1.

## Estimating the mean $\mu$ when the standard deviation $\sigma$ 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 $\sigma$ on the length of the confidence interval? Why does this make sense?
The smaller the value of $\sigma$, the smaller the value of the margin of error and the shorter the length of the confidence interval. This makes sense because a lower $\sigma$ implies that the $x$ distribution is less spread from the mean $\mu$. 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 $\sigma$ 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 $\sigma$ 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 $\mu$ and the same $\sigma$ 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 $\mu$ when the standard deviation $\sigma$ 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 $\sigma$ 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 $\sigma$. 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 $\mu$ 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 $\mu$ 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 $\mu$ 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 $\mu$ 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.