NAME: ANSWER KEY

SECTION 8.1 Exercises "What is sampling distribution?"

For Exercises 1 to 4, identify the population, the parameter, the sample, and the statistic in each setting.

1. Stop smoking! A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted nine months later. It turned out that 210 (21%) of the sampled individuals had not smoked over the past six months.

Population: ALL THE PEOPLE WHO SIGNED A CARD SAYING THEY INTENDED TO QUIT SMOKING

PARAMETER: PROPORTION OF THE POPULATION (ALL SIGNED CARD) WHO QUIT SMOKING denoted by "p"

SAMPLE: RANDOM SAMPLE OF 1000 PEOPLE SIGNED CARD

STATISTIC: $\hat{p} = .21$

Sample proportion who quit

STATISTIC IN APPROPRIATE SETTING

3. Hot turkey Tom is cooking a large turkey breast for a holiday meal. He wants to be sure that the turkey is safe to eat, which requires a minimum internal temperature of 165°F. Tom uses a thermometer to measure the temperature of the turkey meat at four randomly chosen points. The minimum reading in the sample is 170°F.

Population: ALL THE TURKEY MEAT

PARAMETER: MINIMUM TEMPERATURE

Sample: 4 RANDOMLY CHOSEN points ON THE TURKEY

STATISTIC: Sample MINIMUM = 170°F

For each boldface number in Exercises 5 to 6 (1) state whether it is a parameter or a statistic and (2) use appropriate notation to describe each number; for example, p = 0.65.

5. Get your bearings A large container of ball bearings has mean diameter 2.5003 centimeters (cm). This is within the specifications for acceptance of the container by the purchaser. By chance, an inspector chooses 100 bearings from the container that have mean diameter 2.5009 cm. Because this is outside the specified limits, the container is mistakenly rejected.

6. Florida voters Florida has played a key role in recent presidential elections. Voter registration records show that 41% of Florida voters are registered as Democrats. To test a random digit dialing device, you use it to call 250 randomly chosen residential telephones in Florida. Of the registered voters contacted, 33% are registered Democrats.

#5
$\mu = 2.5003$ cm is a parameter
$\bar{x} = 2.5009$ cm is a statistic

#6
$p = .41$ is a parameter
$\hat{p} = .33$ is a statistic

8.1 CONT

#10+12 Work Here ↓

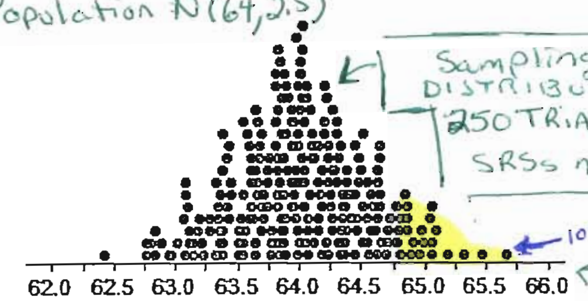
10. Tall girls According to the National Center for Health Statistics, the distribution of heights for 16-year-old females is modeled well by a Normal density curve with mean $\mu = 64$ inches and standard deviation $\sigma = 2.5$ inches. To see if this distribution applies at their high school, an AP Statistics class takes an SRS of 20 of the 300 16-year-old females at the school and measures their heights. What values of the sample mean $\bar{x}$ would be consistent with the population distribution being $N(64, 2.5)$? To find out, we used Fathom software to simulate choosing 250 SRSs of size $n = 20$ students from a population that is $N(64, 2.5)$. The figure below is a dotplot of the sample mean height $\bar{x}$ of the students in the sample.

population

Sample of students

Sampling distribution

Population N(64,2.5)



EACH DOT IS AN $\bar{x}$ and there are 250 dots

(a) Is this the sampling distribution of $\bar{x}$? Justify your answer.

(b) Describe the distribution. Are there any obvious outliers?

(c) Suppose that the average height of the 20 girls in the class's actual sample is $\bar{x} = 64.7$. What would you conclude about the population mean height μ for the 16-year-old females at the school? Explain.

10A THIS IS NOT AN EXACT SAMPLING DISTRIBUTION, BECAUSE THAT WOULD REQUIRE A VALUE OF $\bar{x}$ FOR EVERY POSSIBLE SAMPLE OF SIZE 20. HOWEVER, IT IS AN APPROXIMATION OF THE SAMPLING DISTRIBUTION THAT WE CREATED USING SIMULATION (USING A SIMULATION TOOL CALLED FATHOM).

10B THE SAMPLING DISTRIBUTION OF SAMPLE MEANS ($\bar{x}$) IS Centered at 64 and is reasonably symmetric and bell-shaped. Values vary from about 62.5 to 65.7. There do not appear to be any outliers.

Could say: approx. normal distribution

10C IF WE FOUND THE SAMPLE MEAN WAS 64.7in, we would likely conclude that the population mean height for females at this school could be 64. In our simulation, we found values of 64.7 or larger in about 10% of the samples.

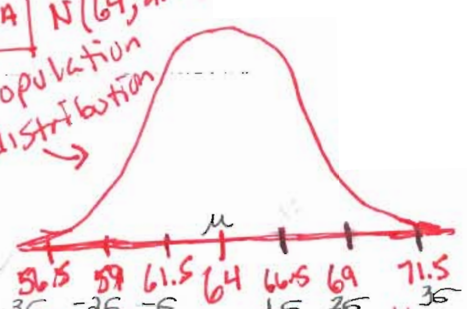
12. Tall girls Refer to Exercise 10.

(a) Make a graph of the population distribution.

(b) Sketch a possible graph of the distribution of sample data for the SRS of size 20 taken by the AP Statistics class.

12A N(64,2.5)

Population distribution



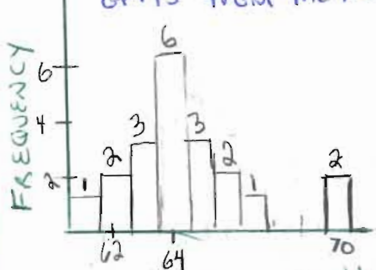
Height of 16 year old females (inches)

(remember 68-95-99.7 Rule)

12B

Sample Distribution HISTOGRAM

20 Randomly Selected Girls from the HS



HEIGHT 16 yr old females (in)

← No single correct answer. This is a histogram for the random sample of 20 girls. My distribution is basically symmetric with 2 tall girls that skewed the mean to 64.7in

8.1 CONT

17 **IRS audits** The Internal Revenue Service plans to examine an SRS of individual federal income tax returns from each state. One variable of interest is the proportion of returns claiming itemized deductions. The total number of tax returns in each state varies from over 15 million in California to about 240,000 in Wyoming.

(a) Will the sampling variability of the sample proportion change from state to state if an SRS of 2000 tax returns is selected in each state? Explain your answer.

(b) Will the sampling variability of the sample proportion change from state to state if an SRS of 1% of all tax returns is selected in each state? Explain your answer.

Tax Returns Range
MIN - WY - 240,000
MAX - CA - 15 Million

10% rule met
samples < 24,000

Remember sampling variability is determined primarily by the size of the random samples taken.
• Larger samples give smaller spreads
• and the size of the population does NOT matter as long as the population is 10 times larger than the sample.

AND remembering this, you should be able to answer these Questions

17A For SRS's of 2,000, The variability of the sample proportion will be (approx.) the same for all states because every state is more than 10 times the 2,000 sample size.

17B The variability for a 1% SRS will be different among states because the sample size for WY will be 2,400 and the sample size for CA will be 15,000; So CA's variability will be much smaller than WY.

18 **Predict the election** Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

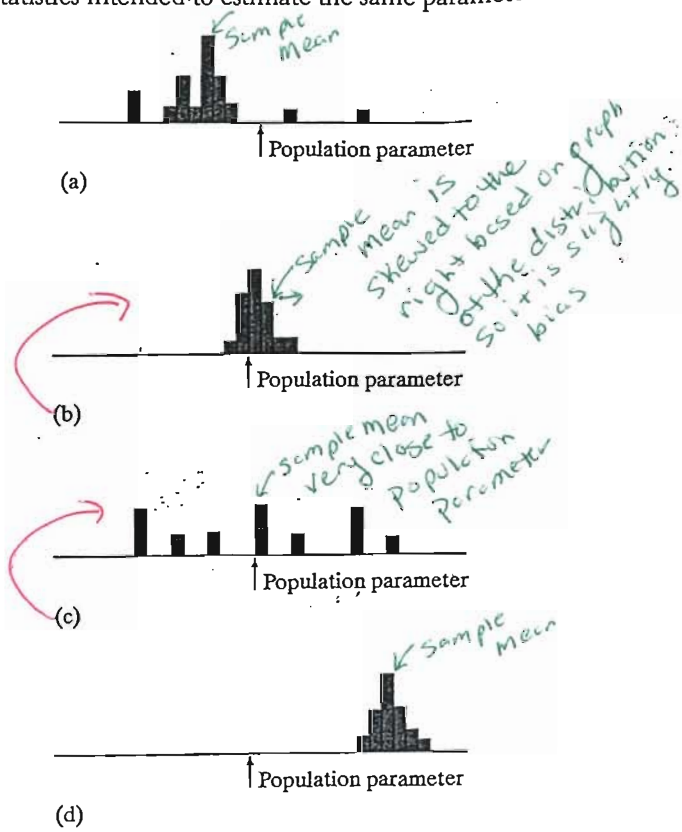
(a) Does the larger random sample reduce the bias of the poll result? Explain.

(b) Does it reduce the variability of the result? Explain.

18A A larger sample does not reduce the bias of a poll result. IF THE SAMPLING TECHNIQUE RESULTS IN BIAS, SIMPLY INCREASING THE SAMPLE SIZE WILL NOT REDUCE THE BIAS.

18B A larger sample will reduce the variability of the result. More people means more information, which means less variability.

19. **Bias and variability** The figure below shows histograms of four sampling distributions of different statistics intended to estimate the same parameter.



(a) Which statistics are unbiased estimators? Justify your answer.

(b) Which statistic does the best job of estimating the parameter? Explain.

19A Graph (C) shows an unbiased estimator because the mean of the distribution ~~the~~ is very close to the population parameter.

19B Graph (b) shows the statistic that does the best job at estimating the (population) parameter. Even though it is slightly bias, the statistic has very little variability.

---

20. **A sample of teens** A study of the health of teenagers plans to measure the blood cholesterol levels of an SRS of 13- to 16-year-olds. The researchers will report the mean $\bar{x}$ from their sample as an estimate of the mean cholesterol level μ in this population.

(a) Explain to someone who knows no statistics what it means to say that $\bar{x}$ is an unbiased estimator of μ.

(b) The sample result $\bar{x}$ is an unbiased estimator of the population mean μ no matter what size SRS the study chooses. Explain to someone who knows no statistics why a large random sample gives more trustworthy results than a small random sample.

20A IF WE CHOOSE MANY SAMPLES, THE AVERAGE OF MEANS OF THESE SAMPLES (called the $\bar{x}$'s) WILL BE VERY CLOSE TO THE TRUE MEAN OF THE ENTIRE POPULATION (called: μ). IN OTHER WORDS, THE SAMPLE DISTRIBUTION OF THE SAMPLE MEAN ($\bar{x}$) IS CENTERED AT THE POPULATION MEAN (μ)

20B A LARGER SAMPLE WILL GIVE MORE INFORMATION AND THEREFORE, MORE PRECISE RESULTS. THE VARIABILITY IN THE DISTRIBUTION OF THE SAMPLE AVERAGE ($\bar{x}$) DECREASES AS THE SAMPLE SIZE INCREASES.

SECTION 8.2 Exercises

28. The **candy machine** Suppose a large candy machine has 15% orange candies. Use Figure 7.13 to help answer the following questions.

(a) Would you be surprised if a sample of 25 candies ~~from the machine contained 8 orange~~ candies (that's 32% orange)? How about 5 orange candies (20% orange)? Explain.

IF YOU LOOK AT THE SAMPLING DISTRIBUTION IN THE FIRST GRAPH You can easily answer this Question.

- We would be surprised to get 8 orange (32%) because there were very few simulctions for this case
- We would not be surprised to get 5 orange (20%) because this was very close to the center.

(b) Which is more surprising: getting a sample of 25 candies in which 32% are orange or getting a sample of 50 candies in which 32% are orange? Explain.

WE WOULD BE SURPRISED TO FIND 32% ORANGE IN EITHER CASE SINCE NEITHER SIMULATION HAD MANY SAMPLES WITH 32% ORANGE.

HOWEVER, IT WOULD BE EVEN RARER WITH THE LARGER SAMPLE SIZE OF 50 (BECAUSE LARGER SAMPLE SIZES HAVE SMALLER VARIABILITY.

Sample Distributions of Sample Means ($\bar{x}$)

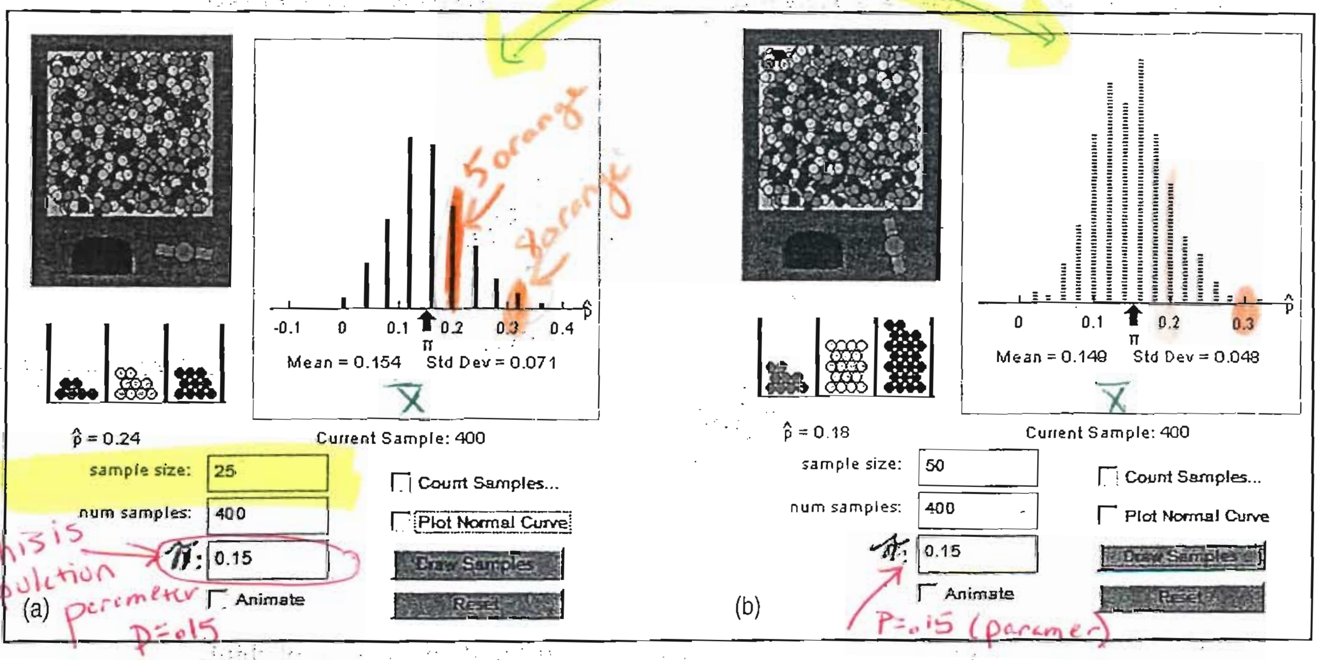


FIGURE 7.13 The result of taking 400 SRSs of (a) size $n = 25$ and (b) size $n = 50$ candies from a large candy machine in which 15% of the candies are orange. The dotplots show the approximate sampling distribution of $\hat{p}$ in each case.

8.2 CONT

29

9. **The candy machine** Suppose a large candy machine has 45% orange candies. Imagine taking an SRS of 25 candies from the machine and observing the sample proportion $\hat{p}$ of orange candies.

(a) What is the mean of the sampling distribution of $\hat{p}$? Why? : $\mu_{\hat{p}}$ – make sure to use correct notation.

(b) Find the standard deviation of the sampling distribution of $\hat{p}$. Check to see if the 10% condition is met. The 10% rule ensures independence

(c) Is the sampling distribution of $\hat{p}$ approximately Normal? Check to see if the Normal condition is met.

(d) If the sample size were 50 rather than 25, how would this change the sampling distribution of $\hat{p}$?

population parameter $p = .45$

A The mean of the sampling distribution is the same as the population proportion so $\mu_{\hat{p}} = p = .45$

B 10% CONDITION – IS MET BECAUSE IT IS VERY LIKELY THERE ARE MORE THAN 250 CANDIES.

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.45)(.55)}{25}} \approx .0995$$

C THE SAMPLING DISTRIBUTION OF $\hat{p}$ IS APPROXIMATELY NORMAL BECAUSE THE NORMAL CONDITIONS ARE MET

$np = 25(.45) = 11.25 \geq 10$ ✓

$n(1-p) = 25(.55) = 13.75 \geq 10$ ✓

D $n = 50$ — $\mu_{\hat{p}} = .45$ stay same

BUT $\sigma_{\hat{p}} = \sqrt{\frac{(.45)(.55)}{50}} \approx .0704$ (decreases)

In Exercises 33 and 34, explain why you cannot use the methods of this section to find the desired probability.

33. **Hispanic workers** A factory employs 3000 unionized workers, of whom 30% are Hispanic. The 15-member union executive committee contains 3 Hispanics. What would be the probability of 3 or fewer Hispanics if the executive committee were chosen at random from all the workers?

NOTE: Using the binomial model will give the exact probability, Using sampling distribution of $\hat{p}$ Give an approx. probability.

33 The normal condition is NOT met here: $n = 15$ $p = .3$

$np = 15(.3) = 4.5 < 10$

So how could you find the $P(X \leq 3)$?

Let X = the number of Hispanics in the sample

X has an approx. BINOMIAL Dist $B(15, .3)$

$P(X \leq 3) = \text{binomcdf}(15, .3, 3) = .2969$

34. **Studious athletes** A university is concerned about the academic standing of its intercollegiate athletes. A study committee chooses an SRS of 50 of the 316 athletes to interview in detail. Suppose that 40% of the athletes have been told by coaches to neglect their studies on at least one occasion. What is the probability that at least 15 in the sample are among this group?

- X = # ATHLETES TOLD..

$B(50, .4)$

$P(X \geq 15) = 1 - P(X \leq 14)$

$= 1 - .0539 = .946$

About 95%

34 population size $= N = 316$

sample size of SRSs $= n = 50$

* The 10% condition to ensure independence is NOT met. The sample of 50 is more than 10% of the population (316).

8.2 CONT

35 Do you drink the cereal milk? A *USA Today* Poll asked a random sample of 1012 U.S. adults what they do with the milk in the bowl after they have eaten the cereal. Of the respondents, 67% said that they drink it. Suppose that 70% of U.S. adults actually drink the cereal milk. Let $\hat{p}$ be the proportion of people in the sample who drink the cereal milk.

(a) What is the mean of the sampling distribution of $\hat{p}$? Why?

(b) Find the standard deviation of the sampling distribution of $\hat{p}$. Check to see if the 10% condition is met.

(c) Is the sampling distribution of $\hat{p}$ approximately Normal? Check to see if the Normal condition is met.

(d) Find the probability of obtaining a sample of 1012 adults in which 67% or fewer say they drink the cereal milk. Do you have any doubts about the result of this poll?

$p = .70$ = population proportion

$\hat{p}$ = proportion in sample who drink milk

A $\mu_{\hat{p}} = .70$ The mean of the sampling distribution is the same as the population proportion

B $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.7)(.3)}{1012}} \approx .0144$

The population (all US adults) is clearly at least 10 times as large as the sample (1012); so the 10% condition is met.

C The sampling distribution for $\hat{p}$ is approximately normal since

$np = 1012(.7) = 708.4 \geq 10$ ✓

$nq = 1012(.3) = 303.6 \geq 10$ ✓

35D $\boxed{P(\hat{p} \leq .67) \approx 2\%}$

use normal model $N(.7, .0144)$

either: use

① normalcdf(-1E99, .67, .7, .0144) =

.0186

ONLY 2%

② OR USE Z SCORE

$Z = \frac{.67 - .7}{.0144} = -2.08$

P(Z < -2.08) = .0188

normalcdf(-1E99, -2.08, 0, 1)

* This is a fairly unusual result if 70% of the population actually drink the cereal

37 Do you drink the cereal milk? What sample size would be required to reduce the standard deviation of the sampling distribution to one-half the value you found in Exercise 35(b)? Justify your answer.

Since the standard deviation is found by dividing by $\sqrt{n}$ and we want 1/2 the std dev, so

$\frac{1}{2}\left(\frac{\sigma}{\sqrt{N}}\right) = \frac{\sigma}{\sqrt{4n}}$

So to get 1/2 the std dev we would multiply the sample size by 4

4 * 1012 to

Need a sample size of 4,048

8.2 CONT

41 **On-time shipping** Your mail-order company advertises that it ships 90% of its orders within three working days. You select an SRS of 100 of the 5000 orders received in the past week for an audit. The audit reveals that 86 of these orders were shipped on time.

(a) If the company really ships 90% of its orders on time, what is the probability that the proportion in an SRS of 100 orders is as small as the proportion in your sample or smaller? Follow the four-step process.

(b) A critic says, "Aha! You claim 90%, but in your sample the on-time percentage is lower than that. So the 90% claim is wrong." Explain in simple language why your probability calculation in (a) shows that the result of the sample does not refute the 90% claim.

Check CONDITIONS:

$\hat{p} = 86/100 = .86$ from sample in audit

$p = .90$ population parameter

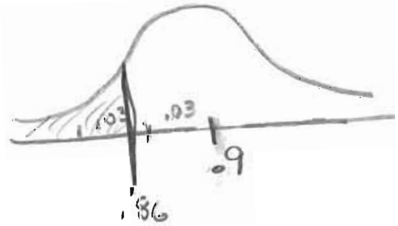
$n = 100$

$\mu_{\hat{p}} = .90$ $\sigma_{\hat{p}} = \sqrt{\frac{(.9)(.1)}{100}} = .03$

- 10% Condition is clearly met taking a sample 100/5000
- Normal conditions met

$np = 100(.9) = 90 \geq 10$ ✓

$nq = 100(.1) = 10 \geq 10$ ✓



(A) $\boxed{P(\hat{p} \leq .86) = .0912}$

$N(.9, .03)$

normalcdf(-1E99, .86, .9, .03)

Conclusion: There is a 9.12% chance that we would get a sample in which 86% or fewer of the orders were shipped within 3 working days.

(B) Getting a sample proportion at or below .86 is not an unlikely event (9.12%). The sample results are lower than the 90% company advertised but the sample was so small that such a difference could arise by chance.

SECTION 8.3A Exercises

49. **Songs on an iPod** David's iPod has about 10,000 songs. The distribution of the play times for these songs is heavily skewed to the right with a mean of 225 seconds and a standard deviation of 60 seconds. Suppose we choose an SRS of 10 songs from this population and calculate the mean play time $\bar{x}$ of these songs. What are the mean and the standard deviation of the sampling distribution of $\bar{x}$? Explain.

Mean of sampling distribution of Sample means $= \mu_{\bar{x}} = 225$ seconds

Std dev of sampling distribution of Sample means $= \sigma_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\sqrt{n}} = \frac{60}{\sqrt{10}}$

$\sigma_{\bar{x}} = 18.974$ seconds

Explanation:

BECAUSE OF THE Central Limit THEOREM (CLT), choosing random samples(sess), the mean and standard deviations of the statistic sample mean $(\bar{x})$ does not depend on the shape of the distribution of individual play times

51. **Songs on an iPod** Refer to Exercise 49. How many songs would you need to sample if you wanted the standard deviation of the sampling distribution of $\bar{x}$ to be 30 seconds? Justify your answer.

IF WE WANT $\sigma_{\bar{x}} = 30$, THEN WE NEED TO SOLVE THE FOLLOWING EQUATION FOR n:

$$30 = \frac{60}{\sqrt{N}} \longrightarrow \frac{\cancel{30}\sqrt{n}}{\cancel{30}} = \frac{60}{30} \longrightarrow \sqrt{n}^2 = 2^2$$

$\boxed{n = 4}$

8.3A Cont

53 **Larger sample** Suppose that the blood cholesterol level of all men aged 20 to 34 follows the Normal distribution with mean $\mu = 188$ milligrams per deciliter (mg/dl) and standard deviation $\sigma = 41$ mg/dl.

(a) Choose an SRS of 100 men from this population. What is the sampling distribution of $\bar{x}$?

(b) Find the probability that $\bar{x}$ estimates μ within ± 3 mg/dl. (This is the probability that $\bar{x}$ takes a value between 185 and 191 mg/dl.) Show your work.

(c) Choose an SRS of 1000 men from this population. Now what is the probability that $\bar{x}$ falls within ± 3 mg/dl of μ? Show your work. In what sense is the larger sample "better"?

$\rightarrow N\overset{(\mu_x,\ \sigma_x)}{(188,\ 41)}$

(a) mean of the sampling distribution of sample means $(\bar{x})$ =

$\mu_{\bar{x}} = \mu_x = 188$ mg/dl

the std dev of $\bar{x}$ =

$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{41}{\sqrt{100}} = 4.1$ mg/dl

(b) $P(185 \le \bar{x} \le 191) = .5346$

normalcdf(185, 191, 188, 4.1)

The probability of about 53% make sense because +/− 1 STD DEV WOULD BE ABOUT 68% (68-95-99.7 rule)

$\sigma_{\bar{x}} = 4.1$

(c) $\mu_{\bar{x}} = 188$ $\quad \sigma_{\bar{x}} = \frac{41}{\sqrt{1000}} = 1.30$

$P(185 \le \bar{x} \le 191) =$ normalcdf(185, 191, 188, 1.3)

$= .9789$

* The larger sample size is better since it is more likely to produce a sample mean $(\bar{x})$ WITHIN 3 mg/dl of the population mean.

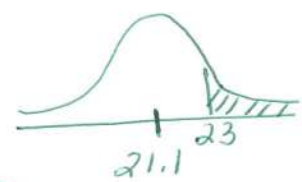
56 **ACT scores** The composite scores of individual students on the ACT college entrance examination in 2009 followed a Normal distribution with mean 21.1 and standard deviation 5.1.

(a) What is the probability that a single student randomly chosen from all those taking the test scores 23 or higher? Show your work.

(b) Now take an SRS of 50 students who took the test. What is the probability that the mean score $\bar{x}$ of these students is 23 or higher? Show your work.

(a) Let X = the ACT score of a randomly selected test taker

$N(21.1,\ 5.1)$



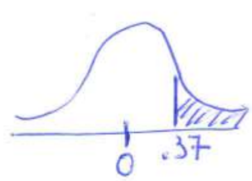
$P(x \ge 23) = .3547$

normalcdf (23, 1E99, 21.1, 5.1)

remember (ALPHA) (EE)

* Make sure to understand how to do probabilities with Z SCORES.

$Z = \frac{x - \mu}{\sigma} = \frac{23 - 21.1}{5.1} = .37$



$P(z > .37) = .3557$

normalcdf(.37, 1E99, 0, 1)

(b) $\mu_{\bar{x}} = 21.1$ $\quad \sigma_{\bar{x}} = \frac{5.1}{\sqrt{50}} = .7212$

10% condition ok since there are more than 500 ACT test takers.

$P(\bar{x} > 23) = .0043$

normalcdf(23, 1E99, 21.1, .7212)

* NOTE $Z = \frac{23 - 21.1}{.7212} = 2.63$

almost 3 std dev from $\bar{x}$

SECTION 8.3B

57. What does the CLT say? Asked what the central limit theorem says, a student replies, "As you take larger and larger samples from a population, the histogram of the sample values looks more and more Normal." Is the student right? Explain your answer.

No! The histogram of the samples values will look like the distribution of the population, whatever it might happen to be.

The CLT says that the histogram of the sampling distribution of sample means (from many large samples) will look more and more Normal

59. Songs on an iPod Refer to Exercise 49.

(a) Explain why you cannot safely calculate the probability that the mean play time $\bar{x}$ is more than 4 minutes (240 seconds) for an SRS of 10 songs.

(b) Suppose we take an SRS of 36 songs instead. Explain how the central limit theorem allows us to find the probability that the mean play time is more than 240 seconds. Then calculate this probability. Show your work.

(a) Since the distribution of the play times of the population of songs is heavily skewed to the right, a sample size of 10 will not be large enough for the Normal approximation to be appropriate (samples must be 30 or more)

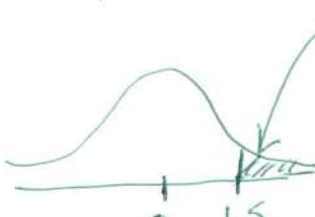
(b) With a sample size of 36 (more than 30), we now have enough observations in our sample for the CLT to apply. AND can find the probability and use the normal distribution

$P(\bar{X} > 240) =$

find
$\mu_{\bar{x}} = 225$

$\sigma_{\bar{x}} = \frac{60}{\sqrt{36}} = 10$

$z = \frac{240 - 225}{10} = 1.5$



$P(z > 1.5) = .0668$

normalcdf (1.5, E99, 0, 1)

8.3B CONT

60. **Lightning strikes** The number of lightning strikes on a square kilometer of open ground in a year has mean 6 and standard deviation 2.4. (These values are typical of much of the United States.) The National Lightning Detection Network (NLDN) uses automatic sensors to watch for lightning in a random sample of 10 one-square-kilometer plots of land.
(a) What are the mean and standard deviation of $\bar{x}$, the sample mean number of strikes per square kilometer?
(b) Explain why you cannot safely calculate the probability that $\bar{x} < 5$ based on a sample of size 10.
(c) Suppose the NLDN takes a random sample of $n = 50$ square kilometers instead. Explain how the central limit theorem allows us to find the probability that the mean number of lightning strikes per square kilometer is less than 5. Then calculate this probability. Show your work.

(a) $\bar{X}$ = mean number of strikes per square kilometer

$\mu_{\bar{x}} = 6$ strikes/km²

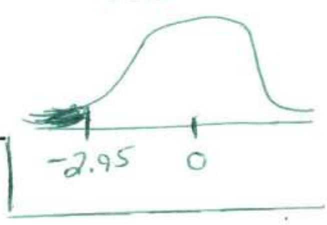
$\sigma_{\bar{x}} = \frac{2.4}{\sqrt{10}} = .7589$ strikes/km²

(b) We cannot calculate the probability because we do not know the shape of the distribution of the number of lightning strikes. If we were told the population is Normal then we could do it.

(c) With a sample size of 50, the CLT assures us the Normal approximation is valid for the sampling distribution of $\bar{X}$. $\mu_{\bar{x}} = 6$ $\sigma_{\bar{x}} = \frac{2.4}{\sqrt{50}} = .3394$

$P(\bar{x} < 5) = P\left(z < \frac{5-6}{.3394}\right) = P(z < -2.95) = .0016$

normalcdf(−E99, −2.95, 0, 1)



63. **More on insurance** An insurance company knows that in the entire population of homeowners, the mean annual loss from fire is $\mu = \$250$ and the standard deviation of the loss is $\sigma = \$300$. The distribution of losses is strongly right-skewed: many policies have \$0 loss, but a few have large losses. If the company sells 10,000 policies, can it safely base its rates on the assumption that its average loss will be no greater than \$275? Follow the four-step process.

Population $\mu = \$250$ $\sigma = \$300$

Conditions

① 10% Condition is met assuming the company has 100,000+ policies.

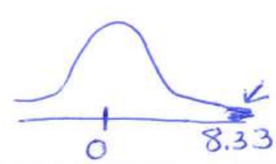
② Since sample size is large (10,000 ≥ 30), we can safely use the Normal distribution as an approximation for the sampling distribution of $\bar{X}$.

$\mu_{\bar{x}} = \$250$

$\sigma_{\bar{x}} = \frac{300}{\sqrt{10{,}000}} = \$3$

$P(\bar{x} > \$275) = P(z > 8.33) \approx 0$

$z = \frac{275-250}{3} = 8.33$



Conclusion: It is very, very, unlikely that the company would have an average loss of more than \$275.

8.3 B CONT

Exercises 69 to 72 refer to the following setting. In the language of government statistics, you are "in the labor force" if you are available for work and either working or actively seeking work. The unemployment rate is the proportion of the labor force (not of the entire population) who are unemployed. Here are data from the Current Population Survey for the civilian population aged 25 years and over in a recent year. The table entries are counts in thousands of people.

Highest education	Total population	In labor force	Employed	UNEMPLOYED
Didn't finish high school	27,669	12,470	11,408	1,062
High school but no college	59,860	37,834	35,857	1,977
Less than bachelor's degree	47,556	34,439	32,977	1,462
College graduate	51,582	40,390	39,293	1,097
	186,667	125,133		

69. **Unemployment (1.1)** Find the unemployment rate for people with each level of education. How does the unemployment rate change with education?

70. **Unemployment (5.1)** What is the probability that a randomly chosen person 25 years of age or older is in the labor force? Show your work. ~ 67%

71. **Unemployment (5.3)** If you know that a randomly chosen person 25 years of age or older is a college graduate, what is the probability that he or she is in the labor force? Show your work. ~ 78%

72. **Unemployment (5.3)** Are the events "in the labor force" and "college graduate" independent? Justify your answer.

These are review questions on probability

(69) Calculate unemployment for each education Level

P(unemployed | didn't finish HS) = $\frac{1062}{12470} = .0852$

P(unemployed | HS but no college) = $\frac{1977}{37834} = .0523$

P(unemployed | less than BS) = $\frac{1462}{34439} = .0425$

P(unemployed | college grad) = $\frac{1097}{40,390} = .0272$

Conclusion unemployment rates decrease with education level

$\boxed{70}$ P(in labor force) = $\frac{125133}{186667} = .6704$

$\boxed{71}$ P(in labor force | college grad) = $\frac{40,390}{51,582} = .7830$

$\boxed{72}$ NOT INDEPENDENT SINCE P(in labor force) ≠ P(in labor force | college grad)

.6704 ≠ .7830