Detailed Solutions to Selected Problems

Chapter 1

1 For rating points, the population is all television-equipped households. For share, the population is all households watching TV at that particular time-slot.

3a The population is all high school seniors in the United States. The sample is the 2000 students who participated in the survey.

4a This is a **parameter** because the school registrar should be able to average the GPA of All seniors, not just a sample

4b Since it is really impossible to get a yes/no answer from every Vermont teenager this quote must have come from sample data and it is therefore a statistic

4c This is a **statistic** because it comes from a sample

4d While the Bureau of Labor and Statistics do conduct very good sampling practices, these figures are based on sample data. Therefore this is a statistic.

6a Answer: Quantitative, Continuous, Ratio
With this type of data, differences are meaningful and so are ratios, for example a car that gets 40 mpg gets twice as many mpg's as a car that gets 20 mpg. Even though zero does not have much meaning, it is not arbitrary. This is a ratio level of measurement.

6c Answer: Qualitative, Nominal

6e Answer: Quantitative, Discrete, Interval
This data is quantitative (numerical year value), and discrete (years only come in whole numbers). This is one of those cases where differences are meaningful but ratios are not (the other common case is temperatures in Fahrenheit or Celsius). As such, the level of measurement is interval.

7 Today, money is more of a number on a computer, and as such could be considered continuous because you can have numerous decimal places. Technically, however, money does have a lowest denomination - that is the penny and the variable is discrete.

8a Neither. Males have a 20/100 chance of being selected and females have a 20/200 chance. Everyone does not have the same probability of being selected so it is not random.

8c This is simple random. All students have the same chance of being selected (random) and no sample of size 20 has a better or worse chance of being selected (simple random).

9a Answer: Cluster, Random
This is a cluster sample, if a certain team (cluster) is chosen, then everyone on that team is in the sample. It is a random sample because all players have a 2 in 30 chance of being in the sample.

9c Answer: Stratified, Probably Not Random
This is a stratified sample with the strata being male and female. It is probably not random unless there are an equal number of males and females in the collection of all my students.

9d Answer: Convenience, Not Random
This is a convenience sample. The surveyor only selects those that happen to come out when the current survey is completed. The participants are determined by when they walk out of the theater. This is not random.

10a This is a little slanted because it doesn't tell if the number of cell phone users who get brain cancer is any greater than the number of non-cell phone users.

11 The tax rate did double. It did increase by 5 percentage points. Since the difference is 5%, that is 100% of the original tax rate. So, the tax rate increased by 100%. The only false statement is (b) *The tax rate increased by 200%.*

13 The difference in pricing is $20. With respect to the jeans, this is 20/100 = .20 or a 20% difference. With respect to the khakis, this is 20/80 = .25 or a 25% difference. So the two correct answers are
The jeans cost 25% more than the khakis.
The khakis cost 20% less than the jeans.

15 In a 10 pound bag of Doug's there are 9.5 pounds of actual grass seed. In a 10 pound bag of generic there are 8 pounds of grass seed. So there is 1.5 more pounds of grass seed in a 10 pound bag. In terms of a percentage of the amount in the generic brand, this is 1.5/8 = .1875. So the answer is
<u>Doug's grass seed contains 18.75 % more actual grass seeds than the generic brand.</u>
This is greater than the 15% more grass seed than you might first expect.

Chapter 2

1

- mean: $$\bar{x} = \frac{4+8+4+6}{4} = \frac{22}{4} = \mathbf{5.5}$$
- median: First order them: 4, $\boxed{4,\ 6}$, 8 and take the average of the middle two $= \frac{4+6}{2} = \mathbf{5}$.
- mode: The most frequently occurring value is **4**.
- range: Max - Min = 8 - 4 = **4**.
- sample variance: See chart below $s^2 = \mathbf{3.7}$
- sample standard deviation: See chart below $s = \mathbf{1.9}$

x	$x-\bar{x}$	$(x-\bar{x})^2$
4	4 - 5.5 = -1.5	2.25
4	4 - 5.5 = -1.5	2.25
6	6 - 5.5 = 0.5	0.25
8	8 - 5.5 = 2.5	6.25
		11

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{11}{3}} =$$
$$\sqrt{3.6667} = 1.9148 \to 1.9$$

3

- mean: $$\bar{x} = \frac{-1.5+2.8+3.4-3.5+7.6-12.1}{6} = \frac{-3.3}{6} \approx -0.55$$
- median:
  Ordered: -12.1, -3.5, $\boxed{-1.5,\ 2.8}$,3.4, 7.6 and take the average of the middle 2 $= \frac{-1.5+2.8}{2} = \frac{1.3}{2} = \mathbf{0.65}$.
- mode: There is no value that occurs more than once $\to$ **no mode**.

- range: Max - Min = 7.6 - (-12.1) = **19.7**.
- sample variance: See chart below $s^2 = \mathbf{47.25}$
- sample standard deviation: See chart below $s = \mathbf{6.87}$

x	$x - \bar{x}$	$(x - \bar{x})^2$
-12.1	-12.1 - (-.55) = -11.55	133.4025
-3.5	-3.5 - (-.55) = -2.95	8.7025
-1.5	-1.5 - (-.55) = -0.95	0.9025
2.8	2.8- (-.55) = 3.35	11.2225
3.4	3.4- (-.55) = 3.95	15.6025
7.6	7.6- (-.55) = 8.15	66.4225
		236.255

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{236.255}{5}} =$$
$$\sqrt{47.251} = 6.873936 \to 6.87$$

5 If you look at the group with less than 5 years experience, the female average is \$24/hr and the male average is \$23/hr. If you look at the group with more than five years experience, the female average is \$32/hr and the male average is \$31/hr. In both sub-groups, the females have a higher average. The *lurking variable* is experience. Most of the people who have more than 5 years of experience are men, and those people get paid more money. It is always a little dangerous to average across categorical data and can lead to very misleading conclusions.

7a This would change everything except the median and mode. The mean would increase as would the standard deviation and variance.

7b The mean, median, and mode would remain the same. However the standard deviation and variance would increase.

7c The mean and median would drop, there would be no mode, and the standard deviation and variance would increase.

9a $z = \dfrac{65.5 - 69.3}{2.8} = -1.36$. Not unusual.

9b $z = \dfrac{70.2 - 69.3}{2.8} = 0.32$. Not unusual.

9c $z = \dfrac{74.0 - 69.3}{2.8} = 1.68$. Not unusual.

9d $z = \dfrac{78.0 - 69.3}{2.8} = 3.11$. Unusual.

11a The z-score for her height with respect to top female models is $z = \dfrac{71.0 - 70}{2.2} = 0.45$. So, her height is not unusual with respect to top female models.

11b The z-score for her height with respect to U.S. women is $z = \dfrac{71.0 - 64}{2.8} = 2.50$. So, her height is unusual with respect to U.S. women.

11c The z-score for her weight with respect to top female models is $z = \dfrac{115 - 115}{18} = 0.00$. So, her weight is not unusual with respect to top female models.

11d The z-score for her weight with respect to U.S. women is $z = \dfrac{115 - 145}{32} = -0.94$. So, her weight is not unusual with respect to U.S. women.

13 Gisele's z-score with respect to top models is $z = \dfrac{71.0 - 70}{2.2} = 0.45$. Tom's z-score with respect to NFL quarterbacks is $z = \dfrac{76.0 - 76.5}{1.8} = -0.28$. So with respect to their peers, Gisele is taller.

16a Yes. The z-score for the number of eggs from this particular Wolf spider is $z = \frac{500-302}{48} = 4.125$ This makes the number of eggs from this particular Wolf spider very unusual.

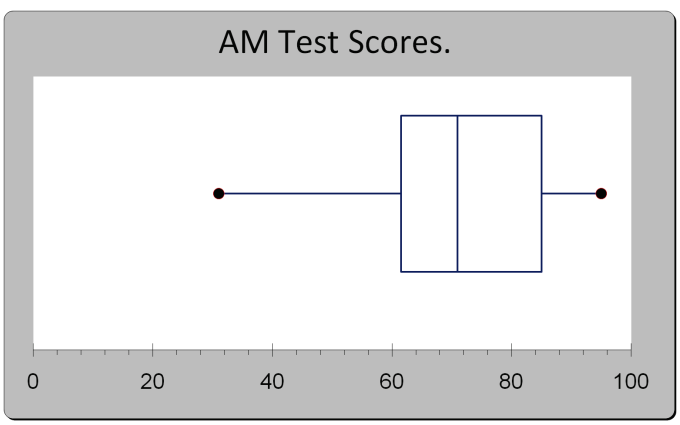
16b Perhaps my estimate was wrong. I didn't actually count the spiders and it's quite possible that when you see 300 baby spiders, you tend to over-estimate. Also, I'm not a spider expert and so it is possible that this was some other breed of spider. Or, perhaps this was a very unusual Wolf spider that laid an extraordinary number of eggs.

17(a)i $i = (90/100) \cdot 22 = 19.8 \to 20$. So P_{90} is the 20'th score $= \boxed{92}$.

17(a)ii

5-number summary		
min	**31**	
Q_1	**61**	$i = 5.5 \to 6$
Q_2	**71**	median
Q_3	**87**	$i = 16.5 \to 17$
max	**95**	

17(a)iii



17c The morning section has a lower median (71 versus 78.5) and has the lowest minimum score (31 versus 45). The different P_{90} scores suggests it is harder to get into the top 10% of my PM section. So, the AM section seems to do worse than the PM section (Sleepy Students?). However, the middle 50% of the scores in both sections are about the same at 61 to 87 (AM) and 63 to 87 (PM). So the difference is not as drastic as you might first think.

18a Sam's GPA is calculated as a weighted average by completing the table below.

Credits (w)	Letter Grade	Numerical Grade (x)	$w \cdot x$
3	B	3.0	9.0
1	A	4.0	4.0
3	C	2.0	6.0
6	D	1.0	6.0
3	A	4.0	12.0
16			37

$$\text{GPA} = \frac{\sum(w \cdot x)}{\sum w} = \frac{37}{16} = \mathbf{2.31}$$

19 This requires a weighted average where the number of students are the weights (w's) and the class averages are the values (x's):

$$\bar{x} = \frac{\sum(w \cdot x)}{\sum w} = \frac{(8 \cdot 88) + (16 \cdot 74) + (30 \cdot 72)}{8 + 16 + 30} = \frac{4048}{54} = 74.923 \approx \mathbf{75}$$

20a

Days	Transaction	balance (x)	# days (w)	$w \cdot x$
1-6	remaining balance	\$1200	6	7200.00
7-10	\$400 purchase	\$1600	4	6400.00
11-20	\$300 purchase	\$1900	10	19000.00
21-30	\$1000 payment	\$900	10	9000.00
	totals		30	\$41,600.00

average daily balance:

$$\bar{x} = \frac{\sum(w \cdot x)}{\sum w} = \frac{41600}{30} \approx \$1386.67$$

21a Within the extremely obese category, the mean weight loss by the exercise plan was 3 pounds greater than the diet plan. Within the moderately obese category, the mean weight loss by the exercise plan was again 3 pounds greater than diet plan. So, the **exercise plan** seems to be more effective.

21b Exercise Only - Weighted Average: $\bar{x} = \dfrac{\sum(w \cdot x)}{\sum w} = \dfrac{(22 \cdot 5) + (16 \cdot 25)}{30} = \mathbf{17}$

Diet Only - Weighted Average: $\bar{x} = \dfrac{\sum(w \cdot x)}{\sum w} = \dfrac{(19 \cdot 25) + (13 \cdot 5)}{30} = \mathbf{18}$

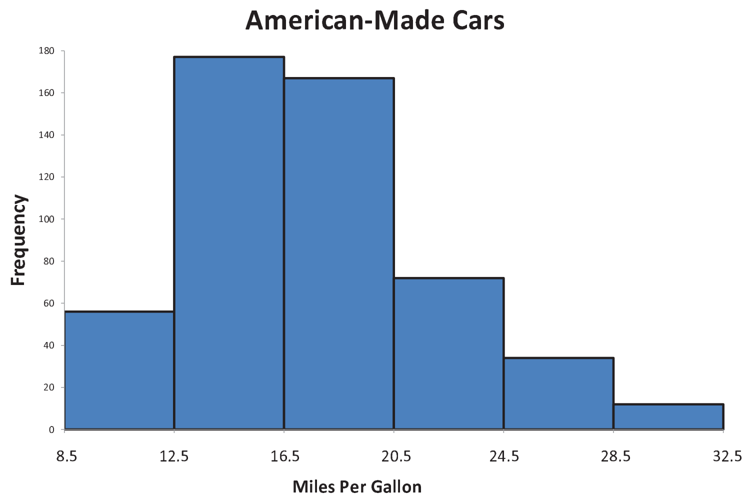
Now, it seems as the **diet plan** is better.

21c The *diet plan* had a greater number of extremely obese participants, and that category lost the most weight.

Chapter 3

1a The class midpoints are 10.5, 14.5, 18.5 …, 30.5. The class boundaries are 12.5, 16.5, 20.5, …, 28.5. The class width is 4.

1b Below is the histogram for the frequency distribution for MPG of American made cars.



1c Below is the relative frequency table for American made cars. Each relative frequency is found by taking the actual frequency and dividing it by 518.

American-Made Cars

MPG	Relative Frequency
9 - 12	10.8%
13- 16	34.2%
17 - 20	32.2%
21 - 24	13.9%
25 - 28	6.6%
29 - 32	2.3%

1d Below is the cumulative and relative-cumulative frequency distributions for American made cars.

American-Made Cars

MPG	Cumulative Frequency	Relative Cumulative Frequency
less than 12.5	56	10.8%
less than 16.5	233	45.0%
less than 20.5	400	77.2%
less than 24.5	472	91.1%
less than 28.5	506	97.7%
less than 32.5	518	100.0%

1e

MPG	Frequency (f)	Class Midpoint (x)	f · x
9 - 12	56	10.5	588.0
13- 16	177	14.5	2566.5
17 - 20	167	18.5	3089.5
21 - 24	72	22.5	1620.0
25 - 28	34	26.5	901.0
29 - 32	12	30.5	366.0
Totals	$\sum f = 518$		$\sum(f \cdot x) = 9131$

The estimate of the mean is

$$\bar{x} = \frac{\sum(f \cdot x)}{\sum f} = \frac{9131}{518} = \mathbf{17.6}$$

4 The distribution is a little more normal than the others. The average seems to be centered somewhere between 16 and 20 MPG. This puts the German-made cars right between the American and Japanese-made cars in terms of fuel efficiency. More detailed data would be needed to make a better comparison.

5 The mean score from a frequency table is found by completing the chart below:

Score	Frequency (f)	Class Midpoint (x)	$f \cdot x$
60-64	8	62	496
65-69	4	67	268
70-74	2	72	144
75-79	1	77	77
80-84	1	82	82
Totals	$\sum f = 16$		$\sum(f \cdot x) = 1067$

So, $\bar{x} = \dfrac{\sum(f \cdot x)}{\sum f} = \dfrac{1067}{16} = 66.6875 \to \boxed{66.7}$.

7a This should result in a bimodal distribution because men and women would have different average sizes.

7b I would expect this distribution to be fairly normal

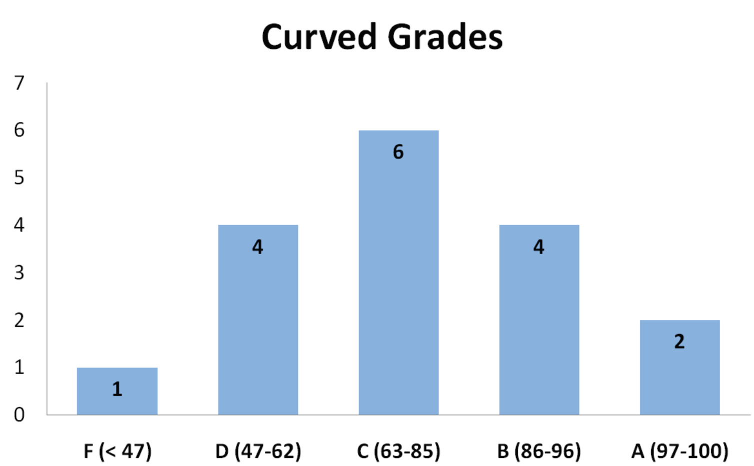
7c I would expect this distribution to be skewed left to account for the children's sizes.

7d I would expect this distribution to be fairly uniform.

10 The best I could do is indicated in in the bar-graph to the right.

Winners (6):
3 D's are now C's
3 F's are now D's

Losers (5):
2 B's are now C's
3 A's are now B's



This is hardly fair to the 5 people who's grade dropped due to the curve.

12a Pirates were huge in the early 1900's while mutants were barely mentioned. Around 1945 (radiation days) mutant popularity sky-rocketed passing pirates around 1970. Around 1980 mutants started to diminish and pirates started to recover. Around 2003 (release of *Pirates of the Caribbean* and headline news about real pirates off the coast of Africa) pirates resumed their popularity over mutants.

13 In the years 2007 to the beginning of 2010, there was a trend where concerns for the environment decreased as concerns for energy increased. This could be due to the increasing cost of oil, an economic downturn, and uneasy relationships with oil-producing nations. In April of 2010, the Deepwater Horizon oil spill occurred. This spill is the largest offshore spill in U.S. history and clearly changed peoples opinions as to the priorities regarding the environment and energy. However, that change was short-lived and by March of 2011 the opinions were back to where they were before the crisis. In 2012, there is a pretty even split of opinion. **Trick:** The differences are exaggerated by starting the y-axis at 30% instead of zero.

Chapter 4

1a The sample space is {bb, bg, gb, gg}.

1b There are 4 possible outcomes and two of these have one boy and one girl. So the probability of having one boy and one girl is $\frac{2}{4} = 0.5$.

1c There are 4 possible outcomes and three of these will have at least one girl. So the probability of having at least one girl is $\frac{3}{4} = 0.75$.

1d There are 4 possible outcomes and only one has no girls in it. So the probability is $\frac{1}{4} = 0.25$.

3a There are 1,000 tickets and only 1 is the grand prize ticket. So your probability is $1/1000 = 0.001$.

3b There are 1,000 tickets and 8 of these will produce some type of prize. So your probability is $8/1000 = 0.008$

3c It is tricky because any one ticket could fall into more than one category. For example, your ticket could be a grand prize winner **and** a small prize winner.

5a $923/1567 = 0.589$.

5b Based on the historical data and using the relative frequency approach to estimating a probability, the estimated probability would again be 0.589.

7 $P(\text{next one sold fails}) \approx \dfrac{24}{4000} = \mathbf{0.006}$

9a The prediction was correct $102 + 205 = 307$ times and it was wrong $18+40 = 58$ times.

9b There are 365 days and the prediction was correct 307 times so the probability is $307/365 = 0.841$.

9c Using the historical data and the relative frequency approach to estimating probabilities, the estimated probability is $307/365 = 0.841$.

10a This not a conditional probability. The prediction was correct 307 times out of 365. So the probability is $\frac{307}{365} \approx 0.841$.

10b Since we are given that it was predicted to rain, there are only 120 options in our sample space. Of these, it rained in 102 of these cases. So the probability is $\frac{102}{120} = 0.850$.

10c Since we are given that it was predicted not to rain, there are only 245 options in our sample space. Of these, it did not rain 205 of these cases. So the probability is $\frac{205}{245} \approx 0.837$.

10d Comparing the two previous answers, the forecast is better at predicting rain because there is an 85.0% success rate at this but only an 83.7% success rate at predicting no rain.

12a If you know that he has no aces then there are 4 aces left in the deck which now only has 44 cards in it. So, the probability that you get an ace is $4/44 = .0909$.

13a Since you are already holding four cards, this is a conditional probability. There are 48 cards left but only 9 hearts (because you have 4 of them). Therefore $\mathrm{P}(\heartsuit) = 9/48 = \mathbf{0.1875}$.

13b Since you are already holding four cards, this is a conditional probability. There are 48 cards left and you would be happy with a 2 or a 7 of any suit. Since there are four 2's and four 7's left in the deck, there are eight ways to get what you want and $P(\text{straight}) = 8/48 \approx \mathbf{0.167}$.

14a Mutually exclusive, you can not roll a 6 and a 2 on a single roll.

14d Not mutually exclusive. It is possible (and likely) that a vegetarian meal will contain vegetables.

15a There are two ways you can do this. You can get a 1 on the red die and a 2 on the white **or** a 2 on the red die and a 1 on the white. These events are mutually exclusive. So, $P(R1 \text{ and } W2 \text{ or } R2 \text{ and } W1) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36} \approx \mathbf{0.0556}$.

15b You could use the addition rule and count all the different ways to roll a total not equal to 3. However, it is a lot easier to use the compliments rule. If $A =$ a total that is not 3, then $\bar{A} =$ a total of 3. You found $P(\bar{A})$ in the last problem. So, $P(A) = 1 - P(\bar{A}) = 1 - \frac{2}{36} = \frac{34}{36} \approx \mathbf{0.944}$.

16 Here is the same table with the totals calculated:

	Has a ski pass at					
	Bolton	Stowe	Smuggler's	Sugarbush	No Pass	**Totals**
Has a Car	18	12	10	30	25	95
Does not have a Car	24	5	4	12	60	105
Totals	42	17	14	42	85	200

16a There are 42 students with a pass at Bolton, 17 at Stowe, 14 at Smuggler's, 42 at Sugarbush, and 85 with no pass. Totalling these numbers you get 200. Since there are only 200 students all together you can conclude that none of them have more than one pass otherwise your total would have been greater than 200.

16b Since these are mutually exclusive events (from the previous problem), P(Stowe **or** Sugarbush) = P(Stowe) + P(Sugarbush) = $17/200 + 42/200 = 59/200 = \mathbf{0.295}$.

16c Since these these are not mutually exclusive, P(car **or** Sugarbush) = P(car) + P(Sugarbush) - P(car **and** Sugarbush) = 95/200 + 42/200 - 30/200 = 107/200 = **0.535**.

16d If A = car or ski pass, then $\bar{A}$ = no car **and** no pass. $P(\bar{A}) = 60/200$, so $P(A) = 1 - 60/200 = 140/200 =$ **0.700**.

18a Since the drawing is done without replacement, the probability of getting a Jack on the second draw is affected by whether or not you got a Jack on the first draw. Thus, the events are **dependent**.

18b If you have no idea how much of the store's milk supply is spoiled, then finding the first one is spoiled increases the probability that the second one is spoiled. If you somehow know exactly how many gallons are spoiled and how many are not, then selecting a spoiled one might decrease the probability that the second one is spoiled (there is one fewer spoiled gallon remaining when you pick the second one). Either way, the events are **dependent**.

18c Recognizing that this is a fair coin, getting *heads* on the first toss does not change the probability of getting *heads* on the second toss. The events are **independent**.

19a We use the multiplication rule for independent events: P(Jack then Jack) = P(Jack) · P(Jack) = 4/52 * 4/52 = 0.00592.

19b We use the multiplication rule for independent events: P(Jack then Queen) = P(Jack) · P(Queen) = 4/52 · 4/52 = 0.00592.

19c We use the multiplication rule for independent events: P(Jack then Heart) = P(Jack) · P(Heart) = 4/52 * 13/52 = 1/52 = 0.0192.

21b The day of the first person does not matter. The probability that the second person has the same day is 1/7 and the probability that the third person has the same day is 1/7 and the same goes for the fourth person. So, the probability that all 4 have were born on the same day of the week is $\left(\frac{1}{7}\right)^3 = 0.00292$.

22a There are two ways for this to happen. The first could be red and the second yellow **or** the first is yellow and the second is red. Since these are mutually exclusive events, we add the probabilities. P({red and yellow} **or** { yellow and red}) $= \frac{3}{12}\frac{5}{12} + \frac{5}{12}\frac{3}{12} = 0.208$.

22c This is a natural extension of the multiplication rule with dependent events. P(yellow and yellow and yellow and yellow and yellow) $= \frac{5}{12}\frac{4}{11}\frac{3}{10}\frac{2}{9}\frac{1}{8} = 0.00126$.

23a Since we are told these are two different senators, the sampling is done without replacement and the events are dependent. So, P(male republican **and** male republican) $= \frac{42}{100}\frac{41}{99} = 0.174$

23b Since we are told these are two different senators, the sampling is done without replacement and the events are dependent. So, P(democrat **and** republican) $= \frac{51}{100}\frac{47}{99} = 0.242$

24a The probability of getting a spark plug that is not defective is 1 - 0.02 = 0.98.

24b The probability that all 4 spark plugs are not defective is $(0.98)^4 = 0.922$.

24c The complement of *at least one defective* is none defective or all are defect-free. The probability that all are defect free was found in part b as 0.922. So the probability that at least one is defective is 1-.922 = 0.078.

26 The complement of the event *at least one* is none. In this case, none means that that the lie detector accurately detects all 10 lies. The probability of this is $(0.95)^{10} = 0.599$. But this is the probability that all 10 lies are detected. If this doesn't happen then at least one lie went undetected. So the probability of at least one lie going undetected is 1 - 0.599 = 0.401.

28a You need all three to have jumper-cables so the probability is $(.25)^3 = 0.015625 \approx \mathbf{0.0156}$

28b Let A = at least one has jumper cables. Then $\bar{A}$ = none have jumper cables.
$P(\bar{A}) = (.75)^3$, so $P(A) = 1 - (.75)^3 = 0.578125 \approx \mathbf{0.578}$.

28c You don't need all three to have jumper-cables. You only need one or, more precisely, at least one of them, to have jumper-cables. So the second probability is more relevant.

Chapter 5

1a The probabilities are all between 0 and 1, and they sum to 1, so this is a probability distribution.

1b The mean value is $\mu = \sum(x \cdot P(x))$ (see table below) and the mean is **2**.

x = # of heads	$P(x)$	$x \cdot P(x)$
0	1/16	0
1	4/16	4/16
2	6/16	12/16
3	4/16	12/16
4	1/16	4/16
sum (Σ)		32/16

1c From the last problem we see that the expected value is just the mean of the probability distribution = **2** heads. You probably could have guessed this without all the math.

3a Each probability is between 0 and 1, and the probabilities add to 1, so this is a probability distribution.

3b From the table below, you can see that the expected value is -1450/200 = -7.25. So the expected value of this raffle to me is **-$7.25**.

Outcomes	value = x	$P(x)$	$x \cdot P(x)$
Win Grand Prize	190	1/200	190/200
Win a Second Prize	90	2/200	180/200
Win a Third Prize	40	3/200	120/200
Win Nothing	-10	194/200	-1940/200
Sum (Σ)			E = -1450/200

5 If you had bought the warranty, your cost would have been $160. If you don't buy it, you have to determine your expected cost.

Outcomes	cost = x	$P(x)$	$x \cdot P(x)$
the phone doesn't fail	150	0.95	142.50
the phone fails	300	0.05	15.00
Sum (Σ)			E = 157.50

Since the expected cost of \$157.50 without the warranty is less than \$160 with the warranty, you made the right decision to not buy the warranty. When you spend the extra \$10 for the warranty you are really getting \$2.50 worth of security or hassle-free replacement. This could be thought of as the profit the company makes from selling the warranty.

7a This is probably **not a binomial probability distribution** because there is more than one type of response. Either there is more than two candidates or some people might not have voted for anybody.

7b This **does not result in a binomial probability distribution** because there is not a fixed number of trials.

9a Using the table with n = 10, x = 6, and p = .8, I get the probability = **0.088**

9b This value of n is not in the table so you have to do it some other way.

$$\begin{aligned}
\frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x} &= \frac{17!}{2!\,15!} \cdot (.8)^{15} \cdot (.2)^2 \\
&= \frac{17 \cdot 16 \cdot 15!}{2 \cdot 1 \cdot 15!} \cdot (.8)^{15} \cdot (.2)^2 \\
&= 136 \cdot (.8)^{15} \cdot (.2)^2 = 0.1914030 \rightarrow \mathbf{0.191}
\end{aligned}$$

9c Using the table with n = 15 and p = 0.2, sum the results for all values of x that are less than 4: = .250 + .231 + .132 + .035 = **0.648**.

9d Using the table with n = 8 and p = .4 and using complements: sum the results for x values of 0 and 1 (0.017 + 0.090 = 0.107), and subtract this answer from 1: 1 - 0.107 = **0.893**. If you had summed all of the probabilities for $x \geq 2$ you would get **0.894**. Either answer is fine.

10 This is a binomial probability with $x = 4$, $n = 10$, and $p = 0.25$ (there is a probability of 13/52 that you get a heart on any given draw). Unfortunately the value $p = 0.25$ is not in the table so we have to use the formula.

$$\begin{aligned}
\frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x} &= \frac{10!}{6!\,4!} \cdot (.25)^4 \cdot (.75)^6 \\
&= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{6! \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot (.25)^4 \cdot (.75)^6 \\
&= 210 \cdot (.25)^4 \cdot (.75)^6 = 0.145998 \rightarrow \mathbf{0.146}.
\end{aligned}$$

12 This is a binomial probability with $n = 10$ and $p = 0.2$ so we can use the table. The probability that $x \geq 5 = 0.026 + 0.006 + 0.001 + * + * + * \approx$ **0.033**. Since this value is less than 0.05 we conclude that 5 undetected lies would be an **unusually large** number.

15a This is a binomial probability with $n = 130$ and $p = 0.75$. The mean from such a distribution would be $\mu = n \cdot p = 130 \cdot 0.75 =$ **97.5** people. Yes, decimals are valid in describing the mean from such a distribution.

15b The standard deviation is given by $\sigma = \sqrt{npq} = \sqrt{130 \cdot 0.75 \cdot 0.25} = 4.9371044 \to \mathbf{4.9}$.

15c First, $n \cdot p = 97.5$ and $n \cdot q = 32.5$ which are both greater than 5 (good). Now, $z = \frac{x-\mu}{\sigma} = \frac{85-97.5}{4.9} \approx -2.55$ which is less than -2. So, 85 is an **unusual** number of survivors in groups of 130. It might be worth checking to see if this particular hospital is doing as much as it can to help patients with this type of cancer.

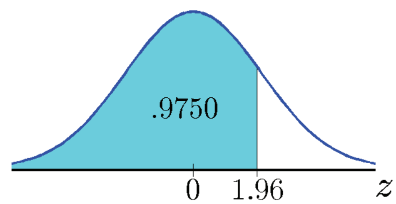
17a This is a binomial probability with $n = 124$ and $p = 0.05$. The mean from such a distribution would be $\mu = n \cdot p = 124 \cdot 0.05 = \mathbf{6.2}$ TV's.

17b First, $n \cdot p = 6.2$ and $n \cdot q = 117.8$ which are both greater than 5 (good).
To get the z-score of 16, we need the mean and standard deviation.
We have the mean. We need the standard deviation: $\sigma = \sqrt{n\,p\,q} = \sqrt{(124)(.05)(.95)} \approx 2.4$.
Now, $z = \frac{x-\mu}{\sigma} = \frac{16-6.2}{2.4} \approx 4.1$ which is way above 2.
So, this is **very unusual** if the 5% value that you got from the manufacturer is correct.

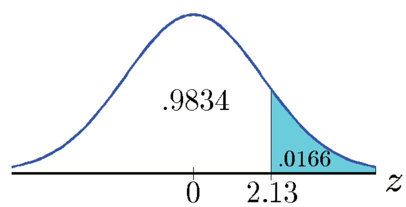
17c One, maybe you got very unlucky. Two, the 5% figure could be wrong. Maybe, the manufacturer can tell the good ones from the bad ones and sells all the bad ones to you.

Chapter 6

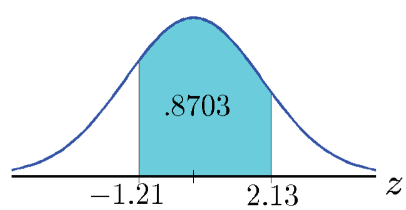
1a (i) $P(z < 1.96)$
Because of the $<$ (less than) sign, we are looking for the area to the left of 1.96. This is found straight from the z-table by using the row for 1.9 and the column for 0.06.
$\mathbf{P(z < 1.96) = 0.9750}$.



(ii) $P(z \geq 2.13)$
Because of the $\geq$ (greater than or equal to) sign, we are looking for area to the right of 2.13. This is found by getting the area to the left of 2.13 (from the z-table) and subtracting it from 1.
$\mathbf{P(z \geq 2.13) = 1 - P(z < 2.13) = 1 - 0.9834 = 0.0166}$

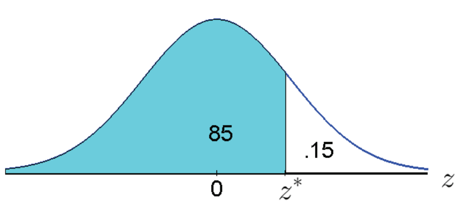


(iii) $P(-1.21 < z < 2.13)$
Because we are looking for the area between two z-scores we use
$P(z_1 < z < z_2) = P(z < z_2) - P(z < z_1)$
$P(-1.21 < z < 2.13) = P(z < 2.13) - P(z < -1.21)$
$P(-1.21 < z < 2.13) = 0.9834 - 0.1131 = \mathbf{0.8703}$



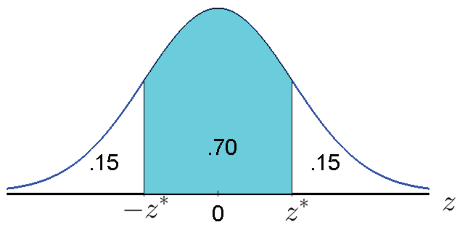
2a First, the area under a probability density curve = 1, so the area to the right of a given value is equal to one minus the area to the left. Second, since $z \leq z^*$ and $z > z^*$ are complementary events then the sum of the two probabilities must be one.

2b



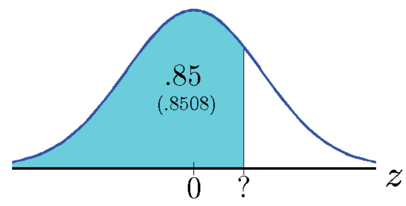
Since the mean of z-distribution = 0
then $P(z < 0) = 0.5$
Since $P(z < z^*) > 0.5$ then $z^* > 0$.
So, z^* is positive.

2c

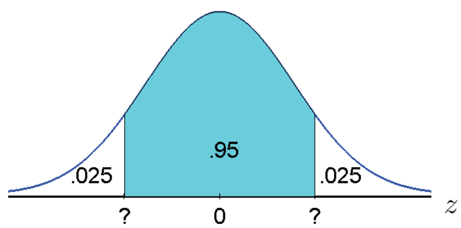


Since $P(z < z^*) = 0.85$ then $P(z > z^*) = 0.15$.
By symmetry $P(z < -z^*) = 0.15$.
So the area in the two tails combined is 0.30.
Therefore, the area between the two tails is **0.70**.

3a Here we are looking for a z-score so that the area below the curve to the left of this z-score is 0.85. So we look for 0.85 INSIDE the z-table. The closest value is 0.8508 corresponding to a z-score of **1.04**. (If you used software you should get 1.036.)



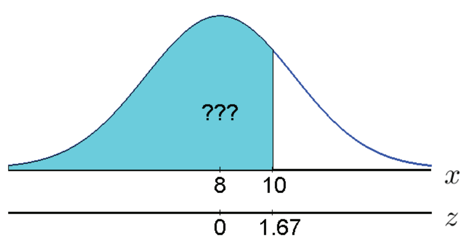
4a



Here we want the middle 95% or area = 0.95
So each tail contains half of 0.05 or 0.025.
Look for 0.025 INSIDE the z-table
Get a corresponding z-score of -1.96.
The left z-value is -1.96.
By symmetry, the right z-value is 1.96.

So the z-scores between **-1.96 and 1.96** constitute the middle 95% of the z-scores.

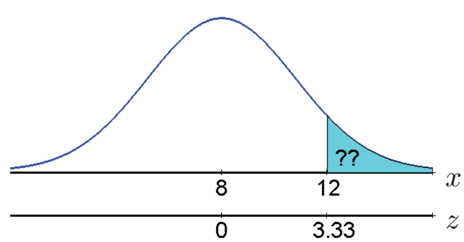
5a



If $x = 10$, then $z = \frac{x-\mu}{\sigma} = \frac{10-8}{1.2} \approx 1.67$

$$\begin{aligned} P(x < 10) &= P(z < 1.67) \\ &= \mathbf{0.9525} \quad \text{from } z\text{-table} \end{aligned}$$

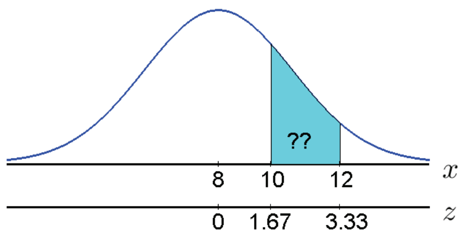
5b



If $x = 12$, then $z = \frac{x-\mu}{\sigma} = \frac{12-8}{1.2} \approx 3.33$

$$\begin{aligned} P(x > 12) &= 1 - P(x < 12) \\ &= 1 - P(z < 3.33) \\ &= 1 - .9996 \quad \text{from } z\text{-table} \\ &= \mathbf{0.0004} \end{aligned}$$

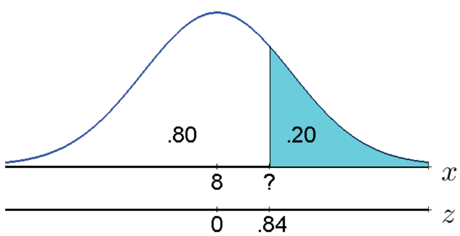
5c



If $x = 10$, then $z = \frac{x-\mu}{\sigma} = \frac{10-8}{1.2} \approx 1.67$
If $x = 12$, then $z = \frac{x-\mu}{\sigma} = \frac{12-8}{1.2} \approx 3.33$

$$\begin{aligned} P(10 < x < 12) &= P(x < 12) - P(x < 10) \\ &= P(z < 3.33) - P(z < 1.67) \\ &= 0.9996 - .9525 \quad \text{from } z\text{-table} \\ &= \mathbf{0.0471} \end{aligned}$$

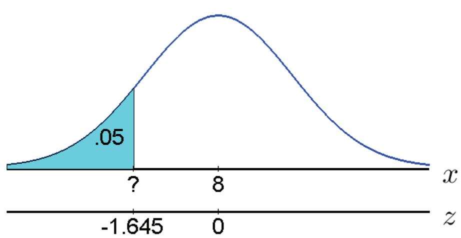
5d



We want 80% to the left and 20% right.
Look for 0.8000 INSIDE the z-table
The closest value is 0.7995
corresponding to $z = 0.84$.
x-value $= \mu + z\,\sigma = 8 + 0.84 \cdot 1.2 = \mathbf{9.008}$

So any potato that weigh more than about 9 ounces should be saved for the farmer's market.

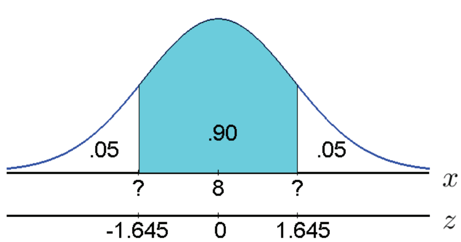
5e



We want 5% to the left and 95% right.
Look for 0.0500 INSIDE the z-table.
The corresponding z-value is $z =$ -1.645.
x-value $= \mu + z\,\sigma = 8 - 1.645 \cdot 1.2 = \mathbf{6.026}$

So any potato that weighs less than about 6 ounces should be saved for the potato launcher.

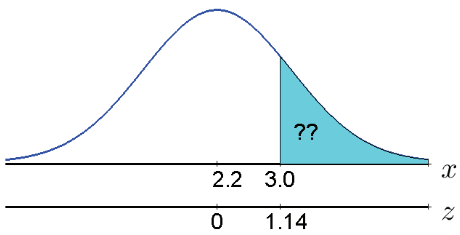
5f



Want 90% in the middle and 5% in each tail.
Look for 0.0500 INSIDE the z-table.
The left z-value is $z =$ -1.645.
By symmetry, the right z-value is 1.645.
If $z = -1.645$, then
$x = \mu + z\,\sigma = 8 - 1.645 \cdot 1.2 = \mathbf{6.03}$
If $z = 1.645$, then
$x = \mu + z\,\sigma = 8 + 1.645 \cdot 1.2 = \mathbf{9.97}$

The weights of the middle 90% of the potatoes fall **between 6.03 ounces and 9.97** ounces.

7a

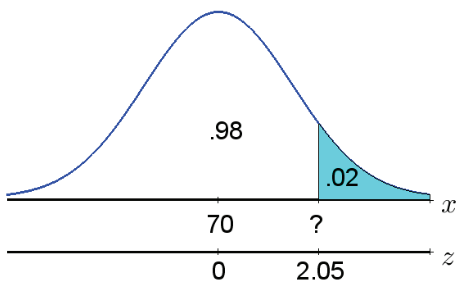


If $x = 76$, then $z = \frac{x-\mu}{\sigma} = \frac{76-70}{4} = 1.50$

$$\begin{aligned} P(x > 76) &= P(z > 1.50) \\ &= 1 - P(z < 1.50) \\ &= 1 - 0.9332 \quad \text{from } z\text{-table} \\ &= \mathbf{0.0668} \end{aligned}$$

So, about 6.68 or 7% of the cars are traveling faster than you.

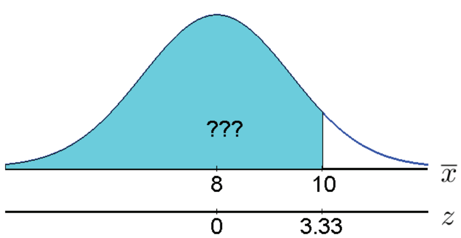
7b



You only want to ticket the top 2% (1/50)
You want 2% to the right and 98% to the left.
Look for 0.98 INSIDE the z-table.
The closest is 0.9798.
The corresponding z-value is $z = 2.05$.
x-value $= \mu + z\,\sigma = 70 + 2.05 \cdot 4 = \mathbf{78.2}$

Therefore, you should stop those cars traveling faster than 78.2 mph.

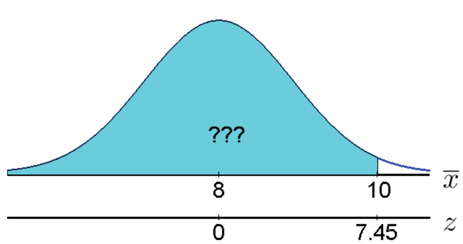
9a



If $\bar{x} = 10$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{10-8}{1.2/\sqrt{4}} \approx 3.33$

$$\begin{aligned} P(\bar{x} < 10) &= P(z < 3.33) \\ &= \mathbf{0.9996} \quad \text{from } z\text{-table} \end{aligned}$$

9b

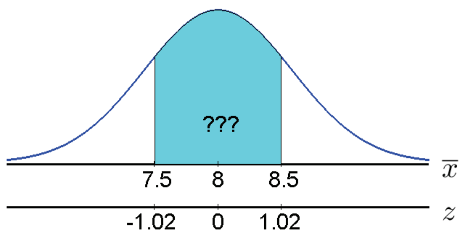


If $\bar{x} = 10$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{10-8}{1.2/\sqrt{20}} \approx 7.45$

$$\begin{aligned} P(\bar{x} < 10) &= P(z < 7.45) \\ &= \mathbf{0.9999} \quad \text{from } z\text{-table} \end{aligned}$$

In actuality the probability is greater than this but still < 1.

9c

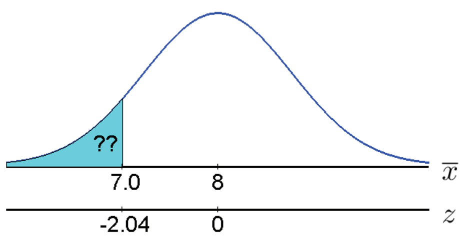


If $\bar{x} = 7.5$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{7.5-8}{1.2/\sqrt{6}} \approx -1.02$
If $\bar{x} = 8.5$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{8.5-8}{1.2/\sqrt{6}} \approx 1.02$

$$\begin{aligned} P(7.5 < \bar{x} < 8.5) &= P(-1.02 < z < 1.02) \\ &= P(z < 1.02) - P(z < -1.02) \\ &= 0.8461 - 0.1539 \quad \text{from } z\text{-table} \\ &= \mathbf{0.6922} \end{aligned}$$

Approximately 69% of these bags should have a mean potato weight of 7.5 to 8.5 ounces.

9d The mean weight of the potatoes in your bag is $42/6 = 7.0$ ounces. This is a mean weight of 1 ounce below the claimed mean. So I am already feeling a little cheated. How cheated? Find the probability of getting a mean less than the one I got.

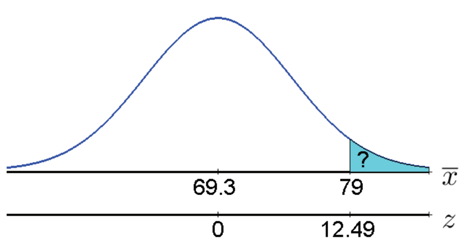


If $\bar{x} = 7.0$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{7-8}{1.2/\sqrt{6}} \approx -2.04$

$$\begin{aligned} P(\bar{x} < 7.0) &= P(z < -2.04) \\ &= \mathbf{0.0207} \quad \text{from } z\text{-table} \end{aligned}$$

As such, there is only a 2% chance of getting a randomly selected bag that weighs less than or equal to my bag. Now, I'm feeling extremely unlucky or cheated.

11a

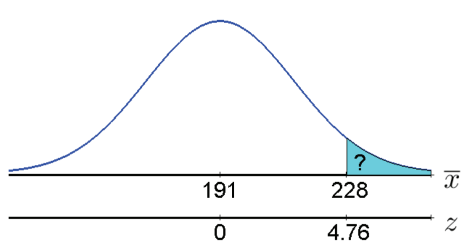


If $\bar{x} = 79.0$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{79.0-69.3}{2.8/\sqrt{13}} \approx 12.49$

$$\begin{aligned} P(\bar{x} > 79) &= P(z > 12.49) \\ &= 1 - P(z < 12.49) \\ &= 1 - 0.9999 \quad \text{from } z\text{-table} \\ &= \mathbf{0.0001} \end{aligned}$$

Actually, the probability is **much** smaller than this.

11b

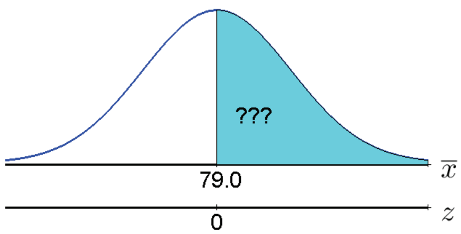


If $\bar{x} = 228$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{228-191}{28/\sqrt{13}} \approx 4.76$

$$\begin{aligned} P(\bar{x} > 228) &= P(z > 4.76) \\ &= 1 - P(z < 4.76) \\ &= 1 - 0.9999 \quad \text{from } z\text{-table} \\ &= \mathbf{0.0001} \end{aligned}$$

Actually, the probability is smaller than this.

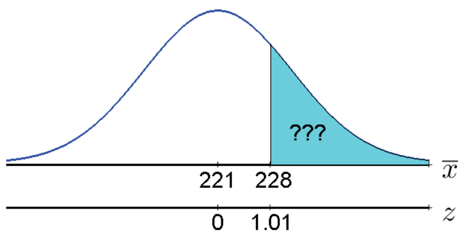
11c



If $\bar{x} = 79.0$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{79.0-79.0}{2.1/\sqrt{13}} = 0$

$$\begin{aligned} P(\bar{x} > 79) &= P(z > 0) \\ &= 1 - P(z < 0) \\ &= 1 - 0.5000 \quad \text{from } z\text{-table} \\ &= \mathbf{0.5000} \end{aligned}$$

11d



If $\bar{x} = 228$, then $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{228-221}{25/\sqrt{13}} \approx 1.01$

$$\begin{aligned} P(\bar{x} > 228) &= P(z > 1.01) \\ &= 1 - P(z < 1.01) \\ &= 1 - 0.8438 \quad \text{from } z\text{-table} \\ &= \mathbf{0.1562} \end{aligned}$$

11e With respect to U.S. men, the team is crazy tall and very heavy. With respect to NBA players, the Lakers have a mean height that is perfectly normal (equal to the population mean) while the mean weight is somewhat high. If 13 NBA players were randomly selected, there is only about a 16% chance that the collection would have a mean weight greater than that of the L.A. Lakers.

13 If the actual satisfaction rate is 48%, you want to calculate the probability of getting a sample 220 employees with 85 or fewer satisfied.

1. If $p = .48$ and $n = 220$ then

$$\mu = n \cdot p = 220 \cdot (.48) = 105.6 \qquad \text{and} \qquad \sigma = \sqrt{n\,p\,q} = \sqrt{220(.48)(.52)} = 7.4$$

.

2. Now, let $z^* = \dfrac{x^* - \mu}{\sigma} = \dfrac{85 - 105.6}{7.4} \approx -2.78$
3. And then, $P(x \leq 85) \approx P(z < -2.78) = \mathbf{0.0027}$
4. **Conclusion:** This is a very unusual number of satisfied employees. In random samples of size 220 you can expect less than 0.3% of those samples to contain 85 or fewer satisfied employees. Your group fell into this category. Their unusually low satisfaction rate is probably not due to random variation but more likely some outside influence.

15 If the actual uninsured rate is 16.6%, you want to calculate the probability of getting a sample 250 patients with 50 or more of them uninsured.

1. If $p = .166$ and $n = 250$ then

$$\mu = n \cdot p = 250 \cdot (.166) = 41.5 \qquad \text{and} \qquad \sigma = \sqrt{n\,p\,q} = \sqrt{250(.166)(.834)} = 5.9$$

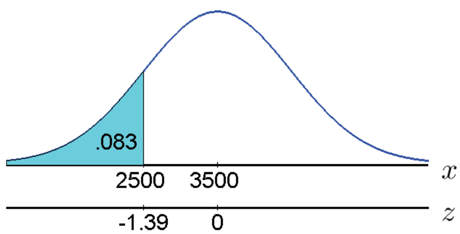
.

2. Now, let $z^* = \dfrac{x^* - \mu}{\sigma} = \dfrac{50 - 41.5}{5.9} \approx 1.44$

3. And then, $P(x \geq 50) \approx P(z \geq 1.44) = 1 - P(z \leq 1.44) = 1 - 0.9251 = \mathbf{0.0749}$

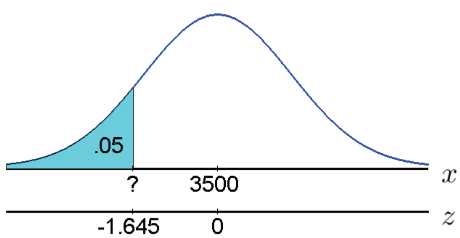
4. **Conclusion:** This is not particularly unusual. In random samples of size 250 you can expect about 7.5% of those samples would contain 50 or more uninsured. Your group fell into this category. Using the 5% cut-off rule for unusual, this is not an unusually large number of uninsured patients.

20a If the area to the left of 2500 is 8.3% or 0.083, find the z-score by looking for 0.083 INSIDE the z-table. The closest value is 0.0823 corresponding to a z-value of -1.39.



If $x = 2500$, then $z = \frac{\bar{x}-\mu}{\sigma} = \frac{2500-3500}{\sigma} \approx -1.39$
Now we solve this last equation for σ
$\sigma = \frac{2500-3500}{-1.39} = 719.42446 \approx \mathbf{719}$
So the standard deviation is approximately 719 grams.

20b



Want 0.05 inside the left tail.
Look for 0.0500 INSIDE the z-table.
The corresponding z-value is -1.645.
So, $x = \mu + z \cdot \sigma = 3500 - 1.645 \cdot 719 \approx \mathbf{2,317}$

So the new definition of a **low birth-weight** would be one that is less than **2,317** grams.

Chapter 7

1a The point estimate is the sample mean of 6.20 hours.

1b The margin of error is given by $E = z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ where $z_{\alpha/2} = 1.96$ (found in the z-table). With a population standard deviation $\sigma = 1.25$ and sample size $n = 50$, the margin of error is $E = 1.96\frac{1.25}{\sqrt{50}} = 0.3465$. The upper and lower bounds on μ are given by $\bar{x} \pm E$ which yields $\mathbf{5.85 < \mu < 6.55}$. **So, we are 95% confident that the mean number of hours of sleep for all college students is between 5.85 and 6.55 hours per day.**

1c Not Quite. Since 6.5 is in our confidence interval we can't be 95% confident that the true population mean is less than 6.5 hours.

1d Use the formula: $n = \left[\dfrac{z_{\alpha/2}\,\sigma}{E}\right]^2$. We want the margin of error to be no more than 0.2 hours so we set $E = 0.2$, and $z_{\alpha/2}$ at the 90% confidence level is 1.645. Now, $n = \left[\frac{1.645 \cdot 1.25}{0.2}\right]^2 = 105.7$. So we will need at least 106 college students in our survey.

1e Use the formula; $n = \left[\frac{z_{\alpha/2}\,\sigma}{E}\right]^2$. We want the margin of error to be no more than 0.1 hours so we set $E = 0.1$, and $z_{\alpha/2}$ at the 90% confidence level is 1.645. Now, $n = \left[\frac{1.645 \cdot 1.25}{0.1}\right]^2 = 422.8$. So we will need at least 423 college students in our survey.

1f Use the formula; $n = \left[\frac{z_{\alpha/2}\,\sigma}{E}\right]^2$. We want the margin of error to be no more than 0.1 hours so we set $E = 0.1$, and $z_{\alpha/2}$ at the 99% confidence level is 2.575. Now, $n = \left[\frac{2.575 \cdot 1.25}{0.1}\right]^2 = 1036.04$. So we will need at least 1037 college students in our survey.

3 Here we consider the middle 98% of the z-scores in a standard normal distribution. Therefore, there is 0.01 area in each tail (this equals $\alpha/2$ when $\alpha = 0.02$). Specifically, there is 0.01 area inside the upper tail and hence 0.99 to the left of the desired z-value. So, we look for 0.99 INSIDE the z-table. The closest value is 0.9901 corresponding to $z = 2.33$. So, $z_{\alpha/2} = \mathbf{2.33}$.

5a The point estimate is the sample proportion of $\hat{p} = 12/80 = \mathbf{0.15}$.

5b The margin of error is given by $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ where $z_{\alpha/2} = 1.96$, $\hat{p} = 0.15$, $\hat{q} = 1 - \hat{p} = 0.85$, and $n = 80$. So, $E = 1.96\sqrt{\frac{.15\ .85}{80}} = 0.0782$ and the upper and lower bounds on p are given by $\hat{p} \pm E$ which yields $\mathbf{0.072 < p < 0.228}$. So, Carl can be 95% confident that the proportion of all ears of corn with worms is between 0.072 and 0.228.

5c The margin of error is given by $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ where $z_{\alpha/2} = 2.575$, $\hat{p} = 0.15$, $\hat{q} = 1 - \hat{p} = 0.85$, and $n = 80$. So, $E = 2.575\sqrt{\frac{.15\ .85}{80}} = 0.1028$ and the upper and lower bounds on p are given by $\hat{p} \pm E$ which yields $\mathbf{0.047 < p < 0.253}$. So, Carl can be 99% confident that the proportion of all ears of corn with worms is between 0.047 and 0.253.

5d We want to use the formula $n = \frac{\left[z_{\alpha/2}\right]^2\ \hat{p}\,\hat{q}}{E^2}$. For a 99% confidence level, $z_{\alpha/2} = 2.575$. We use $\hat{p} = 0.15$ from the last problem. Finally if we want the estimate to be in error by no more than 2 percentage points, we set $E = 0.02$. Now, $n = \frac{[2.575]^2\ 0.15\ 0.85}{(0.02)^2} = 2113.5$, so he will need to sample at least **2114** ears of corn. That's a lot of corn.

5e We want to use the formula $n = \frac{\left[z_{\alpha/2}\right]^2\ 0.25}{E^2}$. For a 99% confidence level, $z_{\alpha/2} = 2.575$. If we want the estimate to be in error by no more than 2 percentage points, we set $E = 0.02$. Now, $n = \frac{[2.575]^2\ 0.25}{(0.02)^2} = 4144.1$, so he will need to sample at least **4145** ears of corn. That's even more corn.

7a The point estimate is the sample proportion of $\hat{p} = 59/100 = \mathbf{0.59}$.

7b The margin of error is given by $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ where $z_{\alpha/2} = 2.575$, $\hat{p} = 0.59$, $\hat{q} = 1 - \hat{p} = 0.41$, and $n = 100$. So, $E = 2.575\sqrt{\frac{(.59)(.41)}{100}} = 0.1266$ and the upper and lower bounds on p are given by $\hat{p} \pm E$ which yields $\mathbf{0.4634 < p < 0.7166}$. So, we can be 99% confident that the proportion of heads in **all** tosses of this token is between 0.463 and 0.717.

7c No. Since 0.50 is within the confidence interval limits, you can't rule out the possibility that the true proportion is 0.50 which would mean the token is fair.

7d The margin of error is given by $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}\,\hat{q}}{n}}$ where $z_{\alpha/2} = 1.645$, $\hat{p} = 0.59$, $\hat{q} = 1 - \hat{p} = 0.41$, and $n = 100$. So, $E = 1.645\sqrt{\dfrac{(.59)(.41)}{100}} = 0.0809$ and the upper and lower bounds on p are given by $\hat{p} \pm E$ which yields $\mathbf{0.5091 < p < 0.6709}$. So, we can be 90% confident that the proportion of heads in **all** tosses of this token is between 0.509 and 0.671.

7e Yes. Since 0.50 is below the lower bound of the confidence interval you are 90% confident the true proportion of heads is above 0.50 and that the coin is not fair.

7f We want to use the formula $n = \dfrac{\left[z_{\alpha/2}\right]^2\ \hat{p}\,\hat{q}}{E^2}$. For a 99% confidence level, $z_{\alpha/2} = 2.575$. We use $\hat{p} = 0.59$ from the last problem. Finally if we want the estimate to be in error by no more 0.04, we set $E = 0.04$. Now, $n = \frac{[2.575]^2\ (0.59)\ (0.41)}{(0.04)^2} = 1002.47$, so I would need to toss this token at least **1003** times.

7g Now, use the formula $n = \dfrac{\left[z_{\alpha/2}\right]^2\ 0.25}{E^2}$. Now, $n = \frac{[2.575]^2\ 0.25}{(0.04)^2} = 1036.035$, so I would need to toss this token at least **1037** times.

9a The margin of error is given by $E = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ where $t_{\alpha/2} = 1.976$ (found in the t-table with 145 degrees of freedom because 149 is not in the table). With a sample standard deviation $s = 2,500$ and sample size $n = 150$, the margin of error is $E = 1.976\frac{2500}{\sqrt{150}} = 403$. The upper and lower bounds on μ are given by $\bar{x} \pm E$ which yields $\mathbf{23,597 < \mu < 24,403}$. **So, you are 95% confident that the mean debt for Vermont college students is between $23,597 and $24,403.**

9b Since we are 95% confident that the mean debt for Vermont students is greater than $23,597, we are at least 95% confident that the mean debt for Vermont students is greater than $21,000.

11a The point estimate for the population mean is the sample mean $\bar{x} = 25.2$ pounds.

11b The margin of error is given by $E = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ where $t_{\alpha/2} = 1.729$ (found in the t-table with degrees of freedom $= 19$ and a confidence level of 90%). With a sample standard deviation $s = 4.5$ and sample size $n = 20$, the margin of error is $E = 1.729\frac{4.5}{\sqrt{20}} = 1.740$. The upper and lower bounds on μ are given by $\bar{x} \pm E$ which yields $\mathbf{23.5 < \mu < 26.9}$. **So, we are 90% confident that the mean weight of all Chinook Salmon in the Columbia River is between 23.5 and 26.9 pounds.**

11c This confidence interval is calculated exactly like the previous one except that the critical value of t is given as $t_{\alpha/2} = 2.093$ (found in the t-table with degrees of freedom $= 19$ and a confidence level of 95%) which yields a margin of error of $E = 2.106$ and a confidence interval of $\mathbf{23.1 < \mu < 27.3}$. **So, we are 95% confident that the mean weight of all Chinook Salmon in the Columbia River is between 23.1 and 27.3 pounds.**

11d This is a close one. I am 95% confident that the mean weight of the salmon is is between 23.1 and 27.3 pounds but I am 99% confident that the mean weight is between 22.3 and 28.1 pounds. As such, **I am 95% confident that the mean weight is greater than 23 pounds but I am not 99% confident of this result.**

11e We can use the t-distribution if the sample size is greater than 30 or the population is normally distributed. We assumed the latter right from the start - though it is difficult to know for sure whether this is true.

13a Use the t-distribution. The sample size is small but the population distribution is normal. Since σ is unknown, use the t-distribution.

13b Neither. You would use the z-distribution (as for all population proportion confidence intervals) but the number of successes is too small, so you can't do anything with this.

13c Use the z-distribution. You can actually use either distribution because σ and s are known. However, the z-distribution gives you a better confidence interval.

13d Use the z-distribution. Always use the z-distributions for population proportions provided the number of successes and failures are both greater than 5.

13e Neither. Here the sample size is too small and the population distribution is not normal, so you can't use either distribution.

15 The sample mean will fall in the middle of the two given bounds. $\bar{x} = \frac{12.4+13.2}{2} = \mathbf{12.8}$. The margin of error E is the distance from the sample mean to either of the bounds. An easy way to find this is to take the upper bound minus the lower bound and divide this by 2. So, $E = \frac{13.2-12.4}{2} = \mathbf{0.4}$.

17a The margin of errors are given by given by

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} \quad \text{or} \quad E = t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Regardless of the formula, if the confidence level increases the critical value of $z_{\alpha/2}$ or $t_{\alpha/2}$ will increase and **the margin of error will increase**. You should have observed this phenomenon in the answers to some of the problems in this homework set.

17b In this case, the critical values of $t_{\alpha/2}$ or $z_{\alpha/2}$ will decrease and **the margin of error will decrease**.

17c The margin of errors are given by given by

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} \quad \text{or} \quad E = t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

In all three cases, if n gets bigger, then E gets smaller so **the margin of error will decrease** (provided the confidence level and sample statistics do not change).

Chapter 8

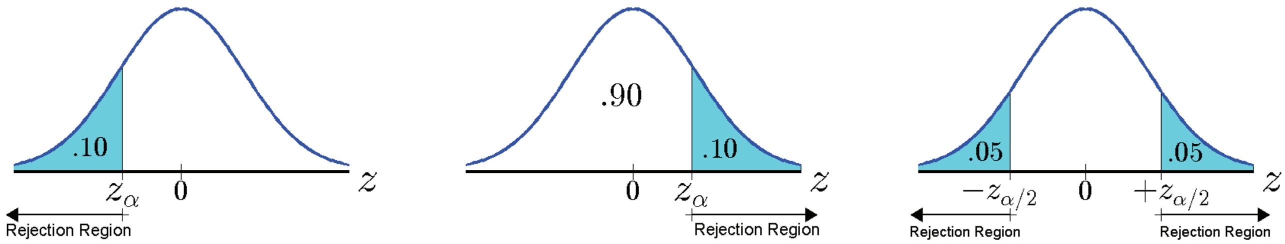
1a (a) Claim: $\mu = 12$, $H_o : \mu = 12$, $H_1 : \mu \neq 12$. (b) μ is the mean volume of all 12 ounce cans of soda. (c) The null hypothesis supports the claim. (d) This would result in a two-tailed test.

1b (a) Claim: $\mu > 40,000$, $H_o : \mu = 40,000$, $H_1 : \mu > 40,000$. (b) μ is the mean daily flow rate of oil. (c) The alternate hypothesis supports the claim. (d) This would result in a right-tailed test.

1c (a) Claim: $p > 0.5$, $H_o : p = 0.5$, $H_1 : p > 0.5$. (b) p is the proportion of all people who have a strong dislike for statistics. (c) The alternate hypothesis supports the claim. (d) This would result in a right-tailed test.

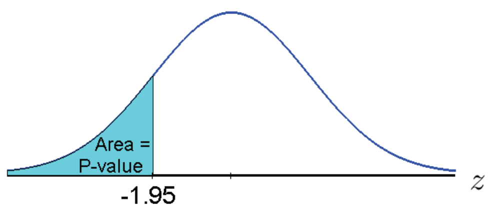
3 You can get all of the critical values from the small table below the z-table. Or,

(A) For this left-tailed test you look for 0.10 INSIDE the z-table. The closest value is 0.1003 corresponding to $z = -1.28$. (B) For this right-tailed test you use the positive z-score from part (A). (C) For this two-tailed test you look for $\alpha/2 = 0.05$ INSIDE the z-table and get a corresponding z-value of -1.645. Since this is a two-tailed test you use both the positive and negative versions of this z-value.



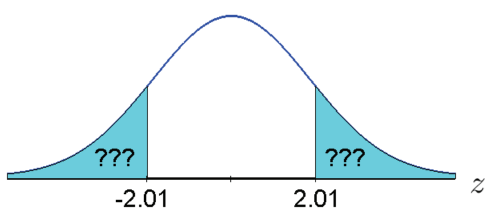
$\mathbf{z}_\alpha = \mathbf{-1.28}$ $\mathbf{z}_\alpha = \mathbf{1.28}$ $\mathbf{z}_{\alpha/2} = \mathbf{\pm 1.645}$

5a Since H_1 has a $<$ symbol, this is a **left-tailed test**.



$$\begin{aligned} \text{P-value} &= P(z < z_{\hat{p}}) \\ &= P(z < -1.95) \\ &= \mathbf{0.0256} \quad \text{from the } z\text{-table} \end{aligned}$$

5b Since H_1 has a $\neq$ symbol, this is a **two-tailed test.**



$$\begin{aligned} \textbf{P-value} &= P(\textbf{getting a more extreme } z\textbf{-score}) \\ &= P(z < -2.01) + P(z > 2.01) \\ &= 0.0222 + .0222 \quad \textbf{from } z\textbf{-table} \\ &= \mathbf{0.0444} \end{aligned}$$

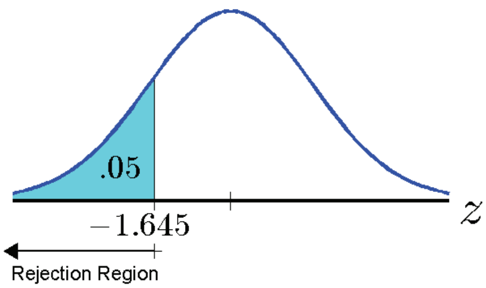
7a You would end up with $H_o : \mu = 12$ and $H_1 : \mu < 12$. If you reject the null hypothesis, the data supports your claim and your conclusion would be something like: **The sample data supports the claim that the mean volume of all *12 ounce* cans of Fizzy Pop is actually less than 12 ounces.**

7b They would end up with $H_o : p = .5$ and $H_1 : p > .5$. If you fail to reject the null hypothesis, the data does not necessarily their claim and your conclusion would be something like: **There is not enough sample data to support the claim that most *12 ounce* cans of Fizzy Pop contain more than 12 ounces.**

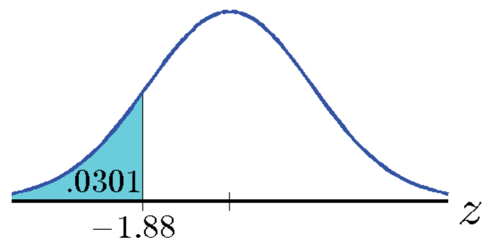
7c The null hypothesis would be $\mu = 72$ (which supports your claim) and the alternate hypothesis would be $\mu \neq 72$. If you reject the null hypothesis you would then *accept* the alternate hypothesis and your conclusion would be something like: *There is enough data to justify rejection of the claim that the average speed of cars going down this stretch of highway is 72 mph.*

9 A Type I error would be one in which you conclude that the Prius gets more than 43 mpg when in fact it doesn't. The probability of this type of error would be α (which is small). A type II error would be one in which you conclude there is not enough data to confirm that it gets more that 43 mpg's when in fact it does.

11a Claim: $p < 0.15$, $H_o : p = 0.15$, $H_1 : p < 0.15$. This is a left-tailed test.
The test statistic is $z_{\hat{p}} = \frac{\hat{p}-p}{\sqrt{\frac{p\,q}{n}}} = \frac{0.075-0.15}{\sqrt{\frac{0.15\;0.85}{80}}} = \mathbf{-1.88}$.



Critical Value Method:
Put 0.05 inside the left tail.
Look for 0.05 INSIDE the z-table
or use the small table below it.
Get $z_\alpha = -1.645$
Reject H_o because the test statistic
(-1.88) is in the rejection region.



P-Value Method:

$$\begin{aligned} \text{P-Value} &= P(z < z_{\hat{p}}) \\ &= P(z < -1.88) \\ &= \mathbf{0.0301} \quad \text{from } z\text{-table} \end{aligned}$$

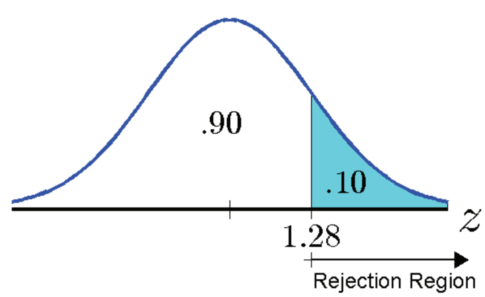
Since P-value $< \alpha$, you **Reject** H_o**.**

Concluding Statement: The data supports the claim that less than 15% of Carl's corn has worms.

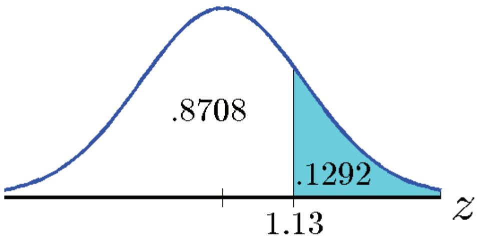
11b All hypotheses, the test statistic, and the P-value remain the same as part (a). However, the critical value of z is now $\mathbf{z_\alpha = -2.33}$. Using the Critical-Value Method (the test statistic is not in the rejection region) or the P-value method (the P-value of the test statistic is greater than α), you **fail to reject the null hypothesis**. The concluding statement goes something like: **There is not enough data to support the claim that less than 15% of Carl's corn has worms.**

11c Because in the first case (a) we used a 0.05 significance level and in the second case (b) we used a 0.01 significance level. The bigger α (significance level) the easier it is to reject the null hypothesis. Ie. In part (a) we were more willing to make a mistake in our conclusion.

13 Because I am claiming that most cars ...., the claim is: $p > 0.50$, $H_o : p = 0.50$, $H_1 : p > 0.50$.
This is a right-tailed test.
The sample proportion is $\hat{p} = 29/50 = 0.58$.
The test statistic is $z_{\hat{p}} = \frac{\hat{p}-p}{\sqrt{\frac{p\,q}{n}}} = \frac{0.58-0.50}{\sqrt{\frac{0.50\;0.50}{50}}} = \mathbf{1.13}$.



Critical Value Method:
Put 0.10 inside the right tail.
Look for 0.90 INSIDE the z-table
or use the small table below it. Get $z_\alpha = 1.28$
The test statistic (1.13) is Not in the rejection region.
Fail to Reject H_o**.**



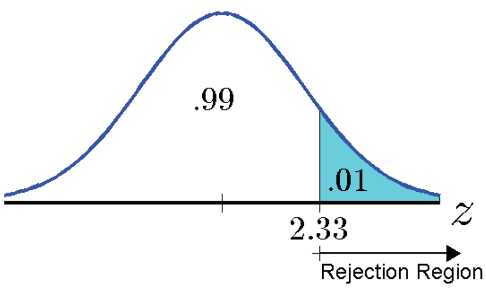
P-Value Method:

$$\begin{aligned} \text{P-Value} &= P(z > z_{\hat{p}}) \\ &= P(z > 1.13) \\ &= 1 - P(z < 1.13) \\ &= 1 - 0.8708 = \mathbf{0.1292} \quad \text{from } z\text{-table} \end{aligned}$$

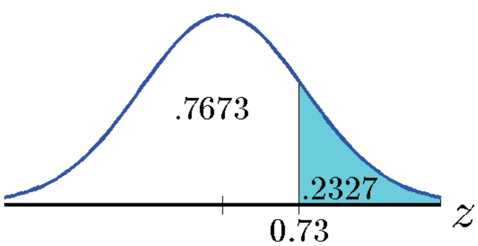
Since P-value $> \alpha$, you **Fail to Reject H_o.**

Concluding Statement: There is not enough data to support the claim that most cars are going more than 5 mph over the posted speed limit.

15 Because he is claiming that more than a quarter of ...., the claim is $p > 0.25$.
Then, $H_o : p = 0.25$, $H_1 : p > 0.25$, where p represents the proportion of all his email that is spam.
This is a right-tailed test. The sample proportion is $\hat{p} = 12/40 = 0.30$.
The test statistic is $z_{\hat{p}} = \frac{\hat{p}-p}{\sqrt{\frac{p\,q}{n}}} = \frac{0.30-0.25}{\sqrt{\frac{0.25\,0.75}{40}}} = \mathbf{0.730}$.



Critical Value Method:
Put 0.01 inside the right tail.
Look for 0.99 INSIDE the z-table
or use the small table below it. Get $z_\alpha = 2.33$
The test statistic (0.730) is **not** in the rejection region.
Fail to Reject H_o.



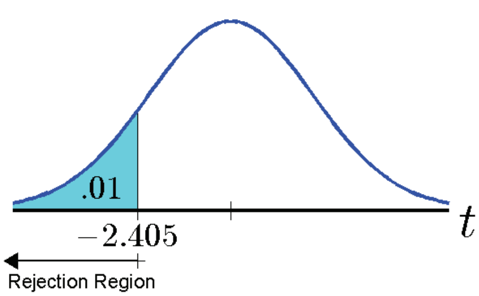
P-Value Method:

$$\begin{aligned} \text{P-Value} &= P(z > z_{\hat{p}}) \\ &= P(z > 0.73) \\ &= 1 - P(z < 0.73) \\ &= 1 - 0.7673 = \mathbf{0.2327} \quad \text{from } z\text{-table} \end{aligned}$$

Since P-value $> \alpha$, you **fail to Reject H_o.**

Concluding Statement: There is not enough data to support the claim that more than a quarter of Larry's email is spam.

17 Claim: $\mu < 7$, $H_o : \mu = 7$, $H_1 : \mu < 7$. This is a left-tailed test.
The test statistic is $t_{\bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{6.2-7}{0.75/\sqrt{50}} = \mathbf{-7.54}$.

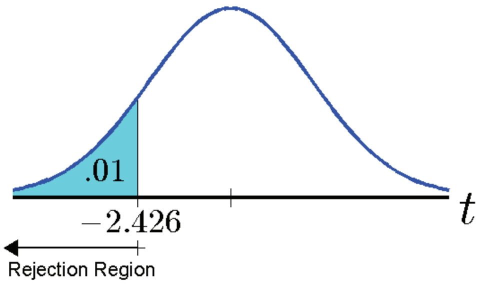


Critical Value Method:
Area in one tail $= 0.01$. d.f. $= 49$.
Get 2.405 from the t-table.
So, $t_\alpha = -2.405$ because this is a left-tailed test.
The test statistic (-7.54) is in the rejection region.
Reject H_o.

P-value Method: Here we must use software. Using Excel's command `TDIST(7.54,49,1)`, you get a P-value $= 4.9 \times 10^{-10}$. This is much smaller than α so you **Reject** H_o.

Using the Critical-Value Method (the test statistic is in the rejection region) or the P-value method (the P-value of the test statistic is less than α), you **reject the null hypothesis**. The concluding statement goes something like: **The data supports the claim that the mean amount of sleep by college students is less than the over-all average of 7 hours.**

19a Claim: $\mu < 2$, $H_o : \mu = 2$, $H_1 : \mu < 2$. This is a left-tailed test.
The test statistic is $t_{\bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{1.78-2}{0.75/\sqrt{40}} = \mathbf{-1.86}$.



Critical Value Method:
Area in one tail = 0.01. d.f. = 39.
Get 2.426 from the t-table.
So, $t_\alpha = -2.426$ because this is a left-tailed test.
Test statistic (-1.86) is in not in the rejection region.
Fail to Reject H_o.

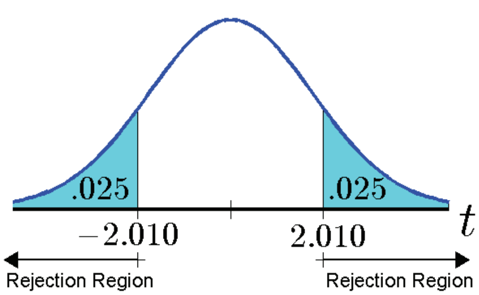
P-value Method: Here we must use software. Using Excel's command `TDIST(1.86,39,1)`, you get a P-value = 0.0352. This is larger than α so you **Fail to Reject** H_o.

Using the Critical-Value Method (the test statistic is not in the rejection region) or the P-value method (the P-value of the test statistic is larger than α), you **fail to reject the null hypothesis**. The concluding statement goes something like: **There is not enough data to support the claim that the average assembly time is less than 2 hours.**

19b The test procedure and test statistic remain the same as in part (a). However, the critical value is now -1.685 instead of -2.426. This puts the test statistic (-1.86) in the rejection region and you reject the null hypothesis. If you use the P-value method, the P-value remains the same at 0.0352 which now less than $\alpha = .05$ and you reject the null hypothesis. Based on either method, you now reject the null hypothesis and conclude something like: **The sample data supports the claim that the mean assembly time is less than 2 hours.**

19c In the first case ($\alpha = 0.01$) you were more cautious with your conclusions. You are willing to be wrong about 1% of the time. As such, you don't conclude that the data is sufficient. In the second case ($\alpha = 0.05$), you are willing to be wrong about 5% of the time. With this extra willingness to be wrong, you now accept the data as being sufficient enough to support your claim. It's all about how willing you are to support potentially false conclusions.

21 Claim: $\mu = 515$, $H_o : \mu = 515$, $H_1 : \mu \neq 515$. This is a two-tailed test.
The test statistic is $t_{\bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{508-515}{35/\sqrt{50}} = \mathbf{-1.41}$.



Critical Value Method:
Area in two tails = 0.05. d.f. = 49
Get 2.010 from the t-table.
So, $t_\alpha = \pm 2.010$ because this is a two-tailed test.
The test statistic (-1.41) is not in the rejection region.
Fail to Reject H_o.

P-value Method: Here we must use software. Using Excel's command `TDIST(1.41,49,2)`, you get a P-value = 0.165. This is not smaller than α so you **Fail to Reject** H_o.

Using the Critical-Value Method (the test statistic is not in the rejection region) or the P-value method (the P-value of the test statistic is not less than α), you **fail to reject the null hypothesis**. The concluding statement goes something like: **There is not enough data to reject the claim that there is no difference in mean SAT scores between those that did not study and the national average.** In other words, **there is not a significant difference based on this sample data.**

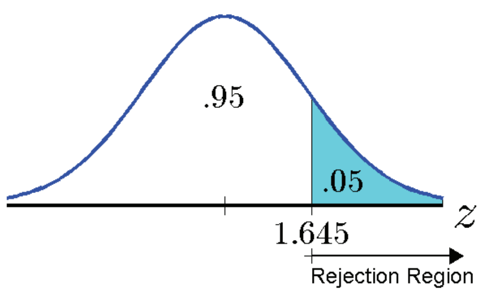
23 (1) This is a claim about a mean.
Let $\mu =$ the mean number of Facebook friends for all College Students.
Claim: $\mu > 254$, $H_o : \mu = 254$, $H_1 : \mu > 254$. This is a right-tailed test.
The test statistic is $t_{\bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$.

(2) This is a claim about a proportion.
Let $p =$ the proportion of all college students with more than 254 friends.
Claim: $p > .5$, $H_o : p = .5$, $H_1 : p > .5$. This is a right-tailed test.
The test statistic would be $z_{\hat{p}} = \frac{\hat{p}-p}{\sqrt{\frac{p\,q}{n}}}$.

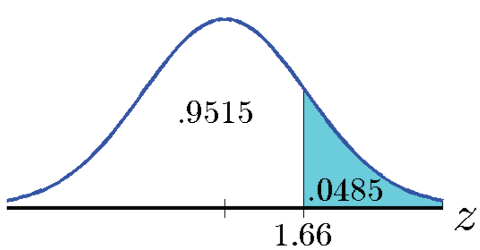
(3) The first is a claim about a mean, the second a claim about a proportion. Hypotheses change form, and test statistics are calculated differently. In the first case the critical value comes from the t-distribution, in the second case it comes from the z-distribution.

(4) If the distribution of Facebook friends was normal (where the mean and median are equal) then you could reasonably expect both of them to be true if one is true. However, any deviation from normal could easily lead to one of them being supported and the other not. A few college students with a whole lot of friends could swing your sample mean well above 254 with only a small proportion having more than 254.

25a Claim: $\mu > 515$, $H_o : \mu = 515$, $H_1 : \mu > 515$. This is a right-tailed test.
The test statistic is $z_{\bar{x}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{534-515}{100/\sqrt{76}} = \mathbf{1.66}$.



<u>Critical Value Method</u>:
Put 0.05 inside the right tail.
Look for 0.95 INSIDE the z-table
or use the small table below it. Get $z_\alpha = 1.645$
The test statistic (1.66) is in the rejection region.
Reject H_o.



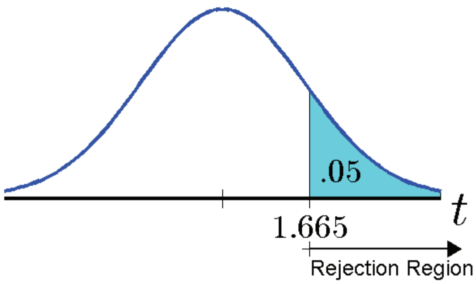
<u>P-Value Method</u>:

$$\begin{aligned} \text{P-Value} &= P(z > z_{\bar{x}}) \\ &= P(z > 1.66) \\ &= 1 - P(z < 1.66) \\ &= 1 - 0.9515 = \mathbf{0.0485} \quad \text{from } z\text{-table} \end{aligned}$$

Since the P-value $< \alpha$, you **reject** H_o.

Concluding Statement: The sample data supports the claim that the mean math SAT score for students taking the prep course is higher than the national average.

25b Claim: $\mu > 515$, $H_o : \mu = 515$, $H_1 : \mu > 515$. This is a right-tailed test.
The test statistic is $t_{\bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{534-515}{100/\sqrt{76}} = \mathbf{1.66}$.



Critical Value Method:
Area in one tail = 0.05. d.f. = 75
Get 1.665 from the t-table.
So, $t_\alpha = 1.665$ because this is a right-tailed test.
The test statistic (1.66) is not in the rejection region.
Fail to Reject H_o.

P-value Method: Here we must use software. Using Excel's command `TDIST(1.66,75,1)`, you get a P-value = 0.0505. This is just barely greater than α (0.05) so you **Fail to Reject** H_o.

Concluding Statement: There is not enough sample data to support the claim that the mean math SAT score for students taking the prep course is higher than the national average.

25c They disagree because the z-distribution and the t-distribution for a sample of size 76 are not quite the same. Specifically, the critical values at the 0.05 significance level are different. For the z-distribution, the critical value is 1.645 but the critical value for the t-distribution is 1.665. Unfortunately, the test statistic fell in between these two values resulting in different conclusions. This does not happen very often but it is quite possible when you have a smaller sample.

Chapter 9

1a (1) $\boxed{\text{Claim: } \mu_d > 0}$ $\boxed{H_o\text{: } \mu_d = 0}$ $\boxed{H_1\text{: } \mu_d > 0}$

(2) **Test Statistic:** $t_{\bar{d}} = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \dfrac{2.5 - 0}{\frac{3.5}{\sqrt{10}}} \approx \boxed{2.26}$

(3) **Fail to Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Put 0.01 into one tail with 9 degrees of freedom: $t_\alpha = 2.821$, and the test statistic is **not** in the rejection region.

(3b) **P-Value:** Using software you get a P-value $\approx$ 0.025 (or 0.024 if you're using raw data) which is larger than α.

(4) **Conclusion:** At the 0.01 significance level, the there is not enough data to support the claim that older brothers have a higher IQ than younger brothers.

1b Everything remains the same except in this case the critical value of t is now 1.833 and the test statistic is in the rejection region. Likewise, the P-value does not change but α is now 0.05 and the P-value is less than α. As such, we **reject** the null hypothesis.
(4) **Conclusion:** At the 0.05 significance level, the data supports the claim that older brothers have higher IQ's than younger brothers.

1c Since we don't have the standard deviation of the differences we don't know for sure what the test statistic will be. However, if we assume the same standard deviation as that in our small study we would get a test statistic of
$t_{\bar{d}} = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \dfrac{2.3 - 0}{\frac{3.5}{\sqrt{100000}}} \approx \boxed{208}$
Regardless of our level of significance (or the actual standard deviation), this will end up in the rejection region and we would be able to support the claim quite comfortably.

1d Notice the IQ's are given from lowest to highest for the older brothers. The scores for the younger brothers also trend in this direction. As such, there seems to be a correlation between the IQ's of the brothers. We determine whether there is a significant correlation in the chapter on correlation and regression.

3a (1) $\boxed{\text{Claim: } \mu_d > 30}$ $\boxed{H_o\text{: } \mu_d = 30}$ $\boxed{H_1\text{: } \mu_d > 30}$

(2) **Test Statistic:** $t_{\bar{d}} = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \dfrac{32 - 30}{\frac{14}{\sqrt{200}}} \approx \boxed{2.02}$

(3) **Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Putting 0.10 into one tail with 199 degrees of freedom: $t_\alpha = 1.287$ (software: 1.286) and the test statistic is in the rejection region.

(3b) **P-Value:** Using software you get a P-value $\approx .0224$ which is smaller than α.

(4) **Conclusion:** At the 0.10 significance level, the sample data supports the claim that retaking the SAT increases the score on average by more than 30 points.

3b The test goes about the same way only the critical value is now 2.351 (software 2.345) and the test statistic is not in the rejection region. Also, the P-value is not going to change and now it is larger than α. So, either way, you now fail to reject the null hypothesis and at the 0.01 significance level, you can not support the claim.

5a (1) $\boxed{\text{Claim: } \mu_1 - \mu_2 > 0}$ $\boxed{H_o\text{: } \mu_1 - \mu_2 = 0}$ $\boxed{H_1\text{: } \mu_1 - \mu_2 > 0}$

(2) **Test Statistic:**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(591.6 - 523.0) - 0}{\sqrt{\frac{1375.1}{8} + \frac{992.8}{11}}} \approx \boxed{4.238}$$

(3) **Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Put 0.01 into one tail with 14 degrees of freedom (given in the problem) yields $t_\alpha \approx 2.624$, and the test statistic lies in the rejection region.

(3b) **P-Value:** Using software you get a P-value ≈ 0.000413 which is smaller than α.

(4) **Conclusion:** The sample data supports the claim that students who study music in high school have a higher average Math SAT score than those who do not.

5b Here, the game is to come up with δ so that we can support the claim (we'll say at the 0.05 confidence level) that $\mu_1 - \mu_2 > \delta$. You really have to do this by trial and error so some type of software is needed. By trial and error I found that $\delta = 40$ is about the best we can do.
(1) $\boxed{\text{Claim: } \mu_1 - \mu_2 > 40}$ $\boxed{H_o\text{: } \mu_1 - \mu_2 = 40}$ $\boxed{H_1\text{: } \mu_1 - \mu_2 > 40}$

(2) **Test Statistic:**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(591.6 - 523.0) - 40}{\sqrt{\frac{1375.1}{8} + \frac{992.8}{11}}} \approx \boxed{1.768}$$

(3) **Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Put 0.05 into one tail with 14 degrees of freedom (given in the problem) yields $t_\alpha \approx 1.761$, and the test statistic just makes it inside the rejection region.

(3b) **P-Value:** Using software you get a P-value ≈ 0.0494 which is smaller than 0.05.

(4) **Conclusion:** The sample data supports the claim that students who study music in high school score on average more than 40 points better than those who do not study music in high school.

7 (1) Here we'll let x_1 be the PM sections and x_2 be the AM sections because the average from the PM sample is greater than the AM sample.

Claim: $\mu_1 - \mu_2 > 0$	H_o: $\mu_1 - \mu_2 = 0$	H_1: $\mu_1 - \mu_2 > 0$

(2) **Test Statistic:**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(75.1 - 71.2) - 0}{\sqrt{\frac{277.5}{30} + \frac{250.3}{22}}} \approx \boxed{0.859}$$

(3) **Fail Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Put 0.05 into one tail with 47 degrees of freedom (given in the table) yields $t_\alpha \approx 1.678$, and the test statistic does not lie in the rejection region.

(3b) **P-Value:** Using software you get a P-value ≈ 0.197 which is greater than α.

(4) **Conclusion:** At the 0.05 significance level, there is not sufficient evidence to support the claim that the average for all students in the PM sections is greater than the AM sections.

9 (1) We are testing the claim that the averages are different

Claim: $\mu_1 - \mu_2 \neq 0$	H_o: $\mu_1 - \mu_2 = 0$	H_1: $\mu_1 - \mu_2 \neq 0$

and this is a two-tailed test.

(2) **Test Statistic:**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(80.1 - 76.9) - 0}{\sqrt{\frac{127.0}{22} + \frac{92.9}{28}}} \approx \boxed{1.061}$$

(3) **Fail Reject** H_0 because of either of the following.

(3a) **Critical Value of** t**:** Put 0.05 into two tails with 41 degrees of freedom (given in the table) yields $t_\alpha \approx \pm 2.020$, and the test statistic does not lie in the rejection region.

(3b) **P-Value:** Using software you get a P-value ≈ 0.295 which is greater than α.

(4) **Conclusion:** At the 0.05 significance level, there is not sufficient evidence to support the claim that the average from Prof Smith's section was **significantly different** from Prof Jones' section. While it is true that Prof. Smith's section did better last year, there is a good chance this is merely a result of random variation. We would expect a difference at least this great about 30% of the time by chance alone.

11a (1) Let p_1 be the proportion of young adults that had to move back in with their parents in 2012 and p_2 be the proportion from 2000. We are testing the claim that $p_1 > p_2$ or $p_1 - p_2 > 0$.

$\boxed{\text{Claim: } p_1 - p_2 > 0}$ $\boxed{H_o\text{: } p_1 - p_2 = 0}$ $\boxed{H_1\text{: } p_1 - p_2 > 0}$

This is a right-tailed test because of the $>$ sign in the alternate hypothesis.

(2) **Test Statistic** using equation (9.4):
Here, $\delta_p = 0$ representing the hypothesized difference in population proportions and the standard error (SE) is given in the table.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \delta_p}{SE} = \frac{(0.24010 - 0.15789) - 0}{0.01634} \approx 5.0309 \to \boxed{5.03}$$

(3) **Reject** H_0 because of either of the following.
(3a) **Critical Value of** z**:** Putting 0.05 into the right tail yields $z_\alpha = 1.645$, and the test statistic lies deep in the rejection region.
(3b) **P-Value:** Using the z-table, the right-tailed P-value of $z = 5.03$ is $(1-0.9999) = 0.0001$ which is smaller than α. If using software (with no rounding), you get a P-value $\approx 2.46 \cdot 10^{-7}$.

(4) **Conclusion:** At the 0.05 significance level, the data supports the claim that a greater proportion of all young adults had to move back in with their parents in 2012 than in 2000.

11b Yes. At the 0.01 significance level, the critical value of z is 2.33 and the test statistic still in the rejection region. Also, the P-value remains the same and is still much smaller than α.

13a (1) Let p_1 be the proportion of wins at home and p_2 be the proportion of wins on the road. We are testing the claim that $p_1 > p_2$ or $p_1 - p_2 > 0$.

$\boxed{\text{Claim: } p_1 - p_2 > 0}$ $\boxed{H_o\text{: } p_1 - p_2 = 0}$ $\boxed{H_1\text{: } p_1 - p_2 > 0}$

This is a right-tailed test because of the $>$ sign in the alternate hypothesis.

(2) **Test Statistic** using equation (9.4):
Here, $\delta_p = 0$ representing the hypothesized difference in population proportions and the standard error (SE) is given in the table.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \delta_p}{SE} = \frac{(0.65854 - 0.43902) - 0}{0.10990} \approx 1.9974 \to \boxed{2.00}$$

(3) **Reject** H_0 because of either of the following.
(3a) **Critical Value of** z**:** Putting 0.05 into the right tail yields $z_\alpha = 1.645$, and the test statistic lies in the rejection region.
(3b) **P-Value:** Using the z-table, the right-tailed P-value of $z = 2.00$ is $(1-0.9772) = 0.0228$ which is smaller than α. If using software (with no rounding), you get a P-value of 0.0229.

(4) **Conclusion:** At the 0.05 significance level, the data supports the claim that the proportion of wins at home is significantly greater than on the road. Specifically, it is unlikely that this difference is merely due to random variation (though it could be).

13b No. At the 0.01 significance level, the critical value of z is 2.33 and the test statistic is not in the rejection region. Also, the P-value remains the same which is now greater than α.

15a (1) Let p_1 be the proportion of all murders committed with a gun in Texas and let p_2 be the proportion of all murders committed with a gun in New York. We are testing the claim that $p_1 > p_2$ or $p_1 - p_2 > 0$.

$\boxed{\text{Claim: } p_1 - p_2 > 0}$ $\boxed{H_o\text{: } p_1 - p_2 = 0}$ $\boxed{H_1\text{: } p_1 - p_2 > 0}$

This is a right-tailed test because of the > sign in the alternate hypothesis.

(2) **Test Statistic** using equation (9.4):

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \delta_p}{SE} = \frac{(0.64187 - 0.57494) - 0}{0.02289} \approx 2.9243 \rightarrow \boxed{2.92}$$

Here, $\delta_p = 0$ representing the hypothesized difference in population proportions and the standard error (SE) was given in the table.

(3) **Reject** H_0 because of either of the following.
(3a) **Critical Value of z:** Putting 0.05 into the right tail yields $z_\alpha = 1.645$, and the test statistic lies in the rejection region.
(3b) **P-Value:** Using the z-table, the right-tailed P-value of $z = 2.92$ is $(1 - 0.9982) = 0.0018$ which is less than α. If using software (with no rounding), you get a P-value of 0.0017.

(4) **Conclusion:** The data supports the claim that the proportion of murders committed with a gun was significantly greater in Texas than New York in 2011. Specifically, it is unlikely that this difference was merely due to random variation (though it could be).

15b Yes. The critical value at the 0.01 significance level is 2.33 and the test statistic is still in the rejection region. Or, the P-value is still ≈ 0.0018 which is still less than α. So you would still reject the null hypothesis and the conclusion remains the same.

Chapter 10

2a Yes, the correlation coefficient is $r = -0.89$ and $|r| > 0.456$ as required by table 4 for $n = 19$. Therefore, we conclude that there is a significant linear correlation.

2b About 79.2% because $r^2 = 0.7921$.

2c Here we put $x = 8$ into the regression equation $\hat{y} = -102.61\,x + 1490.2$ and get $\hat{y} = -102.61 \cdot 8 + 1490.2 = 669.32$. So you would expect to sell about **669** items at this price.

2d Here we put $x = 15$ into the regression equation $\hat{y} = -102.61\,x + 1490.2$ and get $\hat{y} = -102.61 \cdot 15 + 1490.2 = -48.95$. Since you can't sell a negative number of items, you would expect to sell zero items at this price.

2e The slope represents the change in demand for every one dollar increase in price. Ie, for every dollar increase in price you would expect to sell 102.61 fewer items.

2f The y-intercept (1490.2) represents the number you would expect to sell if the price was 0$. It is meaningful in the sense that you can not expect to sell more than this number of items no matter what you charge.

2g The slope of the regression equation gives the marginal demand. So you would expect the demand to lower by 102.61 items for every dollar increase in price.

2h Here, it seems as though price causes the demand to change. As such, price would be the causative variable and demand (# sold) would be the response variable. However, the demand would depend on many other factors as well.

2i It would not change the strength of the correlation because r is independent of which variable you call x and which you call y. As such, you would still get a negative correlation. However, the regression equation would change because now y would be the price and x would be the demand (# sold).

2j If there was a fixed supply then as demand increases, you would expect the price to increase.

5a Using the regression equation for Model Year and Miles we would expect a 2002 Honda Civic to have $\hat{y} = -7.987(2002) + 16,109 \approx 119$ thousand miles. So, yes, 143 thousand is more than you would expect.

5b Even though we don't have the year it was made, we can still use the regression equation for Miles and Price. Doing this yields $\hat{y} = -79.2(84) + 15,853 \approx$ \$9,200. So, yes, \$6000 is a good price, but you can probably bet that the car is pretty old or in bad shape.

5(c)i $\hat{y} = 1056.2\ (2004) - 2,108,670 = 7,954.50$. So an estimated reasonable price would be about **\$7,955**.

5(c)ii $\hat{y} = -79.2\ (140) + 15,853 = 4,765$. So an estimated reasonable price would be about **\$4,765**.

5(c)iii $\hat{y} = 716.9(2004) - 42.5(140) - 1,424,349 = 6,368.60$. So an estimated reasonable price would be about **\$6,369.**

5d Well the car has more miles on it than would be expected from a 2004 vehicle. This makes the first estimate (based purely on year) too high. The car is newer than most cars with this many miles so the second estimate is too low. The third estimate considers both year and mileage to produce the best estimate. Here is one good reason to take another course in stats.

6a About **74%**, because $0.86^2 = 0.7396$.

6b About **58%**, because $0.76^2 = 0.5776$.

6c About **34%**, because $0.58^2 = 0.3364$.

6d This is tricky business and quite controversial. The most outstanding comparison is that the correlation between identical twins raised apart is greater than the correlation between non-identical twins raised together. This would suggest that nature plays a greater role in determining IQ than nurture. However, nurture must play a role because identical twins raised together have a higher correlation than those raised apart.

6e The sample size is missing. It is hard to determine the significance of the correlation coefficient without knowing the sample size.

8a The scatterplot should look like the one in the text.

8b Using software you should get a correlation coefficient of $r = 0.929$. This is significant because $|r|$ is greater than the critical value for $n = 12$ of 0.576 (found in Table 4). If calculating the P-value with software you should get 0.0000122 which is significant by any measure.

8c You should get the regression equation $\hat{y} = 122.3x - 133.66$.
The slope is 122.3 and the y-intercept is -133.66.

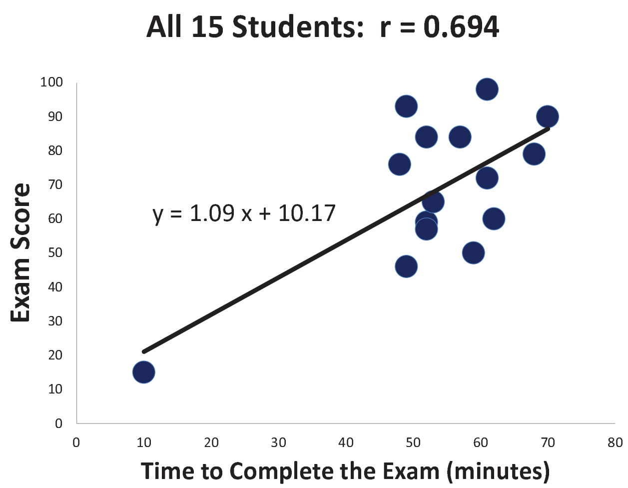
8d Using $\hat{y} = 122.3x - 133.66$ with $x = 5$ you get $\hat{y} = 122.3(5) - 133.66 = 477.84 \approx \mathbf{478}$. So you would expect the supply to be around 478 units. Notice this is greater than either of the values found in the table for this price.

8e Using $\hat{y} = 122.3x - 133.66$ with $x = 1.00$ you get $\hat{y} = 122.3(1.00) - 133.66 = -11.36 \approx \mathbf{-11}$. So you would expect the supply to be around -11 units. While this doesn't make sense, it suggests that if the price gets to be around \$1.00, the supply will disappear. Maybe this is close to the production cost.

8f The slope is 122.3. So for every one dollar increase in price, you can expect the quantity supplied to increase by about 122.3 units.

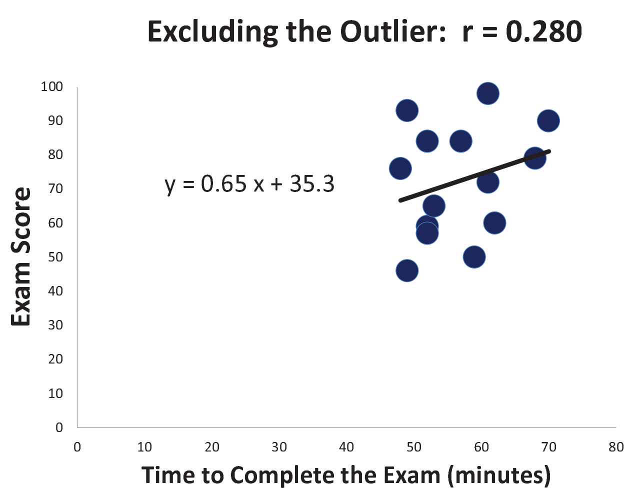
8g The y-intercept is -133.66. Since the supply can never be negative this is not particularly meaningful except to say that production will cease before the price hits zero dollars. This makes sense right? There will be some cost to producing the items and you would expect production to stop if the price goes below this value.

10a



Correlation Coefficient: $r = 0.694$.
This is a **significant** correlation
because $|r| > 0.514$ (Table 4).
Regression Equation: $\hat{y} = 1.09x + 10.17$

10b

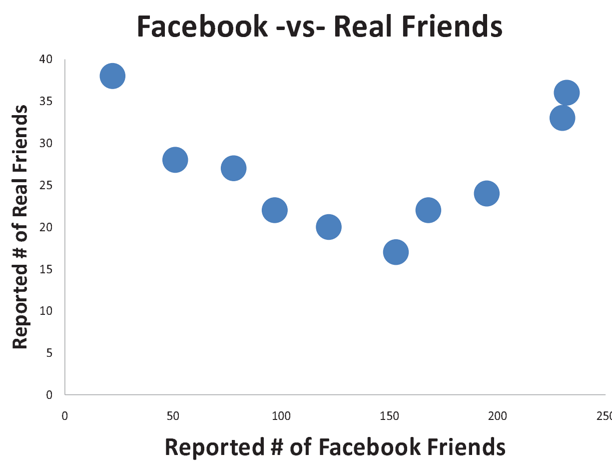


Correlation Coefficient: $r = 0.280$.
This is **not** a significant correlation.
because $|r| < 0.532$ (Table 4).
Regression Equation: $\hat{y} = 0.65x + 35.3$

10c The outlier (10 minutes, 15 test score) created a significant correlation. Additionally it gave a much greater slope to the regression equation.

10d When using all 15 data pairs (including the outlier) you might think that those students who finished sooner did better on the exam. When you exclude this one person you can see the correlation between finishing time and performance is not significant.

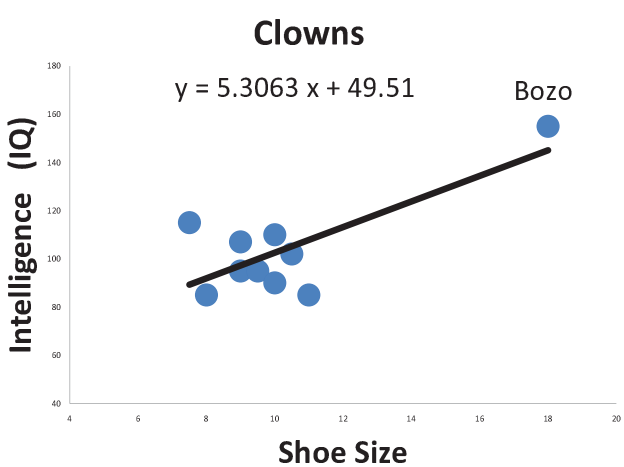
12a There does appear to be an association which is **nonlinear**. It seems that those people who report very few or very many Facebook friends report to have a greater number of real-world friends. Those people who report a medium number of Facebook friends report fewer real-world friends.

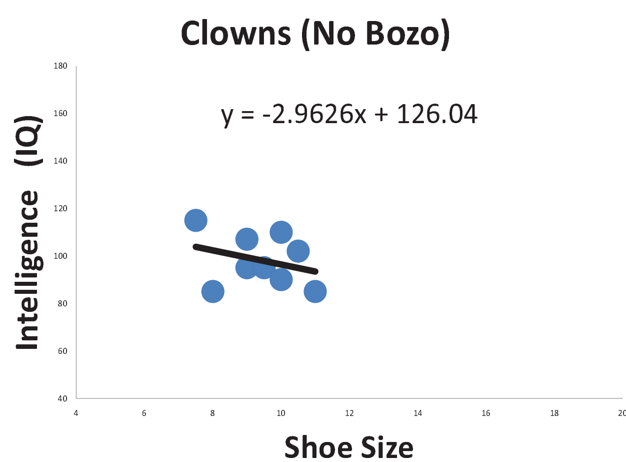


12b The linear correlation coefficient is $r = -0.028$. This does not constitute a significant linear correlation (need $|r| > 0.632$).

12c First, the data does not appear to be linearly related. Second, the correlation is not significant so there is no reason to believe that the regression equation is valid.

13a Here are the scatter plots with and without Bozo. Regression equation and line depicted on the graph.





Chapter 11

1 Preliminary Information: $\boxed{n = 100}$ $\boxed{k = 4}$ $\boxed{\text{d.f.} = 3}$ $\boxed{\alpha = 0.05}$

1. **Claim:** The correct answers are not evenly distributed.
   $\mathbf{H_o}$: $p_1 = p_2 = p_3 = p_4 = 1/4 = 0.25$ (All answers have the same probability of appearing.)
   $\mathbf{H_1}$: The probabilities are not all equal 0.25.

2. **Calculate the test statistic.**

   (a) The assumed probabilities come from H_o and the expected frequencies have been calculated based on $E_i = n \cdot p_i$ and put in the chart below.

   (b) The test statistic (χ^2) is calculated with the aid of the table below.

(i)	Correct Answer	Observed Frequency O_i	Assumed Probability p_i (from H_o)	Expected Frequency $E_i = n \cdot p_i$	$\frac{(O_i - E_i)^2}{E_i}$
(1)	A	12	0.25	25	$\frac{(12-25)^2}{25} = 6.760$
(2)	B	24	0.25	25	$\frac{(24-25)^2}{25} = 0.040$
(3)	C	31	0.25	25	$\frac{(31-25)^2}{25} = 1.440$
(4)	D	33	0.25	25	$\frac{(33-25)^2}{20} = 2.560$
	Total	$n = 100$			$\chi^2 = 10.800$

3. **Reject** H_o because of either of the following.

   (a) The critical value from Table 5 (page 293) with 3 degrees of freedom at the 0.05 significance level is 7.815 and the test statistic (10.800) is in the rejection region.

   (b) The P-value (using software) is about 0.0129 which is smaller than α.

4. **Concluding Statement:** There is sufficient evidence to conclude that the distribution of outcomes does not fit the expected distribution from the null hypothesis. Specifically, the data supports the claim that correct answers are not evenly distributed.

4 Preliminary Information: $\boxed{n = 200}$ $\boxed{k = 6}$ $\boxed{\text{d.f.} = 5}$ $\boxed{\alpha = 0.05}$

1. There is no claim but that does not alter the null and alternate hypotheses.
   $\mathbf{H_o}$: $p_1 = 0.3,\ p_2 = 0.2,\ p_3 = 0.2,\ p_4 = 0.1,\ p_5 = 0.1,\ p_6 = 0.1$
   $\mathbf{H_1}$: The probabilities are not all equal to their respective values assumed in H_o

2. **Calculate the test statistic.**

   (a) The assumed probabilities come from H_o and the expected frequencies have been calculated based on $E_i = n \cdot p_i$ and put in the chart below.

   (b) The test statistic (χ^2) is calculated with the aid of the table below.

(i)	Color	Observed Frequency O_i	Assumed Probability p_i (from H_o)	Expected Frequency $E_i = n \cdot p_i$	$\dfrac{(O_i - E_i)^2}{E_i}$
(1)	Brown	67	0.3	60	$\frac{(67-60)^2}{60} = 0.817$
(2)	Yellow	35	0.2	40	$\frac{(35-40)^2}{40} = 0.625$
(3)	Red	44	0.2	40	$\frac{(44-40)^2}{40} = 0.400$
(4)	Orange	18	0.1	20	$\frac{(18-20)^2}{20} = 0.200$
(5)	Green	26	0.1	20	$\frac{(26-20)^2}{20} = 1.800$
(6)	Blue	10	0.1	20	$\frac{(10-20)^2}{20} = 5.000$
	Total	$n = 200$			$\chi^2 = 8.842$

3. **Fail to reject** H_o because of either of the following.

   (a) The critical value from Table 5 (page 293) with 5 degrees of freedom at the 0.05 significance level is 11.070 and the test statistic (8.842) is not in the rejection region.

   (b) The P-value (using software) is about 0.116 which is larger than α.

4. **Concluding Statement:** There is not enough evidence to conclude that the distribution of outcomes does not fit the expected distribution from the null hypothesis. Specifically, there is not enough evidence to conclude that the distribution of colors does not fit what the manufacturer claims.

7 Preliminary Information: $\boxed{n = 3000}$ $\boxed{\text{d.f} = (2-1)(2-1) = 1}$ $\boxed{\alpha = 0.01}$.

1. There is no specific claim but this does not affect our null and alternate hypotheses.
   $\mathbf{H_o}$: The variables (vaccine and flu) are independent.
   $\mathbf{H_1}$: The variables are dependent.

2. **Calculate the test statistic.**

   (a) The expected frequencies (E_i's) are calculated using equation (11.3) with the totals from the original contingency table. These values are placed in the table of expected frequencies below.

$$E_i = \frac{\text{(Row Total)(Column Total)}}{\text{Table Total}}$$

Expected Frequencies (E_i's)

	Got Vaccine	No Vaccine
Got Flu	$\frac{(63)(1500)}{3000} = 31.5$	$\frac{(63)(1500)}{3000} = 31.5$
No Flu	$\frac{(2937)(1500)}{3000} = 1468.5$	$\frac{(2937)(1500)}{3000} = 1468.5$

   (b) The test statistic is calculated by equation (11.2):

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(18-31.5)^2}{31.5} + \frac{(45-31.5)^2}{31.5} + \frac{(1482-1468.5)^2}{1468.5} + \frac{(1455-1468.5)^2}{1468.5} \approx \boxed{11.820}.$$

3. **Reject** H_o because of either of the following.

(a) The critical value from Table 5 (page 293) with 1 degree of freedom at the 0.01 significance level is 6.635 and the test statistic (11.820) is in the rejection region.

(b) The P-value (using software) is about 0.000586 which is smaller than α.

4. **Concluding Statement:** The evidence suggests that the variables are dependent. Specifically, the probability of getting the flu is dependent upon whether or not the person is vaccinated.

9 Preliminary Information: $\boxed{n = 40}$ $\boxed{\text{d.f} = (2-1)(2-1) = 1}$ $\boxed{\alpha = 0.05}$.

1. There is no specific claim but this does not affect our null and alternate hypotheses.
   $\mathbf{H_o}$: The variables (wins/losses and with/without Joe) are independent.
   $\mathbf{H_1}$: The variables are dependent.

2. **Calculate the test statistic.**

   (a) The expected frequencies (E_i's) are calculated using equation (11.3) with the totals from the original contingency table. These values are placed in the table of expected frequencies below.

$$E_i = \frac{(\text{Row Total})(\text{Column Total})}{\text{Table Total}}$$

Expected Frequencies (E_i's)

	Wins	Losses
Home	$\frac{(24)(21)}{40} = 12.6$	$\frac{(24)(19)}{40} = 11.4$
Visitor	$\frac{(16)(21)}{40} = 8.4$	$\frac{(16)(19)}{40} = 7.6$

   (b) The test statistic is calculated by equation (11.2):

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(15-12.6)^2}{12.6} + \frac{(9-11.4)^2}{11.4} + \frac{(6-8.4)^2}{8.4} + \frac{(10-7.6)^2}{7.6} \approx \boxed{2.4060}.$$

3. **Fail to reject** H_o because of either of the following.

   (a) The critical value from Table 5 (page 293) with 1 degree of freedom at the 0.05 significance level is 3.841 and the test statistic (2.4060) is not in the rejection region.

   (b) The P-value (using software) is about 0.1209 which is greater than α.

4. **Concluding Statement:** There is not enough evidence to conclude that the variables are dependent. Specifically, the outcome of the games (win/loss) is not significantly dependent on whether or not Joe played.

11 Preliminary Information: $\boxed{n = 897}$ $\boxed{\text{d.f} = (2-1)(3-1) = 2}$ $\boxed{\alpha = 0.05}$.

1. There is no specific claim but this does not affect our null and alternate hypotheses.
   $\mathbf{H_o}$: The variables (party affiliation and opinion on the gun law) are independent.
   $\mathbf{H_1}$: The variables are dependent.

2. **Calculate the test statistic.**

(a) The expected frequencies (E_i's) are calculated using equation (11.3) with the totals from the original contingency table.

$$E_i = \frac{\text{(Row Total)(Column Total)}}{\text{Table Total}}$$

Expected Frequencies (E_i's)

	Republican	Independent	Republican
Vote For	249.8	247.3	249.8
Vote Against	50.2	49.7	50.2

(b) The test statistic is calculated by equation (11.2):

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(219 - 249.8)^2}{249.8} + \frac{(240 - 247.3)^2}{247.3} + \frac{(288 - 249.8)^2}{249.8} + \frac{(81 - 50.2)^2}{50.2} + \frac{(57 - 49.7)^2}{49.7} + \frac{(12 - 50.2)^2}{50.2} \approx \boxed{58.924}.$$

3. **Reject** H_o because of either of the following.

   (a) The critical value from Table 5 (page 293) with 2 degrees of freedom at the 0.05 significance level is 5.991 and the test statistic (58.924) is deep in the rejection region.

   (b) The P-value (using software) is about 1.6×10^{-13} which is much smaller than α.

4. **Concluding Statement:** The evidence suggests that the variables are dependent. Specifically, there is a dependent relationship between opinion on background checks for all gun purchases and party affiliation. Notice, all of the parties have a strong majority in favor of the law. The extent of that majority is what causes the dependence.

13 Preliminary Information: $\boxed{n = 132}$ $\boxed{\text{d.f} = (2-1)(5-1) = 4}$ $\boxed{\alpha = 0.05}$.

1. There is no specific claim but this does not affect our null and alternate hypotheses.
   $\mathbf{H_o}$: The variables (grade and section) are independent.
   $\mathbf{H_1}$: The variables are dependent.

2. **Calculate the test statistic.**

   (a) The expected frequencies (E_i's) are calculated using equation (11.3) with the totals from the original contingency table.

$$E_i = \frac{\text{(Row Total)(Column Total)}}{\text{Table Total}}$$

Expected Frequencies (E_i's)

	A	B	C	D	F
AM	9.7	14.5	15.5	14.1	10.2
PM	10.3	15.5	16.5	14.9	10.8

   (b) The test statistic is calculated by equation (11.2):

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(5 - 9.7)^2}{9.7} + \frac{(11 - 14.5)^2}{14.5} + \ldots + \frac{(11 - 14.9)^2}{14.9} + \frac{(7 - 10.8)^2}{10.8} \approx \boxed{11.045}.$$

3. **Reject** H_o because of either of the following.

(a) The critical value from Table 5 (page 293) with 4 degrees of freedom at the 0.05 significance level is 9.488 and the test statistic (11.045) is in the rejection region.

(b) The P-value (using software) is about 0.0261 which is smaller than α.

4. **Concluding Statement:** The evidence suggests that the variables are dependent. Specifically, there is a significant dependent relationship between grades and the section of the course.

15a 1. **Claim:** There is a difference in mean wait-times between the registers.

- $\mathbf{H_o}$: $\mu_1 = \mu_2 = \mu_3$
- $\mathbf{H_1}$: At least one of the means is different from the others.

2. The test statistic and the P-value are given, $F = 3.285$ and the P-value $= 0.0574$.

3. Reject H_o because the P-value (.0574) is less than α (0.10).

4. **Concluding Statement:** There is sufficient evidence to conclude that the population means are not equal. Specifically, at the 0.10 significance level, the data supports the claim that there is a difference in mean wait-times between the registers.

15b Yes. At the 0.05 significance, the P-value (.0574) is now larger than α and we fail to reject the null hypothesis. Specifically, at the 0.05 significance level, there is not enough evidence to support the claim that there is a difference in mean wait-times between the registers.

17a 1. **Claim:** There is a difference in mean number of clients served per hour for these employees.

- $\mathbf{H_o}$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
- $\mathbf{H_1}$: At least one of the means is different from the others.

2. The test statistic and the P-value are given, $F = 3.556$ and the P-value $= 0.0274$.

3. Reject H_o because the P-value (0.0274) is less than α (0.05).

4. **Concluding Statement:** There is sufficient evidence to conclude that the population means are not equal. Specifically, at the 0.05 significance level, the data supports the claim that there is a difference in mean number of clients served per hour for these employees.

17b Yes. The P-value (0.0274) is now greater than α and you fail to reject the null hypothesis. Specifically, at the 0.01 significance level, there is not enough evidence to support the claim that there is a difference in mean number of clients served per hour for these employees.

19a 1. **Claim:** There is a difference in mean number of defects between production days.

- $\mathbf{H_o}$: $\mu_{monday} = \mu_{tuesday} = \mu_{wednesday} = \mu_{thursday} = \mu_{friday}$
- $\mathbf{H_1}$: At least one of the means is different from the others.

2. The test statistic is $F = 12.054$ and the P-value $= 1.6 \times 10^{-8}$ (0.000000016).

3. Reject H_o because the P-value is much less than α.

4. **Concluding Statement:** There is sufficient evidence to conclude that the population means are not equal. Specifically, at the 0.05 significance level, the data supports the claim that there is a difference in mean number of defects between production days.

19b No. The P-value is still much smaller than α.