Chapter 20 – Comparing Groups

Section 20.1

1. Canada.

$$SE\left(\hat{p}_{Can}-\hat{p}_{Am}\right)=\sqrt{\frac{\hat{p}_{Can}\hat{q}_{Can}}{n_{Can}}+\frac{\hat{p}_{Am}\hat{q}_{Am}}{n_{Am}}}=\sqrt{\frac{\left(\frac{192}{960}\right)\left(\frac{768}{960}\right)}{960}+\frac{\left(\frac{170}{1250}\right)\left(\frac{1080}{1250}\right)}{1250}}=0.0161$$

2. Non-profits.

$$SE\left(\hat{p}_{Non}-\hat{p}_{For}\right)=\sqrt{\frac{\hat{p}_{Non}\hat{q}_{Non}}{n_{Non}}+\frac{\hat{p}_{For}\hat{q}_{For}}{n_{For}}}=\sqrt{\frac{\left(\frac{377}{422}\right)\left(\frac{45}{422}\right)}{422}+\frac{\left(\frac{431}{518}\right)\left(\frac{87}{518}\right)}{518}}=0.0223$$

3. Canada, deux.

We are 95% confident that, based on these data, the proportion of foreign-born Canadians is between 3.24% and 9.56% more than the proportion of foreign-born Americans.

4. Non-profits, part 2.

We are 95% confident that, based on these data, the proportion of people who are “highly satisfied” working at non-profits is between 1.77% and 10.50% higher than the proportion of people who are “highly satisfied” working at for-profit companies.

5. Canada, trois.

If we were to take repeated samples of these sizes of Canadians and Americans, and compute two-proportion confidence intervals, we would expect 95% of the intervals to contain the true difference in the proportions of foreign-born citizens.

6. Non-profits, part 3.

If we were to take repeated samples of these sizes of people who work at non-profits and for-profits, and compute two-proportion confidence intervals, we would expect 95% of the intervals to contain the true difference in the proportions of those who are highly satisfied.

Section 20.2

7. Canada, encore.

We must assume the data were collected randomly and that the Americans selected are independent of the Canadians selected. Both assumptions should be met. Also, for both groups, we have at least 10 national-born and foreign-born citizens and the sample sizes are less than 10% of the population sizes. All conditions for inference are met.

8. Non-profits, again.

We have random samples and we must assume the samples were collected independently of each other. This should be met. For both groups, we have at least 10 people who are highly satisfied and 10 people who are not. Finally, the sample sizes are less than 10% of the population sizes. All conditions for inference are met.

Section 20.3

9. Canada, test.

a) $\hat{p}_{Can} - \hat{p}_{Am} = \dfrac{192}{960} - \dfrac{170}{1250} = 0.064$

b) $z = \dfrac{\hat{p}_{Can} - \hat{p}_{Am}}{\sqrt{\dfrac{\hat{p}_{Can}\hat{q}_{Can}}{n_{Can}} + \dfrac{\hat{p}_{Am}\hat{q}_{Am}}{n_{Am}}}} = \dfrac{\dfrac{192}{960} - \dfrac{170}{1250}}{\sqrt{\dfrac{\left(\frac{192}{960}\right)\left(\frac{768}{960}\right)}{960} + \dfrac{\left(\frac{170}{1250}\right)\left(\frac{1080}{1250}\right)}{1250}}} \approx 3.964$

(Using $SE_{pooled}\left(\hat{p}_{Can} - \hat{p}_{Am}\right)$, $z \approx 4.03$)

c) Since the P-value is < 0.001, which is less than $\alpha = 0.05$, reject the null hypothesis. There is very strong evidence that the proportion of foreign-born citizens is different in Canada than it is in the United States. According to this data, the proportion of foreign-born Canadians is likely to be the higher of the two.

10. Non-profits test.

a) $\hat{p}_{Non} - \hat{p}_{For} = \dfrac{377}{422} - \dfrac{431}{518} \approx 0.061$

b) $z = \dfrac{\hat{p}_{Non} - \hat{p}_{For}}{\sqrt{\dfrac{\hat{p}_{Non}\hat{q}_{Non}}{n_{Non}} + \dfrac{\hat{p}_{For}\hat{q}_{For}}{n_{For}}}} = \dfrac{\dfrac{377}{422} - \dfrac{431}{518}}{\sqrt{\dfrac{\left(\frac{377}{422}\right)\left(\frac{45}{422}\right)}{422} + \dfrac{\left(\frac{431}{518}\right)\left(\frac{87}{518}\right)}{518}}} \approx 2.755$

(Using $SE_{pooled}\left(\hat{p}_{Non} - \hat{p}_{For}\right)$, $z \approx 2.691$)

c) Since the P-value = 0.007 is less than $\alpha = 0.05$, reject the null hypothesis. There is strong evidence that the proportion of highly satisfied workers at non-profits is higher than the proportion of highly satisfied workers at for-profits. These data suggest that the proportion higher at non-profits than at for-profits.

Section 20.4

11. Cost of shopping.

We must assume the samples were random or otherwise independent of each other. We also assume that the distributions are roughly normal, so it would be a good idea to check a histogram to make sure there isn't strong skewness or outliers.

12. Athlete ages.

We must assume the samples are unbiased and independent of each other. These conditions seem met. We also assume that the distributions are roughly normal, so it would be a good idea to check a histogram to make sure there isn't strong skewness or outliers.

13. Cost of shopping, again.

We are 95% confident that the mean purchase amount at Walmart is between \$1.85 and \$14.15 less than the mean purchase amount at Target.

14. Athlete ages, again.

We are 95% confident that the mean age of MLB players is between 0.41 years younger and 3.09 years older than the mean age of NFL players.

Section 20.5

15. Cost of shopping, once more.

The difference is –\$8 with an SE of 3.115, so the *t*-stat is –2.569. With 162.75 (or 163) df, the P-value is 0.011 which is less than 0.05. Reject the null hypothesis that the means are equal. There is evidence that the mean purchase amounts at the two stores are not the same. These data suggest that the mean purchase amount at Target is lower than the mean purchase price at Walmart.

16. Athlete ages, ninth inning.

The difference is 1.34 years with an SE of 0.8711 so the *t*-stat is 1.538. With 51.83 (or 52) df, the P-value is 0.065 which is greater than 0.05. Fail to reject the null hypothesis that the means are equal. There is no evidence of a difference in mean age of MLB and NFL players.

Section 20.7

17. Cost of shopping, yet again.

The *t*-statistic is –2.561 using the pooled estimate of the standard deviation, 3.124. There are 163 df so the P-value is still 0.011. We reach the same conclusion as before. Because the sample standard deviations are nearly the same and the sample sizes are large, the pooled test is essentially the same as the two-sample *t*-test.

18. Athlete ages, overtime.

The *t*-statistic is now 1.555 using the pooled estimate of the standard deviation. There are 60 df so the P-value is now 0.125. We reach the same conclusion as before. Because the sample standard deviations are nearly the same, the pooled test is nearly the same as the two-sample *t*-test.

19. Cost of shopping, once more.

No. The two-sample test is almost always the safer choice. In this case, the variances are likely to be quite different. The purchase prices of Italian sports cars are much higher and may be more variable than the domestic prices. They should use the two-sample t-test.

20. Athletes, extra innings.

No. The two-sample test is almost always the safer choice. In this case, the variances are likely to be quite different, since third grade Little League players will all be nearly the same age. They should use the two-sample *t*-test.

Chapter Exercises.

21. Online social networking.

It is very unlikely that samples would show an observed difference this large if, in fact, there was no real difference between the proportion of American adults who visited Facebook on a daily basis in 2013 and the proportion of American adults who visited Facebook on a daily basis 2010.

22. Science news.

If, in fact, there is no difference between the percentage 18-29-year-olds who read newspapers and the percentage of 30-49-year-olds who read newspapers, then it's not unusual to observe a difference of 4 percentage points by sampling.

23. Revealing information.

This test is not appropriate for these data, since the responses are not from independent groups, but are from the same individuals. The independent samples condition has been violated.

24. Regulating access.

The 790 parents are a subset of the 935 parents, so the two groups are not independent. This violates the independent samples condition.

25. Gender gap.

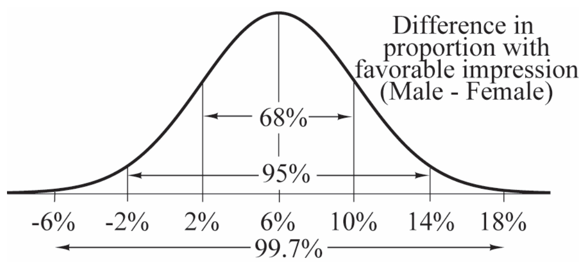
a) This is a stratified random sample, stratified by gender.

b) We would expect the difference in proportions in the sample to be the same as the difference in proportions in the population, with the percentage of respondents with a favorable impression of the candidate 6 percentage points higher among males.

c) The standard deviation of the difference in proportions is

$$SD(\hat{p}_M - \hat{p}_F) = \sqrt{\frac{\hat{p}_M \hat{q}_M}{n_M} + \frac{\hat{p}_F \hat{q}_F}{n_F}} = \sqrt{\frac{(0.59)(0.41)}{300} + \frac{(0.53)(0.47)}{300}} \approx 4\%.$$

d)



e) The campaign could certainly be misled by the poll. According to the model, a poll showing little difference could occur relatively frequently. That result is only 1.5 standard deviations below the expected difference in proportions.

26. Buy it again?

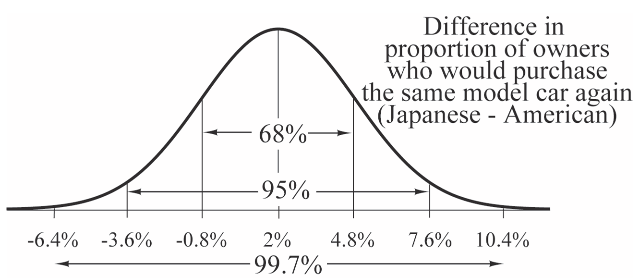
a) This is a stratified random sample, stratified by country of origin of the car.

b) We would expect the difference in proportions in the sample to be the same as the difference in proportions in the population, with the percentage of respondents who would purchase the same model again 2 percentage points higher among owners of Japanese cars than among owners of American cars.

c) The standard deviation of the difference in proportions is

$$SD(\hat{p}_J - \hat{p}_A) = \sqrt{\frac{\hat{p}_J \hat{q}_J}{n_J} + \frac{\hat{p}_A \hat{q}_A}{n_A}} = \sqrt{\frac{(0.78)(0.22)}{450} + \frac{(0.76)(0.24)}{450}} \approx 2.8\%.$$

d)



e) The magazine could certainly be misled by the poll. According to the model, a poll showing greater satisfaction among owners of American cars could occur relatively frequently. That result is a little more than one standard deviation below the expected difference in proportions.

27. Arthritis.

a) **Randomization condition:** Americans age 65 and older were selected randomly.
Independent groups assumption: The sample of men and the sample of women were drawn independently of each other.
Success/Failure condition: $n\hat{p}$ (men) = 411, $n\hat{q}$ (men) = 601, $n\hat{p}$ (women) = 535, and $n\hat{q}$ (women) = 527 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $$(\hat{p}_F - \hat{p}_M) \pm z^* \sqrt{\frac{\hat{p}_F\hat{q}_F}{n_F} + \frac{\hat{p}_M\hat{q}_M}{n_M}} = \left(\tfrac{535}{1062} - \tfrac{411}{1012}\right) \pm 1.960\sqrt{\frac{\left(\frac{535}{1062}\right)\left(\frac{527}{1062}\right)}{1062} + \frac{\left(\frac{411}{1012}\right)\left(\frac{601}{1012}\right)}{1012}} = (0.055,\ 0.140)$$

c) We are 95% confident that the proportion of American women age 65 and older who suffer from arthritis is between 5.5% and 14.0% higher than the proportion of American men the same age who suffer from arthritis.

d) Since the interval for the difference in proportions of arthritis sufferers does not contain 0, there is strong evidence that arthritis is more likely to afflict women than men.

28. Graduation.

a) **Randomization condition:** Assume that the samples are representative of all recent graduates.
Independent groups assumption: The sample of men and the sample of women were drawn independently of each other.
Success/Failure condition: The samples are very large, certainly large enough for the methods of inference to be used.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $$(\hat{p}_F - \hat{p}_M) \pm z^* \sqrt{\frac{\hat{p}_F\hat{q}_F}{n_F} + \frac{\hat{p}_M\hat{q}_M}{n_M}} = (0.881 - 0.849) \pm 1.960\sqrt{\frac{(0.881)(0.119)}{12{,}678} + \frac{(0.849)(0.151)}{12{,}460}}$$
$$= (0.024,\ 0.040)$$

c) We are 95% confident that the proportion of 24-year-old American women who have graduated from high school is between 2.4 and 4.0 percentage points higher than the proportion of American men the same age who have graduated from high school.

d) Since the interval for the difference in proportions of high school graduates does not contain 0, there is strong evidence that women are more likely than men to complete high school.

29. Pets.

a) $$SE(\hat{p}_{Herb} - \hat{p}_{None}) = \sqrt{\frac{\hat{p}_{Herb}\hat{q}_{Herb}}{n_{Herb}} + \frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}} = \sqrt{\frac{\left(\frac{473}{827}\right)\left(\frac{354}{827}\right)}{827} + \frac{\left(\frac{19}{130}\right)\left(\frac{111}{130}\right)}{130}} = 0.035$$

b) **Randomization condition:** Assume that the dogs studied were representative of all dogs.
Independent groups assumption: The samples were drawn independently of each other.
Success/Failure condition: $n\hat{p}$ (herb) = 473, $n\hat{q}$ (herb) = 354, $n\hat{p}$ (none) = 19, and $n\hat{q}$ (none) = 111 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$(\hat{p}_{Herb} - \hat{p}_{None}) \pm z^* \sqrt{\frac{\hat{p}_{Herb}\hat{q}_{Herb}}{n_{Herb}} + \frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}} = \left(\tfrac{473}{827} - \tfrac{19}{130}\right) \pm 1.960\sqrt{\frac{\left(\frac{473}{827}\right)\left(\frac{354}{827}\right)}{827} + \frac{\left(\frac{19}{130}\right)\left(\frac{111}{130}\right)}{130}}$$
$$= (0.356,\ 0.495)$$

29. (continued)

c) We are 95% confident that the proportion of pets with a malignant lymphoma in homes where herbicides are used is between 35.6 and 49.5 percentage points higher than the proportion with lymphoma in homes where no pesticides are used.

30. Carpal Tunnel.

a) $$SE\left(\hat{p}_{Surg}-\hat{p}_{Splint}\right)=\sqrt{\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}+\frac{\hat{p}_{Splint}\hat{q}_{Splint}}{n_{Splint}}}=\sqrt{\frac{(0.80)(0.20)}{88}+\frac{(0.48)(0.52)}{88}}=0.068$$

b) **Randomization condition:** It's not clear whether or not this study was an experiment. If so, assume that the subjects were randomly allocated to treatment groups. If not, assume that the subjects are representative of all carpal tunnel sufferers.
Independent groups assumption: The improvement rates of the two groups are not related.
Success/Failure condition: $n\hat{p}$ (surg) = (88)(0.80) = 70, $n\hat{q}$ (surg) = (88)(0.20) = 18, $n\hat{p}$ (splint) = (88)(0.48) = 42, and $n\hat{q}$ (splint) = (88)(0.52) = 46 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Surg}-\hat{p}_{Splint}\right)\pm z^*\sqrt{\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}+\frac{\hat{p}_{Splint}\hat{q}_{Splint}}{n_{Splint}}}=\left(\tfrac{70}{88}-\tfrac{42}{88}\right)\pm 1.960\sqrt{\frac{\left(\frac{70}{88}\right)\left(\frac{18}{88}\right)}{88}+\frac{\left(\frac{42}{88}\right)\left(\frac{46}{88}\right)}{88}}=(0.184,\ 0.452)$$

c) We are 95% confident that the proportion of patients who show improvement in carpal tunnel syndrome with surgery is between 18.4 and 45.2 percentage points higher than the proportion who show improvement with wrist splints.

31. Prostate cancer.

a) This is an experiment. Men were randomly assigned to have surgery or not.

b) **Randomization condition:** Men were randomly assigned to treatment groups.
Independent groups assumption: The survival rates of the two groups are not related.
Success/Failure condition: $n\hat{p}$ (no surgery) = 31, $n\hat{q}$ (no surgery) = 317, $n\hat{p}$ (surgery) = 16, and $n\hat{q}$ (surgery) = 331 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{None}-\hat{p}_{Surg}\right)\pm z^*\sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}+\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}}=\left(\tfrac{31}{348}-\tfrac{16}{347}\right)\pm 1.960\sqrt{\frac{\left(\frac{31}{348}\right)\left(\frac{317}{348}\right)}{348}+\frac{\left(\frac{16}{347}\right)\left(\frac{331}{347}\right)}{347}}$$
$$=(0.006,\ 0.080)$$

c) We are 95% confident that the survival rate of patients who have surgery for prostate cancer is between 0.6 and 8.0 percentage points higher than the survival rate of patients who do not have surgery. Since the interval is completely above zero, there is evidence that surgery may be effective in preventing death from prostate cancer.

32. Race and smoking 2015.

a) **Randomization condition:** We will assume that these samples were random.
Independent groups assumption: The samples were taken randomly, so the groups are independent.
Success/Failure condition: $n\hat{p}$ (white) = (3607)(0.273) = 985, $n\hat{q}$ (white) = (3607)(0.727) = 2622, $n\hat{p}$ (black) = (485)(0.472) = 229, and $n\hat{q}$ (black) = (485)(0.528) = 256 are all greater than 10, so the samples are both large enough.

32. (continued)

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Black}-\hat{p}_{White}\right)\pm z^*\sqrt{\frac{\hat{p}_{Black}\hat{q}_{Black}}{n_{Black}}+\frac{\hat{p}_{White}\hat{q}_{White}}{n_{White}}}$$

$$=(0.472-0.273)\pm 1.645\sqrt{\frac{(0.472)(0.528)}{485}+\frac{(0.273)(0.727)}{3607}}=(0.160,\ 0.238)$$

b) We are 90% confident that the smoking rate for blacks is between 16.0 and 23.8 percentage points higher than the smoking rate for whites in New Jersey. We can use this interval to test the hypothesis that there is no difference in smoking rates between blacks and whites in New Jersey. Since the interval does not contains zero, there is evidence of a difference in smoking rates based on race.

c) The interval had 90% confidence, so $\alpha=(1-0.90)=0.10$ for the two-sided test.

33. Ear infections.

a) **Randomization condition:** The babies were randomly assigned to the two treatment groups.
Independent groups assumption: The groups were assigned randomly, so the groups are not related.
Success/Failure condition: $n\hat{p}$ (vaccine) = 333, $n\hat{q}$ (vaccine) = 2122,
$n\hat{p}$ (none) = 499, and $n\hat{q}$ (none) = 1953 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $$\left(\hat{p}_{None}-\hat{p}_{Vacc}\right)\pm z^*\sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}+\frac{\hat{p}_{Vacc}\hat{q}_{Vacc}}{n_{Vacc}}}=\left(\frac{499}{2452}-\frac{333}{2455}\right)\pm 1.960\sqrt{\frac{\left(\frac{499}{2452}\right)\left(\frac{1953}{2452}\right)}{2452}+\frac{\left(\frac{333}{2455}\right)\left(\frac{2122}{2455}\right)}{2455}}$$
$$=(0.047,\ 0.089)$$

c) We are 95% confident that the proportion of unvaccinated babies who develop ear infections is between 4.7 and 8.9 percentage points higher than the proportion of vaccinated babies who develop ear infections. The vaccinations appear to be effective, especially considering the 20% infection rate among the unvaccinated. A reduction of 5% to 9% is meaningful.

34. Anorexia.

a) **Randomization condition:** The women were randomly assigned to the groups.
Independent groups assumption: The groups were assigned randomly, so the groups are not related.
Success/Failure condition: $n\hat{p}$ (Prozac) = 35, $n\hat{q}$ (Prozac) = 14, $n\hat{p}$ (placebo) = 32, and $n\hat{q}$ (placebo) = 12 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $$\left(\hat{p}_{Proz}-\hat{p}_{Plac}\right)\pm z^*\sqrt{\frac{\hat{p}_{Proz}\hat{q}_{Proz}}{n_{Proz}}+\frac{\hat{p}_{Plac}\hat{q}_{Plac}}{n_{Plac}}}=\left(\frac{35}{49}-\frac{32}{44}\right)\pm 1.960\sqrt{\frac{\left(\frac{35}{49}\right)\left(\frac{14}{49}\right)}{49}+\frac{\left(\frac{32}{44}\right)\left(\frac{12}{44}\right)}{44}}=(-0.20,\ 0.17)$$

c) We are 95% confident that the proportion of women taking Prozac deemed healthy is between 20 percentage points lower and 17 percentage points higher than the proportion women taking a placebo. Prozac does not appear to be effective, since 0 is in the confidence interval. There is no evidence of a difference in the effectiveness of Prozac and a placebo.

35. Another ear infection.

a) H_0: The proportion of vaccinated babies who get ear infections is the same as the proportion of unvaccinated babies who get ear infections. $(p_{Vacc} = p_{None} \text{ or } p_{Vacc} - p_{None} = 0)$

H_A: The proportion of vaccinated babies who get ear infections is the lower than the proportion of unvaccinated babies who get ear infections. $(p_{Vacc} < p_{None} \text{ or } p_{Vacc} - p_{None} < 0)$

b) Since 0 is not in the confidence interval, reject the null hypothesis. There is evidence that the vaccine reduces the rate of ear infections.

c) If we think that the vaccine really reduces the rate of ear infections and it really does not reduce the rate of ear infections, then we have committed a Type I error.

d) Babies would be given ineffective vaccines.

36. Anorexia again.

a) H_0: The proportion of women taking Prozac who are deemed healthy is the same as the proportion of women taking the placebo who are deemed healthy. $(p_{Prozac} = p_{Placebo} \text{ or } p_{Prozac} - p_{Placebo} = 0)$

H_A: The proportion of women taking Prozac who are deemed healthy is greater than the proportion of women taking the placebo who are deemed healthy. $(p_{Prozac} > p_{Placebo} \text{ or } p_{Prozac} - p_{Placebo} > 0)$

b) Since 0 is in the confidence interval, fail to reject the null hypothesis. There is no evidence that Prozac is an effective treatment for anorexia.

c) If we think that Prozac is not effective and it is, we have committed a Type II error.

d) We might overlook a potentially helpful treatment.

37. Teen smoking.

a) This is a prospective observational study.

b) H_0: The proportion of teen smokers among the group whose parents disapprove of smoking is the same as the proportion of teen smokers among the group whose parents are lenient about smoking.

$(p_{Dis} = p_{Len} \text{ or } p_{Dis} - p_{Len} = 0)$

H_A: The proportion of teen smokers among the group whose parents disapprove of smoking is different than the proportion of teen smokers among the group whose parents are lenient about smoking.

$(p_{Dis} \neq p_{Len} \text{ or } p_{Dis} - p_{Len} \neq 0)$

c) **Randomization condition:** Assume that the teens surveyed are representative of all teens.
Independent groups assumption: The groups were surveyed independently.
Success/Failure condition: $n\hat{p}$ (disapprove) = 54, $n\hat{q}$ (disapprove) = 230, $n\hat{p}$ (lenient) = 11, and $n\hat{q}$ (lenient) = 30 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE(\hat{p}_{Dis} - \hat{p}_{Len}) = \sqrt{\frac{\hat{p}_{Dis}\hat{q}_{Dis}}{n_{Dis}} + \frac{\hat{p}_{Len}\hat{q}_{Len}}{n_{Len}}} = \sqrt{\frac{\left(\frac{54}{284}\right)\left(\frac{230}{284}\right)}{284} + \frac{\left(\frac{11}{41}\right)\left(\frac{30}{41}\right)}{41}} = 0.0730.$$

d) The observed difference between the proportions is 0.190 – 0.268 = – 0.078.

$$z = \frac{\hat{p}_{Dis} - \hat{p}_{Len}}{SE(\hat{p}_{Dis} - \hat{p}_{Len})} = \frac{-0.078}{0.0730} \approx -1.07;$$ Since the P-value = 0.28 is high, we fail to reject the null hypothesis. There is little evidence to suggest that parental attitudes influence teens' decisions to smoke.

e) If there is no difference in the proportions, there is about a 28% chance of seeing the observed difference or larger by natural sampling variation.

37. (continued)

f) If teens' decisions about smoking *are* influenced, we have committed a Type II error.

g) Since the conditions have already been satisfied in part (c), we will find a two-proportion z-interval.

$$\left(\hat{p}_{Dis} - \hat{p}_{Len}\right) \pm z^* \sqrt{\frac{\hat{p}_{Dis}\hat{q}_{Dis}}{n_{Dis}} + \frac{\hat{p}_{Len}\hat{q}_{Len}}{n_{Len}}} = \left(\tfrac{54}{284} - \tfrac{11}{41}\right) \pm 1.960 \sqrt{\frac{\left(\frac{54}{284}\right)\left(\frac{230}{284}\right)}{284} + \frac{\left(\frac{11}{41}\right)\left(\frac{30}{41}\right)}{41}} = (-0.221,\ 0.065)$$

h) We are 95% confident that the proportion of teens whose parents disapprove of smoking who will eventually smoke is between 22.1 percentage points lower and 6.5 percentage points higher than for teens with parents who are lenient about smoking.

i) We expect 95% of random samples of this size to produce intervals that contain the true difference between the proportions.

38. Depression.

a) This is a prospective observational study.

b) H_0: The proportion of cardiac patients without depression who died within the 4 years is the same as the proportion of cardiac patients with depression who died during the same time period.

$\left(p_{None} = p_{Dep} \text{ or } p_{None} - p_{Dep} = 0\right)$

H_A: The proportion of cardiac patients without depression who died within the 4 years is the less than the proportion of cardiac patients with depression who died during the same time period.

$\left(p_{None} < p_{Dep} \text{ or } p_{None} - p_{Dep} < 0\right)$

c) **Randomization condition:** Assume that the cardiac patients followed by the study are representative of all cardiac patients.
Independent groups condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (no depression) = 67, $n\hat{q}$ (no depression) = 294, $n\hat{p}$ (depression) = 26, and $n\hat{q}$ (depression) = 63 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE\left(\hat{p}_{None} - \hat{p}_{Dep}\right) = \sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}} + \frac{\hat{p}_{Dep}\hat{q}_{Dep}}{n_{Dep}}} = \sqrt{\frac{\left(\frac{67}{361}\right)\left(\frac{294}{361}\right)}{361} + \frac{\left(\frac{26}{89}\right)\left(\frac{63}{89}\right)}{89}} \approx 0.052366.$$

d) The observed difference between the proportions is: 0.1856 – 0.2921 = – 0.1065.

$$z = \frac{\hat{p}_{None} - \hat{p}_{Dep}}{SE\left(\hat{p}_{None} - \hat{p}_{Dep}\right)} = \frac{-0.1065}{0.052366} \approx -2.03;$$ Since the P-value = 0.021 is low, we reject the null hypothesis. There is evidence to suggest that the proportion of non-depressed cardiac patients who die within 4 years is less than the proportion of depressed cardiac patients who do.

e) If there is no difference in the proportions, we will see an observed difference this large or larger only about 2.1% of the time by natural sampling variation.

f) If cardiac patients without depression don't actually have a lower proportion of deaths than those with depression, then we have committed a Type I error.

38. (continued)

g) Since the conditions have already been satisfied in part (c), we will find a two-proportion z-interval.

$$\left(\hat{p}_{None} - \hat{p}_{Dep}\right) \pm z^* \sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}} + \frac{\hat{p}_{Dep}\hat{q}_{Dep}}{n_{Dep}}} = \left(\tfrac{67}{361} - \tfrac{26}{89}\right) \pm 1.960 \sqrt{\frac{\left(\frac{67}{361}\right)\left(\frac{294}{361}\right)}{361} + \frac{\left(\frac{26}{89}\right)\left(\frac{63}{89}\right)}{89}}$$
$$= (-0.209, -0.004)$$

h) We are 95% confident that the proportion of cardiac patients who die within 4 years is between 0.4 and 20.9 percentage points higher for depressed patients than for non-depressed patients.

i) We expect 95% of random samples of this size to produce intervals that contain the true difference between the proportions.

39. Birthweight.

a) H_0: The proportion of low birthweight is the same. $\left(p_{Exp} = p_{Not} \text{ or } p_{Exp} - p_{Not} = 0\right)$

H_A: The proportion of low birthweight is higher for women exposed to soot and ash. $\left(p_{Exp} > p_{Not} \text{ or } p_{Exp} - p_{Not} > 0\right)$

Randomization condition: Assume the women are representative of all women.
Independent groups assumption: The groups don't appear to be associated, with respect to soot and ash exposure, but all of the women were in New York. There may be a confounding variable explaining any relationship between exposure and birthweight.
Success/Failure condition: $n\hat{p}$ (Exposed) = 15, $n\hat{q}$ (Exposed) = 167, $n\hat{p}$ (Not) = 92, and $n\hat{q}$ (Not) = 2208 are all greater than 10. All of the samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE\left(\hat{p}_{Exp} - \hat{p}_{Not}\right) = \sqrt{\frac{\hat{p}_{Exp}\hat{q}_{Exp}}{n_{Exp}} + \frac{\hat{p}_{Not}\hat{q}_{Not}}{n_{Not}}} = \sqrt{\frac{\left(\frac{15}{182}\right)\left(\frac{167}{182}\right)}{182} + \frac{\left(\frac{92}{2300}\right)\left(\frac{2208}{2300}\right)}{2300}} \approx 0.020790.$$

The observed difference between the proportions is: $\left(\frac{15}{182}\right) - \left(\frac{92}{2300}\right) = 0.04242$

$$z = \frac{\hat{p}_{Exp} - \hat{p}_{Not}}{SE\left(\hat{p}_{Exp} - \hat{p}_{Not}\right)} = \frac{0.04242}{0.020790} \approx 2.04;$$ Since the P-value = 0.021 is low, we reject the null hypothesis.

There is strong evidence that the proportion of low birthweight babies is higher in the women exposed to soot and ash after the World Trade Center attacks.

b) $$\left(\hat{p}_{Exp} - \hat{p}_{Not}\right) \pm z^* \sqrt{\frac{\hat{p}_{Exp}\hat{q}_{Exp}}{n_{Exp}} + \frac{\hat{p}_{Not}\hat{q}_{Not}}{n_{Not}}} = \left(\tfrac{15}{182} - \tfrac{92}{2300}\right) \pm 1.960 \sqrt{\frac{\left(\frac{15}{182}\right)\left(\frac{167}{182}\right)}{182} + \frac{\left(\frac{92}{2300}\right)\left(\frac{2208}{2300}\right)}{2300}} = (0.002, 0.083)$$

We are 95% confident that the proportion of low birthweight babies is between 0.2 and 8.3 percentage points higher for mothers exposed to soot and ash after the World Trade Center attacks, than the proportion of low birthweight babies for mothers not exposed.

40. Politics and sex.

a) H_0: The proportion of voters in support of the candidate is the same before and after news of his extramarital affair got out. $(p_B = p_A \text{ or } p_B - p_A = 0)$

H_A: The proportion of voters in support of the candidate has decreased after news of his extramarital affair got out. $(p_B > p_A \text{ or } p_B - p_A > 0)$

Randomization condition: Voters were randomly selected.
Independent groups assumption: Since the samples were random, the groups are independent.
Success/Failure condition: $n\hat{p}$ (before) = (630)(0.54) = 340,
$n\hat{q}$ (before) = (630)(0.46) = 290, $n\hat{p}$ (after) = (1010)(0.51) = 515, and
$n\hat{q}$ (after) = (1010)(0.49) = 505 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE(\hat{p}_B - \hat{p}_A) = \sqrt{\frac{\hat{p}_B\hat{q}_B}{n_B} + \frac{\hat{p}_A\hat{q}_A}{n_A}} = \sqrt{\frac{\left(\frac{340}{630}\right)\left(\frac{290}{630}\right)}{630} + \frac{\left(\frac{515}{1010}\right)\left(\frac{505}{1010}\right)}{1010}} \approx 0.0253329.$$

The observed difference between the proportions is: $\left(\frac{340}{630}\right) - \left(\frac{515}{1010}\right) = 0.02978155$

$z = \dfrac{\hat{p}_B - \hat{p}_A}{SE(\hat{p}_B - \hat{p}_A)} = \dfrac{0.02978155}{0.0253329} \approx 1.18$; Since the P-value = 0.120 is fairly high, we fail to reject the null hypothesis. There is little evidence of a decrease in the proportion of voters in support of the candidate after the news of his extramarital affair got out.

b) No evidence of a decrease in the proportion of voters in support of the candidate was found. If there is actually a decrease, and we failed to notice, that's a Type II error.

41. Mammograms.

a) H_0: The proportion of deaths from breast cancer is the same for women who never had a mammogram as for women who had mammograms. $(p_N = p_M \text{ or } p_N - p_M = 0)$

H_A: The proportion of deaths from breast cancer is greater for women who never had a mammogram than for women who had mammograms. $(p_N > p_M \text{ or } p_N - p_M > 0)$

Randomization condition: Assume the women are representative of all women.
Independent groups assumption: We must assume that these groups are independent.
Success/Failure condition: $n\hat{p}$ (never) = 196, $n\hat{q}$ (never) = 30,369,
$n\hat{p}$ (mammogram) = 153, and $n\hat{q}$ (mammogram) = 29,978 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}(\hat{p}_N - \hat{p}_M) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M}} = \sqrt{\frac{\left(\frac{349}{60{,}696}\right)\left(\frac{60{,}347}{60{,}696}\right)}{30{,}565} + \frac{\left(\frac{349}{60{,}696}\right)\left(\frac{60{,}347}{60{,}696}\right)}{30{,}131}} \approx 0.000614.$$

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is 0.006413 – 0.005078 = 0.001335.

41. (continued)

$$z = \frac{\hat{p}_N - \hat{p}_M}{SE(\hat{p}_N - \hat{p}_M)} = \frac{0.001335}{0.000614} \approx 2.17;$$ Since the P-value = 0.0148 is low, we reject the null hypothesis.

There is strong evidence that the breast cancer mortality rate is higher among women that have never had a mammogram. However, the large sample sizes involved may have yielded a result that is statistically significant, but not practically significant. These data suggest a difference in mortality rate of only about 0.1 percentage points.

b) If there is actually no difference in the mortality rates, we have committed a Type I error.

42. Mammograms redux.

a) H_0: The proportion of deaths from breast cancer is the same for women who never had a mammogram as for women who had mammograms. $(p_N = p_M \text{ or } p_N - p_M = 0)$

H_A: The proportion of deaths from breast cancer is greater for women who never had a mammogram than for women who had mammograms. $(p_N > p_M \text{ or } p_N - p_M > 0)$

Randomization condition: Assume the women are representative of all women.
Independent groups assumption: We must assume that these groups are independent.
Success/Failure condition: $n\hat{p}$ (never) = 66, $n\hat{q}$ (never) = 21,129, $n\hat{p}$ (mammogram) = 63, and $n\hat{q}$ (mammogram) = 21,025 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated using the pooled proportion

by $$SE_{\text{pooled}}(\hat{p}_N - \hat{p}_M) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M}} = \sqrt{\frac{\left(\frac{129}{42{,}283}\right)\left(\frac{42{,}154}{42{,}283}\right)}{21{,}195} + \frac{\left(\frac{129}{42{,}283}\right)\left(\frac{42{,}154}{42{,}283}\right)}{21{,}088}}$$

≈ 0.000536.

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is 0.003114 – 0.002987 = 0.000127.

$$z = \frac{\hat{p}_N - \hat{p}_M}{SE(\hat{p}_N - \hat{p}_M)} = \frac{0.000127}{0.000536} \approx 0.24;$$ Since the P-value = 0.4068 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of breast cancer deaths for women who have never had a mammogram is greater than the proportion of deaths from breast cancer for women who underwent screening by mammogram.

b) If the proportion of deaths from breast cancer for women who have not had mammograms is actually greater than the proportion of deaths from breast cancer for women who have had mammograms, we have committed a Type II error.

43. Pain.

a) **Randomization condition:** The patients were randomly selected AND randomly assigned to treatment groups. If that's not random enough for you, I don't know what is!
Success/Failure condition: $n\hat{p}$ (A) = 84, $n\hat{q}$ (A) = 28, $n\hat{p}$ (B) = 66, and $n\hat{q}$ (B) = 42 are all greater than 10, so both samples are large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of patients who may get relief from medication A.

43. (continued)

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{84}{112}\right) \pm 1.960 \sqrt{\frac{\left(\frac{84}{112}\right)\left(\frac{28}{112}\right)}{112}} = (67.0\%,\ 83.0\%)$$

We are 95% confident that between 67.0% and 83.0% of patients with joint pain will find medication A to be effective.

b) Since the conditions were met in part (a), we can use a one-proportion *z*-interval to estimate the percentage of patients who may get relief from medication B.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{66}{108}\right) \pm 1.960 \sqrt{\frac{\left(\frac{66}{108}\right)\left(\frac{42}{108}\right)}{108}} = (51.9\%,\ 70.3\%)$$

We are 95% confident that between 51.9% and 70.3% of patients with joint pain will find medication B to be effective.

c) The 95% confidence intervals overlap, which might lead one to believe that there is no evidence of a difference in the proportions of people who find each medication effective. However, if one was led to believe that, one should proceed to part (d).

d) Most of the conditions were checked in part (a). We only have one more to check:
Independent groups assumption: The groups were assigned randomly, so there is no reason to believe there is a relationship between them.

Since the conditions have been satisfied, we will find a two-proportion *z*-interval.

$$\left(\hat{p}_A - \hat{p}_B\right) \pm z^* \sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}} = \left(\tfrac{84}{112} - \tfrac{66}{108}\right) \pm 1.960 \sqrt{\frac{\left(\frac{84}{112}\right)\left(\frac{28}{112}\right)}{112} + \frac{\left(\frac{66}{108}\right)\left(\frac{42}{108}\right)}{112}} = (0.017,\ 0.261)$$

We are 95% confident that the proportion of patients with joint pain who will find medication A effective is between 1.70 and 26.1 percentage points higher than the proportion of patients who will find medication B effective.

e) The interval does not contain zero. There is evidence that medication A is more effective than medication B.

f) The two-proportion method is the proper method. By attempting to use two, separate, confidence intervals, you are adding standard deviations when looking for a difference in proportions. We know from our previous studies that *variances* add when finding the standard deviation of a difference. The two-proportion method does this.

44. Gender gap.

a) **Randomization condition:** The poll was probably random, although not specifically stated.
Success/Failure condition: $n\hat{p}$ (men) = 246, $n\hat{q}$ (men) = 227, $n\hat{p}$ (women) = 235, and $n\hat{q}$ (women) = 287 are all greater than 10, so both samples are large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of men who may vote for the candidate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.52) \pm 1.960 \sqrt{\frac{(0.52)(0.48)}{473}} = (47.5\%,\ 56.5\%)$$

We are 95% confident that between 47.5% and 56.5% of men may vote for the candidate.

44. (continued)

b) Since the conditions were met in part (a), we can use a one-proportion z-interval to estimate the percentage of women who may vote for the candidate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.45) \pm 1.960\sqrt{\frac{(0.45)(0.55)}{522}} = (40.7\%,\ 49.3\%)$$

We are 95% confident that between 40.7% and 49.3% of women may vote for the candidate.

c) The 95% confidence intervals overlap, which might make you think that there is no evidence of a difference in the proportions of men and women who may vote for the candidate. However, if you think that, don't delay! Move on to part (d).

d) Most of the conditions were checked in part (a). We only have one more to check:
Independent groups assumption: There is no reason to believe that the samples of men and women influence each other in any way.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$(\hat{p}_M - \hat{p}_W) \pm z^* \sqrt{\frac{\hat{p}_M \hat{q}_M}{n_M} + \frac{\hat{p}_W \hat{q}_W}{n_W}} = (0.52 - 0.45) \pm 1.960\sqrt{\frac{(0.52)(0.48)}{473} + \frac{(0.45)(0.55)}{522}} = (0.008,\ 0.132)$$

We are 95% confident that the proportion of men who may vote for the candidate is between 0.8 and 13.2 percentage points higher than the proportion of women who may vote for the candidate.

e) The interval does not contain zero. There is evidence that the proportion of men may vote for the candidate is greater than the proportion of women who may vote for the candidate.

f) The two-proportion method is the proper method. By attempting to use two, separate, confidence intervals, you are adding standard deviations when looking for a difference in proportions. We know from our previous studies that *variances* add when finding the standard deviation of a difference. The two-proportion method does this.

45. Convention bounce.

Randomization condition: The polls were conducted randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (after) = 705, $n\hat{q}$ (after) = 795, $n\hat{p}$ (before) = 735, and $n\hat{q}$ (women) = 765 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$(\hat{p}_{After} - \hat{p}_{Before}) \pm z^* \sqrt{\frac{\hat{p}_{After}\hat{q}_{After}}{n_{After}} + \frac{\hat{p}_{Before}\hat{q}_{Before}}{n_{Before}}} = (0.49 - 0.47) \pm 1.960\sqrt{\frac{(0.49)(0.51)}{1500} + \frac{(0.47)(0.53)}{1500}}$$
$$= (-0.016,\ 0.056)$$

We can be 95% confident that the proportion of likely voters who favored John Kerry was between 1.6 percentage points lower and 5.6 percentage points higher after the convention, compared to before the convention. Since zero is contained in the interval, it is plausible that there was no difference in Kerry support. The poll showed no evidence of a convention bounce.

46. Stay-at-home dads.

a) **Randomization condition:** We will assume that the polls were conducted randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (black) = 11, $n\hat{q}$ (black) = 150, $n\hat{p}$ (Latino) = 20, and $n\hat{q}$ (Latino) = 338 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Black} - \hat{p}_{Latino}\right) \pm z^* \sqrt{\frac{\hat{p}_{Black}\hat{q}_{Black}}{n_{Black}} + \frac{\hat{p}_{Latino}\hat{q}_{Latino}}{n_{Latino}}} = \left(\frac{11}{161} - \frac{20}{358}\right) \pm 1.960\sqrt{\frac{\left(\frac{11}{161}\right)\left(\frac{150}{161}\right)}{161} + \frac{\left(\frac{20}{358}\right)\left(\frac{338}{358}\right)}{358}}$$
$$= (-0.033,\ 0.058)$$

We are 95% confident that the percentage of black men who are stay-at-home dads is between 3.3 percentage points lower and 5.8 percentage points higher than the percentage of Latino men who are stay-at-home dads.

b) The margin of error is higher for the interval we constructed because the sample size is smaller than the entire *Time* magazine survey. Our interval only involved black and Latino men.

47. Sensitive men.

H_0: The proportion of 18-24-year-old men who are comfortable talking about their problems is the same as the proportion of 25-34-year old men. $\left(p_{Young} = p_{Old} \text{ or } p_{Young} - p_{Old} = 0\right)$

H_A: The proportion of 18-24-year-old men who are comfortable talking about their problems is higher than the proportion of 25-34-year old men. $\left(p_{Young} > p_{Old} \text{ or } p_{Young} - p_{Old} > 0\right)$

Randomization condition: We assume the respondents were chosen randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (young) = 80, $n\hat{q}$ (young) = 49, $n\hat{p}$ (old) = 98, and $n\hat{q}$ (old) = 86 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation using the pooled proportion estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{Young} - \hat{p}_{Old}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_Y} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_O}} = \sqrt{\frac{\left(\frac{178}{313}\right)\left(\frac{135}{313}\right)}{129} + \frac{\left(\frac{178}{313}\right)\left(\frac{135}{313}\right)}{184}} \approx 0.05687.$$

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is: 0.620 – 0.533 = 0.087.

$$z = \frac{\hat{p}_{Young} - \hat{p}_{Old}}{SE\left(\hat{p}_{Young} - \hat{p}_{Old}\right)} = \frac{0.087}{0.05687} \approx 1.54;$$ Since the P-value = 0.0619 is high, we fail to reject the null hypothesis. There is little evidence that the proportion of 18-24-year-old men who are comfortable talking about their problems is higher than the proportion of 25-34-year-old men who are comfortable. *Time* magazine's interpretation is questionable.

48. Carbs.

H_0: The proportion of U.S. adults who actively avoid carbohydrates in their diet is the same now as it was in 2002. $\left(p_{Now} = p_{2002} \text{ or } p_{Now} - p_{2002} = 0\right)$

H_A: The proportion of U.S. adults who actively avoid carbohydrates in their diet is greater now than it was in 2002. $\left(p_{Now} > p_{2002} \text{ or } p_{Now} - p_{2002} > 0\right)$

Randomization condition: We assume the respondents were chosen randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (now) = 271, $n\hat{q}$ (now) = 734, $n\hat{p}$ (2002) = 201, and $n\hat{q}$ (2002) = 804 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation using the pooled proportion estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{Now} - \hat{p}_{2002}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Now}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2002}}} = \sqrt{\frac{\left(\frac{472}{2010}\right)\left(\frac{1538}{2010}\right)}{1005} + \frac{\left(\frac{472}{2010}\right)\left(\frac{1538}{2010}\right)}{1005}} \approx 0.0189.$$

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is: 0.27 – 0.20 = 0.07.

$z = \dfrac{\hat{p}_{Now} - \hat{p}_{2002}}{SE\left(\hat{p}_{Now} - \hat{p}_{2002}\right)} = \dfrac{0.07}{0.0189} \approx 3.70$; Since the P-value = 0.0001 is low, we reject the null hypothesis. There is strong evidence that the percentage of U.S. adults who actively try to avoid carbs has increased since 2002.

49. Food preference.

H_0: The proportion of people who agree with the statement is the same in rural and urban areas. $\left(p_{Urban} = p_{Rural} \text{ or } p_{Urban} - p_{Rural} = 0\right)$

H_A: The proportion of people who agree with the statement differs in rural and urban areas. $\left(p_{Urban} \neq p_{Rural} \text{ or } p_{Urban} - p_{Rural} \neq 0\right)$

Randomization condition: The respondents were chosen randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (urban) = 417, $n\hat{q}$ (urban) = 229, $n\hat{p}$ (rural) = 78, and $n\hat{q}$ (rural) = 76 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation using the pooled proportion estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{Urban} - \hat{p}_{Rural}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Urban}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Rural}}} = \sqrt{\frac{\left(\frac{495}{800}\right)\left(\frac{305}{800}\right)}{646} + \frac{\left(\frac{495}{800}\right)\left(\frac{305}{800}\right)}{154}} \approx 0.0436.$$

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is $\frac{417}{646} - \frac{78}{154} \approx 0.139$.

$z = \dfrac{\hat{p}_{Urban} - \hat{p}_{Rural}}{SE\left(\hat{p}_{Urban} - \hat{p}_{Rural}\right)} = \dfrac{0.139}{0.436} \approx 3.19$; Since the P-value = 0.0014 is low, reject the null hypothesis. There is evidence that the proportion of people who agree with the statement is not the same in urban and rural areas. These data suggest that the proportion is higher in urban areas than in rural areas.

50. Fast food.

H$_0$: The proportion of people who agree with the statement is the same for the two age groups. $\left(p_{>35} = p_{\le 35} \text{ or } p_{>35} - p_{\le 35} = 0\right)$
H$_A$: The proportion of people who agree with the statement is different for the two age groups. $\left(p_{>35} \ne p_{\le 35} \text{ or } p_{>35} - p_{\le 35} \ne 0\right)$

Randomization condition: The respondents were chosen randomly.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (>35) = 246, $n\hat{q}$ (>35) = 143, $n\hat{p}$ (≤ 35) = 197, and $n\hat{q}$ (≤ 35) = 214 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation using the pooled proportion estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{>35} - \hat{p}_{\le 35}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{>35}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{\le 35}}} = \sqrt{\frac{\left(\frac{443}{800}\right)\left(\frac{357}{800}\right)}{389} + \frac{\left(\frac{443}{800}\right)\left(\frac{357}{800}\right)}{411}} \approx 0.0352.$$

We could use the unpooled standard error. There is no practical difference.

The observed difference between the proportions is $\frac{246}{389} - \frac{197}{411} \approx 0.153$.

$$z = \frac{\hat{p}_{>35} - \hat{p}_{\le 35}}{SE_{\text{pooled}}\left(\hat{p}_{>35} - \hat{p}_{\le 35}\right)} = \frac{0.153}{0.0352} \approx 4.35;$$ Since the P-value = 0.00001 is low, reject the null hypothesis.

There is evidence that the proportion of people who agree with the statement is not the same for the two age groups. These data suggest that a greater percentage of those 35 or older will say they avoid fast foods.

51. Hot dogs.

Yes, the 95% confidence interval would contain 0. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Meat} - \mu_{Beef}$.

52. Washers.

Yes, the 95% confidence interval would contain 0. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Top} - \mu_{Front}$.

53. Hot dogs, second helping.

a) Plausible values for $\mu_{Meat} - \mu_{Beef}$ are all negative, so the mean fat content is probably higher for beef hot dogs.

b) The fact that the confidence interval does not contain 0 indicates that the difference is significant.

c) The corresponding alpha level is 10%.

54. Second load of wash.

a) Plausible values for $\mu_{Top} - \mu_{Front}$ are all negative, so the mean cycle time is probably higher for front loading machines.

b) The fact that the confidence interval does not contain 0 indicates that the difference is significant.

c) The corresponding alpha level is 2%.

55. Hot dogs, last one.

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. Confidence intervals 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.

56. Third load of wash.

a) False. The confidence interval is about means, not about individual loads.

b) False. The confidence interval is about means, not about individual loads.

c) False. Confidence intervals 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.

57. Learning math.

a) The margin of error of this confidence interval is (11.427 – 5.573)/2 = 2.927 points.

b) The margin of error for a 98% confidence interval would have been larger. The critical value of t^* is larger for higher confidence levels. We need a wider interval to increase the likelihood that we catch the true mean difference in test scores within our interval. In other words, greater confidence comes at the expense of precision.

c) We are 95% confident that the mean score for the CPMP math students will be between 5.573 and 11.427 points higher on this assessment than the mean score of the traditional students.

d) Since the entire interval is above 0, there is strong evidence that students who learn with CPMP will have higher mean scores is algebra than those in traditional programs.

58. Stereograms.

a) We are 90% confident that the mean time required to "fuse" the image for people who receive no information or verbal information only will be between 0.55 and 5.47 seconds longer than the mean time required to "fuse" the image for people who receive both verbal and visual information.

b) Since the entire interval is above 0, there is evidence that viewing the picture of the image helps people "see" the 3D image.

c) The margin of error for this interval is (5.47 – 0.55)/2 = 2.46 seconds.

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

e) A 99% confidence interval would be wider. The critical value of t^* is larger for higher confidence levels. We need a wider interval to increase the likelihood that we catch the true mean difference in test scores within our interval. In other words, greater confidence comes at the expense of precision.

f) The conclusion reached may very well change. A wider interval may contain the mean difference of 0, failing to provide evidence of a difference in mean times.

59. CPMP, again.

a) H_0: The mean score of CPMP students is the same as the mean score of traditional students.
$(\mu_C = \mu_T \text{ or } \mu_C - \mu_T = 0)$
H_A: The mean score of CPMP students is different from the mean score of traditional students.
$(\mu_C \neq \mu_T \text{ or } \mu_C - \mu_T \neq 0)$

b) **Independent groups assumption:** Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to either CPMP or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, we can use a two-sample *t*-test with 583 degrees of freedom (from the computer).

c) If the mean scores for the CPMP and traditional students are really equal, there is less than a 1 in 10,000 chance of seeing a difference as large or larger than the observed difference just from natural sampling variation.

d) Since the P-value < 0.0001, reject the null hypothesis. There is strong evidence that the CPMP students have a different mean score than the traditional students. The evidence suggests that the CPMP students have a higher mean score.

60. CPMP and word problems.

H_0: The mean score of CPMP students is the same as the mean score of traditional students.
$(\mu_C = \mu_T \text{ or } \mu_C - \mu_T = 0)$
H_A: The mean score of CPMP students is different from the mean score of traditional students.
$(\mu_C \neq \mu_T \text{ or } \mu_C - \mu_T \neq 0)$

Independent groups assumption: Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to either CPMP or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 590.05 degrees of freedom (from the approximation formula).

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error:
$$SE(\bar{y}_C - \bar{y}_T) = \sqrt{\frac{32.1^2}{320} + \frac{28.5^2}{273}} \approx 2.489.$$

The observed difference between the mean scores is 57.4 – 53.9 = 3.5.

$t = \dfrac{(\bar{y}_C - \bar{y}_T) - (0)}{SE(\bar{y}_C - \bar{y}_T)} \approx \dfrac{3.5}{2.489} \approx 1.406$; Since the P-value = 0.1602, we fail to reject the null hypothesis. There is no evidence that the CPMP students have a different mean score on the word problems test than the traditional students.

61. Commuting.

a) **Independent groups assumption:** Since the choice of route was determined at random, the commuting times for Route A are independent of the commuting times for Route B.
Randomization condition: The man randomly determined which route he would travel on each day.
Nearly Normal condition: The histograms of travel times for the routes are roughly unimodal and symmetric. (Given)

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 33.1 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_B - \bar{y}_A) \pm t_{df}^{*}\sqrt{\frac{s_B^2}{n_B}+\frac{s_A^2}{n_A}} = (43-40) \pm t_{33.1}^{*}\sqrt{\frac{2^2}{20}+\frac{3^2}{20}} \approx (1.36,\ 4.64)$$

We are 95% confident that Route B has a mean commuting time between 1.36 and 4.64 minutes longer than the mean commuting time of Route A.

b) Since 5 minutes is beyond the high end of the interval, there is no evidence that the Route B is an average of 5 minutes longer than Route A. It appears that the old-timer may be exaggerating the average difference in commuting time.

62. Pulse rates.

a) The boxplots suggest that the mean pulse rates for men and women are roughly equal, but that females' pulse rates are more variable.

b) **Independent groups assumption:** There is no reason to believe that the pulse rates for men and women are related.
Randomization condition: There is no mention of randomness, but we can assume that the researcher chose a representative sample of men and women with regards to pulse rate.
Nearly Normal condition: The boxplots are reasonably symmetric. Let's hope the distributions of the samples are unimodal, too.

The conditions for inference are satisfied, so we can analyze these data using the methods discussed in this chapter.

c) Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 40.2 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 90% confidence.

$$(\bar{y}_M - \bar{y}_F) \pm t_{df}^{*}\sqrt{\frac{s_M^2}{n_M}+\frac{s_F^2}{n_F}} = (72.75-72.625) \pm t_{40.2}^{*}\sqrt{\frac{5.37225^2}{28}+\frac{7.69987^2}{24}} \approx (-3.025,\ 3.275)$$

We are 90% confident that the mean pulse rate for men is between 3.025 points lower and 3.275 points higher than the mean pulse rate for women.

d) Since 0 is in the interval, there is no evidence of a difference in mean pulse rate for men and women. This confirms our answer to part (a).

63. View of the water.

Independent groups assumption: Since the 170 properties were randomly selected, the groups should be independent.
Randomization condition: The 170 properties were selected randomly.
Nearly Normal condition: The boxplots of sale prices are roughly symmetric. The plots show several outliers, but the sample sizes are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 105.48 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

63. (continued)

$$(\bar{y}_W - \bar{y}_N) \pm t_{df}^* \sqrt{\frac{s_W^2}{n_W} + \frac{s_N^2}{n_N}} = (319{,}906.40 - 219{,}896.60) \pm t_{105.48}^* \sqrt{\frac{153{,}303.80^2}{70} + \frac{94{,}627.15^2}{100}}$$
$$\approx (\$59121,\ \$140898)$$

We are 95% confident that waterfront property has a mean selling price that is between $59,121 and $140,898 higher, on average, than non-waterfront property.

64. New construction.

Independent groups assumption: Since the 200 properties were randomly selected, the groups should be independent.
Randomization condition: The 200 properties were selected randomly.
Nearly Normal condition: The boxplots of sale prices are roughly symmetric. The plots show several outliers, but the sample sizes are large. The Central Limit Theorem allows us to proceed.

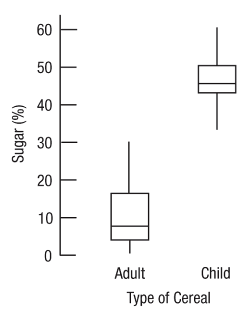
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 197.8 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{New} - \bar{y}_{Old}) \pm t_{df}^* \sqrt{\frac{s_{New}^2}{n_{New}} + \frac{s_{Old}^2}{n_{Old}}} = (267{,}878.10 - 201{,}707.50) \pm t_{197.8}^* \sqrt{\frac{93{,}302.18^2}{100} + \frac{96{,}116.88^2}{100}}$$
$$\approx (\$39754,\ \$92587)$$

We are 95% confident that newly-constructed property has a mean selling price that is between $39,754 and $92,587 higher, on average, than property that not newly-constructed.

65. Sugar in cereal.

Independent groups assumption: The percentage of sugar in the children's cereals is unrelated to the percentage of sugar in adult's cereals.
Randomization condition: It is reasonable to assume that the cereals are representative of all children's cereals and adult cereals, in regard to sugar content.
Nearly Normal condition: The boxplots of sugar content are shown below. Both distributions appear at least roughly symmetric, with no outliers. Adult cereal sugar content is skewed to the right, but the sample sizes are of reasonable size. The Central Limit Theorem allows us to proceed.



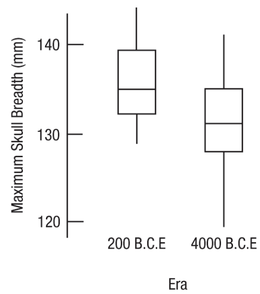
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 42 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_C - \bar{y}_A) \pm t_{df}^* \sqrt{\frac{s_C^2}{n_C} + \frac{s_A^2}{n_A}} = (46.8 - 10.1536) \pm t_{42}^* \sqrt{\frac{6.41838^2}{19} + \frac{7.61239^2}{28}} \approx (32.49,\ 40.80)$$

We are 95% confident that children's cereals have a mean sugar content that is between 32.49% and 40.80% higher than the mean sugar content of adult cereals.

66. Egyptians.

a) **Independent groups assumption:** The skull breadth of Egyptians in 4000 B.C.E is independent of the skull breadth of Egyptians almost 4 millennia later!
Randomization condition: It is reasonable to assume that the skulls measured have skull breadths that are representative of all Egyptians of the time.
Nearly Normal condition: The boxplots of skull breadth are shown below. Both distributions appear at least roughly symmetric, with no outliers.



b) Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 54 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

$$(\bar{y}_{200} - \bar{y}_{4K}) \pm t_{df}^{*} \sqrt{\frac{s_{200}^{2}}{n_{200}} + \frac{s_{4K}^{2}}{n_{4K}}} = (135.633 - 131.367) \pm t_{54}^{*} \sqrt{\frac{4.03846^2}{30} + \frac{5.12925^2}{30}} \approx (1.88,\ 6.66)$$

We are 95% confident that Egyptian males in 200 B.C.E. had a mean skull breadth between 1.88 and 6.66 mm larger than the mean skull breadth of Egyptian males in 4000 B.C.E.

c) Since the interval is completely above 0, there is evidence that the mean breadth of males' skulls has changed over this time period. The evidence suggests that the mean skull breadth has increased.

67. Reading.

H_0: The mean reading comprehension score of students who learn by the new method is the same as the mean score of students who learn by traditional methods. $(\mu_N = \mu_T \text{ or } \mu_N - \mu_T = 0)$

H_A: The mean reading comprehension score of students who learn by the new method is greater than the mean score of students who learn by traditional methods. $(\mu_N > \mu_T \text{ or } \mu_N - \mu_T > 0)$

Independent groups assumption: Student scores in one group should not have an impact on the scores of students in the other group.
Randomization condition: Students were randomly assigned to classes.
Nearly Normal condition: The stem-and-leaf plots show distributions of scores that are unimodal and symmetric.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 37.3 degrees of freedom (from the approximation formula). We will perform a two-sample t-test. We know the following.

$\bar{y}_N = 51.7222$ $\qquad$ $\bar{y}_T = 41.8182$

$s_N = 11.7062$ $\qquad$ $s_T = 16.5979$

$n_N = 18$ $\qquad$ $n_T = 22$

The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_N - \bar{y}_T) = \sqrt{\frac{11.7062^2}{18} + \frac{16.5979^2}{22}} \approx 4.487.$$

67. (continued)

The observed difference between the mean scores is 51.7222 – 41.8182 ≈ 9.904.

$t = \dfrac{(\bar{y}_N - \bar{y}_T) - (0)}{SE(\bar{y}_N - \bar{y}_T)} \approx \dfrac{9.904}{4.487} \approx 2.207$; Since the P-value = 0.0168 is low, we reject the null hypothesis. There is evidence that the students taught using the new activities have a higher mean score on the reading comprehension test than the students taught using traditional methods.

68. Streams.

a) H_0: Streams with limestone substrates and streams with shale substrates have the same mean pH level. $(\mu_L = \mu_S \text{ or } \mu_L - \mu_S = 0)$

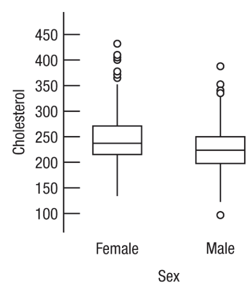
H_A: Streams with limestone substrates and streams with shale substrates have different mean pH levels. $(\mu_L \neq \mu_S \text{ or } \mu_L - \mu_S \neq 0)$

b) **Independent groups assumption:** pH levels from the two types of streams are independent.
Independence assumption: Since we don't know if the streams were chosen randomly, assume that the pH level of one stream does not affect the pH of another stream. This seems reasonable.
Nearly Normal condition: The boxplots provided show that the pH levels of the streams may be skewed (since the median is either the upper or lower quartile for the shale streams and the lower whisker of the limestone streams is stretched out), and there are outliers. However, since there are 133 degrees of freedom, we know that the sample sizes are large. It should be safe to proceed.

c) Since the P-value ≤ 0.0001 is low, we reject the null hypothesis. There is strong evidence that the streams with limestone substrates have mean pH levels different than those of streams with shale substrates. The limestone streams are less acidic on average.

69. Cholesterol and gender.

a) Boxplots are shown below.



b) Since there is a great deal of overlap between the cholesterol levels for men and women, it does not appear that one gender has higher average levels of cholesterol.

c) H_0: The mean cholesterol level is the same for females and males. $(\mu_F = \mu_M \text{ or } \mu_F - \mu_M = 0)$

H_A: The mean cholesterol level is different for females and males. $(\mu_F \neq \mu_M \text{ or } \mu_F - \mu_M \neq 0)$

Independent groups assumption: Female cholesterol levels should not affect male cholesterol
Randomization condition: Participants in the Framingham survey should be representative of males and females in general.
Nearly Normal condition: The distributions of cholesterol levels are both generally unimodal and symmetric, though each has a large number of outliers.

69. (continued)

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 1401 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test. We know the following.

$\bar{y}_F = 242.70963$	$\bar{y}_M = 225.87892$
$s_F = 48.505208$	$s_M = 42.023646$
$n_F = 737$	$n_M = 669$

The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_F - \bar{y}_M) = \sqrt{\frac{48.505208^2}{737} + \frac{42.023646^2}{669}} \approx 2.4149704.$$

The observed difference between the mean scores is 242.70963 – 225.87892 ≈ 16.83071.

$t = \dfrac{(\bar{y}_F - \bar{y}_M) - (0)}{SE(\bar{y}_F - \bar{y}_M)} \approx \dfrac{16.83071}{2.4149704} \approx 6.97$; Since the P-value < 0.0001 is very low, we reject the null hypothesis. There is strong evidence that the mean cholesterol levels for males and females are different. These data suggest that females have significantly higher mean cholesterol.

$$(\bar{y}_F - \bar{y}_M) \pm t_{df}^{*}\sqrt{\frac{s_F^2}{n_F} + \frac{s_M^2}{n_M}} = (242.70963 - 225.87892) \pm t_{1401}^{*}\sqrt{\frac{48.505208^2}{737} + \frac{42.023646^2}{669}} \approx (12.1,\ 21.6)$$

We are 95% confident that the interval 12.1 to 21.6 contains the true mean difference in the cholesterol levels of females and males. In other words, we are 95% confident that females have a mean cholesterol level between 12.1 and 21.6 mg/dL higher than the mean cholesterol level of men.

d) Our conclusion is different than our conclusion in part (b). The evidence of a difference in mean cholesterol level is much stronger than it appeared in the boxplot. The large sample sizes make us very confident in our predictions of the means.

e) Removing the outliers would not change our conclusion. Without the outliers, the standard error of the means would be smaller, since the cholesterol levels would be less variable, and the means would not change much. The value of the test statistic would not change that much.

70. Memory.

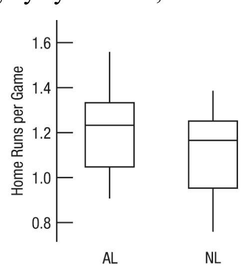
a) If there is no difference between ginkgo and the placebo, there is a 93.74% chance of seeing a difference as large as or larger than that observed, just from natural sampling variation.

b) There is no evidence based on this study that ginkgo biloba improves memory, as the difference in mean memory score was not significant and, in fact, negative.

c) If we fail to notice the effectiveness of ginkgo biloba, we have committed a Type II error.

71. Home runs 2016.

a) The boxplots of the average number of home runs hit at the ballparks in the two leagues are shown below. Both distributions appear at least roughly symmetric, with no outliers.



71. (continued)

b) $\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 1.21624 \pm t_{14}^{*}\left(\frac{0.17736}{\sqrt{15}}\right) \approx (1.12, 1.31)$

We are 95% confident that the mean number of home runs hit per game in American League stadiums is between 1.12 and 1.31.

c) The average of 1.26 home runs hit per game in Coors Field is not unusual. It isn't an outlier on the boxplot, and isn't even the highest average in the National League.

d) If you attempt to use two confidence intervals to assess a difference in means, you are actually adding standard deviations. But it's the variances that add, not the standard deviations. The two-sample difference of means procedure takes this into account.

e) $(\bar{y}_A - \bar{y}_N) \pm t_{df}^{*}\sqrt{\frac{s_A^2}{n_A} + \frac{s_N^2}{n_N}} = (1.21624 - 1.09495) \pm t_{27.55}^{*}\sqrt{\frac{0.17736^2}{15} + \frac{0.20179^2}{15}} \approx (-0.021, 0.263)$

f) We are 95% confident that the mean number of home runs in American League stadiums is between 0.021 home runs lower and 0.263 home runs higher than the mean number of home runs in National League stadiums.

g) Since the interval contains 0, there is no evidence of a difference in the mean number of home runs hit per game in the stadiums of the two leagues.

72. Hard water by region.

a) H_0: The mean mortality rate is the same for towns North and South of Derby. $(\mu_N = \mu_S \text{ or } \mu_N - \mu_S = 0)$

H_A: The mean mortality rate is different for towns North and South of Derby. $(\mu_N \neq \mu_S \text{ or } \mu_N - \mu_S \neq 0)$

Independent groups assumption: The towns were sampled independently.
Independence assumption: Assume that the mortality rates are in each town are independent of the mortality rates in the others.
Nearly Normal condition: We don't have the actual data, so we can't look at histograms of the distributions, but the samples are fairly large. It should be okay to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 53.49 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error

$SE(\bar{y}_N - \bar{y}_S) = \sqrt{\frac{138.470^2}{34} + \frac{151.114^2}{27}} \approx 37.546.$

The observed difference between the mean scores is 1631.59 – 1388.85 = 242.74.

$t = \frac{(\bar{y}_N - \bar{y}_S) - (0)}{SE(\bar{y}_N - \bar{y}_S)} \approx \frac{242.74}{37.546} \approx 6.47$; Since the P-value = 3.2×10^{-8} is low, we reject the null hypothesis.

There is strong evidence that the mean mortality rate different for towns north and south of Derby. There is evidence that the mortality rate north of Derby is higher.

b) Since there is an outlier in the data north of Derby, the conditions for inference are not satisfied, and it is risky to use the two-sample *t*-test. The outlier should be removed, and the test should be performed again. Without the actual data, we are not able to do this. The test without the outlier would *probably* help us reach the same conclusion, but there is no way to be sure.

73. Job satisfaction.

A two-sample *t*-procedure is not appropriate for these data, because the two groups are not independent. They are before and after satisfaction scores for the same workers. Workers that have high levels of job satisfaction before the exercise program is implemented may tend to have higher levels of job satisfaction than other workers after the program as well.

74. Summer school.

A two-sample *t*-procedure is not appropriate for these data, because the two groups are not independent. They are before and after scores for the same students. Students with high scores before summer school may tend to have higher scores after summer school as well.

75. Sex and violence.

a) Since the P-value = 0.136 is high, we fail to reject the null hypothesis. There is no evidence of a difference in the mean number of brands recalled by viewers of sexual content and viewers of violent content.

b) H_0: The mean number of brands recalled is the same for viewers of sexual content and viewers of neutral content. $\left(\mu_S = \mu_N \text{ or } \mu_S - \mu_N = 0\right)$

H_A: The mean number of brands recalled is different for viewers of sexual content and viewers of neutral content. $\left(\mu_S \neq \mu_N \text{ or } \mu_S - \mu_N \neq 0\right)$

Independent groups assumption: Recall of one group should not affect recall of another.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: The samples are large.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 214 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error $SE(\bar{y}_S - \bar{y}_N) = \sqrt{\frac{1.76^2}{108} + \frac{1.77^2}{108}} \approx 0.24.$

The observed difference between the mean scores is 1.71 – 3.17 = – 1.46.

$t = \frac{(\bar{y}_S - \bar{y}_N) - (0)}{SE(\bar{y}_S - \bar{y}_N)} \approx \frac{-1.46}{0.24} \approx -6.08$; Since the P-value = 5.5×10^{-9} is low, we reject the null hypothesis.

There is strong evidence that the mean number of brand names recalled is different for viewers of sexual content and viewers of neutral content. The evidence suggests that viewers of neutral ads remember more brand names on average than viewers of sexual content.

76. Ad campaign.

a) We are 95% confident that the mean number of ads remembered by viewers of shows with violent content will be between 1.6 and 0.6 lower than the mean number of brand names remembered by viewers of shows with neutral content.

b) If they want viewers to remember their brand names, they should consider advertising on shows with neutral content, as opposed to shows with violent content.

77. Hungry?

H_0: The mean number of ounces of ice cream people scoop is the same for large and small bowls.
$\left(\mu_{big} = \mu_{small} \text{ or } \mu_{big} - \mu_{small} = 0\right)$
H_A: The mean number of ounces of ice cream people scoop is the different for large and small bowls.
$\left(\mu_{big} \neq \mu_{small} \text{ or } \mu_{big} - \mu_{small} \neq 0\right)$

Independent groups assumption: The amount of ice cream scooped by individuals should be independent.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: Assume that this condition is met.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 34 degrees of freedom (from the approximation formula). Perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_{big} - \bar{y}_{small}) = \sqrt{\frac{2.91^2}{22} + \frac{1.84^2}{26}} \approx 0.7177 \text{ oz.}$$

The observed difference between the mean amounts is 6.58 – 5.07 = 1.51 oz.

$t = \dfrac{\left(\bar{y}_{big} - \bar{y}_{small}\right) - (0)}{SE\left(\bar{y}_{big} - \bar{y}_{small}\right)} \approx \dfrac{1.51}{0.7177} \approx 2.104$; Since the P-value of 0.0428 is low, we reject the null hypothesis.
There is strong evidence that the that the mean amount of ice cream people put into a bowl is related to the size of the bowl. People tend to put more ice cream into the large bowl, on average, than the small bowl.

78. Thirsty?

H_0: The mean number of milliliters of liquid people pour when asked to pour a "shot" is the same for highballs and tumblers. $\left(\mu_{tumbler} = \mu_{highball} \text{ or } \mu_{tumbler} - \mu_{highball} = 0\right)$
H_A: The mean number of milliliters of liquid people pour when asked to pour a "shot" is different for highballs and tumblers. $\left(\mu_{tumbler} \neq \mu_{highball} \text{ or } \mu_{tumbler} - \mu_{highball} \neq 0\right)$

Independent groups assumption: The amount of liquid poured by individuals should be independent.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: Assume that this condition is met.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 194 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_{tumbler} - \bar{y}_{highball}) = \sqrt{\frac{17.9^2}{99} + \frac{16.2^2}{99}} \approx 2.4264 \text{ ml.}$$

The observed difference between the mean amounts is 60.9 – 42.2 = 18.7 ml.

$t = \dfrac{\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right) - (0)}{SE\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right)} \approx \dfrac{18.7}{2.4264} \approx 7.71$; Since the P-value (less than 0.0001) is low, we reject the null hypothesis. There is strong evidence that the mean amount liquid people pour into a glass is related to the shape of the glass. People tend to pour more, on average, into a small, wide tumbler than into a tall, narrow highball glass.

79. Swim the Lake 2016 revisited.

a) $(\bar{y}_F - \bar{y}_M) \pm t^*_{df}\sqrt{\frac{s_F^2}{n_F}+\frac{s_M^2}{n_M}} = (1262.08 - 1226.29) \pm t^*_{39.6}\sqrt{\frac{254.701^2}{36}+\frac{368.022^2}{25}} \approx (-135.99, 207.57)$

(This interval was constructed using the summary statistics presented. If you used the full data set, the interval will vary slightly due to rounding.)

If the assumptions and conditions are met, we can be 95% confident that the interval –135.99 to 207.57 minutes contains the true difference in mean crossing times between females and males. Because the interval includes zero, we cannot be confident that there is any difference at all.

b) H_0: The mean crossing time is the same for females and males. $(\mu_F = \mu_M \text{ or } \mu_F - \mu_M = 0)$

H_A: The mean crossing time is different for females and males. $(\mu_F \neq \mu_M \text{ or } \mu_F - \mu_M \neq 0)$

We will check the conditions in part (c), but for now we will model the sampling distribution of the difference in means with a Student's *t*-model, with 39.6 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error $\sqrt{\frac{254.701^2}{36}+\frac{368.022^2}{25}} \approx 84.968373$.

The observed difference between the mean times is 1262.08 – 1226.29 = 35.79 minutes.

$t = \frac{(\bar{y}_F - \bar{y}_M)-(0)}{SE(\bar{y}_F - \bar{y}_M)} \approx \frac{35.79}{84.968373} \approx 0.4212$; Since the P-value = 0.676 is high, we fail to reject the null hypothesis. There is no evidence that the mean Lake Ontario crossing times are different for females and males.

c) **Independent groups assumption:** The times from the two groups are likely to be independent of one another, provided that these were all individual swims.
Randomization condition: The times are not a random sample from any identifiable population, but it is likely that the times are representative of times from swimmers who might attempt a challenge such as this. Hopefully, these times were recorded from different swimmers.
Nearly Normal condition: The boxplots show two high outliers for the men and some skewness for both. Removing the outliers may make the difference in times larger, but there is no justification for doing so. The histograms are unimodal; but somewhat skewed to the right.

We are reluctant to draw any conclusions about the difference in mean times it takes females and males to swim the lake. The sample is not random, we have no way of knowing if it is representative, and the data are skewed with some outliers.

80. Still swimming.

a) $(\bar{y}_F - \bar{y}_M) \pm t^*_{df}\sqrt{\frac{s_F^2}{n_F}+\frac{s_M^2}{n_M}} = (1208.80 - 1226.29) \pm t^*_{37.16}\sqrt{\frac{215.473^2}{30}+\frac{368.022^2}{25}} \approx (-186.57, 151.59)$

We are 95% confident that the interval –186.57 to 151.59 minutes contains the true difference in mean crossing times between females and males. Because the interval includes zero, we cannot be confident that there is any difference at all.

80. (continued)

b) $(\bar{y}_F - \bar{y}_M) \pm t^*_{df}\sqrt{\frac{s_F^2}{n_F} + \frac{s_M^2}{n_M}} = (1208.80 - 1140.57) \pm t^*_{47.57}\sqrt{\frac{215.473^2}{30} + \frac{214.722^2}{23}} \approx (-51.63, 188.09)$

We are 95% confident that the interval –51.63 to 188.09 minutes contains 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. Even giving men the benefit of the doubt, there is no evidence that the means are any difference. However, since this is not a random sample, we should be cautious in making any conclusions.

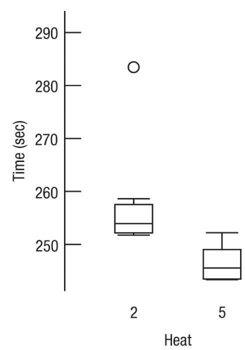
c) It is reasonable to assume that the same swimmer crossing the lake on two different occasions might perform similarly, so these are not all independent events. That could reduce the variability within each of the groups, reducing the standard error. We don't place much faith in our earlier analyses.

81. Running heats London.

H_0: The mean time to finish is the same for heats 2 and 5. $(\mu_2 = \mu_5 \text{ or } \mu_2 - \mu_5 = 0)$

H_A: The mean time is not the same for heats 2 and 5. $(\mu_2 \neq \mu_5 \text{ or } \mu_2 - \mu_5 \neq 0)$

Independent groups assumption: The two heats were independent.
Randomization condition: Runners were randomly assigned.
Nearly Normal condition: The boxplots show an outlier in the distribution of times in heat 2. We will perform the test twice, with and without the outlier.



Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 8.39 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error

$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{3.3618249^2}{6} + \frac{2.0996325^2}{6}} \approx 1.6181.$

The observed difference between mean times is 53.2267 – 53.1483 = 0.0784.

$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)} \approx \frac{0.0784}{1.6181} \approx 0.048$; Since the P-value = 0.9625 is high, we fail to reject the null hypothesis.

There is no evidence that the mean time to finish differs between the two heats.

Without the outlier, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 8.79 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error:

$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{1.4601267^2}{5} + \frac{2.0996325^2}{6}} \approx 1.0776.$

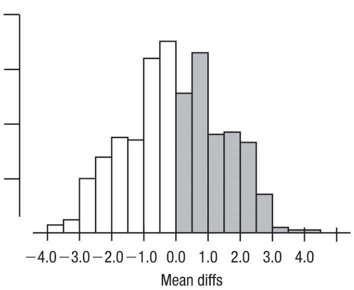
81. (continued)

The observed difference between mean times is 51.962 – 53.1483 = –1.1863.

$t = \dfrac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)} \approx \dfrac{-1.1863}{1.0776} \approx -1.10$; Since the P-value = 0.3001 is high, we fail to reject the null hypothesis.

There is no evidence that the mean time to finish differs between the two heats.

Using the randomization test, 500 simulated differences are plotted to the right. The actual difference of 0.07833 is almost exactly in the middle of the distribution of simulated differences. The simulated P-value is 0.952, which means that 95.2% of the simulated differences were at least 0.7833 in absolute value; there is no evidence of a difference in mean times between heats.

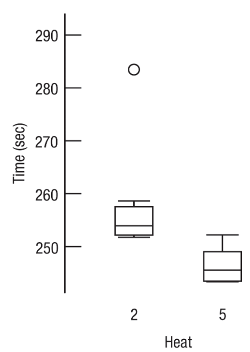


82. Swimming heats London.

H_0: The mean time to finish is the same for heats 2 and 5. $(\mu_2 = \mu_5 \text{ or } \mu_2 - \mu_5 = 0)$

H_A: The mean time is not the same for heats 2 and 5. $(\mu_2 \neq \mu_5 \text{ or } \mu_2 - \mu_5 \neq 0)$

Independent groups assumption: The two heats were independent.
Randomization condition: Swimmers were not randomly assigned, but if we consider these heats to be representative of seeded heats, we may be able to generalize the results.
Nearly Normal condition: The boxplots of the times in each heat show distributions that are reasonably symmetric, but there is one high outlier in heat 2. We will perform the test twice, with and without the outlier.



Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 8.41 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error

$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\dfrac{10.6634^2}{8} + \dfrac{3.4062^2}{8}} \approx 3.9578.$

82. (continued)

The observed difference between the mean times is 257.8 – 246.425 = 11.375.

$t = \dfrac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)} \approx \dfrac{11.375}{3.9578} \approx 2.874$; Since the P-value = 0.0196, we reject the null hypothesis. There is strong evidence that the mean time to finish differs between the two heats. In fact, the mean time in heat two was higher than the mean time in heat five.

Without the outlier, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 12.82 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

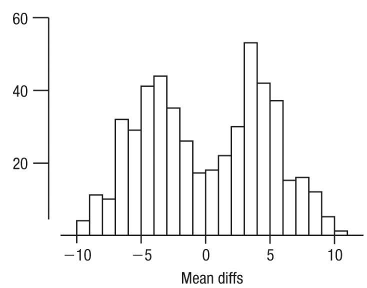
The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{2.6177^2}{7} + \frac{3.4062^2}{8}} \approx 1.5586.$$

The observed difference between the mean times is 254.129 – 246.425 = 7.704.

$t = \dfrac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)} \approx \dfrac{7.704}{1.5586} \approx 4.943$; Since the P-value = 0.0003, we reject the null hypothesis. There is strong evidence that the mean time to finish differs between the two heats. In fact, the mean time in heat two was higher than the mean time in heat five.

Using the randomization test, 500 simulated differences are plotted to the right. and the observed difference of 11.375 is beyond the right edge of the histogram. The simulated P-value is 0, which means that none of the simulated differences were at least 11.375 in absolute value. There is strong evidence of a difference in mean times between heats.



83. Tees.

H_0: The mean ball velocity is the same for regular and Stinger tees. $(\mu_S = \mu_R \text{ or } \mu_S - \mu_R = 0)$

H_A: The mean ball velocity is higher for the Stinger tees. $(\mu_S > \mu_R \text{ or } \mu_S - \mu_R > 0)$

Assuming the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 7.03 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_S - \bar{y}_R) = \sqrt{\dfrac{0.41^2}{6} + \dfrac{0.89^2}{6}} \approx 0.4000$.

The observed difference between the mean velocities is 128.83 – 127 = 1.83.

$t = \dfrac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)} \approx \dfrac{1.83}{0.4000} \approx 4.57$; Since the P-value = 0.0013, we reject the null hypothesis. There is strong evidence that the mean ball velocity for stinger tees is higher than the mean velocity for regular tees.

84. Golf again.

H_0: The mean distance is the same for regular and Stinger tees. $(\mu_S = \mu_R \text{ or } \mu_S - \mu_R = 0)$

H_A: The mean distance is greater for the Stinger tees. $(\mu_S > \mu_R \text{ or } \mu_S - \mu_R > 0)$

Assuming the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 9.42 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_S - \bar{y}_R) = \sqrt{\frac{2.76^2}{6} + \frac{2.14^2}{6}} \approx 1.426.$

The observed difference between mean distances is 241 – 227.17 = 13.83.

$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)} \approx \frac{13.83}{1.426} \approx 9.70$; ;Since the P-value <0.0001, we reject the null hypothesis. There is strong evidence that the mean distance for Stinger tees is higher than the mean distance for regular tees.

85. Music and memory.

a) H_0: The mean memory test score is the same for those who listen to Mozart as it is for those who listen to rap music. $(\mu_M = \mu_R \text{ or } \mu_M - \mu_R = 0)$

H_A: The mean memory test score is greater for those who listen to Mozart than it is for those who listen to rap music. $(\mu_M > \mu_R \text{ or } \mu_M - \mu_R > 0)$

Independent groups assumption: The groups are not related with regards to memory score.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: We don't have the actual data. We will assume that the distributions of the populations of memory test scores are Normal.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 45.88 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$SE(\bar{y}_M - \bar{y}_R) = \sqrt{\frac{3.19^2}{20} + \frac{3.99^2}{29}} \approx 1.0285.$

The observed difference between the mean number of objects remembered is 10.0 – 10.72 = – 0.72.

$t = \frac{(\bar{y}_M - \bar{y}_R) - (0)}{SE(\bar{y}_M - \bar{y}_R)} \approx \frac{-0.72}{1.0285} \approx -0.70$; Since the P-value = 0.7563 is high, we fail to reject the null hypothesis. There is no evidence that the mean number of objects remembered by those who listen to Mozart is higher than the mean number of objects remembered by those who listen to rap music.

b) $(\bar{y}_M - \bar{y}_N) \pm t_{df}^* \sqrt{\frac{s_M^2}{n_M} + \frac{s_N^2}{n_N}} = (10.0 - 12.77) \pm t_{19.09}^* \sqrt{\frac{3.19^2}{20} + \frac{4.73^2}{13}} \approx (-5.351, -0.189)$

We are 90% confident that the mean number of objects remembered by those who listen to Mozart is between 0.189 and 5.352 objects lower than the mean of those who listened to no music.

86. Rap.

a) H_0: The mean memory test score is the same for those who listen to rap as it is for those who listen to no music. $\left(\mu_R = \mu_N \text{ or } \mu_R - \mu_N = 0\right)$

H_A: The mean memory test score is lower for those who listen to rap than it is for those who listen to no music. $\left(\mu_R < \mu_N \text{ or } \mu_R - \mu_N < 0\right)$

Independent groups assumption: The groups are not related in regards to memory score.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: We don't have the actual data. We will assume that the distributions of the populations of memory test scores are Normal.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 20.00 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error $SE(\bar{y}_R - \bar{y}_N) = \sqrt{\frac{3.99^2}{29} + \frac{4.73^2}{13}} \approx 1.5066.$

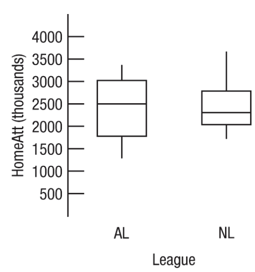
The observed difference between the mean number of objects remembered is 10.72 – 12.77 = – 2.05.

$t = \frac{\left(\bar{y}_R - \bar{y}_N\right) - (0)}{SE\left(\bar{y}_R - \bar{y}_N\right)} \approx \frac{-2.05}{1.5066} \approx -1.36$; Since the P-value = 0.0944 is high, we fail to reject the null hypothesis. There is little evidence that the mean number of objects remembered by those who listen to rap is lower than the mean number of objects remembered by those who listen to no music.

b) We did not conclude that there was a difference in the number of items remembered.

87. Attendance 2016 revisited.

a) Boxplots of Home Attendance by League are shown below.



b) A difference of 60,000 is very near the center of the distribution of simulated differences. The vast majority of the simulated differences are more than 60,000 in absolute value, so the difference is not statistically significant.

88. Hard water revisited.

None of the simulated differences were anywhere near 242.7 in absolute value. The simulated P-value is 0. This difference is highly statistically significant.

89. Cholesterol and gender II.

Using a randomization test with 500 simulated differences, the actual difference of -16.8 is far from the simulated values. The P-value is 0. In fact, none of the simulated differences are even close to an absolute value of 16.8. This means that there is strong evidence of a difference in mean cholesterol levels between men and women.

90. Memory II.

Using the randomization test, 500 simulated differences of $\mu_G - \mu_P$ are plotted. Values less than the actual difference of -0.99 are highlighted. There are 30 values below –0.99 and 470 above. It is not really necessary to calculate a P-value because the mean number of items recalled was actually lower for the gingko biloba group. This is a one-sided test because if gingko biloba improved memory, we would expect a positive difference, so the simulated P-value is 0.94. That means 94.0% of the simulated differences are greater than the actual difference from the experiment. This means there is no evidence that gingko biloba improves memory.