. 1, 2, 4
C. Finding the p-value	Section 10.2 # 7-11
D. Interpreting the results	Section 10.1 # 23-33
E. Some other part	

Chapter 11: 2-Sample T-Test

1. Can you "prime" your memory by thinking about an intellectual topic? A Dutch research study randomly assigned 400 students to two different groups. Group 1 was asked to spend 5 minutes thinking about what it would mean to be a professor. Group 2 was asked to think about soccer players. Then each student was asked 42 questions from Trivial Pursuit.

The 200 students in group 1 had a mean score of 23.4, with a standard deviation of 4.1. The 200 students in group 2 had a mean of 17.9, with a standard deviation of 3.9 (Sullivan 2013, p. 562).

a. Find a 95% confidence interval for $\mu_1 - \mu_2$.
Let μ_1 represent the mean Trivial Pursuit score for all people who are "primed" by thinking about being a professor. Let μ_2 represent the mean Trivial Pursuit score for all people who are "primed" by thinking about soccer players. (Notice that these are hypothetical populations. μ_1 represents what the mean *would* be if everyone in the Dutch population thought about being a professor and then tried to answer these 42 Trivial Pursuit questions.)

The variable is number of questions correct, which is quantitative. We are interested in comparing the population **mean** based on two samples, and there is no relationship or pairing between the samples. So, we should check the assumptions for a 2-sample t-interval. (They are the same assumptions as for a 2-sample t-test.)

- The sample size in each group is 200, which is much larger than 30. This tells us that the sample means will be approximately distributed (according to the Central Limit Theorem). So, the 2-sample t-interval is appropriate.

Use Stat -> Tests -> 2-Samp T Int, and enter the statistics, as shown in Table A.5.

Table A.5. Summary Statistics for Priming Example.

$\bar{x}_1 = 23.4$	$\bar{x}_2 = 17.9$
$Sx_1 = 4.1$	$Sx_2 = 3.9$
$n_1 = 200$	$\bar{n}_2 = 200$

We want a 95% confidence interval, so enter C-Level: .95.

We don't have any particular reason to think that the standard deviation would be the same in both populations, so we leave *Pooled* set to No.

This gives (4.7134, 6.2866) as the 95% confidence interval for $\mu_1 - \mu_2$.

b. Interpret the 95% confidence interval.
$\mu_1 - \mu_2$ is the difference between the mean scores in the populations. So, we are 95% confident that the mean score among people who are "primed" by thinking about being a professor is between 4.71 and 6.29 points **higher** than the mean score among people who are primed by thinking about soccer players.

Another interpretation is that if we repeated this experiment many times and computed a 95% confidence interval each time, about 95% of the intervals would contain the true difference between the means.

c. Is it plausible that the population means of the two groups are the same?
If $\mu_1 = \mu_2$, then we would have $\mu_1 - \mu_2 = 0$. But 0 is not in the confidence interval we found in part b. This indicates that 0 is not a plausible value for $\mu_1 - \mu_2$ (although it is *possible*). So, at a 95% confidence level, it is not plausible that $\mu_1 = \mu_2$.

This indicates that if we did a hypothesis test of

$$H_0: \mu_1 = \mu_2$$

versus

$$H_a: \mu_1 \neq \mu_2,$$

we would get a p-value less than .05.

Interpretation Question 1
Before making a 2-sample t-interval, we should check that both samples…

A. Come from normally distributed populations
B. Are large
C. Both A and B
D. Either A or B

Interpretation Question 2
In your opinion, is the difference between mean scores in this study of practical relevance?

A. Yes
B. No

(Ask students to explain their opinions. They may say that this is a large effect relative to the effort required for the treatment, so it is relevant; that a difference of 4.7-6.3 points might not be enough to help a person win the game—especially if they only get about half the questions right—so it is not relevant; or they may question the practical importance of trying to improve one's score in Trivial Pursuit at all. Affirm students' opinions; note that different researchers may disagree about the practical relevance of results, depending on the goals of the study. Point out that the goal of this study is to investigate the psychological phenomenon of priming, and that further studies in this area could investigate priming in more important contexts, such as statistics exams.)

2. To help traffic flow on freeways, engineers sometimes install "ramp metering" lights that require cars to wait for a short time before entering the freeway. But does ramp metering actually work?

Engineers in Minneapolis measured the speed of a systematic random sample of 15 cars on a freeway on a Monday at 6 pm with the ramp meters on. The following Monday, they turned off the ramp meters and took a second sample. There were no holidays or other factors that caused them to believe that the traffic on those two Mondays would be different.
The data are shown in Table A.6 (Sullivan 2013, p. 562).

Table A.6. Speed of cars on a freeway.

Ramp Meters On		
28	48	56
38	31	25
43	46	50
35	55	40
42	26	47

Ramp Meters Off		
24	26	42
34	37	31
47	38	17
29	23	40
37	52	41

Is there evidence that traffic flows faster when ramp meters are on? Use $\boldsymbol{\alpha = 0.10}$.
Speed of a vehicle is a quantitative variable, so it makes sense to use the mean speed to measure whether traffic flows faster. So our hypotheses are

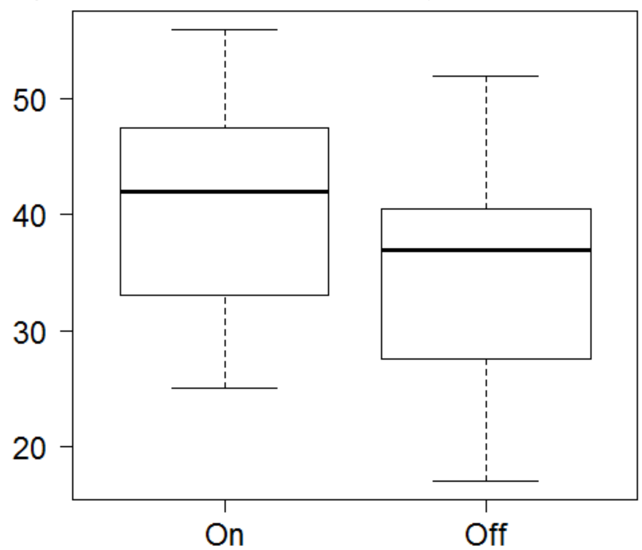
$$H_0: \mu_{On} = \mu_{Off}$$
$$H_a: \mu_{On} > \mu_{Off}$$

where μ_{On} is the mean speed of all cars on this freeway, on a Monday at 6 pm, when the ramp meter is on and μ_{Off} is the mean speed of all cars on this freeway, on a Monday at 6 pm, when the ramp meter is off.

We are interested in comparing two means, both estimated from samples, and there was no pairing between the samples. So, we should check the assumptions of a 2-sample t-test. The samples are small ($n_1 = 15$ and $n_2 = 15$), so we need to check whether the individual observations are approximately normal.

A boxplot (Figure A.12) shows that both samples are roughly symmetric, without outliers.

Figure A.12. Boxplot of Freeway Speeds.



Speeds are measured in miles per hour. "On" and "off" refer to whether ramp meters were turned on.

This suggests that a normal approximation is reasonable, so we can use a 2-sample t-test.

Here, we weren't told the sample mean or standard deviation. Fortunately, your calculator can calculate these for you as part of the 2-sample t-test. Start by entering the data for "Ramp Meters On" into L1, and the data for "Ramp Meters Off" into L2.

- Then press Stat -> Tests -> 2-Samp T Test.
- We entered the data into lists, so highlight Inpt: Data with the cursor and press enter. This tells the calculator that the input is the data; we're not going to type in summary statistics like $\bar{x}$.
- Set List1 to the name of the list where you stored the "Ramp Meters On" data (usually L1).

  - You can type the names of lists by pressing 2nd and then a number key. For example, 2nd -> 1 gives “L1”.
- Similarly, List2 should be the name of the list where you stored the “Ramp Meters Off” data (usually L2).
- Freq1 and Freq2 should be 1, because we observed 1 car with each speed in L1 and L2.
- Our alternative hypothesis was

$$H_a\colon \mu_{On} > \mu_{Off},$$

  so choose $\mu_1 : > \mu_2$.
- Leave Pooled: No.

This gives a p-value of .04886.

We have

$$.04886 < .10$$

So the p-value is less than α, the level of significance → Reject H_0.

Conclusion: There is enough evidence to claim that the mean speed is higher when the ramp meters are on.

Interpretation Question 3
How do we recognize when a 2-sample t-test might be appropriate?

A. Comparing two means, both estimated from samples
B. No pairing between samples
C. A and B
D. A or B

Interpretation Question 4
When should you choose Inpt: Data?

A. When the sample size is small
B. When you entered the data as lists
C. When comparing two means estimated from samples

Interpretation Question 5
In your opinion, what is the hardest part of this kind of problem?

A. Deciding whether a 2-sample t-test is appropriate
B. Writing the hypotheses
C. Computing the p-value
D. Interpreting the p-value
E. Some other part

(Affirm students’ experience that these problems can be challenging. Give students the opportunity to ask questions about the aspects that they find hardest.)

Chapter 12: Chi-squared Test of Independence

The manager of our city's Goodwill wants to know if the distribution of shoppers' incomes is different at her store compared to a Goodwill in a neighboring city. She surveys a random sample of 241 shoppers at our city's Goodwill and 218 shoppers at a neighboring city's Goodwill, and finds the data in Table A.7 (Adapted from Moore et al. 2009, p. 554):

Table A.7. Example Data from Goodwill.

Income	Our City	Neighboring	**Total**
< $10,000	70	62	**132**
$10,000 to $19,999	52	63	**115**
$20,000 to $24,999	69	50	**119**
$25,000 to $34,999	22	19	**41**
$35,000 to $44,999	27	23	**50**
$45,000 to $54,999	1	0	**1**
$55,000 to $64,999	0	1	**1**
Total	**241**	**218**	**459**

a. Determine what kind of test to check assumptions for. State the hypotheses.

"City" is a categorical variable with 2 categories. "Income" could potentially be a quantitative variable, but in this case the data were collected using a multiple-choice survey, so in our data, income is a categorical variable with 7 categories. We have 2 categorical variables, and one of them has more than 2 categories.

We want to know whether the distribution of shoppers' incomes is different between the two stores. This is the same as wanting to know whether income is **associated** with the city. So, we should check assumptions for a χ^2 Test of Independence.

The hypotheses are
H_0: Among all shoppers at Goodwill in the two cities, a person's income category is *independent* of which city they shop in.
H_a: Among all shoppers at Goodwill in the two cities, a person's income category is *associated with* which city they shop in.

Another way we could phrase this would be
H_0: Among all shoppers at Goodwill in in the two cities, a person's income category is *unrelated* to which city they shop in. Or, there is *no difference* in the distribution (probabilities) of income categories, based on which city people shop in.
H_a: The distribution of income categories is *different* among all shoppers at Goodwill in our city, compared to all shoppers at Goodwill in the neighboring city.

Interpretation Question 1
Why do we treat Income as a categorical variable in this example?

A. The data were presented in a table.
B. The other variable, City, is categorical.
C. We only have information about income ranges, not specific numerical values of people's incomes.

Interpretation Question 2
If income is *associated* with city, it means that…

A. Income and city data were gathered on the same sample of people.
B. The percentage of people in a certain income group is different for our city vs. the neighboring city.
C. Within a particular city, not all income groups include the same percentages of people.

b. Check the assumptions for the test. If applicable, combine categories.

Both of the assumptions for the χ^2 Test of Independence relate to the expected counts in each cell of the table. So, we need to compute the expected counts, if H_0 were true.

The formula for the expected count in a particular cell is

$$Expected\ count = \frac{(row\ sum) * (column\ sum)}{total\ sample\ size}$$

(**Where does this formula come from?** It's based on the formula for the expected value of a binomial random variable, from section 6.2! It also relies on the method for computing probabilities from a table that we saw in sections 4.4 and 5.4. Because it's computed assuming that H_0 is true—so income and city are independent—it also uses the Multiplication Rule for Independent Events from section 5.3. Email me if you're curious about the details!)

For example, the expected count in the "Our City, < \$10,000" cell is $\frac{132*241}{459} = 69.3072$. The rest of the expected values are as shown in Table A.8.

Table A.8. Expected Counts.

Income	Our City	Neighboring
< $10,000	69.3072	62.6928
$10,000 to $19,999	60.3813	54.6187
$20,000 to $24,999	62.4815	56.5185
$25,000 to $34,999	21.5272	19.4728
$35,000 to $44,999	26.2527	23.7473
$45,000 to $54,999	0.5251	0.4749
$55,000 to $64,999	0.5251	0.4749

- Just like in the χ^2 Goodness of Fit test, you should not round off the expected counts.
- **Hint to help check your work:** The row sums and column sums of the table of expected counts should be the same as the row sums and column sums of the table of observed data (possibly with a small amount of rounding error).

Checking assumptions: The χ^2 Test of Independence assumes that

- All expected counts are ≥ 1, and
- Not more than 20% of the cells have expected counts < 5.

Neither of these assumptions are true in this problem. The last 2 rows contain cells with expected counts < 1. Also, 4 out of the 14 cells, or 28.57% of the cells, have expected counts < 5.

Fortunately, we can fix the problem so the assumptions hold, by combining categories. When the categories have an order, like the income categories here, it usually makes the most sense to combine adjacent categories. For example, we could combine the category "$45,000 to $54,999" with "$55,000 to $64,999" to get one category of "$45,000 to $64,999".

When we combine 2 categories, we sum the expected counts in each category. So the new expected count matrix would be as shown in Table A.9.

Table A.9. Expected Counts With Last Two Categories Combined.

Income	Our City	Neighboring
< $10,000	69.3072	62.6928
$10,000 to $19,999	60.3813	54.6187
$20,000 to $24,999	62.4815	56.5185
$25,000 to $34,999	21.5272	19.4728
$35,000 to $44,999	26.2527	23.7473
$45,000 to $64,999	**1.0502**	**.9498**

But we still have one expected count < 1. So, combine the categories "$35,000 to $44,999" and "$45,000 to $64,999" to get the expected count matrix shown in Table A.10.

Table A.10. Expected Counts With Last Three Categories Combined.

Income	Our City	Neighboring
< $10,000	69.3072	62.6928
$10,000 to $19,999	60.3813	54.6187
$20,000 to $24,999	62.4815	56.5185
$25,000 to $34,999	21.5272	19.4728
$35,000 to $64,999	**27.3029**	**24.6971**

Now both assumptions are satisfied, so we can proceed with the χ^2 Test of Independence. We need to make the categories in the table of observed data match the categories in the table of expected counts by adding them together, as shown in Table A.11.

Table A.11. Observed Counts With Last Three Categories Combined.

Income	Our City	Neighboring
< $10,000	70	62
$10,000 to $19,999	52	63
$20,000 to $24,999	69	50
$25,000 to $34,999	22	19
$35,000 to $64,999	28	24

c. Find the degrees of freedom, test statistic, and p-value for the test.

The number of degrees of freedom for the χ^2 Test of Independence is

$$(number\ of\ rows - 1) * (number\ of\ columns - 1).$$

We determine degrees of freedom based on the table *after* categories are combined, if necessary. So, the test statistic for this problem has $(5 - 1) * (2 - 1) = 4 * 1 = 4$ degrees of freedom.

The test statistic can be found either by hand or on a calculator. Here's how to do it on a calculator:

- Press 2nd -> Matrix (same key as x^{-1}) -> Edit -> [A]. Enter the observed data table.
- Press 2nd -> Matrix (same key as x^{-1}) -> Edit -> [B]. Enter the expected counts table.
- Go to Stat -> Tests -> χ^2-Test.
    - Enter Observed: [A]
    - Expected: [B]

- Calculate.

The test statistic is $\chi^2 = 3.9553$. The p-value is 0.4121. We also see that df = 4, which agrees with the answer we found above.

Interpretation Question 3

Where does the formula

$$Expected\ count = \frac{(row\ sum) * (column\ sum)}{total\ sample\ size}$$

come from?

A. *Methods that we know about from chapters 4, 5, and 6.*
B. It is totally unrelated to anything we've learned before.

Interpretation Question 4

How did we decide which categories to combine?

A. Combine categories with the smallest expected counts.
B. Combine the categories that maximize the test statistic.
C. *When the categories are ordered, combine adjacent categories.*

d. State your conclusion in the context of the problem. Use $\alpha = 0.05$.

The p-value is greater than α, so we should retain H_0. There is **not** enough evidence to claim that among all shoppers at Goodwill in the two cities, a person's income category is associated with which city they shop in.

In other words, there is not enough evidence to claim that the distribution of incomes is different among all shoppers at Goodwill in our city, compared to all shoppers at Goodwill in the neighboring city.

Note: It would be **incorrect** to say, "There is enough evidence to claim that income is independent of city," or "The distribution of incomes is the same between the two cities." This would be accepting H_0, which we never do. The most we can say is that it is *plausible* that income could be independent of the city. It is also possible that if we gathered a larger sample size, we would find more evidence to claim that income and city were associated.

Interpretation Question 5
In your opinion, what is the hardest part about using the χ^2 Test of Independence?

A. Recognizing when to apply it
B. Computing expected counts
C. Checking assumptions and combining categories
D. Interpreting the results

(Affirm students' experience that these problems can be challenging. Give students the opportunity to ask questions about the aspects that they find hardest.)