24. a) Random sample; less than 10% of all high-school students; many more than 10 successes and 10 failures. 90% confidence interval for proportion who cheat: 72.9% to 75.1%.
b) Based on this information, we are 90% confident that the proportion of all high-school students who have cheated at least once on a test is between 72.9% and 75.1%.
c) 90% of all such random samples will produce confidence intervals that contain the true proportion of students who cheat.
d) Wider. More confidence means a larger margin of error.

25. a) Yes. $0.8(60) = 48, 0.2(60) = 12$. Both are ≥ 10.
b) 0.834
c) Higher. Bigger sample means smaller standard deviation for $\hat{p}$.
d) Answers will vary. For $n = 500$, the probability is 0.997.

26. a) H_0: Progress is on track, $p = 0.20$. H_A: Progress is off track, $p > 0.20$.
b) Random samples; less than 10% of all high-school students; many more than 10 failures and successes.
c) 0.0008
d) If the proportion were 20%, the probability of seeing a value as high as (or higher than) 23% in a random sample this large is about 0.0008.
e) By 2006, the rate of cigarette smoking in high-school students was higher than the target of 20%.
f) Type I

27. a) 54.4 to 62.5%
b) Based on this study, with 95% confidence the proportion of Crohn's disease patients who will respond favorable to infliximab is between 54.4% and 62.5%.
c) 95% of all such random samples will produce confidence intervals that contain the true proportion of patients who respond favorably.

28. No. The probability of 30% or more is very small: 6.1×10^{-5}.

29. At least 423, assuming that p is near 50%.

30. a) An experiment
b) A one-sided test, since they are interested only in a decrease in percentage needing repairs.
c) Deciding the additive reduces the number of repairs needed when there really is no difference.
d) Deciding the additive makes no difference when it really does reduce the number of repairs needed.
e) Type II
f) Given that the two groups received roughly the same use and care, yes. They can't necessarily claim it will work for all cars, only cars similar to their fleet.

31. a) Random sample (?); certainly less than 10% of all preemies and normal babies; more than 10 failures and successes in each group. 1.7% to 16.3% greater for normal-birth weight children.
b) Since 0 is not in the interval, there is evidence that preemies have a lower high school graduation rate than children of normal birth weight.
c) Type I, since we rejected the null hypothesis.

32. a) We are 95% confident that between 11.6% and 16.4% of Texas children wear helmets when biking, roller skating, or skateboarding, based on these data.
b) The data might not be a random sample.
c) About 408, using the previous 14% as $\hat{p}$.

33. a) H_0: The computer is undamaged. H_A: The computer is damaged.
b) 20% of good PCs will be classified as damaged (bad), while all damaged PCs will be detected (good).
c) 3 or more. d) 20%
e) By switching to two or more as the rejection criterion, 7% of the good PCs will be misclassified, but only 10% of the bad ones will, increasing the power from 20% to 90%.

34. a) Increase b) Decrease

35. The null hypothesis is that Bush's disapproval proportion is 66%—the Nixon benchmark. The one-tailed test has a z-value of −2.00, so the P-value is 0.0228. It looks like Bush's May 2007 ratings were better than the Nixon benchmark low.

36. The null hypothesis is that the percentage of students who attain a GPA of at least 3.5 remained 20% in 2000. The sample proportion of 25% is more than four standard deviations above the hypothesized rate, strong evidence the results are not due to chance. This may be an indication of grade inflation.

37. a) The company is interested only in confirming that the athlete is well known.
b) Type I: the company concludes that the athlete is well known, but that's not true. It offers an endorsement contract to someone who lacks name recognition. Type II: the company overlooks a well-known athlete, missing the opportunity to sign a potentially effective spokesperson.
c) Type I would be more likely, Type II less likely.

38. a) Although 27% of the people polled could identify her, her name recognition rate in the whole population could be less than the required 25%.
b) Type II. c) Higher.

39. I am 95% confident that the proportion of U.S. adults who favor nuclear energy is between 7 and 19 percentage points higher than the proportion who would accept a nuclear plant near their area.

40. We're 95% confident that between 46% and 60% of anorexia patients will drop out of treatment programs. However, this wasn't a random sample of all patients; they were assigned a treatment rather than choosing one on their own, and they may have had different experiences if they were not part of an experiment.

CHAPTER 23

1. a) 1.74 b) 2.37 c) 0.0524 d) 0.0889

2. a) 2.36 b) 2.62 c) 0.9829 d) 0.0381

3. Shape becomes closer to Normal; center does not change; spread becomes narrower.

4. The critical value becomes smaller, approaching 1.96.

5. a) The confidence interval is for the population mean, not the individual cows in the study.
b) The confidence interval is not for individual cows.
c) We *know* the average gain in this study was 56 pounds!
d) The average weight gain of all cows does not vary. It's what we're trying to estimate.
e) No. There is not a 95% chance for another sample to have an average weight gain between 45 and 67 pounds. There is a 95% chance that another sample will have its average weight gain within two standard errors of the true mean.

6. a) Nine out of 10 intervals will contain the true mean salary; different samples will produce different intervals.
b) This is correct.
c) The interval is for the population mean, not individual teachers.
d) The interval is for the mean, not individual teachers.
e) The interval addresses only Nevada teachers, not the entire country.

7. a) No. A confidence interval is not about individuals in the population.
b) No. It's not about individuals in the sample, either.
c) No. We know the mean cost for students in the sample was $1196.
d) No. A confidence interval is not about other sample means.
e) Yes. A confidence interval estimates a population parameter.

8. a) No. The confidence interval is not about the years in the sample.
b) No. The confidence interval does not predict what will happen in any one year.

c) No. The confidence interval was not based on a sample of days.
d) Yes. The confidence interval estimates the true mean.
e) No. We know that the mean annual snowfall for the last century was 23".

9. a) Based on this sample, we can say, with 95% confidence, that the mean pulse rate of adults is between 70.9 and 74.5 beats per minute.
b) 1.8 beats per minute
c) Larger

10. a) Based on this sample, we can say, with 95% confidence, that the mean age at which babies begin to crawl is between 29.2 and 31.8 weeks.
b) 1.3 c) Smaller

11. The assumptions and conditions for a t-interval are not met. The distribution is highly skewed to the right and there is a large outlier.

12. The assumptions and conditions for a t-interval are not met. There is one cardholder who spent over $3,000,000 on his card. This made the standard deviation, and hence the SE, huge, resulting in a t-interval too wide to be useful.

13. a) Yes. Randomly selected group; less than 10% of the population; the histogram is not unimodal and symmetric, but it is not highly skewed and there are no outliers, so with a sample size of 52, the CLT says $\bar{y}$ is approximately Normal.
b) (98.06, 98.51) degrees F
c) We are 98% confident, based on the data, that the average body temperature for an adult is between 98.06°F and 98.51°F.
d) 98% of all such random samples will produce intervals containing the true mean temperature.
e) These data suggest that the true normal temperature is somewhat less than 98.6°F.

14. a) The data are a representative sample of less than 10% of all days; $n = 44$ is a large sample.
b) ($122.20, $129.80)
c) We are 90% confident the mean daily income for the parking garage is between $122.20 and $129.80, based on this sample.
d) 90% of all such samples will produce intervals that contain the true mean daily fee.
e) No. The interval is below $130.

15. a) Narrower. A smaller margin of error, so less confident.
b) Advantage: more chance of including the true value. Disadvantage: wider interval.
c) Narrower; due to the larger sample, the SE will be smaller.
d) About 252

16. a) More chance the interval contains the true mean.
b) The interval would be wider.
c) More data would result in a narrower interval at same confidence level.
d) About 99

17. a) (709.90, 802.54)
b) With 95% confidence, based on these data, the speed of light is between 299,709.9 and 299,802.5 km/sec.
c) Normal model for the distribution, independent measurements. These seem reasonable here, but it would be nice to see if the Nearly Normal Condition held for the data.

18. a) 7.9 km/sec
b) Should be narrower. Different mean will change the center of the interval. Larger sample size and smaller standard deviation will reduce the SE. Since the t-critical value is also smaller, the margin of error will be smaller.
c) New interval: (836.72, 868.08) km/sec. This experiment is worse than the first, which included 710.5. This interval is considerably above that value. There may have been a bias in this experiment's measurements.

19. a) Given no time trend, the monthly on-time departure rates should be independent. Though not a random sample, these months should be representative, and they're fewer than 10% of all months. The histogram looks unimodal, but slightly left-skewed; not a concern with this large sample.
b) $80.57 < \mu(\text{OT Departure\%}) < 81.80$
c) We can be 90% confident that the interval from 80.57% to 81.80% holds the true mean monthly percentage of on-time flight departures.

20. a) Given no time trend, the monthly late-arrival rates should be independent. Though not a random sample, these months should be representative, and they're fewer than 10% of all months. The histogram looks unimodal and symmetric.
b) $19.19 < \mu(\text{Late Arrival\%}) < 20.97$
c) We can be 99% confident that the interval from 19.19% to 21.0% holds the true mean monthly percentage of late flight arrivals.

21. The 95% confidence interval lies entirely above the 0.08 ppm limit, evidence that mirex contamination is too high and consistent with rejecting the null. We used an upper-tail test, so the P-value should therefore be smaller than $\frac{1}{2}(1 - 0.95) = 0.025$, and it was.

22. The 90% confidence interval contains the 325 mg limit, so they can't assert that the mean sodium content is less than 325—consistent with not rejecting the null. They used a lower-tail test, so we'd expect the P-value to be more than $\frac{1}{2}(1 - 0.90) = 0.05$, which it was.

23. If in fact the mean cholesterol of pizza eaters does not indicate a health risk, then only 7 of every 100 samples would have mean cholesterol levels as high (or higher) as observed in this sample.

24. If in fact this ball meets the velocity standard, then 34% of all samples of this size would have mean speeds at least as high as was recorded in this sample.

25. a) Upper-tail. We want to show it will hold 500 pounds (or more) easily.
b) They will decide the stands are safe when they're not.
c) They will decide the stands are unsafe when they are in fact safe.

26. a) Two-sided. If they're too big, they won't fit through the vein. If they're too small, they probably won't work well.
b) The catheters are rejected when in fact the diameters are fine, and the manufacturing process is needlessly stopped.
c) Catheters that do not meet specifications are allowed to be produced and sold.

27. a) Decrease α. This means a smaller chance of declaring the stands safe if they are not.
b) The probability of correctly detecting that the stands are capable of holding more than 500 pounds.
c) Decrease the standard deviation—probably costly. Increase the sample size—takes more time for testing and is costly. Increase α—more Type I errors. Increase the "design load" to be well above 500 pounds—again, costly.

28. a) Increase
b) The probability of correctly detecting deviations from 2 mm in diameter.
c) Increase d) Increase the sample size or increase α.

29. a) $H_0: \mu = 23.3; H_A: \mu > 23.3$
b) We have a random sample of the population. Population may not be normally distributed, as it would be easier to have a few much older men at their first marriage than some very young men. However, with a sample size of 40, $\bar{y}$ should be approximately Normal. We should check the histogram for severity of skewness and possible outliers.
c) $(\bar{y} - 23.3)/(s/\sqrt{40}) \sim t_{39}$ d) 0.1447
e) If the average age at first marriage is still 23.3 years, there is a 14.5% chance of getting a sample mean of 24.2 years or older simply from natural sampling variation.
f) We lack evidence that the average age at first marriage has increased from the mean of 23.3 years.

30. a) $H_0: \mu = 26; H_A: \mu < 26$
b) We have a representative sample, fewer than 10% of all trips, and a large enough sample that skewness should not be a problem.
c) $(\bar{y} - 26)/(4.83/\sqrt{50}) \sim t_{49}$ d) 0.0789
e) If the average fuel economy is 26 mpg, the chance of obtaining a sample mean of 25.02 or less by natural sampling variation is 8%.
f) Since $0.05 < P < 0.10$, there is some evidence that the company may not be achieving the fuel economy goal.

31. a) Probably a representative sample; the Nearly Normal Condition seems reasonable. (Show a Normal probability plot or histogram.) The histogram is nearly uniform, with no outliers or skewness.
b) $\bar{y} = 28.78, s = 0.40$ c) (28.36, 29.21) grams
d) Based on this sample, we are 95% confident the average weight of the content of Ruffles bags is between 28.36 and 29.21 grams.
e) The company is erring on the safe side, as it appears that, on average, it is putting in slightly more chips than stated.

32. a) Probably a representative sample; the Nearly Normal Condition seems reasonable. (Show a Normal probability plot or histogram.) The data are fairly symmetric with no apparent outliers.
b) $\bar{y} = 28.98, s = 0.36$ c) (28.61, 29.36) grams
d) With 95% confidence, the average weight of the content of Doritos bags is between 28.61 and 29.36 grams.
e) The company is erring on the safe side, as it appears that, on average, it is putting in slightly more chips than stated.

33. a) Type I; he mistakenly rejected the null hypothesis that $p = 0.10$ (or worse).
b) Yes. These are a random sample of bags and the Nearly Normal Condition is met (Show a Normal probability plot or histogram.); $t = -2.51$ with 7 df for a one-sided P-value of 0.0203.

34. a) He would have made a Type II error.
b) No. We'll consider these races a representative sample, and the Nearly Normal Condition is met (Show your plot.); $t = -0.7646$ with 7 df for a one-sided P-value of 0.2347.

35. a) Random sample; the Nearly Normal Condition seems reasonable from a Normal probability plot. The histogram is roughly unimodal and symmetric with no outliers. (Show plot.)
b) (1187.9, 1288.4) chips
c) Based on this sample, the mean number of chips in an 18-ounce bag is between 1187.9 and 1288.4, with 95% confidence. The *mean* number of chips is clearly greater than 1000. However, if the claim is about individual bags, then it's not necessarily true. If the mean is 1188 and the SD deviation is near 94, then 2.5% of the bags will have fewer than 1000 chips, using the Normal model. If in fact the mean is 1288, the proportion below 1000 will be less than 0.1%, but the claim is still false.

36. a) Random sample. Nearly Normal Condition is reasonable by examining a Normal probability plot. The histogram is roughly unimodal (although somewhat uniform) and symmetric with no outliers. (Show your plot.)
b) (132.0, 183.7) calories
c) The mean number of calories in a serving of vanilla yogurt is between 132 and 183.7, with 95% confidence. We conclude that the diet guide's claim of 120 calories is too low.

37. a) The Normal probability plot is relatively straight, with one outlier at 93.8 sec. Without the outlier, the conditions seem to be met. The histogram is roughly unimodal and symmetric with no other outliers. (Show your plot.)
b) $t = -2.63$, P-value $= 0.0160$. With the outlier included, we might conclude that the mean completion time for the maze is not 60 seconds; in fact, it is less.
c) $t = -4.46$, P-value $= 0.0003$. Because the P-value is so small, we reject H_0. Without the outlier, we see strong evidence that the average completion time for the maze is less than 60 seconds. The outlier here did not change the conclusion.
d) The maze does not meet the "one-minute average" requirement. Both tests rejected a null hypothesis of a mean of 60 seconds.

38. The data value of 102 feet is an outlier. When this is removed, the Normal probability plot is relatively straight. The Nearly Normal Condition seems satisfied. With the outlier removed, the histogram is roughly unimodal and symmetric with no other outliers. $H_0: \mu = 125; H_A: \mu > 125$ feet. With the outlier eliminated, $\bar{y} = 128.89$, $t = 3.29$, P-value $= 0.01$. With a P-value this low, we reject H_0. There is strong evidence to suggest that the tires will not bring the car to a complete stop within 125 feet. On the basis of these data, the company should not adopt the new tread pattern. Only 2 out of the 10 data values were less than the desired 125 feet, and 1 of these was an outlier.

39. a) $287.3 < \mu(\text{Drive Distance}) < 289.9$
b) These data are not a random sample of golfers. The top professionals are (unfortunately) not representative and were not selected at random. We might consider the 2006 data to represent the population of all professional golfers, past, present, and future.
c) The data are means for each golfer, so they are less variable than if we looked at all the separate drives.

40. a) The timeplot shows no pattern, so it seems that the measurements are independent. Although this is not a random sample, an entire year is measured, so it is likely that we have representative values. We certainly have fewer than 10% of all possible wind readings. Both the histogram and Normal probability plot suggest near normality.
b) Testing $H_0: \mu = 8$ mph vs. $H_A: \mu > 8$ mph with 1113 df gives $t = 0.1663$ for a P-value of about 0.44. Even though the observed mean wind speed is over 8 mph, I can't be confident that the true annual mean wind speed exceeds 8 mph. I would not recommend building a turbine at this site.

CHAPTER 24

1. Yes. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Meat} - \mu_{Beef}$.

2. Yes. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Meat} - \mu_{Beef}$.

3. a) Plausible values of $\mu_{Meat} - \mu_{Beef}$ are all negative, so the mean fat content is probably higher for beef hot dogs.
b) The difference is significant. c) 10%

4. a) Plausible values of $\mu_{Top} - \mu_{Front}$ are all negative, so the mean cycle time is probably higher for front-loading machines.
b) The difference is significant. c) 2%

5. a) False. The confidence interval is about means, not about individual hot dogs.
b) False. The confidence interval is about means, not about individual hot dogs.
c) True.
d) False. CI's based on other samples will also try to estimate the true difference in population means; there's no reason to expect other samples to conform to this result.
e) True.

6. a) False. The confidence interval is about means, not about individual machines.
b) False. The confidence interval is about means, not about individual machines.
c) False. CI's based on other samples will also try to estimate the true difference in population means; there's no reason to expect other samples to conform to this result.
d) True. e) True.

7. a) 2.927 b) Larger
c) Based on this sample, we are 95% confident that students who learn Math using the CPMP method will score, on average,

between 5.57 and 11.43 points better on a test solving applied Algebra problems with a calculator than students who learn by traditional methods.

d) Yes; 0 is not in the interval.

8. a) Based on this sample, we are 90% confident that people who receive either no or only verbal information about the image in a stereogram will take between 0.55 and 5.47 seconds longer, on average, to report they saw the image than people who receive both verbal and visual information.

b) Yes, since 0 is not in the interval. c) 2.46

d) 90% of all random samples will produce intervals that contain the true value of the mean difference between the times for these two groups.

e) Wider. More confidence means less precision.

f) Possibly. The wider interval may contain 0.

9. a) $H_0: \mu_C - \mu_T = 0$ vs. $H_A: \mu_C - \mu_T \neq 0$

b) Yes. Groups are independent, though we don't know if students were randomly assigned to the programs. Sample sizes are large, so CLT applies.

c) If the means for the two programs are really equal, there is less than a 1 in 10,000 chance of seeing a difference as large as or larger than the observed difference just from natural sampling variation.

d) On average, students who learn with the CPMP method do significantly worse on Algebra tests that do not allow them to use calculators than students who learn by traditional methods.

10. $H_0: \mu_C - \mu_T = 0$; $H_A: \mu_C - \mu_T \neq 0$. $t = 1.406$, df $= 590.05$, P-value $= 0.1602$. Because of the large P-value, we fail to reject H_0. Based on these samples, there is no evidence of a difference in mean scores on a test of word problems, whether students learned with CPMP or traditional methods.

11. a) (1.36, 4.64)

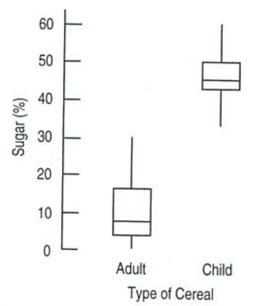
b) No; 5 minutes is beyond the high end of the interval.

12. a) The mean rates are roughly equal, but females are more variable.

b) Yes; boxplots look symmetric.

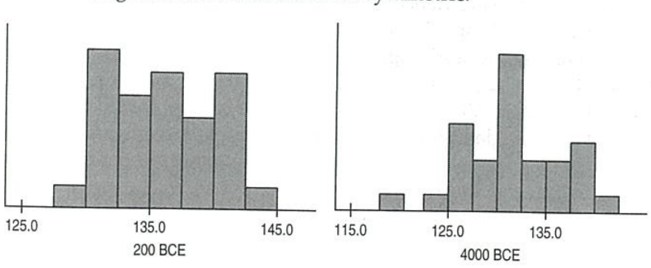
c) (−3.025, 3.275) d) Yes, 0 is in the interval.

13.



Random sample—questionable, but probably representative, independent samples, less than 10% of all cereals; boxplot shows no outliers—not exactly symmetric, but these are reasonable sample sizes. Based on these samples, with 95% confidence, children's cereals average between 32.49% and 40.80% more sugar content than adult's cereals.

14. a) Random sample (we assume), independent samples, histograms look unimodal and symmetric.



b) (1.88, 6.66) mm

c) These data provide evidence that mean maximum skull breadth in Egyptians in 200 B.C.E. was between 1.88 and 6.66 mm larger than that in 4000 B.C.E.

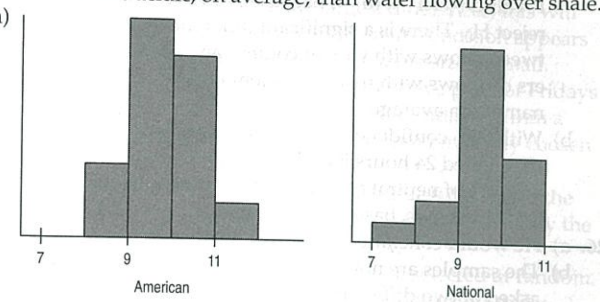
15. $H_0: \mu_N - \mu_C = 0$ vs. $H_A: \mu_N - \mu_C > 0$; $t = 2.207$; P-value $=$ 0.0168; df $= 33.4$. Because of the small P-value, we reject H_0. These data do suggest that new activities are better. The mean reading comprehension score for the group with new activities is significantly (at $\alpha = 0.05$) higher than the mean score for the control group.

16. a) $H_0: \mu_L - \mu_S = 0$ vs. $H_A: \mu_L - \mu_S \neq 0$

b) Don't know if the streams were a random sample or whether they are less than 10% of all Adirondack streams. Boxplots show outliers and shale may be skewed (median is equal to Q1 or Q3), but samples are large.

c) Based on these data, it appears that water flowing over limestone is less acidic, on average, than water flowing over shale.

17. a)



Both are unimodal and reasonably symmetric.

b) Based on these data, the average number of runs in an American League stadium is between 9.36 and 10.23, with 95% confidence.

c) No. The boxplot indicates it isn't an outlier.

d) We want to work directly with the average difference. The two separate confidence intervals do not answer questions about the difference. The difference has a different standard deviation, found by adding variances.

18. a) Males: (18.67, 20.11) pegs; females: (16.95, 18.87) pegs.

b) It suggests that there is no evidence of a difference, but the method is incorrect.

c) (0.29, 2.67) pegs

d) There is evidence of a difference: Males do better, on average, by between 0.29 and 2.67 pegs (with 95% confidence).

e) Two–sample inference

f) If we examine the gap between the means using two separate confidence intervals, we are essentially adding the margins of error, which are based on standard deviations. To be correct, we must instead add variances. That's exactly what happens when we create the confidence interval for the difference in means—the proper approach.

19. a) (−0.18, 0.89)

b) Based on these data, with 95% confidence, American League stadiums average between 0.18 fewer runs and 0.89 more runs per game than National League stadiums.

c) No; 0 is in the interval.

20. a) $H_0: \mu_N - \mu_S = 0$ vs. $H_A: \mu_N - \mu_S \neq 0$. $t = 6.47$, df $= 53.49$, P-value $= 3.2 \times 10^{-8}$. Because the P-value is low, we reject H_0. On the basis of these data, there is clear evidence that mortality rates are different. The mean rate in the north is significantly higher.

b) It will raise $\bar{y}$ for the north, but from looking at the boxplots and the fact that the mean and median are nearly the same, it probably will not change the conclusion of the test.

21. These are not two independent samples. These are before and after scores for the same individuals.

22. These are before and after scores for the same individuals, not independent samples, so we cannot proceed.

23. a) These data do not provide evidence of a difference in ad recall between shows with sexual content and violent content.

b) H_0: $\mu_S - \mu_N = 0$ vs. H_A: $\mu_S - \mu_N \neq 0$. $t = -6.08$, df $= 213.99$, P-value $= 5.5 \times 10^{-9}$. Because the P-value is low, we reject H_0. These data suggest that ad recall between shows with sexual and neutral content is different; those who saw shows with neutral content had higher average recall.

24. a) With 95% confidence, those who watch shows with violent content remembered an average of between 0.6 and 1.6 fewer brand names than those who saw shows with neutral content.

b) If they want viewers to remember their brand, advise that they consider advertising in shows with neutral content in preference to those with violent content. Of course, costs of the ad should be considered.

25. a) H_0: $\mu_V - \mu_N = 0$ vs. H_A: $\mu_V - \mu_N \neq 0$. $t = -7.21$, df $= 201.96$, P-value $= 1.1 \times 10^{-11}$. Because of the very small P-value, we reject H_0. There is a significant difference in mean ad recall between shows with violent content and neutral content; viewers of shows with neutral content remember more brand names, on average.

b) With 95% confidence, the average number of brand names remembered 24 hours later is between 1.45 and 2.41 higher for viewers of neutral content shows than for viewers of sexual content shows, based on these data.

26. a) He would conclude that recall is higher 24 hours later.

b) The samples are not independent. They are the same people asked at two different periods.

c) The first inquiry might influence people.

d) Use two different groups for each type of show. Interview one group immediately, the other 24 hours later.

27. H_0: $\mu_{big} - \mu_{small} = 0$ vs. H_A: $\mu_{big} - \mu_{small} \neq 0$; bowl size was assigned randomly; amount scooped by individuals and by the two groups should be independent. With 34.3 df, $t = 2.104$ and P-value $= 0.0428$. The low P-value leads us to reject the null hypothesis. There is evidence of a difference in the average amount of ice cream that people scoop when given a bigger bowl.

28. H_0: $\mu_{tumbler} - \mu_{highball} = 0$ vs. H_A: $\mu_{tumbler} - \mu_{highball} \neq 0$; glass size was assigned randomly; amount poured by individuals and by the two groups should be independent. With 194 df, $t = 7.71$; P-value < 0.001. The low P-value leads us to reject the null hypothesis and conclude that there is a difference in the amount of liquid that people pour, on average, when given a small wide tumbler as opposed to a tall narrow highball glass.

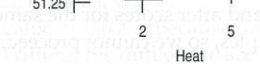
29. a) The 95% confidence interval for the difference is (0.61, 5.39). 0 is not in the interval, so scores in 1996 were significantly higher. (Or the t, with more than 7500 df, is 2.459 for a P-value of 0.0070.)

b) Since both samples were very large, there shouldn't be a difference in how certain you are, assuming conditions are met.

30. a) The observed differences are too large to attribute to chance or natural sampling variation.

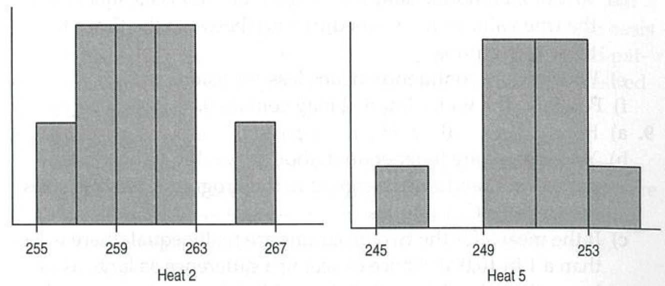
b) Type I c) No. There may be many other factors.

31. Independent Groups Assumption: The runners are different women, so the groups are independent. The Randomization Condition is satisfied since the runners are selected at random for these heats.



Nearly Normal Condition: The boxplots show an outlier, but we will proceed and then redo the analysis with the outlier deleted. When we include the outlier, $t = 0.035$ with a two-sided P-value of 0.97. With the outlier deleted, $t = -1.14$, with P $= 0.2837$. Either P-value is so large that we fail to reject the null hypothesis of equal means and conclude that there is no evidence of a difference in the mean times for runners in unseeded heats.

32. Independent Groups Assumption: The swimmers in the two heats are different women, so the groups are independent. The histograms indicate the Nearly Normal Condition is satisfied.



The Randomization Condition is not strictly satisfied since the swimmers are not selected at random, but if we can consider these heats to be representative of seeded heats, we may be able to generalize the results. With $t = 7.19$ and a P-value less than 0.001, we reject the null hypothesis of equal means. There is strong evidence to suggest that seeded swimming heats have different mean times. It appears that on average times were faster in heat 5.

33. With $t = -4.57$ and a very low P-value of 0.0013, we reject the null hypothesis of equal mean velocities. There is strong evidence that golf balls hit off Stinger tees will have a higher mean initial velocity.

34. With $t = -9.70$ and a P-value less than 0.0001, we reject the null hypothesis of equal mean distances. There is strong evidence that the mean distance traveled will be greater for golf balls hit off Stinger tees. We can be 95% confident that they'll travel an average of between 10.7 and 17.0 yards farther.

35. a) We can be 95% confident that the interval 74.8 ± 178.05 minutes includes the true difference in mean crossing times between men and women. Because the interval includes zero, we cannot be confident that there is any difference at all.

b) Independence Assumption: There is no reason to believe that the swims are not independent or that the two groups are not independent of each other.

Randomization Condition: The swimmers are not a random sample from any identifiable population, but they may be representative of swimmers who tackle challenges such as this.

Nearly Normal Condition: the boxplots show no outliers. The histograms are unimodal; the histogram for men is somewhat skewed to the right. (Show your graphs.)

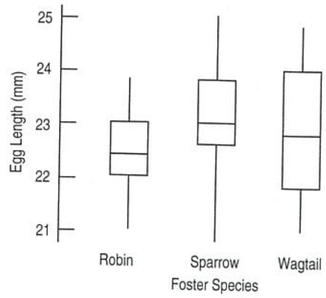
36. a) H_0: $\mu_M - \mu_R = 0$ vs. H_A: $\mu_M - \mu_R > 0$. $t = -0.70$, df $= 45.88$, P-value $= 0.7563$. Because the P-value is so large, we do not reject H_0. These data provide no evidence that listening to Mozart while studying is better than listening to rap.

b) With 90% confidence, the average difference in score is between 0.189 and 5.351 objects more for those who listen to no music while studying, based on these samples.

37. a) H_0: $\mu_R - \mu_N = 0$ vs. H_A: $\mu_R - \mu_N < 0$. $t = -1.36$, df $= 20.00$, P-value $= 0.0945$. Because the P-value is large, we fail to reject H_0. These data show no evidence of a difference in mean number of objects recalled between listening to rap or no music at all.

b) Didn't conclude any difference.

38.



$H_0: \mu_S - \mu_R = 0$ vs. $H_A: \mu_S - \mu_R \neq 0$.
$t = 1.641$, df $= 21.60$, P-value $= 0.115$.

Since $P > 0.05$, fail to reject H_0. There is no evidence of a difference in mean cuckoo egg length between robin and sparrow foster parents.

$H_0: \mu_S - \mu_W = 0$ vs. $H_A: \mu_S - \mu_W \neq 0$.
$t = 0.549$, df $= 26.86$, P-value $= 0.587$.

Since $P > 0.05$, fail to reject H_0. There is no evidence of a difference in mean cuckoo egg length between sparrow and wagtail foster parents.

$H_0: \mu_R - \mu_W = 0$ vs. $H_A: \mu_R - \mu_W \neq 0$.
$t = -1.012$, df $= 23.60$, P-value $= 0.322$.

Since $P > 0.05$, fail to reject H_0. There is no evidence of a difference in mean cuckoo egg length between robin and wagtail foster parents.

In general, we should be wary of doing three t-tests on the same data. Our Type I error rate is not the same for doing three tests as it is for doing one test. However, because none of the tests showed significant differences, this is less of a concern here.

CHAPTER 25

1. a) Randomly assign 50 hens to each of the two kinds of feed. Compare production at the end of the month.
b) Give all 100 hens the new feed for 2 weeks and the old food for 2 weeks, randomly selecting which feed the hens get first. Analyze the differences in production for all 100 hens.
c) Matched pairs. Because hens vary in egg production, the matched-pairs design will control for that.

2. a) Randomly assign half the volunteers to do the puzzles in a quiet room and half to do them with MTV on. Compare times.
b) Randomly assign half the volunteers to do a puzzle in a quiet room and half to do a puzzle with MTV on. Then have each do a puzzle under the other condition. Look at the differences in completion times.
c) Matched pairs. People vary in their ability to do crossword puzzles.

3. a) Show the same people ads with and without sexual images, and record how many products they remember in each group. Randomly decide which ads a person sees first. Examine the differences for each person.
b) Randomly divide volunteers into two groups. Show one group ads with sexual images and the other group ads without. Compare how many products each group remembers.

4. a) Select a random sample of freshmen. Weigh them when college starts in the fall and again when they leave for home in the spring. Examine the difference in weights.
b) Weigh a random sample of freshmen as they enter college in the fall to determine their average weight. Repeat with a new sample of students at the end of the spring semester. Compare the mean weights of the two groups.

5. a) Matched pairs—same cities in different periods.
b) There is a significant difference (P-value $= 0.0244$) in the labor force participation rate for women in these cities; women's participation seems to have increased between 1968 and 1972.

6. a) Two-sample. Clouds are independent of one another.
b) Based on these data, there is some evidence of a difference (P-value $= 0.0538$) in the amount of rain between seeded and unseeded clouds.

7. a) Use the paired t-test because we have pairs of Fridays in 5 different months. Data from adjacent Fridays within a month may be more similar than data from randomly chosen Fridays.
b) We conclude that there is evidence (P-value 0.0212) that the mean number of cars found on the M25 motorway on Friday the 13th is less than on the previous Friday.
c) We don't know if these Friday pairs were selected at random. If these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal Condition appears to be met by the differences, but the sample size is small.

8. a) The paired t-test is appropriate since we have pairs of Fridays in 6 different months. Data from adjacent Fridays within a month may be more similar than data from randomly chosen Fridays.
b) We conclude that there is evidence (P-value 0.0211) that the mean number of admissions to hospitals found on Friday the 13th is more than on the previous Friday.
c) We don't know if these Friday pairs were selected at random. Obviously, if these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal Condition appears to be met by the differences, but the sample size is small.

9. Adding variances requires that the variables be independent. These price quotes are for the same cars, so they are paired. Drivers quoted high insurance premiums by the local company will be likely to get a high rate from the online company, too.

10. Adding variances requires that the variables be independent. The wind speeds were recorded at nearby sites, so they are likely to be both high or both low at the same time.

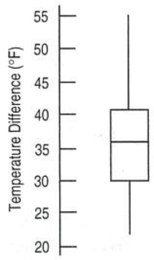
11. a) The histogram—we care about differences in price.
b) Insurance cost is based on risk, so drivers are likely to see similar quotes from each company, making the differences relatively smaller.
c) The price quotes are paired; they were for a random sample of fewer than 10% of the agent's customers; the histogram of differences looks approximately Normal.

12. a) The outliers are particularly windy days, but they were windy at both sites, making the difference in wind speeds less unusual.
b) The histogram and summaries of the differences are more appropriate because these are paired observations and all we care about is which site was more windy.
c) The wind measurements at the same times at two nearby sites are paired. We should be concerned that there might be a lack of independence from one time to the next, but the times were 6 hours apart and the *differences* in speeds are likely to be independent. Although the sample is not random, we can regard a sample this large as generally representative of wind speeds at these sites. The histogram of differences is unimodal, symmetric, and bell-shaped.

13. $H_0: \mu(Local - Online) = 0$ vs. $H_A: \mu(Local - Online) > 0$; with 9 df, $t = 0.83$. With a high P-value of 0.215, we don't reject the null hypothesis. These data don't provide evidence that online premiums are lower, on average.

14. $H_0: \mu(2 - 4) = 0$ vs. $H_A: \mu(2 - 4) \neq 0$; $t = 2.667$ with 1114 df. The P-value of 0.008 is very low, so we reject the null. There's strong evidence that the average wind speed is higher at site 2.

15.



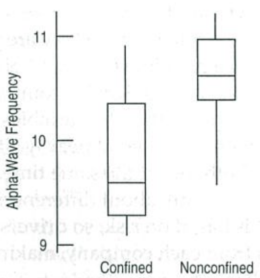
Data are paired for each city; cities are independent of each other; boxplot shows the temperature differences are reasonably symmetric, with no outliers. This is probably not a random sample, so we might be wary of inferring that this difference applies to all European cities. Based on these data, we are 90% confident that the average temperature in European cities in July is between 32.3°F and 41.3°F higher than in January.

16. We are 90% confident that the average women's winning marathon time is between 16.25 and 17.45 minutes higher than the men's based on this sample.

17. Based on these data, we are 90% confident that boys, on average, can do between 1.6 and 13.0 more push-ups than girls (independent samples—not paired).

18. a) H_0: $\mu_{NC} - \mu_C = 0$ vs. H_A: $\mu_{NC} - \mu_C \neq 0$; μ_{NC} is the mean for nonconfined inmates, μ_C is the mean for inmates confined to solitary.

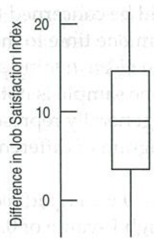
b) Groups are independent of each other, not paired; random assignment to groups, less than 10% of all inmates, boxplot shows no outliers in either group.



c) Two-sample t-test statistic: 3.357, df = 19.6, P-value = 0.0038.

d) Because the P-value is so small, we reject H_0. Solitary confinement makes a difference in mean alpha-wave frequencies; it seems those subjected to confinement have lower frequencies.

19. a) Paired sample test. Data are before/after for the same workers; workers randomly selected; assume fewer than 10% of all this company's workers; boxplot of differences shows them to be symmetric, with no outliers.



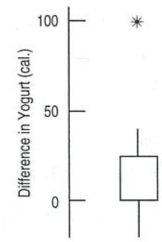
b) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 3.60$, P-value = 0.0029. Because $P < 0.01$, reject H_0. These data show evidence that average job satisfaction has increased after implementation of the program.

c) Type I

20. a) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 1.75$, df = 5, P-value = 0.0699. Because $P > 0.05$, we fail to reject H_0. These data do not provide enough evidence to conclude that the summer school program is worthwhile, at $\alpha = 0.05$.

b) Type II

21. H_0: $\mu_D = 0$ vs. H_A: $\mu_D \neq 0$. Data are paired by brand; brands are independent of each other; fewer than 10% of all yogurts (questionable); boxplot of differences shows an outlier (100) for Great Value:



With the outlier included, the mean difference (Strawberry – Vanilla) is 12.5 calories with a t-stat of 1.332, with 11 df, for a P-value of 0.2098. Deleting the outlier, the difference is even smaller, 4.55 calories with a t-stat of only 0.833 and a P-value of 0.4241. With P-values so large, we do not reject H_0. We conclude that the data do not provide evidence of a difference in mean calories.

22. a) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 4.47$, df = 9, P-value = 0.0008. Because of the very small P-value, we reject H_0. These data provide strong evidence that cars get significantly better mileage, on average, with premium than with regular gasoline.

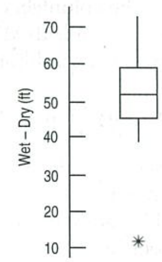
b) (1.18, 2.82) miles per gallon

c) Premium gasoline costs more than regular.

d) $t = 1.25$, df = 17.89, P-value is 0.1144. Would have decided no difference. The variation in the cars' performances is larger than the differences.

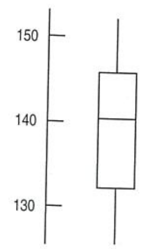
23. a) Cars were probably not a simple random sample, but may be representative in terms of stopping distance; boxplot does not show outliers, but does indicate right skewness. A 95% confidence interval for the mean stopping distance on dry pavement is (131.8, 145.6) feet.

b) Data are paired by car; cars were probably not randomly chosen, but representative; boxplot shows an outlier (car 4) with a difference of 12. With deletion of that car, a Normal probability plot of the differences is relatively straight.



Retaining the outlier, we estimate with 95% confidence that the average braking distance is between 38.8 and 62.6 feet more on wet pavement than on dry, based on this sample. (Without the outlier, the confidence interval is 47.2 to 62.8 feet.)

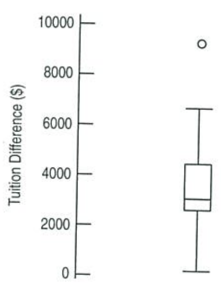
24. a) Not a simple random sample, but most likely representative; stops most likely independent of each other; boxplot is symmetric with no outliers.



Based on these data, with 95% confidence, the average braking distance for these tires on dry pavement is between 133.6 and 145.2 feet.

b) Not simple random samples, but most likely representative; stops most likely independent of each other; less than 10% of all possible wet stops; Normal probability plots are relatively straight. Based on these data, with 95% confidence, the average increase in distance for these tires on wet pavement is between 51.4 and 74.6 feet

25. a) Paired Data Assumption: Data are paired by college. Randomization Condition: This was a random sample of public colleges and universities. 10% Condition: these are fewer than 10% of all public colleges and universities.



Normal Population Assumption: U.C. Irvine seems to be an outlier; we might consider removing it.

b) Having deleted the observation for U.C.-Irvine, whose difference of \$9300 was an outlier, we are 90% confident, based on the remaining data, that nonresidents pay, on average, between \$2615.31 and \$3918.02 more than residents. If we retain the outlier, the interval is (\$2759, \$4409).

c) Assertion is reasonable; with or without the outlier, \$3500 is in the confidence interval.

26. Using a *t*-test for paired differences, $t = -0.86$ and two-tailed $P = 0.396$. With a P-value so high, we fail to reject the null hypothesis of no mean difference. There is no evidence that sexual images in ads affects people's ability to remember the product being advertised.

27. a) 60% is 30 strikes; H_0: $\mu = 30$ vs. H_A: $\mu > 30$. $t = 6.07$, P-value $= 3.92 \times 10^{-6}$. With a very small P-value, we reject H_0. There is very strong evidence that players can throw more than 60% strikes after training, based on this sample.

b) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 0.135$, P-value $= 0.4472$. With such a high P-value, we do not reject H_0. These data provide no evidence that the program has improved pitching in these Little League players.

28. If this group is representative of all students, we can be 95% confident that freshmen gain a mean of between 1.40 and 2.43 pounds during their first 12 weeks at college. That's strong evidence of a weight gain, but it's unlikely that it amounts to 15 pounds for the whole first year.

PART VI REVIEW

1. a) H_0: $\mu_{Jan} - \mu_{Jul} = 0$; H_A: $\mu_{Jan} - \mu_{Jul} \neq 0$. $t = -1.94$, df $= 43.68$, P-value $= 0.0590$. Since $P < 0.10$, reject the null. These data show a significant difference in mean age to crawl between January and July babies.

b) H_0: $\mu_{Apr} - \mu_{Oct} = 0$; H_A: $\mu_{Apr} - \mu_{Oct} \neq 0$. $t = -0.92$; df $= 59.40$; P-value $= 0.3610$. Since $P > 0.10$, do not reject the null; these data do not show a significant difference between April and October with regard to the mean age at which crawling begins.

c) These results are not consistent with the claim.

2. H_0: $\mu_D = 0$; H_A: $\mu_D > 0$. $t = 1.36$; df $= 20$; P-value $= 0.0949$. Because the P-value is high, we do not reject H_0. These data do not show that floral scent improved the average maze completion time between scented and unscented.

3. H_0: $p = 0.26$; H_A: $p \neq 0.26$. $z = 0.946$; P-value $= 0.3443$. Because the P-value is high, we do not reject H_0. These data do not show that the Denver-area rate is different from the national rate in the proportion of businesses with women owners.

4. a) We are 95% confident the average savings in Canada for prescription drugs is between \$77.57 and \$174.43.

b) We are 95% confident that the average savings in Canada for prescription drugs is between 40.1% and 64.2%.

c) Using percents makes the histogram more unimodal and symmetric.

d) Probably would change. The pharmacy may have listed only the 12 drugs with the "best" savings.

5. Based on these data, we are 95% confident that the mean difference in aluminum oxide content is between −3.37 and 1.65. Since the interval contains 0, the means in aluminum oxide content of the pottery made at the two sites could reasonably be the same.

6. We are 95% confident that the proportion of streams in the Adirondacks with shale substrates is between 32.8% and 47.4%.

7. a) H_0: $p_{ALS} - p_{Other} = 0$; H_A: $p_{ALS} - p_{Other} > 0$. $z = 2.52$; P-value $= 0.0058$. With such a low P-value, we reject H_0. This is strong evidence that there is a higher proportion of varsity athletes among ALS patients than those with other disorders.

b) Observational retrospective study. To make the inference, one must assume the patients studied are representative.

8. Paired samples; boxplot shows no strong skewness or outliers. One might wonder how the individuals in the study were selected. We are 95% confident that average percentage of 15-year-old males who have been drunk is between 4.5% and 11.4% more than 15-year-old females for these countries. We cannot infer that these percentages are true for other countries.

9. H_0: $\mu = 7.41$; H_A: $\mu \neq 7.41$. $t = 2.18$; df $= 111$; P-value $= 0.0313$. With such a low P-value, we reject H_0. Assuming that Missouri babies fairly represent the United States, these data suggest that American babies are different from Australian babies in birth weight; it appears American babies are heavier, on average.

10. a) 88.6% b) 82.2%

c) The petition would be certified when there are not enough valid signatures.

d) A correct petition is not certified.

e) H_0: $p = 0.822$; H_A: $p > 0.822$. $z = 7.48$; P-value $= 3.64 \times 10^{-14}$ With such a low P-value, we reject H_0. This sample provides sufficient evidence for certification of the petition.

f) Increase sample size.

11. a) If there is no difference in the average fish sizes, the chance of seeing an observed difference this large just by natural sampling variation is less than 0.1%.

b) If cost justified, feed them a natural diet. c) Type I

12. We have two independent samples, but we don't know how these vehicles were chosen. Even if we consider them representative, the samples are small, and the data for SUVs are skewed to the right. These data are not appropriate for inferences.

13. a) Assuming the conditions are met, from these data we are 95% confident that patients with cardiac disease average between 3.39 and 5.01 years older than those without cardiac disease.