MATH 2510: WEEKS 09 AND 10

JOSEPH TIMMER (EDITED BY COLTON GRAINGER)
OCTOBER 24, 2019

This was prepared for University of Colorado's MATH 2510 (Introduction to Statistics) in Fall 2019. Please send questions, comments, complaints, and corrections to colton.grainger@colorado.edu. Any mistakes are my own. This material corresponds to Chapter 7 of *Understandable Statistics* [Brase and Brase, 2018].

Contents

0.

Lecture Outlines

Confidence Intervals When we can't collect a measurement for every member of a population, how can we determine a population mean?

(1) Without a full set of population data, we can never be 100% certain that we know the population mean, but with certain restrictions applied, we can use the data from a random sample to estimate a population parameter.

(2) Even with a large random sample, the value of the sample mean is usually not exactly equal to the population mean. But, according to the Central Limit Theorem, we can have some expectations on how likely it is that the sample mean falls within a certain interval around the population mean. A confidence interval is an interpretation of precisely this application.

(3) What criterion is required to apply the Central Limit Theorem?

(4) The idea is that we start with a sample statistic (called a point estimate). We then create a margin of error around that point estimate which yields an interval of values that is asserted as one that contains the population parameter (at least with some high level of, but not 100%, certainty)

  (a) So, this looks like (point – error, point + error) or a guess ±error.

  (b) The size of the error depends on what level of certainty we want to assert.

  (c) The most prevalent example of a confidence interval are during elections and predicting election results.

(5) Keep in mind that although we will be asserting that we have an interval that contains the population parameter, there is no indication where within the interval we expect it to lie.

7.1 Estimating μ when σ is known

(1) The first look at confidence intervals assumes that we know σ. (This might seem a bit contrived, because why would we know σ if we don't know μ.)

(2) The TI `Z-INTERVAL` function will compute the confidence interval, either by entering the actual list of sample data or by entering the sample mean. For that reason some of the details of the formulation developed in the reading are not so critical, but understanding what the function does and how to interpret the result is critical.

(3) When reading, pay special attention to
- The formula for E in this case, as it should make sense why this is the correct formula.
- The definition of z_c, as this should also be something that you already know how to compute.
- The interpretation of a confidence interval, as it is easy to misinterpret what the interval found means.

7.2 Estimating μ when σ is unknown

(1) In this case, a seemingly more common case, we don't know σ. So, we must estimate the value of σ too. This translates into a slightly larger margin of error to compensate for the potential error in our guess of σ.

(2) Rather than the standard normal distribution, we use the Student's t-distributions. There is a slightly different distribution for each sample size, but they are all bell-shaped. As n gets larger, the t-distributions approach the standard normal distribution.

(3) The formulation and interpretation of a confidence interval in this case is very similar to that when σ is known, just the distributions from which the critical values are determined has changed.

(4) The `T-INTERVAL` function the TI can do the work for us.

(5) When reading, pay special attention to
- The formula for E (which should look VERY similar to that from section 7.1).
- The formula for degrees of freedom.

(6) If your calculator does not already have an `InvT` function, there is a screencast available about how to create a program in your calculator that does the job.

Internet Visual for Confidence Intervals

(1) What is this whole confidence interval and confidence level thing about? This applet could be effective in confirming what idea of the c-level, also as a way to compare how the size of the interval changes with changes to the variables.

(2) The website is http://www.rossmanchance.com/applets/ConfSim.html

(3) Start with Means, Normal, z with σ.
- Show 1 sample
- Show 10 samples
- Show 100 samples, maybe a few times to show how the number of "good" confidence intervals can vary.
- Change Conf level
- Change n

(4) Reset and look at Means, Normal, t. Note how they can have different lengths. (this is sometimes subtle) Why?

7.3 Confidence Interval for a Proportion

(1) This time, rather than a mean, we are estimating a population proportion, like "What percentage of all college students change their major at least once in their first four years?" which is different than "What is the average number of times a college student changes their major within their first four years?"

(2) The calculator function is `1-PropZInt`.

(3) In the reading watch out for
- The requirement on the sample size, it is more complicated than just $n \geq 30$.
- The formula for E.
- The formulas for finding sample size.
- Interpreting poll results.

7.4 Confidence Intervals for Differences

(1) As our last look at confidence intervals, we look at differences (between two means or between two proportions), as a way to tell if two populations are different.

(2) This subsection refers only to independent samples, but check out the reading for the definitions as it will matter later (when we do tests).

(3) `2-SampZInt`, `2-SampTInt`, and `2-PropZInt` are the calculator functions.

(4) In the reading watch out for
- The degrees of freedom for Tint
- The criteria on sample size (it is again different for p than for μ).
- The interpretation of the confidence interval
    - When the interval contains only negative values
    - When the interval contains only positive values
    - When the interval contains both positive and negative values.

Hypothesis Testing In Chapter 7, we estimated the value of population parameters (mean and proportion) using confidence intervals. Another method of statistical inference is to make decisions concerning the value of a population parameter, which we do in Chapter 8 with hypothesis testing.

(1) Suppose that you roll a regular six-sided die 600 times. About how many times would you expect to see a 4 rolled within those 600?
    (a) If you saw 105 rolls that were 4, would this be surprising...enough to question the fairness of the die?
    (b) What if you saw 595 rolls that were 4, would this be surprising...enough to question the fairness of the die?
    (c) Where would you draw the line between "not so surprising" and "surprising"?

(2) The basic idea in hypothesis testing is to start with an assumption of what "should" happen and to draw a line on what extreme outcomes would be "surprising". If the random sample indicates a "surprising" result, we have evidence to abandon our assumption...if the ransom sample indicates a "not so surprising" result, we do not have adequate evidence to abandon our assumption and we must stick with it.

(3) Note, as with the case of the rolls of the die, the result of the random sample may be very "surprising" (595 of our 600 rolls were 4), but it will never serve as PROOF that our assumption is wrong (as it's possible that this 595/600 happens with a completely fair die).

1.

Estimating the mean μ when the standard deviation σ is known

(1) Suppose that you read in a local newspaper that 45 officials in student services at CU earned an average $\bar{x} = \$50,340$ each year.

   (a) Assume that $\sigma = \$16,920$ for salaries of college officials in student services. Find a 90% confidence interval for the population mean salaries of such personnel. What is the margin of error?

   Using the `ZInterval` function, with $\sigma = 16920$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.90$, the confidence interval is ($46191, \$54489$).

   This implies that the margin of error is $E = 54489 - 50340 = \$4149$.

   The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.645 \dfrac{16920}{\sqrt{45}} = 4149$.

   (b) Assume that $\sigma = \$10,780$ for salaries of college officials in student services. Find a 90% confidence interval for the population mean salaries of such personnel. What is the margin of error?

   Using the `ZInterval` function, with $\sigma = 10780$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.90$, the confidence interval is ($47697, \$52983$).

   This implies that the margin of error is $E = 52983 - 50340 = \$2643$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.645 \dfrac{10780}{\sqrt{45}} = 2643$.

   (c) Assume that $\sigma = \$4830$ for salaries of college officials in student services. Find a 90% confidence interval for the population mean salaries of such personnel. What is the margin of error?

   Using the `ZInterval` function, with $\sigma = 4830$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.90$, the confidence interval is ($49156, \$51524$).

   This implies that the margin of error is $E = 51524 - 50340 = \$1184$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.645 \dfrac{4830}{\sqrt{45}} = 1184$.

   (d) What does this example illustrate about the effect of the size of σ on the length of the confidence interval? Why does this make sense?

   The smaller the value of σ, the smaller the value of the margin of error and the shorter the length of the confidence interval. This makes sense because a lower σ implies that the x distribution is less spread from the mean μ. So, the likelihood that the mean of sample selected at random is a good estimate of the population mean is greater, thereby making the need for error in the confidence interval smaller.

(2) Suppose that you read in a local newspaper that the annual salary of administrators at CU is $\bar{x} = \$50,340$. Assume that σ is known to be $\$18,490$ for college administrators salaries.

   (a) Suppose that the $\bar{x} = \$50,340$ is based on a random sample of $n = 36$ administrators. Find a 90% confidence interval for the population mean annual salary of local college administrators. What is the margin of error?

     Using the `ZInterval` function, with $\sigma = 18490$, $\bar{x} = 50340$, $n = 36$, and C-level $= 0.90$, the confidence interval is ($\$45271, \$55409$).

     This implies that the margin of error is $E = 55409 - 50340 = \$5069$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.64485 \dfrac{18490}{\sqrt{36}} = 5069$.

   (b) Suppose that the $\bar{x} = \$50,340$ is based on a random sample of $n = 64$ administrators. Find a 90% confidence interval for the population mean annual salary of local college administrators. What is the margin of error?

     Using the `ZInterval` function, with $\sigma = 18490$, $\bar{x} = 50340$, $n = 64$, and C-level $= 0.90$, the confidence interval is ($\$46538, \$54142$).

     This implies that the margin of error is $E = 54142 - 50340 = \$3802$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.64485 \dfrac{18490}{\sqrt{64}} = 3802$.

   (c) Suppose that the $\bar{x} = \$50,340$ is based on a random sample of $n = 121$ administrators. Find a 90% confidence interval for the population mean annual salary of local college administrators. What is the margin of error?

     Using the `ZInterval` function, with $\sigma = 18490$, $\bar{x} = 50340$, $n = 121$, and C-level $= 0.90$, the confidence interval is ($\$47575, \$53105$).

     This implies that the margin of error is $E = 53105 - 50340 = \$2765$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.64485 \dfrac{18490}{\sqrt{121}} = 2765$.

   (d) What does this example illustrate about the effect of the sample size on the length of the confidence interval? Why does this make sense?

     The larger the sample size n, the smaller the value of the margin of error and the shorter the length of the confidence interval. This makes sense because a larger sample increases the likelihood that the sample mean is a good estimate of the population mean, thereby making the need for error in the confidence interval smaller.

   (e) What sample size n is necessary for a 90% confidence interval with maximal margin of error $E = 1000$ for the mean annual salary?

     Solving the margin of error formula $E = z_c \dfrac{\sigma}{\sqrt{n}}$ for n yields $n = \left(\dfrac{z_c \sigma}{E} \right)^2 = \left(\dfrac{(1.64485)(18490)}{1000} \right)^2 = 924.967$. So, a sample size of $n = 925$ is necessary.

(3) Suppose that you read in a local newspaper that 45 officials in student services at CU earned an average $\bar{x} = \$50,340$ each year. Assume that σ is known to be $\$16,920$ for salaries of college officials in student services.

(a) Find a 90% confidence interval for the population mean salaries of such personnel. What is the margin of error?

Using the `ZInterval` function, with $\sigma = 16920$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.90$, the confidence interval is ($\$46191, \$54489$).

This implies that the margin of error is $E = 54489 - 50340 = \$4149$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.645 \dfrac{16920}{\sqrt{45}} = 4149$.

(b) Find a 95% confidence interval for the population mean salaries of such personnel. What is the margin of error?

Using the `ZInterval` function, with $\sigma = 16,920$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.95$, the confidence interval is ($\$45396, \$55284$).

This implies that the margin of error is $E = 55284 - 50340 = \$4944$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 1.96 \dfrac{16920}{\sqrt{45}} = 4944$.

(c) Find a 99% confidence interval for the population mean salaries of such personnel. What is the margin of error?

Using the `ZInterval` function, with $\sigma = 16,920$, $\bar{x} = 50340$, $n = 45$, and C-level $= 0.99$, the confidence interval is ($\$43843, \$56837$).

This implies that the margin of error is $E = 56837 - 50340 = \$6497$. The margin of error could also be computed using $E = z_c \dfrac{\sigma}{\sqrt{n}} = 2.576 \dfrac{16,920}{\sqrt{45}} = 6497$.

(d) What does this example illustrate about the effect of the level of confidence on the length of the confidence interval? Why does this make sense?

The lower the level of confidence, the smaller the value of the margin of error and the shorter the length of the confidence interval. This makes sense because with the same expected sampling distribution (based on the same μ and the same σ for the population) to be more confident that our confidence interval will capture the actual mean, this interval must span a wider range.

2.

Estimating the mean μ when the standard deviation σ is unknown

(1) Suppose the random variable x has a mound-shaped, symmetric distribution. Consider a random sample of size $n = 21$, sample mean $\bar{x} = 45.2$, and sample standard deviation $s = 5.3$.

   (a) Use the Student's t distribution to compute the 95% confidence interval.
TInterval: Stats with $\bar{x} = 45.2$, $S_x = 5.3$, $n = 21$, and C-Level $= .95$ yields $(42.787, 47.613)$.

   (b) Now, assume that $\sigma = s$ (so that σ is now known) and use the standard normal distribution (with z_c) to compute the 95% confidence interval.
ZInterval: Stats with $\sigma = 5.3$, $\bar{x} = 45.2$, $n = 21$, and C-Level $= .95$ yields $(42.933, 47.467)$.

   (c) How do the intervals compare? Is one longer than the other? Why does this make sense?
The TInterval yields a longer interval. This makes sense because with the ZInterval more is known about the population (and therefore the sampling distribution), namely σ. When we know less, we need to allow for more potential error in our results accurately representing the population and as a result we stretch that maximal margin of error.

   (d) Now repeat these same steps with $n = 200$. How do the intervals compare now and why should we expect this?
TInterval: Stats with $\bar{x} = 45.2$, $S_x = 5.3$, $n = 21$, and C-Level $= .95$ yields $(44.461, 45.939)$.
ZInterval: Stats with $\sigma = 5.3$, $\bar{x} = 45.2$, $n = 21$, and C-Level $= .95$ yields $(44.465, 45.935)$.
The intervals are much closer in length. We should expect this because the sample size is larger. As the sample size increases, the Student's t distribution approaches the normal curve. So, the results will be more similar.

(2) What percentage of hospitals provide at least some charity care? The following problem is based on information taken from *State Health Care Data: Utilization, Spending, and Characteristics* (American Medical Association). Based on a random sample of hospital reports from the eastern states, the following information was obtained (units in percentage of hospitals providing at least some charity care):

$$57.1,\ 56.2,\ 53.0,\ 66.1,\ 59.0,\ 64.7,\ 70.1,\ 64.7,\ 53.5,\ 78.2$$

   (a) Is it more appropriate to use the Student's t distribution or the standard normal distribution to determine a confidence interval for this data? Assume the percentage of hospital's follows an approximately normal distribution.
Because the standard deviation of the population is not know, the t distribution is more appropriate. If the sample size is large, the results may not vary much, but the t distribution is still more appropriate.

   (b) Using the distribution you determined was more appropriate in part (a), find a 90% confidence interval for the population average μ of the percentage of hospitals providing at least some charity care.
Entering the above data into list L_1, the TInterval : Data option can be used with List $= L_1$, Freq $= 1$, C-Level $= .90$. The result is $(57.612, 66.908)$.

(3) With some interest in running your own candy store and a decent credit rating, you can probably get a bank loan for franchises such as Candy Express, The Fudge Company, Karmel Corn, and Rocky Mountain Chocolate Factory. Startup costs (in thousands of dollars) for a random sample of candy stores are given below (Source: *Entrepreneur Magazine*, Vol.23, No.10). Assume start up costs follow an approximately normal distribution.

$$95, \ 173, \ 129, \ 95, \ 75, \ 94, \ 116, \ 100, \ 85$$

(a) Find a 90% confidence interval for the population average startup costs μ for candy store franchises. Entering the above data into list L_1, the `TInterval :  Data` option can be used with `List` $= L_1$, `Freq` $= 1$, `C-Level` $= .90$. The result is $(88.639, 125.14)$.

(b) What does this confidence interval mean in the context of the problem? We are 90% confident that the interval \$88,639 to \$125,140 is one that contains the average startup cost for all candy store franchises.

(4) From a random sample of $n = 40$ current major league baseball players, a 90% confidence interval for the population mean μ of home run percentages for all current major league baseball players was determined to be 1.93 to 2.65.

(a) What does this imply that the sample mean $\bar{x}$ of home run percentages was? What is E in this case? The sample mean $\bar{x}$ is the center of the confidence interval. So, that implies $\bar{x} = 2.29$. The maximal margin of error E is the distance from $\bar{x}$ to either endpoint of the interval, or equivalently, half the length of the interval. So, that implies $E = 0.36$.

(b) Determine a 99% confidence interval for the population mean μ of home run percentages. (HINT: First, use E to find the value of s, then use n and $\bar{x}$ along with the s.) We know the value of E is 0.36 from above. We also know by formula that $E = t_{0.90} \cdot \frac{s}{\sqrt{n}}$. Note that $t_{0.90} = $ `InvT(0.95, 39)` $= 1.684875$. So, $0.36 = 1.684875 \cdot \frac{s}{\sqrt{40}}$ which implies $1.3513405 = s$. Then `TInterval:  Stats` with $\bar{x} = 2.29$, $S_x = 1.3513405$, $n = 40$, and `C-Level` $= .99$, yields $(1.7114, 2.8686)$. The 99% confidence is longer than the 90% confidence, as we would expect. To make a more confident statement with the same data, we need to stretch that interval out to allow for a bit more error in our sample.

3.

Estimating p in the Binomial Distribution

---

(1) Consider $n = 200$ binomial trials with $r = 80$ successes.

    (a) Is it appropriate to use a normal distribution to approximate the $\hat{p}$ distribution? Explain.

        We estimate p by the sample point estimate $\hat{p} = \frac{r}{n} = \frac{80}{200} = 0.4$. Since $n\hat{p} = (200)(0.4) = 80$ and $n\hat{q} = (200)(0.6) = 120$ are both greater than 5, it is appropriate to use a normal approximation to the binomial.

    (b) Find a 95% confidence interval for the population proportion of successes p.

        Using `1-PropZInt` with $x = 80$, $n = 200$, and `C-Level` $= .95$, we get $(0.3321, 0.4679)$. That is, we are 95% confident that the interval from 33.21% to 46.79% is one that contains the true population proportion.

(2) In a random sample of 519 judges, it was found that 285 introverts.

    (a) Let p represent the proportion of all judges who are introverts. Find a point estimate for p. We estimate p by the sample point estimate $\hat{p} = \frac{r}{n} = \frac{285}{519} \approx 0.5491329$.

    (b) Are the conditions $n\hat{p} > 5$ and $n\hat{q} > 5$ satisfied in this problem?

        $n\hat{p} \approx 285$ and $n\hat{q} \approx 234$, so the conditions are satisfied.

    (c) Find a 99% confidence interval for p. Give a brief interpretation of this interval.

        Using `1-PropZInt` with $x = 285$, $n = 519$, and `C-Level` $= .99$, we get $(0.49287, 0.60539)$.

        That is, we are 99% confident that the interval from 49.287% to 60.539% is one that contains the true population proportion.

(3) A *New York Times*/CBS poll asked the question, "What do you think is the most important problem facing this country today?" Nineteen percent of the respondents answered, "Crime and violence." The margin of sampling error was plus or minus 3 percentage points. Following the convention that the margin of error is based on a 95% confidence interval, find a 95% confidence interval for the percentage of the entire population that would respond "Crime and violence" to the question asked by the pollsters.

    Since we are following the convention that the margin of error creates a 95% confidence interval, that interval would simply be $19 - 3 = 16$ to $19 + 3 = 22$. So, we are 95% confident that the interval 16% to 22% is one that contains the true population percentage.

(4) In a marketing survey, a random sample of 1001 supermarket shoppers revealed that 273 always stock up on an item when they find that item at a real bargain price.

    (a) Let p represent the proportion of all supermarket shoppers who always stock up on an item when they find a real bargain. Find point estimate for p.

        We estimate p by the sample point estimate $\hat{p} = \frac{r}{n} = \frac{273}{1001} \approx 0.2727272$.

    (b) Find a 95% confidence interval for p. Give a brief explanation of this interval.

        Using `1-PropZInt` with $x = 273$, $n = 1001$, and `C-Level` $= .95$, we get $(0.24514, 0.30032)$. That is, we are 95% confident that the interval from 24.514% to 30.032% is one that contains the true population proportion.

(5) The National Council of Small Businesses in interested in the proportion of small businesses that declared Chapter 11 bankruptcy last year. Since there are so many small businesses, the National Council intends to estimate the proportion from a random sample. Let p be the proportion of small businesses that declared Chapter 11 bankruptcy last year.

    (a) If no preliminary sample is take to estimate p, how large a sample is necessary to be 95% sure that a point estimate $\hat{p}$ will be within a distance of 0.10 from p?

        With no preliminary estimate, we essentially use the "safe" guess that $p = \frac{1}{2}$. So, by the sample size formula, $n = \hat{p}(1-\hat{p})\left(\frac{z_c}{E}\right)^2 = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{\texttt{InvNorm}(0.975,0,1)}{0.01}\right)^2 = 96.03647066$. This implies that the sample size must be at least 97 small businesses.

    (b) On the other hand, suppose that in a preliminary random sample of 38 small businesses, it was found that six had declared Chapter 11 bankruptcy. How many small businesses should be included in a sample to be 95% sure that a point estimate $\hat{p}$ will be within a distance of 0.10 from p?

        Since we have a preliminary estimate, we can use it instead. So, by the sample size formula, $n = \hat{p}(1-\hat{p})\left(\frac{z_c}{E}\right)^2 = \left(\frac{6}{38}\right)\left(\frac{32}{38}\right)\left(\frac{\texttt{InvNorm}(0.975,0,1)}{0.01}\right)^2 = 51.07756888$. This implies that a sample size must be at least 52 total small businesses, which would mean 14 more over the 38.

4.
Estimating the Difference Between Two Means or Proportions
---

(1) If a 90% confidence interval for the difference of means $\mu_1 - \mu_2$ contains all positive values, what can we conclude about the relationship between μ_1 and μ_2? What if the interval contains all negative values? What if it contains a mix of positive and negative values?

When the interval contains all positive values, then we are asserting that we are 90% certain the true difference lies in that interval...so, is positive...and therefore $\mu_1 > \mu_2$.

On the other hand, when the interval contains all negative values, then we are asserting that we are 90% certain the true difference lies in that interval...so, is negative...and therefore $\mu_2 > \mu_1$.

When the confidence contains both positive and negative, we cannot assert that there is any difference between μ_1 and μ_2.

(2) Independent random samples of professional football and basketball players gave the following information (references: *Sports Encyclopedia of Pro Football* and *Official NBA Basketball Encyclopedia*). Assume that the weight distributions are mound-shaped and symmetric.

Weights (in lb) of pro football players:
245, 262, 255, 251, 244, 276, 240, 265, 257, 252, 282, 256, 250, 264, 270, 275, 245, 275, 253, 265, 270

Weights (in lb) of pro basketball players:
205, 200, 220, 210, 191, 215, 221, 216, 228, 207, 225, 208, 195, 191, 207, 196, 181, 193, 201

(a) Let μ_1 be the population mean for the weights of professional football players and let μ_2 be the population mean for the weights of professional basketball players. Which distribution (standard normal or Student's t) should we use to compute a confidence interval for $\mu_1 - \mu_2$? Explain.

Because we do not know the population standard deviation for at least one of the groups, we should use the Student's t distribution.

(b) Find a 99% confidence interval for $\mu_1 - \mu_2$.

Using `2-SampTInt` with the `Inpt:  Data` option and `List1:`L_1 (for the 21 football players), `List2:`L_2 (for the 19 basketball players), `Freq1` and `Freq2` both 1, `C-Level:  .99`, and `Pooled:  No`, the interval is $(43.076, 64.583)$.

(c) Does the interval consist of numbers that are all positive? all negative? of different signs? At the 99% confidence level, do professional football players tend to have a higher population mean weight than professional basketball players?

Because the interval contains only positive values, we are 99% confident that the population mean weight of professional football players is greater than the population mean weight of professional basketball players.

(3) Most married couples have two or three personality preferences in common. Myers used a random sample of 375 married couples and found that 132 had three preferences in common. Another random sample of 571 couples showed that 217 had two personality preferences in common. Let p_1 be the population proportion of all married couples who have three personality preferences in common. Let p_2 be the population proportion of all married couples who have two personality preferences in common.

(a) Can a normal distribution be used to approximate the $\hat{p}_1 - \hat{p}_2$ distribution? Explain.

Since $n_1\hat{p}_1 = (375)(\frac{132}{375}) = 132$, $n_1\hat{q}_1 = (375)(\frac{243}{375}) = 243$, $n_2\hat{p}_2 = (571)(\frac{217}{571}) = 217$, and $n_2\hat{q}_2 = (571)(\frac{354}{571}) = 354$ are all greater than 5, we can use the normal approximation.

(b) Find a 90% confidence interval for $p_1 - p_2$.

Using 2-PropZInt with x1:132, n1:375, x2:217, n2:571, and C-Level:.90, the interval is:

$$(-0.0806, 0.02452).$$

(c) Does the interval consist of numbers that are all positive? all negative? of different signs? What does this tell you about the proportion of married couples with three personality preferences in common compared with the proportion of couples with two preferences in common (at the 90% confidence level)? Because the interval contains a mix of positive and negative values. So at the 90% confidence level, there does not appear to be difference in the population proportion of married couples who have three personality preferences in common and those which have only two.

(4) "Parental Sensitivity to Infant Cues: Similarities and Difference Between Mothers and Fathers" by MV Graham reports a study of parental empathy for sensitivity cues and baby temperament (higher scores means more empathy). Let x_1 be a random variable that represents the score of a mother on an empathy test. Let x_2 be the empathy score of a father. A random sample of 32 mothers gave a sample mean of $\bar{x}_1 = 69.44$. Another random sample of 32 fathers gave $\bar{x}_2 = 59$. Assume that $\sigma_1 = 11.69$ and $\sigma_2 = 11.60$.

(a) Which distribution, normal or Student's t, do we use to approximate the $x_1 - x_2$ distribution? Explain. Because we are given the population standard deviations σ_1 and σ_2 and the sample sizes are large enough (without knowing the shape of the x_1 and x_2 distributions), we can use the normal distribution.

(b) Let μ_1 be the population mean of x_1 and let μ_2 be the population mean of x_2. Find a 99% confidence interval for $\mu_1 - \mu_2$. What can we conclude?

Using 2-SampZInt with the Inpt: Stats option and $\sigma 1 : 11.69$, $\sigma 2 : 11.60$, $\bar{x}1 : 69.44$, $n1 : 32$, $\bar{x}2 : 59$, $n2 : 32$, and C-Level:.99, the interval is $(2.9411, 17.939)$.

Because the interval contains only positive values the mothers' mean score appears higher than the fathers' at the 99% confidence level.