CHAPTER 8: ESTIMATING WITH CONFIDENCE

Part 1: Confidence Intervals, The Basics

[redacted]

If we had to give a single number to estimate the value of the mean of a distribution, what would it be?

Point estimator and point estimate:

Open with Mystery Means Activity pg 476

Confidence intervals are what we use when we don't know the population parameter.

Point estimator and point estimate: point estimator is a statistic that provides an estimate of a population parameter. The value of that statistic from a sample is called a point estimate
i.e. estimator = mean, estimator = 240.80
**An ideal point estimator will have no bias and low variability

[redacted]

From Batteries to Smoking

In each of the following settings, determine the point estimator you would use and calculate the value of the point estimate.

a) What proportion p of US high school students smoke? The 2011 Youth Risk Behavioral Survey questioned a random sample of 15, 425 students in grades 9-12. Of these 2792 said they had smoked cigarettes at least one day in the past month.

b) Quality control inspectors want to estimate the mean lifetime of the AA batteries produced in an hour at a factory. They select a random sample of 50 batteries during each hour of production and then drain them under conditions that mimic normal use. Here are the lifetimes in hours of the batteries from one such sample.

16.73	15.60	16.31	17.57	16.14	17.28	16.67	17.28	17.27	17.50	15.46	16.50	16.19
15.59	17.54	16.46	15.63	16.82	17.16	16.62	16.71	16.69	17.98	16.38	17.80	16.61
15.99	15.64	17.20	17.24	16.68	16.55	17.48	15.58	17.61	15.98	16.99	16.93	16.01
17.54	17.41	16.91	16.60	16.78	15.76	17.31	16.50	16.72	17.55	16.46		

c) The same quality control inspectors want to investigate the variability in battery lifetimes by estimating the population variance.

[redacted]

- Answers the question: How close to μ is the sample mean $\bar{x}$ likely to be?
- Several ways to write a confidence interval, but we will mainly use:

Point estimate +/- margin of error → point estimate = number you're using as your center, margin of error = how close we believe our guess is

The confidence level C gives the overall success rate C% of all possible samples that would yield an interval that captures the true parameter value.

To interpret a C% confidence interval for an unknown parameter, we say "We are C% confidence that the interval from ___ to ___ captures the (parameter in context).
To interpret confidence LEVEL: If we say that we are 95% confident in some interval, it is shorthand for "If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the actual parameter value.

[redacted]

Who Will Win the Election?

Two weeks before a presidential election, a polling organization asked a random sample of registered voters the following question: "If the presidential election were held today, would you vote for candidate A or candidate B?" Based on this poll, the 85% confidence interval for the population proportion who favor candidate A is (0.48, 0.54).

Problem:

a) Interpret the confidence interval?

b) What is the point estimate that was used to create the interval? What is the margin of error?

c) Based on this poll, a political reporter claims that the majority of registered voters favor candidate A. Use the confidence interval to evaluate this claim.

A) We are 95% confident that the interval from 0.48 to 0.54 captures the true proportion of all registered voters who favor candidate A in the election.
B) A confidence interval has the form point estimate +- margin of error. Find the middle point between them (p = 0.51) with a margin of error 0.03
C) Any value from 0.48 to 0.54 is a plausible value for the population proportion p that favors candidate A. Because there are plausible values of p less than 0.50, the confidence interval does not give convincing evidence to support the reporter's claim that the majority of registered voters favor candidate A.

PLAUSIBLE MEANS BELIEVABLE. TECHNICALLY ANY VALUE IS POSSIBLE

**Confidence intervals applet pg 482

[redacted]

What's the probability that our 95% confidence interval captures the parameter?

Interpreting a confidence interval and a confidence level

The Pew Internet and American Life Project asked a random sample of 2253 US adults, "Do you even use Twitter or another service to share updates about yourself or to see updates about others?" Of the sample, 19% said "yes." According to Pew, the resulting 95% confidence interval is (0.167, 0.213). Interpret the confidence interval and the confidence level.

It's NOT 95% percent! 95% stands for the long-run probability of a set of intervals containing the population parameter.

Interval: we are 95% confidence that the interval from 0.167 to 0.213 captures the true proportion p of all US adults who use Twitter or another service for updates.
Level: If many samples of 2253 US adults were taken, the resulting confidence interval would capture the true proportion of all US adults who use twitter for about 95% of those samples.

we KNOW the sample proportion is in there. We are trying to guess the population proportion. Based on the sample, we believe that the population parameter is somewhere in there.

[redacted]

How much does the fat content of Brand X hot dogs vary? To find out, researchers measured the fat content (in grams) of a random sample of 10 Brand X hot dogs. A 95% confidence interval for the population standard deviation is 2.84 to 7.55.

1. Interpret the confidence interval.
2. Interpret the confidence level.
3. True or false: The interval from 2.84 to 7.55 has a 95% chance of containing the actual population standard deviation. Justify your answer.

1. We are 95% confident that the interval from 2.84 to 7.44 captures the population standard deviation of the fat content in brand x hot dogs
2. If this sampling process were repeaated many times, approximately 95% of the resulting confidence intervals would capture the population standard deviation of the fat content of Brand X hot dogs.
3. False. Once the interval is calculated, it either contains sigma or it doesn't.

[redacted]

The confidence interval for estimating a population parameter has the form:

Open with Confidence Intervals Applet pg 485

statistic +/- (critical value)(standard deviation of statistic)
Where the statistic we use is the point estimator for the parameter.
**critical value = multiplier that makes the interval wide enough to have the stated capture rate; it depends on both the confidence level C and the sampling distribution of the statistic
**the user chooses the confidence level, and the margin of error follows from said choice (we will get into this more later)
**want high confidence and low margin of error
**greater confidence requires a larger critical value

[redacted]

1. Margin of error depends on the critical value and the standard deviation of the statistic
2. Margin of error gets smaller when
    1. Confidence level decreases
    2. Sample size n increases (remember larger samples give more precise estimates)

[redacted]

In practice, when we don't know mu, we don't know sigma either.
Confidence intervals for means and proportions are the most common tools in everyday use
Our method of calculation assumes that the data come from an SRS of size n from the population of interest. Not all samples are SRs's (might be cluster or stratified) but those require more complex calculations
The margin of error covers only chance variation due to random sampling or random assignment. Undercoverage and nonresponse can lead to additional errors that the margin of error may not account for.
The way in which a survey or experiment is conducted influences the trustworthiness of its results in ways that are not included in the announced margin of error

CHAPTER 8: ESTIMATING WITH CONFIDENCE

Part 2: Estimating a Population Proportion

Open w/ beads activity pg 493

[redacted]

You MUST check the following conditions BEFORE constructing a confidence interval for p

- **Random:**

- **10% Condition:**

- **Large Counts Condition:**

Random: gives us scope for inference to a population or inference about cause/effect; gives us the chance to model chance behavior with a probability distribution
10%: only when we sample without replacement; sampling less than 10% of the population ensures our results are mostly independent
Large counts: Let In large random samples, phat will tend to be close to p. Therefore, check nphat and n(1-phat) $\geq$ 10. These ensures normality.

If one condition is violated there's no point in constructing a confidence interval because the sampling was done poorly

[redacted]

The Beads

A large bucket of several thousand plastic beads has an array of colors. Mr. Starnes' class takes an SRS of beads, finding that 107 of them are red and 144 are not red. The class wants to construct a confidence interval for the true proportion p of red beads in the bucket.

Problem: Check that the conditions for constructing a confidence interval for p are met.

Random: The class took an SRS of 251 beads from the container, so yes.
10%: The class sampled without replacement, so we need to check that there are at least 10(251) = 2510 beads in the population. There are.
Large Counts: $np = 251(107/251) = 107 \geq 10$ and $n(1-p) = 251(144/251) = 144 \geq 10$, therefore the distribution can be normally approximated.
All conditions are met, so it should be safe to construct a confidence interval.

[redacted]

In each of the following settings, check whether the conditions for calculating a confidence interval for the population proportion p are met.

1. An AP Stats class at a large high school conducted a survey. They ask the first 100 students to arrive at school one morning whether or not they slept at least 8 hours the night before. Only 17 students said "yes."

2. A quality control inspector takes a random sample of 25 bags of potato chips from the thousands of bags filled in an hour. Of the bags selected, 3 had too much salt.

1. Not met
2. Met

[redacted]

Once conditions are met, we can move ahead with constructing our confidence intervals. Remember, we do NOT know the population parameters – we are simply estimating our accuracy based on a sample.

We use the general formula:

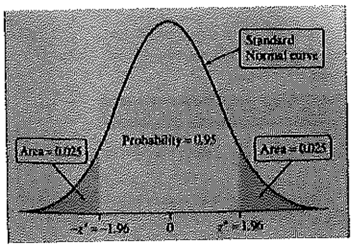
Statistic +/- (critical value)(standard deviation of statistic)

Sample proportion phat = statistic (makes sense if it came from a well designed study)
Standard deviation of the sampling distribution of phat is usually sigma = sqrt(p(1-p)/n), but we don't have p, so we will replace it with phat.

→ doing this turns standard deviation into **standard error** of the sample proportion phat, it describes how close the sample proportion phat will typically be to the population proportion p in repeated SRSs of size n

Will also see them called "one sample z interval for a population proportion

If the large counts condition is met, we can use the Normal curve. To find the critical value for our confidence interval, simply identify the z-score from Table A (or the 68-95-99.7 rule) that will render the desired proportion.



95% confidence interval uses a critical value of ______ because…

But if we look at Table A…

99% confidence =

98% confidence =

90% confidence =

85% confidence =

80% confidence =

2 because of the 68 – 95 – 99.7 rule → 95% of the data is accounted for by ±2 standard deviations
In table A, look for .025 (that's the area to the left and right that's NOT included), and it has a critical value of 1.96 (USE THIS ONE)

To find critical values, subtract level from 100, divide by 2, and look for that value as a decimal in table A (use the negative side only)
99% = 2.575
98% = 2.33
95% = 1.96
90% = 1.645
85% = 1.44
80% = 1.28

[redacted]

Back to the Beads

Recall that an SRS of beads from a bucket rendered 107 red beads and 144 non-red beads.

a) Calculate and interpret a 90% confidence interval for p, the amount of red beads.

b) Mr. Starnes claims that exactly half of the beads in the container are red. Use your results from part (a) to comment on this claim.

a) Recall that confidence interval = phat ± critical value(sqrt(phat(1-phat)/n))

Phat = 107/251 = .426

SE = sqrt(.426(.574)/251) = .0312

Critical value = 1.645

Answer: .426 ± .0513 → (0.3747, .4773)

We are 90% confident that the interval from 0.3747 to .4773 captures the true proportion of red beads in the bucket of beads.

b) The confidence interval in part (a) gives a set of plausible values for the population proportion of red beads. Because 0.5 is NOT contained in the interval, it is not a plausible value for *p* and we have reason to doubt this claim.

[redacted]

Alcohol abuse has been described by college presidents as the umber one problem on campus, and it is an important cause of death in young adults. How common is it? A survey of 10,904 randomly selected US college students collected information on drinking behavior and alcohol-related problems. The researchers defined "frequent binge drinking" as having 5 or more drinks in a row three or more times in the past 2 weeks. According to this definition, 2486 students were classified as frequent binge drinkers.

a) Identify the parameter of interest.
b) Check conditions for constructing a confidence interval for the parameter.
c) Find the critical value for a 99% confidence interval. Show your method. Then calculate the interval.
d) Interpret the interval in context.

a) P = true proportion of all US college students who are binge drinkers
b) Random: met; 10% met; large counts: met
c) Critical value = 2.575 (1-.99/2 = .005, closest to that is 2.575)

Phat = 2486/10904 = .228

SE = .004

Interval: .228 ± .01 → (.218, .238)

d) We are 99% confident that the interval from .218 to .238 is the true proportion of all US college students who are classified as binge drinkers

[redacted]

If a free-response question asks you to construct and interpret a confidence interval, you are expected to do the entire four-step process:

1. **State:** What parameter do you want to estimate, and at what confidence level?
2. **Plan:** Identify the appropriate inference method and check conditions.
3. **Do:** If the conditions are met, perform calculations.
4. **Conclude:** Interpret your interval in the context of the problem.

[redacted]

The Gallup Youth Survey asked a random sample of 439 US teens aged 13 to 17 whether they thought young people should wait to have sex until marriage. Of the sample, 246 said "Yes." Construct and interpret a 95% confidence interval for the proportion of all teens who would say "Yes" if asked this question.

State: We want to estimate the true proportion p of all 13-17 year olds in the US who would say that young people should wait to have sex until marriage with 95% confidence.

Plan: We should use a one-sample z-interval for p if the conditions are met

Random: Gallup surveyed a random sample of US teens

10%: Because Gallup is sampling without replacement, we need to check this condition. There are more than 10(439) – 4390 teens aged 13-17

Large Counts: We check the counts of successes (np = 246) and failures (n(1-p) = 193) are $\geq$ 10 and this is so

The conditions are met, so it is safe to proceed with a confidence interval.

Do: Sample statistic phat = 246/439 = .56

95% confidence interval for p is given by .56 $\pm$ (1.96)(sqrt((0.56)(0.44)/439) = 0.56 $\pm$ .046 → (0.514, 0.606)

Conclude: We are 95% confident that the interval from 0.514 to 0.606 captures the true proportion of 13-17 year olds in the US who would say that teens should wait until marriage to have sex.

**POINT: remember that margins of error only tell us about sampling variability, not about other factors. We have to trust that the teens answered honestly, or that the wording of the question didn't affect answers.

[redacted]

On the TI 84 and above:

- Press STAT, then choose TESTS and 1-PropZInt
- When the screen appears, enter x=246, n=439, and confidence level .95
- Highlight "Calculate" and press ENTER. The interval is reported along with the sample proportion and sample size.

If you're going to use this method, be sure to identify what you plug into the calculator. Don't try to "show your work" by using the formula, because if one or the other is faulty, you will be scored on the worst response!

[redacted]

In planning a study, we may want to choose a sample size that allows us to estimate a population proportion within a given margin of error.

The margin of error (ME) in the confidence interval for p is:

When we have to choose the sample size, we generally also have to estimate $\hat{p}$ (unless, of course, it is given, such as in the form of "previous study"). The ME is largest when $\hat{p}$ = 0.5, so this is generally the best guess.

ME = z* (sqrt(phat(1-phat)/n)) where z* is the standard normal critical value we use based on the confidence level we want
*recognize that as n icnreases, ME will decrease

**solve for n to find the right sample size! But know that whatever you find, n should be greater than that boundary.

[redacted]

A company has received complaints about its customer service. The managers intend to hire a consultant to carry out a survey of customers, but want some idea of the sample size that they will be required to pay for. The critical question is the degree of satisfaction with the company's customer service, measured on a 5-point scale. The company wants to estimate the proportion of customers who are satisfied (a 4 or 5 rating) within 3% at a 95% confidence level. How large of a sample is needed?

Important: ME=3%, confidence = 95% → $z^* = 1.96$

1.96(sqrt(0.5(0.5))/n) ≤ .03
Solve for n...
N ≥ 1067.111
We round up to 1068 respondents to ensure that the margin of error is no more than 3%.

[redacted]

Refer to the previous example about the company's customer satisfaction survey:

1. In the company's prior-year survey, 80% of customers surveyed said they were satisfied. Using this value as a guess for $\hat{p}$ find the sample size needed for a margin of error of 3% at a 95% confidence level.

2. What if the company demands a 99% confidence level instead? Determine how this would affect your answer in question 1.

CHAPTER 8: ESTIMATING WITH CONFIDENCE

Part 3: Estimating a Population Mean

[redacted]

Inference for a population proportion usually arises when we study *categorical* variables. To estimate a population mean, we have to record values of a *quantitative* variable.

We will still use the confidence interval formula:

statistic $\pm$ (critical value)(standard deviation of statistic)

We will call such intervals a "one-sample" z-interval for a populatio mean.

Formula: xbar $\pm$ z*(s_x/sqrt(n))

**The problem here is the population standard deviation is not actually known, so we are estimating it with the sample.

**In a perfect world, we'd know the standard deviation of the population, so this is the answer to that.

Open with Activity pg 509

[redacted]

- When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing. Recall what happens to the shape, center, and spread of the distribution:

- When we don't know σ, we estimate it using the sample standard deviation s_x:

Z = xbar – mu / (sigma/sqrt(n))
* - mu moves it to the left mu units (no change in shape or spread), dividing keeps the mean at 0 and makes SD 1 and leaves shape unchanged. Recall that z is N(0,1).

?? = xbar – mu/ (s_x/sqrt(n))
Exchanging s_x for sigma still renders a normal distribution, but with much larger spread. This is no longer a z-score, it's called a **t-distribution**: still a standard Normal curve with a single peak at 0, but with much more area in the tails

[redacted]

- still a standard Normal curve with a single peak at 0, but with much more area in the tails (draw standard normal vs lots of degrees of freedom example)
- Same interpretation as any standardized score: it says how far away xbar is from its mean mu in standard deviation units
    - Therefore there's more probability in the tails than we're used to seeing
- There is a different t-distribution for each sample size, specified by degrees of freedom (df = n – 1). Degrees of freedom basically come from the idea that every sample size is different so it deserves a different distribution.
    - The greater degrees of freedom, the closer it is to the standard normal curve (z distribution). This is what we want bc it indicates less variability
- T is always for means! Restate formula:

$$t = xbar - mu/(s_x/sqrt(n))$$

- Using Table B for interpretation now. Not all degrees of freedom are shown – when this happens, use the greatest df available that is less than your desired df

[redacted]

Problem: What critical t* from Table B should be used in constructing a confidence interval for the population mean in each of the following settings?

a) A 95% confidence interval based on an SRS of size n=12.
b) A 90% confidence interval from a random sample of 48 observations.

Solution:

A) Df = 12-1=11. 95% confidence means .05 left, .025 on each side. Upper tail probability p at .025 w/ df = 11 has 2.201 critical value
B) Df = 48-1 = 47, highest is 40 on the table so we'll use that one; 90% confidence means .05 left in each tail; p=.05 w/ df=40 has 1.684 critical value

On the calculator:
Press 2nd VARS (Distr) and choose invT(.
For part (a): area: .025, df: 11, Enterx2
For part (b): invT(.05, 47)
**they'll be negative, but remember for confidence intervals is ±

[redacted]

Use Table B to find the critical value t* that you would use for a confidence interval for a population mean μ in each of the following settings. If possible, check your answers with technology (or vice versa).

1. A 96% confidence interval based on a random sample of 22 observations.

2. A 99% confidence interval from an SRS of 71 observations.

[redacted]

- Random:
- 10%:
- Normal/Large Sample:

Random: the data comes from a well-designed random sample or randomized experiment
10%: when sampling without replacement, check that $n \leq 1/10N$
Normal/Large Sample: the population has a Normal distribution or the sample size is large ($n \geq 30$). If the population distribution has unknown shape and $n<30$, use a graph of the same data to assess Normality of the population. Do NOT use t procedures if the graph shows strong skewness or outliers.

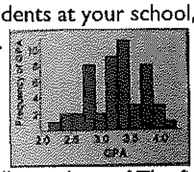
**larger samples improve the accuracy of critical values from the t distributions when the population is NOT normal because:

1. The sampling distribution for xbar for large sample sizes is close to Normal (Central Limit Theorem)
2. As the sample size n grows, the sample standard deviation s_x will give a more accurate estimation of sigma.

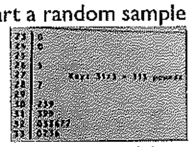
[redacted]

Problem: Determine if we ca safely use a t* critical value to calculate a confidence interval for the population mean in each of the following settings:

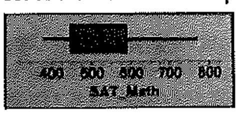
a) To estimate the average GPA of students at your school, you randomly select 50 students from classes you take. The figure below is a histogram of their GPAs.



b) How much force does it take to pull wood apart? The figure below shows a stemplot of the force (in pounds) required to pull apart a random sample of 20 pieces of Douglas fir.



c) Suppose you want to estimate the mean SAT math score at a large high school. The figure below is a boxplot of the SAT math scores for a random sample of 20 students at the school.



a) No – not random sampling
b) No, n<30 and skewed left w/ outliers
c) Yes, even though n<30, mostly normal shape (skew must be significant to not be able to use t)

[redacted]

- There is no accepted rule of thumb for identifying strong skewness. For that reason, you should be able to tell easily if strong skewness is present in a graph of data from a small sample.
- For example, look at the difference between the stemplot and the box and whiskers plot from the previous slide.
  - Compare the difference from the maximum to the median, and from the median to the minimum in either graph

[redacted]

- It's the same process!

Statistic ± critical value • standard deviation of statistic
Xbar ± t* (s_x/sqrt(n))
Where t* is the critical value for the t_{n-1} distribution (degrees of freedom), and the standard error SE_x = s_x/sqrt(n) describes how far xbar will typically be from mu in repeated SRSs of size n.
C% of the area between -t* and t*

[redacted]

Environmentalists, government officials, and vehicle manufacturers are all interested in studying the auto exhaust emissions produced by motor vehicles. The major pollutants in auto exhaust from gasoline engines are hydrocarbons, carbon monoxide, and nitrogen oxides (NOX). Researchers collected data on the NOX levels (in grams/mile) for a random sample of 40 light-duty engines of the same type. The mean NOX reading was 1.2675 and the standard deviation was 0.3332.

a) Construct and interpret a 95% confidence interval for the mean amount of NOX emitted by light-duty engines of this type.

b) The EPA sets a limit of 1.0 grams/mile for average NOX emissions. Are you convinced that this type of engine violates the EPA limit?

Solution:

a) STATE: We want to estimate the true mean amount μ of NOX emitted by all light-duty engines of this type at a 95% confidence level.

PLAN: We should construct a one-sample t interval for μ if the conditions are met

Random: The data come from a random sample of 40 light-duty engines of this type.

10% : We are sampling w/o replacement, so we need to assume that there are at least 10(40) = 400 light-duty engines of this type.

Normal/Large Sample: We don't know whether the population distribution of NOX emissions is Normal, but because the sample size is large ($n=40\geq30$), we should be safe using a t distribution.

DO: For n=40, we have degrees of freedom df = 40-1 = 39. Since there is no corresponding row in Table B for this degree of freedom, we will use the more conservative df=30. At a 95% confidence level, the critical value is t* = 2.042. We are given xbar = 1.2675 and s_x = 0.3332. Therefore, the 95% confidence interval for μ is

$$1.2675 \pm 2.042(0.3332/\text{sqrt}(40)) = 1.2675 \pm .1066 = (1.1609, 1.3741).$$

Technology: invT(.025, 39) = 2.023

Conclude: We are 95% confident that the true population mean of NOX emissions from this type of light-duty engine is between 1.1609 and 1.3741 grams/mile.

b) The interval (1.1609, 1.3741) gives plausible values of the mean NOX emissions for this type of engine. Since the entire interval exceeds 1.0 grams/mile, it appears that this type of engine violates EPA guidelines.

[redacted]

A manufacturer of high-resolution video terminals must control the tension on the mesh of fine wires that lies behind the surface of the viewing screen. Too much tension will tear the mesh, and too little will allow wrinkles. The tension is measured by an electrical device with output readings in millivolts (mV). Some variation is inherent in the production process. Here are the tension readings from a random sample of 20 screens from a single day's production:

269.5	297.0	269.6	283.3	304.8	280.4	233.5	257.4	317.5	[illegible]
264.7	307.7	310.0	343.3	328.1	342.6	338.8	340.1	374.6	[illegible]

Construct and interpret a 90% confidence interval for the mean tension μ of all the screens produced on this day.

State: We want to estimate the true mean tension μ of all the video terminals produced this day with 90% confidence.

Plan: If the conditions are met, we should use a one-sample t interval to estimate μ

Random: We are told that the data come from a random sample of 20 screens produced that day

10%: Because we are sampling w/o replacement, we must assume that at least 10(20) = 200 video terminals were produced this day

Normal/Large sample: Because the same size is small (n=20), we must check whether it's reasonable to believe that the population distribution is Normal. We examine a ________ plot to determine shape of the distribution. *draw thing*

The graph does not show strong skewness or outliers, so we have no reason to doubt the Normality of the population. It is safe to use a t-distribution

DO: Using a calculator, we find xbar = 306.32 mV and s_x = 36.21mV. We use the t distribution with df = 19 to find the critical value. For a 90% confidence level, the critical value is t*=1.729. Therefore, the 90% confidence interval for μ is:

$$306.32 \pm 1.729(36.21/\text{sqrt}(20)) = 306.32 \pm 14 = (292.32, 320.32)$$

Conclude: We are 90% confident that the interval from 292.32 to 320.32 mV captures the true mean tension in the entire batch of video terminals produced that day.

AP TIP: It's not enough to make the graph on your calculator and say what it looks like. YOU MUST SKETCH IT ON YOUR PAPER*

CAN BE DONE WITH TECHNOLOGY: Stat→ TESTS, Tinterval…plug in info you know and it will spit it out, or put in info to L1 and tell it to use Data as the input method

[redacted]

Biologists studying the healing of skin wounds measured the rate at which new cells closed a cut made in the skin of an anesthetized newt. Here are the data from a random sample of 18 newts, measured in micrometers per hour:

29 27 34 40 22 28 14 35 26 35 12 30 23 18 11 22 23 33

Calculate and interpret a 95% confidence interval for the mean healing rate μ.

[redacted]

You can arrange to have both high confidence and a small margin of error by taking enough observations. Recall the margin of error (ME) of the confidence interval is:

We usually just solve for n in this case. The problem here is that we haven't produced the data yet to know s_x AND t* depends on the sample size we choose.

The solution:

$ME = t^*(s_x/\text{sqrt}(n))$

Solution: come up with a reasonable estimate for the population standard deviation sigma from a similar study that was done in the past or from a small-scale pilot study. By pretending sigma is known, then we can use a one-sample z-interval for μ, use the critical z* values, and solve for n to find the sample size we want.

[redacted]

Researchers would like to estimate the mean cholesterol level μ of a particular variety of monkeys that is often used in lab experiments. They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of μ at a 95% confidence level. A previous study involving this variety of monkeys suggests that the standard deviation of cholesterol level is about 5 mg/dl.
Obtaining monkeys for research is time-consuming, expensive, and controversial. What is the minimum number of monkeys the researchers will need to get a satisfactory estimate?

Solution: For a 95% confidence, z* = 1.96. We will use sigma=5 as our best guess for the standard deviation of the mnokeys cholesterol level. Set the expression for ME to be at most 1 and solve for n:

$$1.96(5/\text{sqrt}(n)) \leq 1$$
$$n \geq 96.04$$

Because 96 monkeys would give a slightly larger margin of error than desired, the researchers would need 97 monkeys to estimate the cholesterol levels to their satisfaction.

[redacted]

Administrators at your school want to estimate how much time students spend on homework, on average, during a typical week. They want to estimate μ at the 90% confidence level with a margin of error of at most 30 minutes. A pilot study indicated that the standard deviation of time spend on homework per week is about 154 minutes. How many students need to be surveyed to meet the administrator's goal? Show your work.

Sigma = 154, z* = 1.645
30 ≥ 1.645(154/sqrt(n))
N≥ 71.3
Take 72 students

**CLOSE CHAPTER WITH CASE CLOSED ACTIVITY PG 525