Chapter 21 – Paired Samples and Blocks

Section 21.1

1. Which method?

a) Paired. Each individual has two scores, so they are certainly associated.

b) Not paired. The scores of males and females are independent.

c) Paired. Each student was surveyed twice, so their responses are associated.

2. Which method II?

a) Paired. Each respondent is paired with his or her spouse.

b) Not paired. The respondents in the different treatment groups are not associated.

c) Not paired. Opinions of freshman and sophomores are not associated.

Section 21.2

3. Cars and trucks.

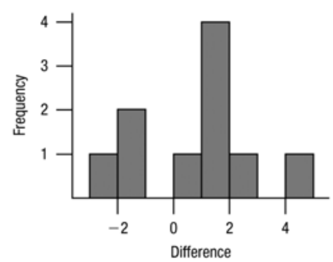
a) No. The vehicles have no natural pairing.

b) Possibly. The data are quantitative and paired by vehicle.

c) The sample size is large, but there is at least one extreme outlier that should be investigated before applying these methods. And there are several values near 0 that should be investigated as well.

4. Vehicle weights II.

Yes. The data are quantitative and paired. A graph of the differences is roughly symmetric.



Section 21.3

5. Cars and trucks again.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 7.37 \pm t_{631}^{*}\left(\frac{2.52}{\sqrt{632}}\right) \approx (7.17,\ 7.57)$$

We are 95% confident that the interval 7.17 mpg to 7.57 mpg captures the true improvements in highway gas mileage compared to city gas mileage.

6. Vehicle weights III.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 0.740 \pm t_{9}^{*}\left(\frac{2.280}{\sqrt{10}}\right) \approx (-1.29,\ 2.77)$$

We are 98% confident that the interval from –1290 pounds to 2770 pounds captures the true mean difference in measured weights.

Section 21.4

7. Blocking cars and trucks.

The difference between fuel efficiency of cars and that of trucks can be large, but isn't relevant to the question asked about highway vs. city driving. Pairing places each vehicle in its own block to remove that variation from consideration.

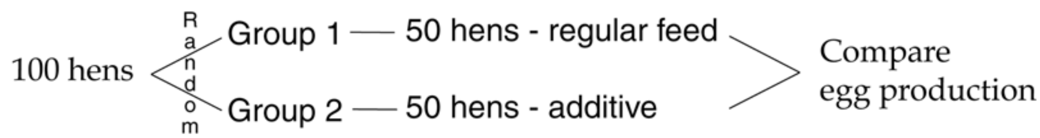
8. Vehicle weights IV.

No. We are not concerned with the measurements of each scale. We want to know the difference between the methods for each truck. So, it is the paired differences that we are concerned with.

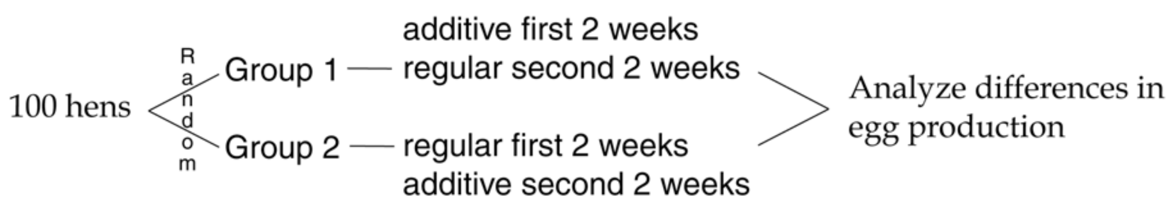
Chapter Exercises

9. More eggs?

a) Randomly assign 50 hens to each of the two kinds of feed. Compare the mean egg production of the two groups at the end of one month.



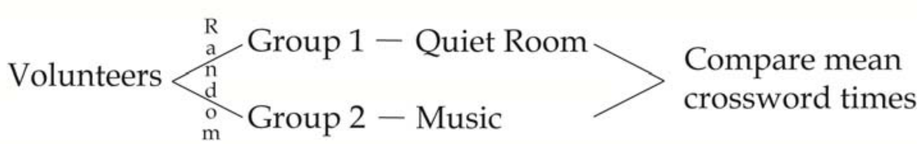
b) Randomly divide the 100 hens into two groups of 50 hens each. Feed the hens in the first group the regular feed for two weeks, then switch to the additive for 2 weeks. Feed the hens in the second group the additive for two weeks, and then switch to the regular feed for two weeks. Subtract each hen's "regular" egg production from her "additive" egg production, and analyze the mean difference in egg production.



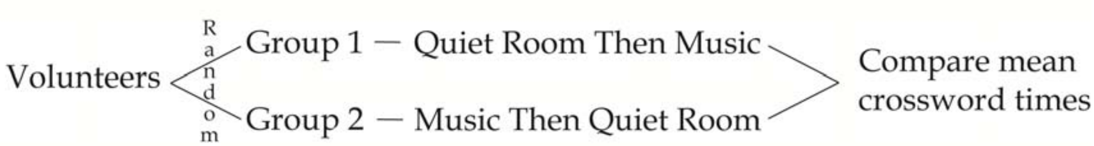
c) The matched pairs design in part (b) is the stronger design. Hens vary in their egg production regardless of feed. This design controls for that variability by matching the hens with themselves.

10. Music.

a) Randomly assign half of the volunteers to do the puzzles in a quiet room, and assign the other half to do the puzzles with music in headphones. Compare the mean time of the group in the quiet room to the mean time of the group listening to music.



b) Randomly assign half of the volunteers to do a puzzle in a quiet room, and assign the other half to do the puzzles with music. Then have each do a puzzle under the other condition. Subtract each volunteer's "quiet" time from his or her "music" time, and analyze the mean difference in times.

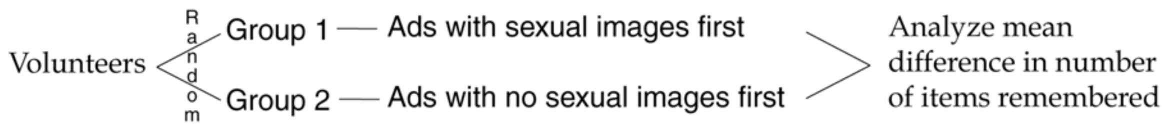


10. (continued)

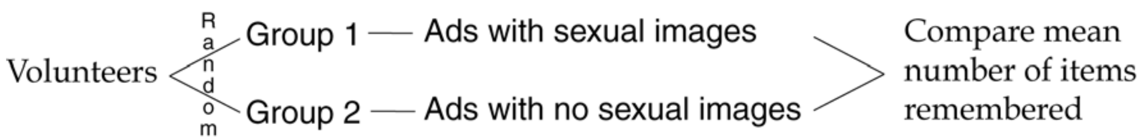
c) The matched pairs design in part (b) is the stronger design. People vary in their ability to do crossword puzzles. This design controls for that variability by matching the volunteers with themselves.

11. Sex sells?

a) Randomly assign half of the volunteers to watch ads with sexual images, and assign the other half to watch ads without the sexual images. Record the number of items remembered. Then have each group watch the other type of ad. Record the number of items recalled. Examine the difference in the number of items remembered for each person.



b) Randomly assign half of the volunteers to watch ads with sexual images, and assign the other half to watch ads without the sexual images. Record the number of items remembered. Compare the mean number of products remembered by each group.



12. Freshman 15?

a) Select a random sample of freshmen. Weigh them when college starts in the fall, and again when they leave for home in the spring. Examine the difference in weight for each student.

b) Select a random sample of freshman as they enter college in the fall to determine their average weight. Select a new random sample of students at the end of the spring semester to determine their average weight. Compare the mean weights of the two groups.

13. Women.

a) The paired *t*-test is appropriate. The labor force participation rate for two different years was paired by city.

b) Since the P-value = 0.0244, there is evidence of a difference in the average labor force participation rate for women between 1968 and 1972. The evidence suggests an increase in the participation rate for women.

14. Cloud seeding.

a) The two-sample *t*-test is appropriate for these data. The seeded and unseeded clouds are not paired in any way. They are independent.

b) Since the P-value = 0.0538, there is some evidence that the mean rainfall from seeded clouds is greater than the mean rainfall from unseeded clouds.

15. Friday the 13th, traffic.

a) The paired *t*-test is appropriate, since we have 10 pairs of Fridays in 5 different months. Data from adjacent Fridays within a month may be more similar than randomly chosen Fridays.

b) Since the P-value = 0.0004, there is evidence that the mean number of cars on the M25 motorway on Friday the 13th is less than the mean number of cars on the previous Friday.

c) We don't know if these Friday pairs were selected at random. The Nearly Normal condition appears to be met by the differences, but the sample size of ten pairs is small. Additionally, mixing data collected at both Junctions 9 and 10 together in one test confounds the test. We don't know if we are seeing a Friday the 13th effect or a Junction effect.

16. Friday the 13th, accidents.

a) The paired *t*-test is appropriate, since we have pairs of Fridays in 6 different months. Data from adjacent Fridays within a month may be more similar than randomly chosen Fridays.

b) Since the P-value = 0.0211, there is evidence that the mean number of admissions to hospitals found on Friday the 13th is more than on the previous Friday.

c) We don't know if these Friday pairs were selected at random. Obviously, if these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal condition appears to be met by the differences, but the sample size of six pairs is small.

17. Online insurance I.

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

18. Wind speed, part I.

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

19. Online insurance II.

a) The histogram would help you decide whether the online company offers cheaper insurance. We are concerned with the difference in price, not the distribution of each set of prices.

b) Insurance cost is based on risk, so drivers are likely to see similar quotes from each company, making the differences relatively smaller.

c) The price quotes are paired. They were for a random sample of fewer than 10% of the agent's customers and the histogram of differences looks approximately Normal.

20. Wind speed, part II.

a) The outliers are particularly windy days, but they were windy at both sites, making the difference in wind speeds less unusual.

b) The histogram and summaries of the differences are more appropriate because they are paired observations and all we care about is which site is windier.

c) The wind measurements at the same times at two nearby sites are paired. We should be concerned that there might be a lack of independence from one time to the next, but the times were 6 hours apart and the *differences* in speeds are likely to be independent. Although not random, we can regard a sample this large as generally representative of wind speed s at these sites. The histogram of differences is unimodal, symmetric and bell-shaped.

21. Online insurance III.

H_0: The mean difference between online and local insurance rates is zero. $(\mu_{Local-Online} = 0)$

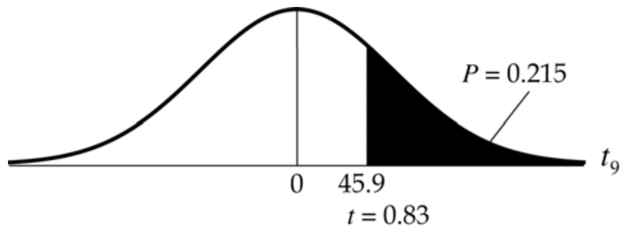
H_A: The mean difference is greater than zero. $(\mu_{Local-Online} > 0)$

Since the conditions are satisfied (in Exercises 17 and 19), the sampling distribution of the difference can be modeled with a Student's *t*-model with 10 – 1 = 9 degrees of freedom, $t_9\left(0, \frac{175.663}{\sqrt{10}}\right)$.

21. (continued)

We will use a paired *t*-test, with $\bar{d} = 45.9$.

$t = \dfrac{\bar{d} - 0}{s_d / \sqrt{n}} = \dfrac{45.9 - 0}{175.663/\sqrt{10}} \approx 0.83$; Since the P-value = 0.215 is high, we fail to reject the null hypothesis. There is no evidence that online insurance premiums are lower on average.



22. Wind speed, part III.

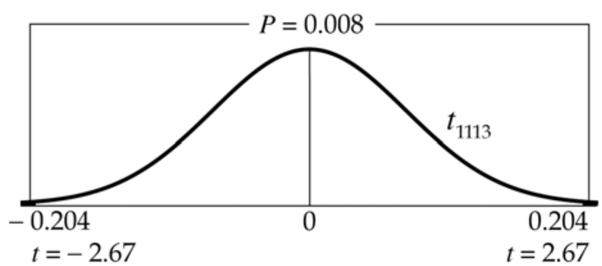
H_0: The mean difference between wind speeds at the two sites is zero. $(\mu_{2-4} = 0)$

H_A: The mean difference between at the two sites is different than zero. $(\mu_{2-4} \neq 0)$

Since the conditions are satisfied (in Exercises 18 and 20), the sampling distribution of the difference can be modeled with a Student's *t*-model with 1114 – 1 = 1113 degrees of freedom, $t_{1113}\left(0, \dfrac{2.551}{\sqrt{1114}}\right)$.

We will use a paired *t*-test, with $\bar{d} = 0.204$.

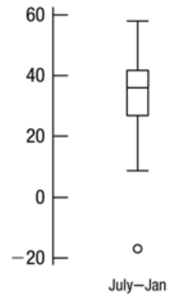
$t = \dfrac{\bar{d} - 0}{s_d / \sqrt{n}} = \dfrac{0.204 - 0}{2.551/\sqrt{1114}} \approx 2.67$; Since the P-value = 0.008 is low, we reject the null hypothesis. There is strong evidence that the average wind speed is higher at site 2.



23. City Temperatures.

a) Paired data assumption: The data are paired by city.
Randomization condition: These cities might not be representative of all cities, so be cautious in generalizing the results.
Normal population assumption: The boxplot of differences between January and July mean temperature is roughly unimodal and symmetric, but shows a low outlier. This is Auckland, New Zealand, a city in the southern hemisphere. Seasons here would be the opposite of the rest of the cities, which are in the northern hemisphere. It should be set aside.

23. (continued)

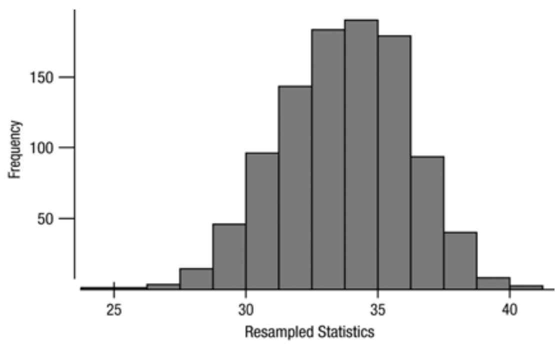


b) With Auckland set aside, the conditions are satisfied. The sampling distribution of the difference can be modeled with a Student's *t*-model with 11 – 1 = 10 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 33.7429 \pm t_{34}^{*}\left(\frac{14.9968}{\sqrt{34}}\right) \approx (30.94,\ 39.53)$$

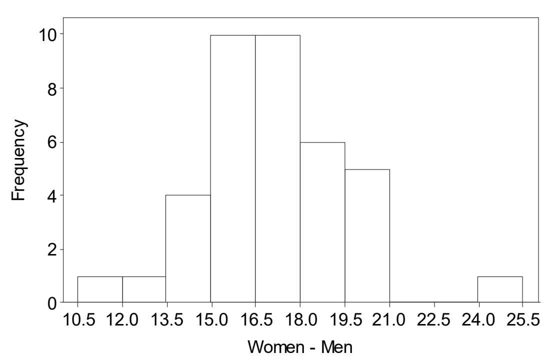
We are 95% confident that the average high temperature in northern hemisphere cities in July is an average of between 30.94° to 39.53° higher than in January.

c) Answers will vary. Based on our simulation, we are 95% confident that the average high temperature in northern hemisphere cities in July is an average of between 28.9° to 38.1° higher than in January.



24. NY Marathon 2016.

a) **Paired data assumption:** The data are paired by year.
Randomization condition: Assume these years, at this marathon, are representative of all differences.
Normal population assumption: The histogram of differences between women's and men's times is roughly unimodal and symmetric, though there is one outlier, in the year 2014. The sample size is large enough for the Central Limit Theorem to apply.



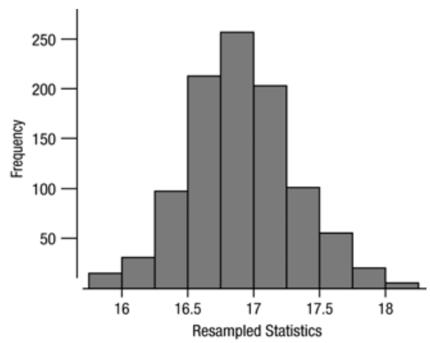
24. (continued)

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 38 – 1 = 37 df. We will find a paired *t*-interval, with 90% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 16.894 \pm t^*_{37}\left(\frac{2.440}{\sqrt{38}}\right) \approx (16.227, 17.563)$$

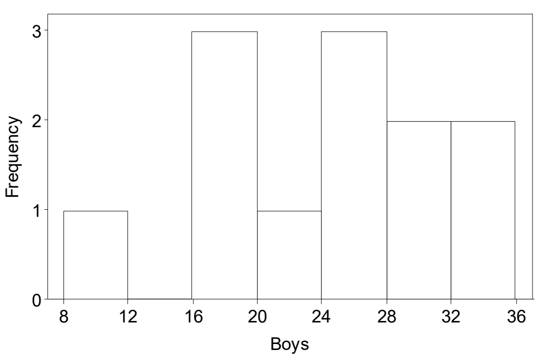
We are 90% confident women's winning marathon times are an average of between 16.227 and 17.563 minutes longer than men's winning times.

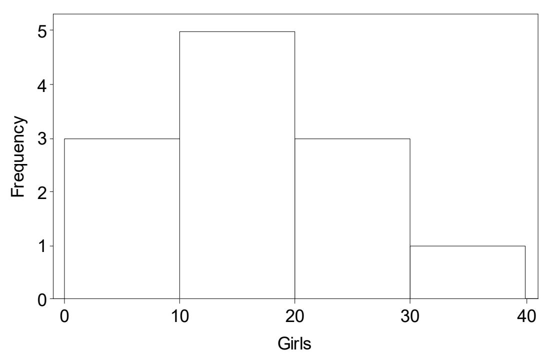
b) Answers will vary. According to a bootstrap interval of 1000 samples, we are 90% confident that women's winning marathon times are an average of between 16.228 and 17.519 minutes longer than men's winning times.



25. Push-ups.

Independent groups assumption: The group of boys is independent of the group of girls.
Randomization condition: Assume that students are assigned to gym classes at random.
Nearly Normal condition: The histograms of the number of push-ups from each group are roughly 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 21 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 90% confidence.

$$(\bar{y}_B - \bar{y}_G) \pm t^*_{df}\sqrt{\frac{s_B^2}{n_B} + \frac{s_G^2}{n_G}} = (23.8333 - 16.5000) \pm t^*_{21}\sqrt{\frac{7.20900^2}{12} + \frac{8.93919^2}{12}} \approx (1.6, 13.0)$$

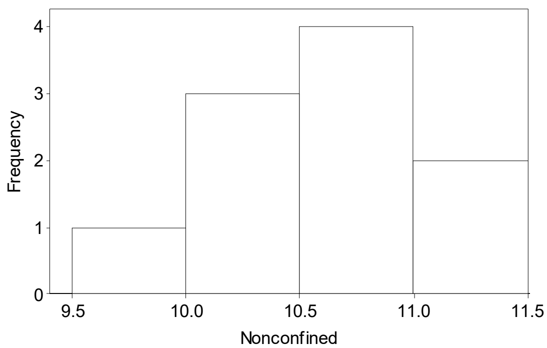
We are 90% confident that, at Gossett High, the mean number of push-ups that boys can do is between 1.6 and 13.0 more than the mean for the girls.

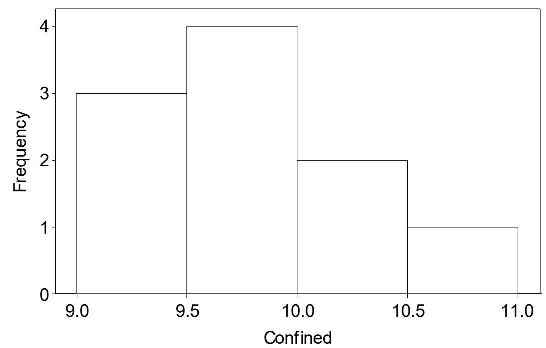
26. Brain waves.

a) H_0: The mean alpha-wave frequency for nonconfined inmates is the same as the mean alpha wave frequency for confined inmates. $(\mu_{NC} = \mu_C$ or $\mu_{NC} - \mu_C = 0)$

H_A: The mean alpha-wave frequency for nonconfined inmates is different from the mean frequency for confined inmates. $(\mu_{NC} \neq \mu_C$ or $\mu_{NC} - \mu_C \neq 0)$

b) **Independent Groups Assumption:** The two groups of inmates were placed under different conditions, solitary confinement and not confined.
Randomization Condition: Inmates were randomly assigned to groups.
Nearly Normal Condition: The histograms of the alpha-wave frequencies are unimodal and symmetric.





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 16.9 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test. We know the following.

$\bar{y}_{NC} = 10.58$ $\quad$ $\bar{y}_C = 9.78$

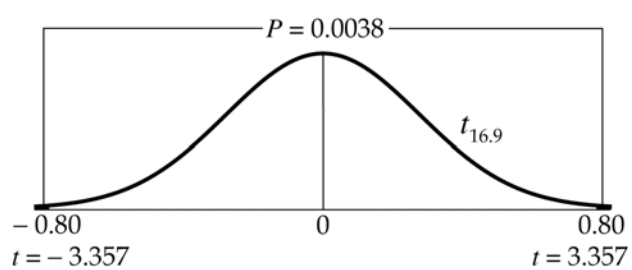
$s_{NC} = 0.458984$ $\quad$ $s_C = 0.597774$

$n_{NC} = 10$ $\quad$ $n_C = 10$

The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_{NC} - \bar{y}_C) = \sqrt{\frac{0.458984^2}{10} + \frac{0.597774^2}{10}} \approx 0.2383.$$

The observed difference between the mean scores is 10.58 – 9.78 ≈ 0.80.



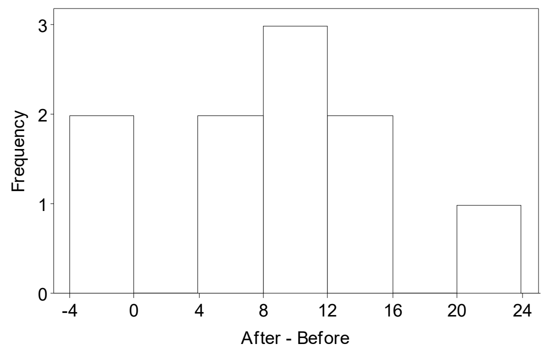
$t = \dfrac{(\bar{y}_{NC} - \bar{y}_{NC}) - (0)}{SE(\bar{y}_{NC} - \bar{y}_{NC})} = \dfrac{0.80}{0.2382} \approx 3.357$; Since the P-value = 0.0038 is low, we reject the null hypothesis. There is evidence the mean alpha-wave frequency is different for nonconfined inmates and confined inmates.

d) The evidence suggests that the mean alpha-wave frequency for inmates subjected to confinement is different than the mean alpha-wave frequency for inmates that are not confined. This experiment suggests that mean alpha-wave frequency is lower for confined inmates.

27. Job satisfaction.

a) Use a paired *t*-test.

Paired data assumption: The data are before and after job satisfaction rating for the same workers.
Randomization condition: The workers were randomly selected to participate.
Nearly Normal condition: The histogram of differences between before and after job satisfaction ratings is roughly unimodal and symmetric.

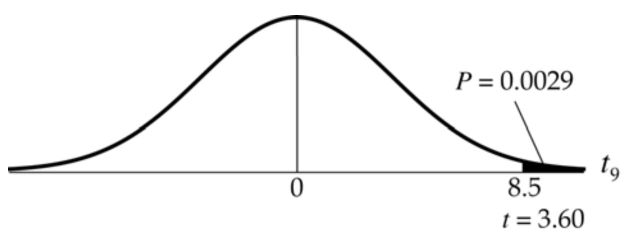


b) H_0: The mean difference in before and after job satisfaction scores is zero, and the exercise program is not effective at improving job satisfaction. $(\mu_d = 0)$

H_A: The mean difference is greater than zero, and the exercise program is effective at improving job satisfaction. $(\mu_d > 0)$

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 10 – 1 = 9 degrees of freedom, $t_9\left(0, \frac{7.47217}{\sqrt{10}}\right)$. We will use a paired *t*-test, with $\bar{d} = 8.5$.

$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{8.5 - 0}{7.47217/\sqrt{10}} \approx 3.60$; Since the P-value = 0.0029 is low, we reject the null hypothesis. There is evidence that the mean job satisfaction rating has increased since the implementation of the exercise program.



c) We concluded that there was an increase job satisfaction rating. If we are wrong, and there actually was no increase, we have committed a Type I error.

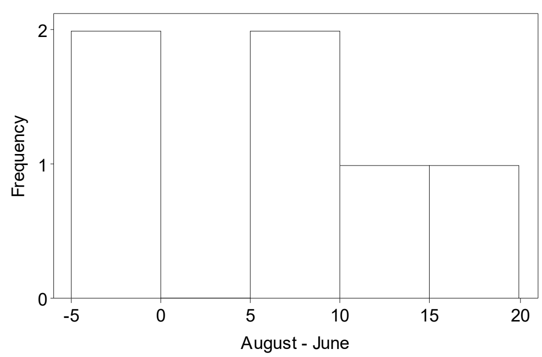
d) The P-value of the sign test is 0.1094, which would lead us to fail to reject the null hypothesis that the median difference was 0. This is a different conclusion than the paired *t*-test. However, since the conditions for the paired *t*-test are met, we should use those results. The *t*-test has more power.

28. Summer school.

a) H_0: The mean difference between August and June scores is zero, and the summer school program is not worthwhile. $(\mu_d = 0)$

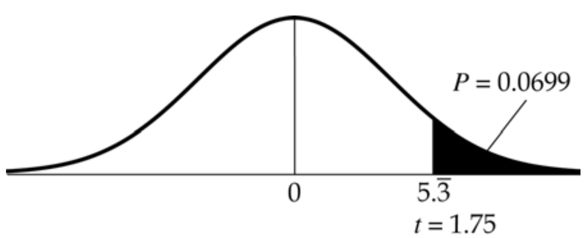
H_A: The mean difference between August and June scores is greater than zero, and the summer school program is worthwhile. $(\mu_d > 0)$

Paired data assumption: The scores are paired by student.
Randomization condition: Assume that these students are representative of students who attend this school in other years.
Normal population assumption: The histogram of differences between August and June scores shows a distribution that could have come from a Normal population.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 6 – 1 = 5 degrees of freedom, $t_5\left(0, \frac{7.44759}{\sqrt{6}}\right)$. We will use a paired *t*-test, with $\bar{d} = 5.\bar{3}$.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{5.\bar{3} - 0}{7.44759 / \sqrt{6}} \approx 1.75$; Since the P-value = 0.0699 is fairly high, we fail to reject the null hypothesis. There is not strong evidence that scores increased on average. The summer school program does not appear worthwhile, but the P-value is low enough that we should look at a larger sample to be more confident in our conclusion.



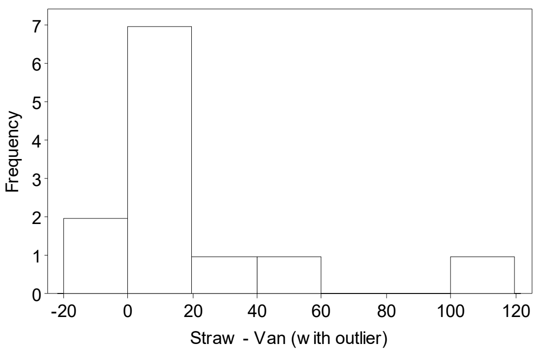
b) We concluded that there was no evidence of an increase. If there actually was an increase, we have committed a Type II error.

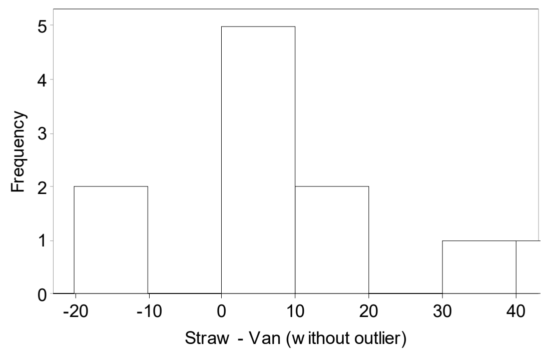
29. Yogurt flavors.

H_0: The mean difference in calories between servings of strawberry and vanilla yogurt is zero. $(\mu_d = 0)$

H_A: The mean difference in calories between servings of strawberry and vanilla yogurt is different from zero. $(\mu_d \neq 0)$

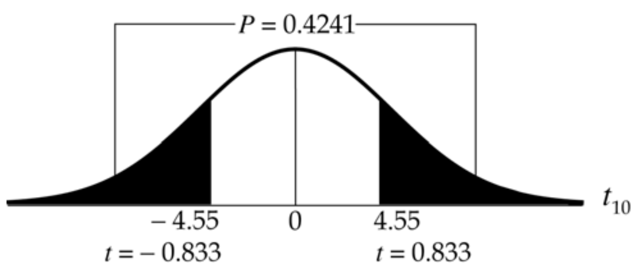
Paired data assumption: The yogurt is paired by brand.
Randomization condition: Assume that these brands are representative of all brands.
Normal population assumption: The histogram of differences in calorie content between strawberry and vanilla shows an outlier, Great Value. When the outlier is eliminated, the histogram of differences is roughly unimodal and symmetric.





When Great Value yogurt is removed, the conditions are satisfied. The sampling distribution of the difference can be modeled with a Student's *t*-model with
11 – 1 = 10 degrees of freedom, $t_{10}\left(0, \frac{18.0907}{\sqrt{11}}\right)$. We will use a paired *t*-test, with $\bar{d} \approx 4.54545.$.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{5.54545 - 0}{18.0907/\sqrt{11}} \approx 0.833$; Since the P-value = 0.4241 is high, we fail to reject the null hypothesis.
There is no evidence of a mean difference in calorie content between strawberry yogurt and vanilla yogurt.

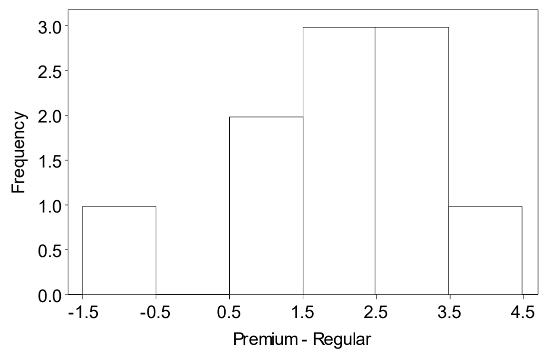


30. Gasoline.

a) H_0: The mean difference in mileage between premium and regular is zero. $(\mu_d = 0)$

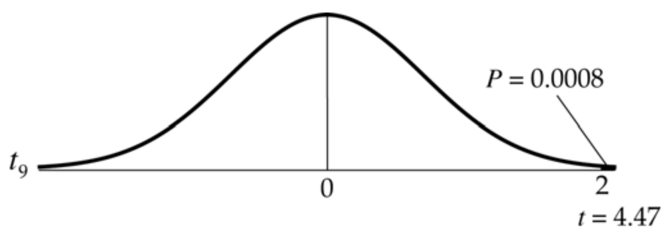
H_A: The mean difference in mileage between premium and regular is greater than zero. $(\mu_d > 0)$

Paired data assumption: The mileage is paired by car.
Randomization condition: We randomized the order in which the different types of gasoline were used in each car.
Normal population assumption: The histogram of differences between premium and regular is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with $10 - 1 = 9$ degrees of freedom, $t_9\left(0, \frac{1.41421}{\sqrt{10}}\right)$. We will use a paired t-test, with $\bar{d} = 2$.

$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{2-0}{1.41421/\sqrt{10}} \approx 4.47$; Since the P-value = 0.0008 is very low, we reject the null hypothesis.
There is strong evidence of a mean increase in gas mileage between regular and premium.



b) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 2 \pm t^*_9\left(\frac{1.41421}{\sqrt{10}}\right) \approx (1.18, 2.82)$

We are 90% confident that the mean increase in gas mileage when using premium rather than regular gasoline is between 1.18 and 2.82 miles per gallon.

c) Premium costs more than regular. This difference might outweigh the increase in mileage.

d) With $t = 1.25$ and a P-value = 0.1144, we would have failed to reject the null hypothesis, and conclude that there was no evidence of a mean difference in mileage. The variation in performance of individual cars is greater than the variation related to the type of gasoline. This masked the true difference in mileage due to the gasoline. (Not to mention the fact that the two-sample test is not appropriate because we don't have independent samples!)

31. Stopping distance.

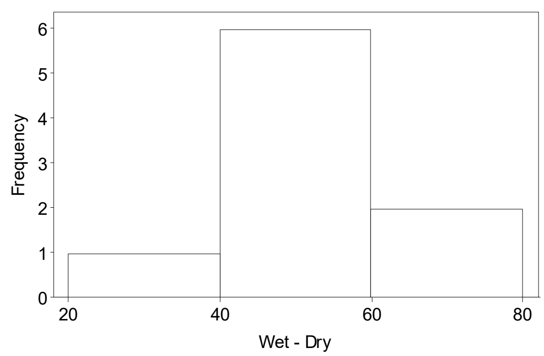
a) Randomization Condition: These cars are not a random sample, but are probably representative of all cars in terms of stopping distance.
Nearly Normal Condition: A histogram of the stopping distances is skewed to the right, but this may just be sampling variation from a Normal population. The "skew" is only a couple of stopping distances. We will proceed cautiously.

The cars in the sample had a mean stopping distance of 138.7 feet and a standard deviation of 9.66149 feet. Since the conditions have been satisfied, construct a one-sample t -interval, with 10 – 1 = 9 degrees of freedom, at 95% confidence.

$$\overline{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 138.7 \pm t_{9}^{*}\left(\frac{9.66149}{\sqrt{10}}\right) \approx (131.8,\ 145.6)$$

We are 95% confident that the mean dry pavement stopping distance for cars with this type of tires is between 131.8 and 145.6 feet. This estimate is based on an assumption that these cars are representative of all cars and that the population of stopping distances is Normal.

b) Paired data assumption: The data are paired by car.
Randomization condition: Assume that the cars are representative of all cars.
Normal population assumption: The difference in stopping distance for car #4 is an outlier, at only 12 feet. After excluding this difference, the histogram of differences is unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 9 – 1 = 8 degrees of freedom. We will find a paired t-interval, with 95% confidence.

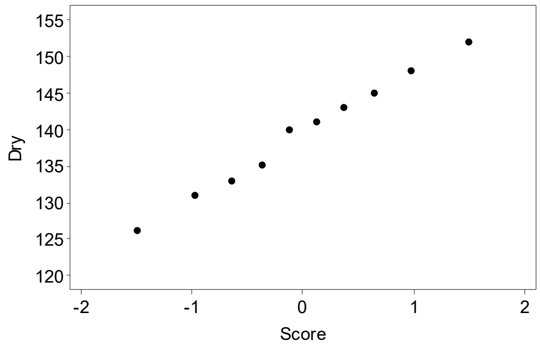
$$\overline{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 55 \pm t_{8}^{*}\left(\frac{10.2103}{\sqrt{9}}\right) \approx (47.2,\ 62.8)$$

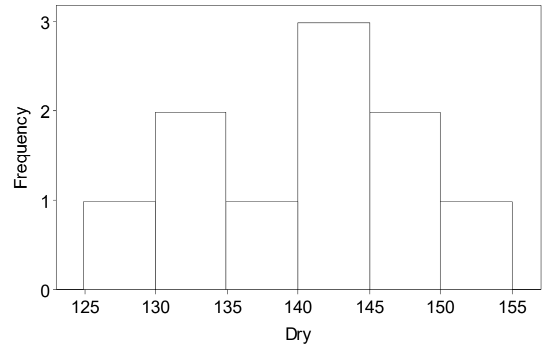
With car #4 removed, we are 95% confident that the mean increase in stopping distance on wet pavement is between 47.2 and 62.8 feet. (If you leave the outlier in, the interval is 38.8 to 62.6 feet, but you should remove it! This procedure is sensitive to outliers!)

32. Stopping distance 60.

a) Randomization Condition: These stops are probably representative of all such stops for this type of car, but not for all cars.
Nearly Normal Condition: A histogram of the stopping distances is roughly unimodal and symmetric and the Normal probability plot looks very straight.

32. (continued)



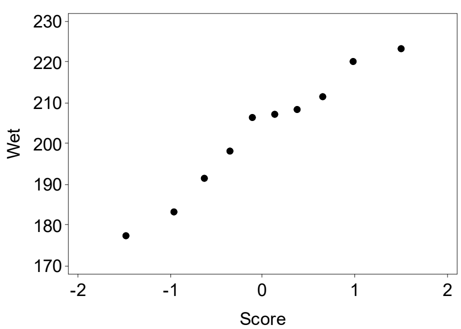


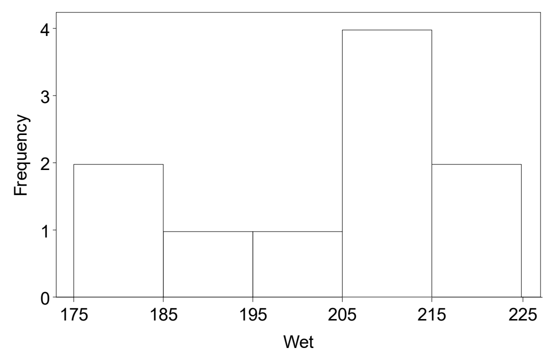
The stops in the sample had a mean stopping distance of 139.4 feet, and a standard deviation of 8.09938 feet. Since the conditions have been satisfied, construct a one-sample t -interval, with 10 – 1 = 9 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 139.4 \pm t_{9}^{*}\left(\frac{8.09938}{\sqrt{10}}\right) \approx (133.6, 145.2)$$

We are 95% confident that the mean dry pavement stopping distance for this type of car is between 133.6 and 145.2 feet.

b) **Independent Groups Assumption:** The wet pavement stops and dry pavement stops were made under different conditions and not paired in any way.
Randomization Condition: These stops are probably representative of all such stops for this type of car, but not for all cars.
Nearly Normal Condition: From part (a), the histogram of dry pavement stopping distances is roughly unimodal and symmetric, but the histogram of wet pavement stopping distances is a bit skewed. Since the Normal probability plot looks fairly straight, we will proceed.





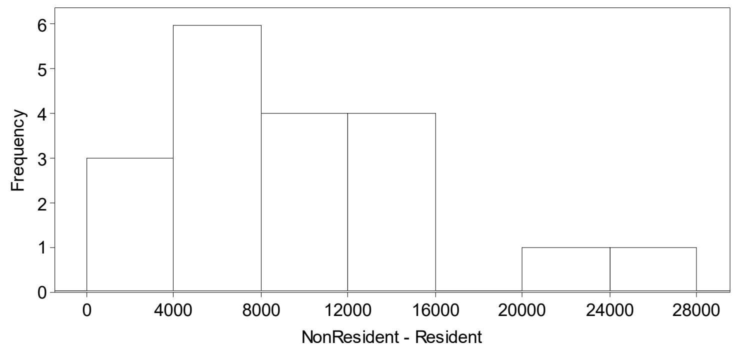
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 13.8 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

$$(\bar{y}_W - \bar{y}_D) \pm t_{df}^{*}\sqrt{\frac{s_W^2}{n_W} + \frac{s_D^2}{n_D}} = (202.4 - 139.4) \pm t_{13.8}^{*}\sqrt{\frac{15.07168^2}{10} + \frac{8.09938^2}{10}} \approx (51.4, 74.7)$$

We are 95% confident that the mean stopping distance on wet pavement is between 51.4 and 74.6 feet longer than the mean stopping distance on dry pavement.

33. Tuition 2016.

a) **Paired data assumption:** The data are paired by college.
Randomization condition: The colleges were selected randomly.
Normal population assumption: The tuition difference for UC Irvine, at \$26,682, is an outlier, as is the tuition difference for New College of Florida at \$23,078. Once these have been set aside, the histogram of the differences is roughly unimodal and slightly skewed to the right. This should be fine for inference in a sample of 17 colleges.



b) Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 17 – 1 = 16 degrees of freedom. We will find a paired *t*-interval, with 90% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 8158.5882 \pm t^*_{16}\left(\frac{3968.3927}{\sqrt{17}}\right) \approx (6478.216,\ 9838.960)$$

With outliers removed, we are 90% confident that the mean increase in tuition for nonresidents versus residents is between about \$6478 and \$9839. (If you left the outliers in your data, the interval is about \$7335 to \$12,501, but you should set them aside! This procedure is sensitive to the presence of outliers!)

c) There is no evidence to suggest that the magazine made a false claim. An increase of \$7000 for nonresidents is contained within our 90% confidence interval. (If the outliers were left in, then \$7000 is actually lower than the interval of plausible values for the mean tuition difference.)

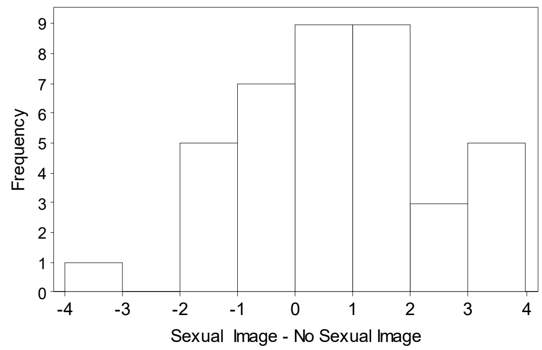
34. Sex sells, part II.

H_0: The mean difference in number of items remembered for ads with sexual images and ads without sexual images is zero. $(\mu_d = 0)$

H_A: The mean difference in number of items remembered for ads with sexual images and ads without sexual images is not zero. $(\mu_d \neq 0)$

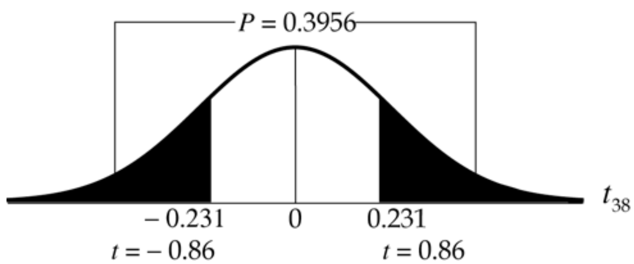
Paired data assumption: The data are paired by subject.
Randomization condition: The ads were in random order.
Normal population assumption: The histogram of differences is roughly unimodal and symmetric.

34. (continued)



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 39 – 1 = 38 degrees of freedom, $t_{38}\left(0, \frac{1.677}{\sqrt{39}}\right)$. We will use a paired *t*-test, with $\bar{d} = 0.231$.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{0.231 - 0}{1.67747/\sqrt{39}} \approx 0.86$; Since the P-value = 0.3956 is high, we fail to reject the null hypothesis.
There is no evidence of a mean difference in the number of objects remembered with ads with sexual images and without.



35. Strikes.

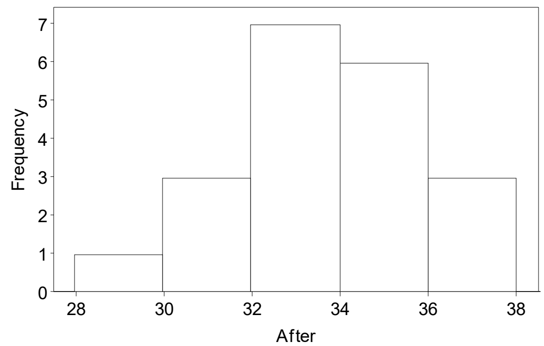
a) Since 60% of 50 pitches is 30 pitches, the Little Leaguers would have to throw an average of more than 30 strikes in order to give support to the claim made by the advertisements.

H_0: The mean number of strikes thrown by Little Leaguers who have completed the training is 30. $(\mu_A = 30)$

H_A: The mean number of strikes thrown by Little Leaguers who have completed the training is greater than 30. $(\mu_A > 30)$

Randomization Condition: Assume that these players are representative of all Little League pitchers.
Nearly Normal Condition: The histogram of the number of strikes thrown after the training is roughly unimodal and symmetric.

35. (continued)



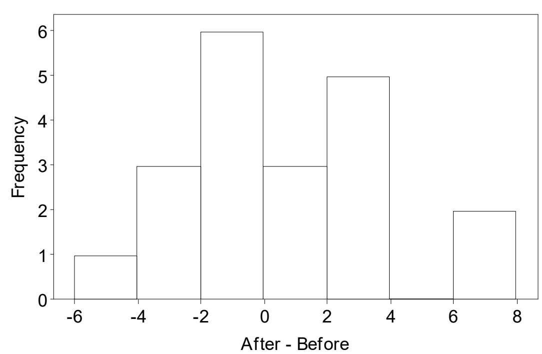
The pitchers in the sample threw a mean of 33.15 strikes, with a standard deviation of 2.32322 strikes. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean number of strikes thrown with a Student's *t*-model, with 20 – 1 = 19 degrees of freedom, $t_{19}\left(30, \frac{2.32322}{\sqrt{20}}\right)$.

We will perform a one-sample *t*-test.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{33.15 - 30}{2.32322/\sqrt{20}} \approx 6.06$; Since the P-value = 3.92×10^{-6} is very low, we reject the null hypothesis. There is strong evidence that the mean number of strikes that Little Leaguers can throw after the training is more than 30. (This test says nothing about the effectiveness of the training; just that Little Leaguers can throw more than 60% strikes on average after completing the training. This might not be an improvement.)

b) H_0: The mean difference in number of strikes thrown before and after the training is zero. $(\mu_d = 0)$
H_A: The mean difference in number of strikes thrown before and after the training is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by pitcher.
Randomization condition: Assume that these players are representative of all Little League pitchers.
Normal population assumption: The histogram of differences is roughly unimodal and symmetric.

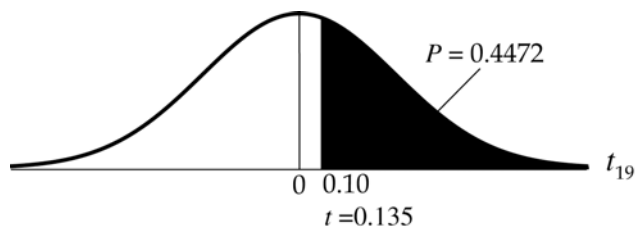


Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 20 – 1 = 19 degrees of freedom, $t_{19}\left(0, \frac{3.32297}{\sqrt{19}}\right)$.

35. (continued)

We will use a paired t-test, with $\bar{d} = 0.1$.

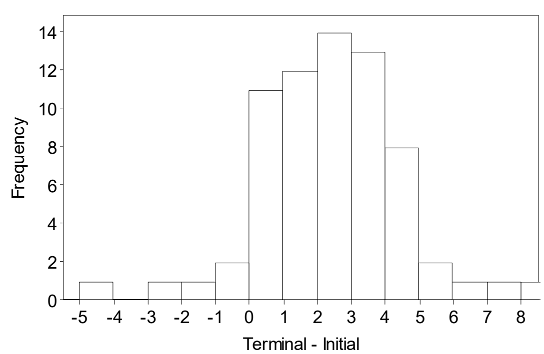
$t = \dfrac{\bar{d} - 0}{s_d / \sqrt{n}} = \dfrac{0.1 - 0}{3.32297 / \sqrt{20}} \approx 0.135$; Since the P-value = 0.4472 is high, we fail to reject the null hypothesis. There is no evidence of a mean difference in number of strikes thrown before and after the training. The training does not appear to be effective.



c) Answers will vary. According to our bootstrap interval, we are 95% confident that the mean number of strikes thrown is between 32.2 and 34.5. This is well above the 30 strikes out of 50 pitches (60%) claimed by the advertisements. There is evidence that Little Leaguers can throw at least 60% strikes after the training program.

36. Freshman 15, revisited.

a) **Paired data assumption:** The data are paired by student.
Randomization condition: The students matched the rest of the class in terms of demographic variables.
Normal population assumption: The histogram of differences is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 68 – 1 = 67 degrees of freedom. We will find a paired t-interval, with 95% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 1.91176 \pm t^*_{67}\left(\frac{2.12824}{\sqrt{68}}\right) \approx (1.40,\ 2.43)$$

We are 95% confident that freshmen at Cornell have a mean weight gain of between 1.40 and 2.43 pounds during the first 12 weeks of college. This interval does not contain zero, so there is evidence of a weight gain among freshmen, although it is quite small. These data certainly do not support the idea of the "freshman 15".

b) Answers will vary. According to our bootstrap interval, we are 95% confident that freshmen at Cornell have a mean weight gain of between 1.39 and 2.43 pounds during the first 12 weeks of college. This interval does not contain zero, so there is evidence of a weight gain among freshmen, although it is quite small. These data certainly do not support the idea of the "freshman 15".

37. Wheelchair marathon 2016.

a) The data are certainly paired. Even if the individual times show a trend of improving speed over time, the differences may well be independent of each other. They are subject to random year-to-year fluctuations, and we may believe that these data are representative of similar races. We don't have any information with which to check the Nearly Normal condition, unless we use the data set instead of the summary statistics provided. At any rate, with a sample of size 40, the Central Limit Theorem will allow us to construct the confidence interval.

b) $\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = -7.27 \pm t_{39}^{*}\left(\frac{33.568}{40}\right) \approx (-18.01, 3.47)$

We are 95% confident that the interval –18.01 to 3.47 minutes contains the true mean time difference for women's wheelchair times and men's running times.

c) The interval contains zero, so we would not reject a null hypothesis of no mean difference at a significance level of 0.05. We are unable to discern a difference between the female wheelchair times and the male running times.

38. Marathon start-up years 2016.

a) The data are certainly paired. Even if the individual times show a trend of improving speed over time, the differences may well be independent of each other. They are subject to random year-to-year fluctuations, and we may believe that these data are representative of similar races. After removing the three initial years, the remaining part of the histogram is a bit skewed and possibly bimodal, but we can probably use paired-*t* methods with caution. The sample size of 37 is still fairly large.

b) $\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = -15.29 \pm t_{36}^{*}\left(\frac{18.14}{\sqrt{37}}\right) \approx (-21.34, -9.24)$

We are 95% confident that female wheelchair marathoners average between 9.24 and 21.34 minutes faster than male runners in races such as this.

c) Since the interval does not contain zero, we would reject the null hypothesis of no difference at a significance level of 0.05. There is strong evidence that female wheelchair marathoners finish faster than male runners, on average.

39. BST.

a) **Paired data assumption:** We are testing the same cows, before and after injections of BST.
Randomization condition: These cows are likely to be representative of all Ayrshires.
Normal population assumption: We don't have the list of individual differences, so we can't look at a histogram. The sample is large, so we may proceed.

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 60 – 1 = 59 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

b) $\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 14 \pm t_{59}^{*}\left(\frac{5.2}{\sqrt{60}}\right) \approx (12.66, 15.34)$

c) We are 95% confident that the mean increase in daily milk production for Ayrshire cows after BST injection is between 12.66 and 15.34 pounds.

d) 25% of 47 pounds is 11.75 pounds. According to the interval generated in part (b), the average increase in milk production is more than this, so the farmer can justify the extra expense for BST.

40. BST II.

Although the data from each herd of cows are paired, we are asked to compare the paired differences from each herd. The herds are independent, so we will use a two-sample *t*-test.

H_0: The mean increase in milk production due to BST is the same for both breeds.
$\left(\mu_{dA} = \mu_{dJ} \text{ or } \mu_{dA} - \mu_{dJ} = 0\right)$
H_A: The mean increase in milk production due to BST is different for the two breeds.
$\left(\mu_{dA} \neq \mu_{dJ} \text{ or } \mu_{dA} - \mu_{dJ} \neq 0\right)$

Independent groups assumption: The cows are from different herds.
Randomization condition: Assume that the cows are representative of their breeds.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the two sets of differences. 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 (actually, the difference in mean differences!) with a Student's *t*-model, with 109.55 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

$$SE(\bar{d}_A - \bar{d}_J) = \sqrt{\frac{5.2^2}{60} + \frac{4.8^2}{52}} \approx 0.945.$$

The observed difference between the mean differences is 14 – 9 = 5.

$t = \dfrac{(\bar{d}_A - \bar{d}_J) - (0)}{SE(\bar{d}_A - \bar{d}_J)} \approx \dfrac{5}{0.945} \approx 5.29$; Since the P-value = 6.4×10^{-7} is very small, we reject the null hypothesis.

There is strong evidence that the mean increase for each breed is different. The average increase for Ayrshires is significantly greater than the average increase for Jerseys.