Chapter 24 – Comparing Means

1. Dogs and calories.

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 $\mu_{Meat} - \mu_{Beef}$.

2. Dogs and sodium.

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 $\mu_{Meat} - \mu_{Beef}$.

3. Dogs and fat.

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%.

4. Washers.

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%.

5. Dogs and fat, second helping.

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 not reason to expect other samples to conform to this result.

e) True.

6. Second load of wash.

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) False. Confidence intervals based on other samples will also try to estimate the true difference in population means. There's not reason to expect other samples to conform to this result.

d) True. **e)** True.

7. 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.

8. 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.

9. CPMP, again.

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

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

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.
10% condition: 312 and 265 are less than 10% of all students.
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.

10. CPMP and word problems.

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

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

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.
10% condition: 320 and 273 are less than 10% of all students.
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: $SE(\bar{y}_C - \bar{y}_T) = \sqrt{\frac{32.1^2}{320} + \frac{28.5^2}{273}} \approx 2.489$.

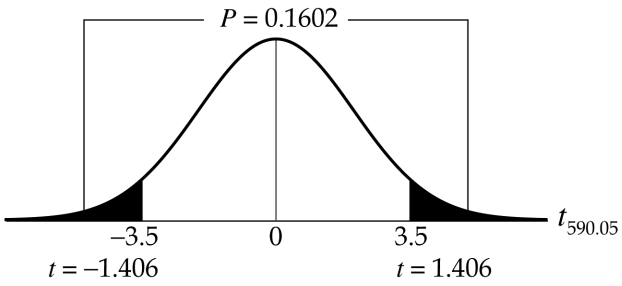
The observed difference between the mean scores is 57.4 – 53.9 = 3.5.

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.

$$t = \frac{(\bar{y}_C - \bar{y}_T) - (0)}{SE(\bar{y}_C - \bar{y}_T)}$$

$$t \approx \frac{3.5}{2.489}$$

$$t \approx 1.406$$



11. 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.

12. 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.

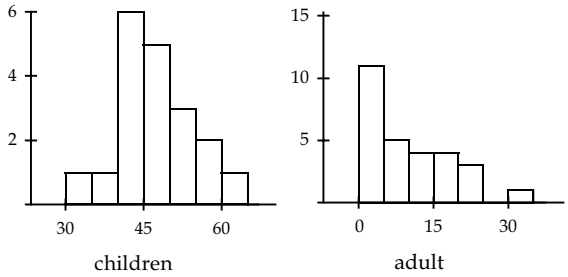
$$(\bar{y}_M - \bar{y}_F) \pm t^*_{df}\sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}} = (72.75 - 72.625) \pm t^*_{40.2}\sqrt{\frac{5.37225^2}{28} + \frac{7.69987^2}{24}} \approx (-3.025,\ 3.275)$$

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.

13. 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 histogram of 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.



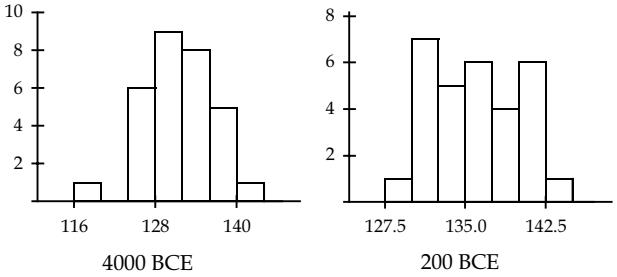
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 42 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

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

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.

14. 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 histograms of skull breadths are both unimodal and symmetric.



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.

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

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.

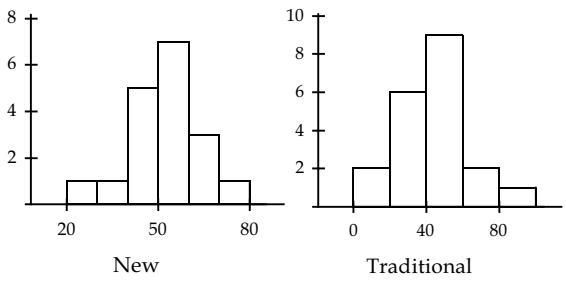
c) Since the interval is completely above 0, there is evidence that the mean breadth of males' skulls has changed over this time period. The evidence suggests that the mean skull breadth has increased.

15. 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 histograms of the scores 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 33 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test. We know:

$\bar{y}_N = 51.7222$ $\quad$ $\bar{y}_T = 41.8$
$s_N = 11.7062$ $\quad$ $s_T = 17.4495$
$n_N = 18$ $\quad$ $n_T = 20$

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

$$SE(\bar{y}_N - \bar{y}_T) = \sqrt{\frac{11.7062^2}{18} + \frac{17.4495^2}{20}} \approx 4.779.$$

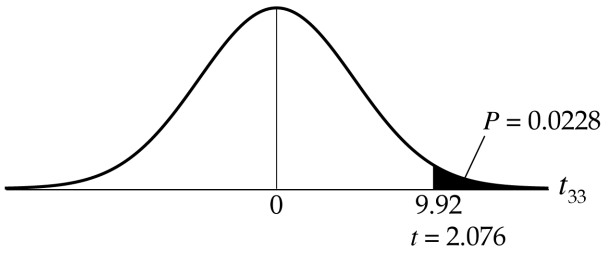
The observed difference between the mean scores is 51.7222 – 41.8 ≈ 9.922.

Since the *P*-value = 0.0228 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.

$$t = \frac{(\bar{y}_N - \bar{y}_T) - (0)}{SE(\bar{y}_N - \bar{y}_T)}$$

$$t \approx \frac{9.922}{4.779}$$

$$t \approx 2.076$$



16. Streams.

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

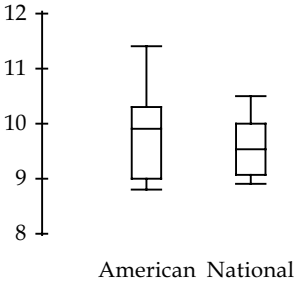
H_A: Streams with limestone substrates and streams with shale substrates have different mean pH levels. $(\mu_L \neq \mu_S \text{ or } \mu_L - \mu_S \neq 0)$

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.

17. Baseball 2006.

a) The boxplots of the average number of runs scored at the ballparks in the two leagues are at the right. Both distributions appear at least roughly symmetric, with roughly the same center, around 9.5 runs. The distribution of average runs appears a bit more spread out for the American League.



b) $\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 9.79286 \pm t^*_{13}\left(\frac{0.757998}{\sqrt{14}}\right) \approx (9.36, 10.23)$

We are 95% confident that the mean number of runs scored per game in American League stadiums is between 9.36 and 10.23.

c) The average of 10.5 runs scored per game in Coors Field is not unusual. It is the highest average in the National League, but by no means an outlier.

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.

18. Handy.

a) Males: $\bar{y}_M \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 19.39 \pm t^*_{49}\left(\frac{2.52}{\sqrt{50}}\right) \approx (18.67, 20.11)$

We are 95% confident that males can place between 18.67 and 20.11 pegs on average.

Females: $\bar{y}_F \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 17.91 \pm t^*_{49}\left(\frac{3.39}{\sqrt{50}}\right) \approx (16.95, 18.87)$

We are 95% confident that females can place between 16.95 and 18.87 pegs on average.

b) It may appear to suggest that there is no difference in the mean number of pegs placed by males and females, but a two-sample *t*-interval should be constructed to assess the difference in mean number of pegs placed.

c) $(\bar{y}_M - \bar{y}_F) \pm t^*_{df}\sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}} = (19.39 - 17.91) \pm t^*_{90.49}\sqrt{\frac{2.52^2}{50} + \frac{3.39^2}{50}} \approx (0.29, 2.67)$

d) We are 95% confident that the mean number of pegs placed by males is between 0.29 and 2.67 pegs higher than the mean number of pegs placed by females.

e) The two-sample *t*-interval is the correct procedure.

f) 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.

19. Double header 2006.

a) $(\bar{y}_A - \bar{y}_N) \pm t^*_{df}\sqrt{\frac{s_A^2}{n_A} + \frac{s_N^2}{n_N}} = (9.79286 - 9.43750) \pm t^*_{23}\sqrt{\frac{0.757998^2}{14} + \frac{0.638618^2}{16}} \approx (-0.18, 0.89)$

b) We are 95% confident that the mean number of runs scored in American League stadiums is between 0.18 runs lower and 0.89 runs higher than the mean number of runs scored in National League stadiums.

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

20. Hard water.

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.

$$t = \frac{(\bar{y}_N - \bar{y}_S) - (0)}{SE(\bar{y}_N - \bar{y}_S)}$$

$$t \approx \frac{242.74}{37.546}$$

$$t \approx 6.47$$

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.

21. 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.

22. 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.

23. 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. $(\mu_S = \mu_N \text{ or } \mu_S - \mu_N = 0)$

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

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.

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.

$$t = \frac{(\bar{y}_S - \bar{y}_N) - (0)}{SE(\bar{y}_S - \bar{y}_N)}$$
$$t \approx \frac{-1.46}{0.24}$$
$$t \approx -6.08$$

24. 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.

25. Sex and violence II.

a) H_0: The mean number of brands recalled is the same for viewers of violent content and viewers of neutral content. $(\mu_V = \mu_N \text{ or } \mu_V - \mu_N = 0)$

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

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 201.96 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:

$$SE(\bar{y}_V - \bar{y}_N) = \sqrt{\frac{1.61^2}{101} + \frac{1.62^2}{103}} \approx 0.226.$$

The observed difference between the mean scores is 3.02 – 4.65 = – 1.63.

$$t = \frac{(\bar{y}_V - \bar{y}_N) - (0)}{SE(\bar{y}_V - \bar{y}_N)}$$

$$t \approx \frac{-1.63}{0.226}$$

$$t \approx -7.21$$

Since the *P*-value = 1.1×10^{-11} is low, we reject the null hypothesis. There is strong evidence that the mean number of brand names recalled is different for viewers of violent content and viewers of neutral content. The evidence suggests that viewers of neutral ads remember more brand names on average than viewers of violent content.

b) $(\bar{y}_N - \bar{y}_S) \pm t^*_{df}\sqrt{\frac{s_N^2}{n_N} + \frac{s_S^2}{n_S}} = (4.65 - 2.72) \pm t^*_{204.8}\sqrt{\frac{1.62^2}{103} + \frac{1.85^2}{106}} \approx (1.456, 2.404)$

We are 95% confident that the mean number of brand names recalled 24 hours later is between 1.46 and 2.40 higher for viewers of shows with neutral content than for viewers of shows with sexual content.

26. Ad recall.

a) He might attempt to conclude that the mean number of brand names recalled is greater after 24 hours.

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

c) A person with high recall right after the show might tend to have high recall 24 hours later as well. Also, the first interview may have helped the people to remember the brand names for a longer period of time than they would have otherwise.

d) Randomly assign half of the group watching that type of content to be interviewed immediately after watching, and assign the other half to be interviewed 24 hours later.

27. Hungry?

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

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

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). We will perform a two-sample *t*-test.

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

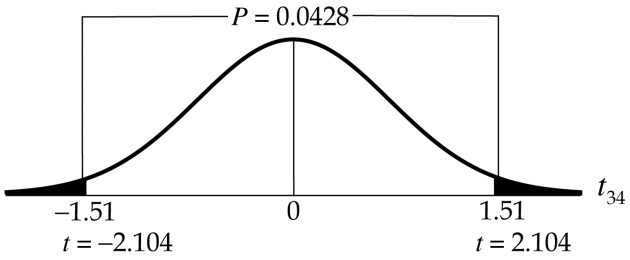
$$SE(\bar{y}_{big} - \bar{y}_{small}) = \sqrt{\frac{2.91^2}{22} + \frac{1.84^2}{26}} \approx 0.7177 \text{ oz.}$$

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

$$t = \frac{\left(\bar{y}_{big} - \bar{y}_{small}\right) - (0)}{SE\left(\bar{y}_{big} - \bar{y}_{small}\right)}$$

$$t \approx \frac{1.51}{0.7177}$$

$$t \approx 2.104$$



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.

28. 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.

$$t = \frac{\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right) - (0)}{SE\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right)}$$

$$t \approx \frac{18.7}{2.4264}$$

$$t \approx 7.71$$

Since the *P*-value (less than 0.0001) is low, we reject the null hypothesis. There is strong evidence that the 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.

29. Lower scores?

a) Assuming that the conditions for inference were met by the NAEP, a 95% confidence interval for the difference in mean score is:

$$(\bar{y}_{1996} - \bar{y}_{2000}) \pm t^*_{df}\sqrt{\frac{s^2_{1996}}{n_{1996}} + \frac{s^2_{2000}}{n_{2000}}} = (150 - 147) \pm 1.960(1.22) \approx (0.61, 5.39)$$

Since the samples sizes are very large, it should be safe to use $z^* = 1.960$ for the critical value of *t*. We are 95% confident that the mean score in 2000 was between 0.61 and 5.39 points lower than the mean score in 1996. Since 0 is not contained in the interval, this provides evidence that the mean score has decreased from 1996 to 2000.

b) Both sample sizes are very large, which will make the standard errors of these samples very small. They are both likely to be very accurate. The difference in sample size shouldn't make you any more certain or any less certain.

However, these results are completely dependent upon whether or not the conditions for inference were met. If, by sampling more students, the NAEP sampled from a different population, then the two years are incomparable.

30. The Internet.

a) The differences that were observed between the group of students with Internet access and those without were too great to be attributed to natural sampling variation.

b) The researchers have incorrectly rejected their null hypothesis of no difference between the groups, committing a Type I error.

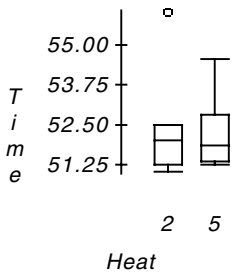
c) There is evidence of an association between Internet access and mean science score, but this does not prove that access to the Internet causes higher scores. There may be other variables involved, such as socioeconomic status or the education level of the parents. We would need results from a controlled experiment to determine cause and effect.

31. Running heats.

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 to finish 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, once with the outlier and once without.



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 10.82 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:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{1.69319^2}{7} + \frac{1.20055^2}{7}} \approx 0.7845.$$

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$

The observed difference between mean times is 52.3557 – 52.3286 = 0.0271.

$$t \approx \frac{0.0271}{0.7845}$$

$$t \approx 0.035$$

Since the *P*-value = 0.97 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 8.83 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:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{0.56955^2}{6} + \frac{1.20055^2}{7}} \approx 0.5099.$$

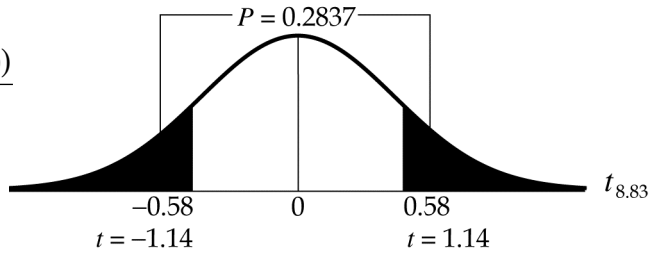
The observed difference between mean times is 51.7467 – 52.3286 = –0.5819.

Since the *P*-value = 0.2837 is high, we fail to reject the null hypothesis. There is no evidence that the mean time to finish differs between the two heats.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$

$$t \approx \frac{-0.58}{0.5099}$$

$$t \approx -1.14$$

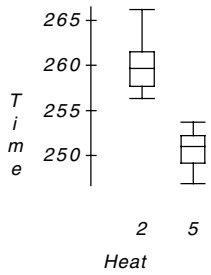


32. Swimming heats.

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 to finish 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.



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 12.62 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:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{3.031^2}{8} + \frac{2.149^2}{8}} \approx 1.3136.$$

The observed difference between the mean times is 260.23 – 250.79 = 9.44.

Since the *P*-value < 0.001, 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.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$

$$t \approx \frac{9.44}{1.3136}$$

$$t \approx 7.19$$

33. 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:

$$SE(\bar{y}_S - \bar{y}_R) = \sqrt{\frac{.41^2}{6} + \frac{.89^2}{6}} \approx 0.4000.$$

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

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.

$$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)}$$

$$t \approx \frac{1.83}{0.4000}$$

$$t \approx 4.57$$

34. 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.

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.

$$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)}$$

$$t \approx \frac{13.83}{1.426}$$

$$t \approx 9.70$$

35. Crossing Ontario.

a) $(\bar{y}_M - \bar{y}_W) \pm t^*_{df}\sqrt{\frac{s_M^2}{n_M} + \frac{s_W^2}{n_W}} = (1196.75 - 1271.59) \pm t^*_{37.67}\sqrt{\frac{304.369^2}{20} + \frac{261.111^2}{22}} \approx (-252.89, 103.21)$

We are 95% confident that the interval –252.89 to 103.20 minutes (–74.84 ± 178.05 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.

b) **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 distributions of times are both unimodal, with no outliers. The distribution of men's times is somewhat skewed to the left.

36. 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 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 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:

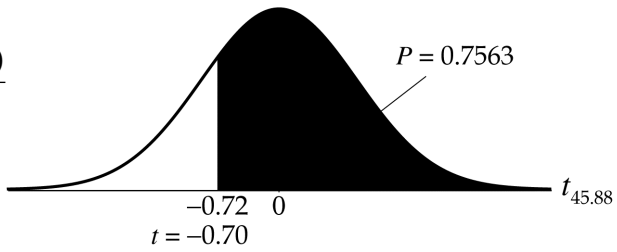
$$SE(\bar{y}_M - \bar{y}_R) = \sqrt{\frac{3.19^2}{20} + \frac{3.99^2}{29}} \approx 1.0285.$$

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

$$t = \frac{(\bar{y}_M - \bar{y}_R) - (0)}{SE(\bar{y}_M - \bar{y}_R)}$$

$$t \approx \frac{-0.72}{1.0285}$$

$$t \approx -0.70$$



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.

37. 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. $(\mu_R = \mu_N$ or $\mu_R - \mu_N = 0)$

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. $(\mu_R < \mu_N$ or $\mu_R - \mu_N < 0)$

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:

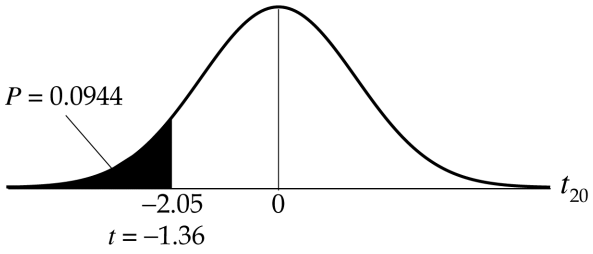
$$SE(\bar{y}_R - \bar{y}_N) = \sqrt{\frac{3.99^2}{29} + \frac{4.73^2}{13}} \approx 1.5066.$$

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

$$t = \frac{(\bar{y}_R - \bar{y}_N) - (0)}{SE(\bar{y}_R - \bar{y}_N)}$$

$$t \approx \frac{-2.05}{1.5066}$$

$$t \approx -1.36$$



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.

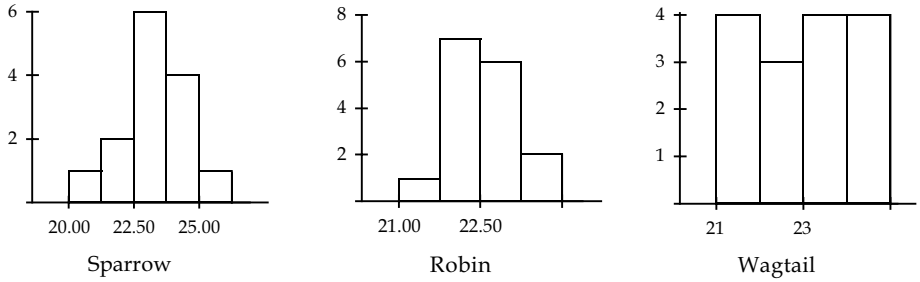
b) We did not conclude that there was a difference in the number of items remembered.

38. Cuckoos.

In order to determine whether the mean length of cuckoo eggs is the same for different species, we will conduct three hypothesis tests.

Independent groups assumption: The eggs were collected from the nests of three different species of bird.
Randomization condition: Assume that the eggs are representative of all cuckoo eggs laid in the nest of the particular species of bird.
10% condition: 14, 16, and 15 are less than 10% of all cuckoo eggs.

Nearly Normal condition: The histograms of the distribution of the lengths of cuckoo eggs found in sparrow and robin nests are unimodal and symmetric. The histogram of the distribution of the lengths of cuckoo eggs found in wagtail nests is uniform, but since there are no outliers and the sample size is not too small, it should be safe to proceed.



1) H_0: The mean length of cuckoo eggs is the same whether the foster parents are sparrows or robins. $(\mu_S = \mu_R \text{ or } \mu_S - \mu_R = 0)$

H_A: The mean length of cuckoo eggs is different, depending on whether the foster parents are sparrows or robins. $(\mu_S \neq \mu_R \text{ or } \mu_S - \mu_R \neq 0)$

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 21.60 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:

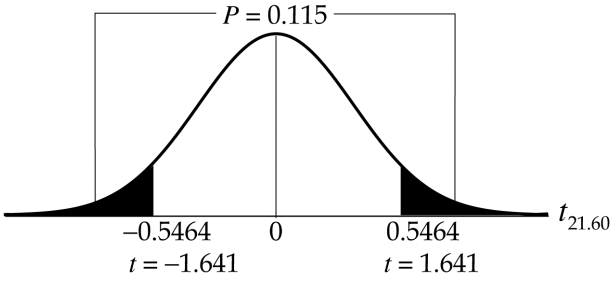
$$SE(\bar{y}_S - \bar{y}_R) = \sqrt{\frac{1.06874^2}{14} + \frac{0.68452^2}{16}} \approx 0.3330.$$

The observed difference between the mean length of the cuckoo eggs is 23.1214 – 22.575 = 0.5464.

$$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)}$$

$$t \approx \frac{0.5464}{0.3330}$$

$$t \approx 1.641$$



Since the *P*-value = 0.115 is high, we fail to reject the null hypothesis. There is little evidence that the mean length of cuckoo eggs is different when the foster parents are sparrows than when they are robins.

2) H_0: The mean length of cuckoo eggs is the same whether the foster parents are sparrows or wagtails. $(\mu_S = \mu_W \text{ or } \mu_S - \mu_W = 0)$

H_A: The mean length of cuckoo eggs is different, depending on whether the foster parents are sparrows or wagtails. $(\mu_S \neq \mu_W \text{ or } \mu_S - \mu_W \neq 0)$

This test is virtually identical in mechanics to the first test. We know:

$\bar{y}_S = 23.1214$	$\bar{y}_W = 22.9033$	$t = 0.549$
$s_S = 1.06874$	$s_W = 1.06762$	$df = 26.86$
$n_S = 14$	$n_W = 15$	$P - Value = 0.587$

Since the *P*-value = 0.587 is high, we fail to reject the null hypothesis. There is little evidence that the mean length of cuckoos eggs is different when the foster parents are sparrows than when they are wagtails.

3) H_0: The mean length of cuckoo eggs is the same whether the foster parents are robins or wagtails. $(\mu_R = \mu_W$ or $\mu_R - \mu_W = 0)$

H_A: The mean length of cuckoo eggs is different, depending on whether the foster parents are robins or wagtails. $(\mu_R \neq \mu_W$ or $\mu_R - \mu_W \neq 0)$

This test is virtually identical in mechanics to the first test. We know:

$\bar{y}_R = 22.575$	$\bar{y}_W = 22.9033$	$t = -1.012$
$s_R = 0.68452$	$s_W = 1.06762$	$df = 23.60$
$n_R = 16$	$n_W = 15$	$P - Value = 0.322$

Since the *P*-value = 0.322 is high, we fail to reject the null hypothesis. There is little evidence that the mean length of cuckoo eggs is different when the foster parents are robins than when they are wagtails.

There is no evidence to suggest a difference in mean length of cuckoo eggs that are laid in the nests of different foster parents. In general, we should be wary of doing three *t*-tests on the same data. Our Type I error is not the same for doing three tests as it is for one test. However, because none of the tests showed significant differences, this is less of a concern here.