Review of Part VI – Inference for Relationships

R6.1. Herbal cancer.

H_0: The cancer rate for those taking the herb is the same as the cancer rate for those not taking the herb. $\left(p_{Herb} = p_{Not} \text{ or } p_{Herb} - p_{Not} = 0\right)$

H_A: The cancer rate for those taking the herb is higher than the cancer rate for those not taking the herb. $\left(p_{Herb} > p_{Not} \text{ or } p_{Herb} - p_{Not} > 0\right)$

R6.2. Birth days.

a) If births are distributed uniformly across all days, we expect the number of births on each day to be $np = (72)\left(\frac{1}{7}\right) \approx 10.29$.

b) H_0: Babies are equally likely to be born on any of the days of the week.
H_A: Babies are not equally likely to be born on any of the days of the week.

Counted data condition: The data are counts.
Randomization condition: Assume these births are representative of all births.
Expected cell frequency condition: Expected counts (all $\frac{72}{7}$) are greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on $7 - 1 = 6$ degrees of freedom. We will use a chi-square goodness-of-fit test.

$$\chi^2 = \frac{(7-10.29)^2}{10.29} + \frac{(17-\frac{72}{7})^2}{\frac{72}{7}} + \frac{(8-\frac{72}{7})^2}{\frac{72}{7}} + \frac{(12-\frac{72}{7})^2}{\frac{72}{7}} + \frac{(9-\frac{72}{7})^2}{\frac{72}{7}} + \frac{(10-\frac{72}{7})^2}{\frac{72}{7}} + \frac{(9-\frac{72}{7})^2}{\frac{72}{7}} \approx 6.56,\ \text{df} = 6;$$

Since the P-value = 0.3639 is high, we fail to reject the null hypothesis. There is no evidence that babies are not equally likely to be born on any of the days of the week.

c) The standardized residuals for Monday and Tuesday are –1.02 and 2.09, respectively. Monday's value is not unusual at all. Tuesday's is borderline high, but we concluded that there is not evidence that births are not uniform. With 7 standardized residuals, it is not surprising that one is large.

d) Some births are scheduled for the convenience of the doctor and/or the mother.

R6.3. Surgery and germs.

a) Lister imposed a treatment, the use of carbolic acid as a disinfectant. This is an experiment.

b) H_0: The survival rate when carbolic acid is used is the same as the survival rate when carbolic acid is not used. $\left(p_C = p_N \text{ or } p_C - p_N = 0\right)$

H_A: The survival rate when carbolic acid is used is greater than the survival rate when carbolic acid is not used. $\left(p_C > p_N \text{ or } p_C - p_N > 0\right)$

Randomization condition: There is no mention of random assignment. Assume that the two groups of patients were similar, and amputations took place under similar conditions, with the use of carbolic acid being the only variable.
Independent samples condition: It is reasonable to think that the groups were not related in any way.
Success/Failure condition: $n\hat{p}$ (carbolic acid) = 34, $n\hat{q}$ (carbolic acid) = 6, $n\hat{p}$ (none) = 19, and $n\hat{q}$ (none) = 16. The number of patients who died in the carbolic acid group is only 6, but the expected number of deaths using the pooled proportion, $n\hat{q}_{\text{pooled}} = (40)(\frac{22}{75}) = 11.7$, so the samples are both large enough.

R6.3. (continued)

Since the conditions have been satisfied, we will perform a two-proportion *z*-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by $SE\left(\hat{p}_C - \hat{p}_N\right) = \sqrt{\frac{\hat{p}_C\hat{q}_C}{n_C} + \frac{\hat{p}_N\hat{q}_N}{n_N}} = \sqrt{\frac{\left(\frac{34}{40}\right)\left(\frac{6}{40}\right)}{40} + \frac{\left(\frac{19}{35}\right)\left(\frac{16}{35}\right)}{35}} \approx 0.10137987.$

The observed difference between the proportions is $\frac{34}{40} - \frac{19}{35} = 0.3071429.$

$z = \frac{(\hat{p}_C - \hat{p}_N) - (0)}{SE(\hat{p}_C - \hat{p}_N)} = \frac{0.3071429 - 0}{0.10137987} \approx 3.03;$ Since the P-value = 0.0012 is low, we reject the null hypothesis. There is strong evidence that the survival rate is higher when carbolic acid is used to disinfect the operating room than when carbolic acid is not used.

c) We don't know whether or not patients were randomly assigned to treatments, and we don't know whether or not blinding was used.

R6.4. Free throws 2017.

a) **Randomization condition:** Assume that these free throws are representative of the free throw ability of these players.
Independent samples condition: The free throw abilities of these two players should be independent.
Success/Failure condition: $n\hat{p}$ (Hardin) = 468, $n\hat{q}$ (Hardin) = 76, $n\hat{p}$ (Westbook) = 425, and $n\hat{q}$ (Westbook) = 92 are all greater than 10, so the samples are both large enough.

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

$$\left(\hat{p}_C - \hat{p}_B\right) \pm z^* \sqrt{\frac{\hat{p}_C\hat{q}_C}{n_C} + \frac{\hat{p}_B\hat{q}_B}{n_B}} = \left(\frac{468}{544} - \frac{425}{517}\right) \pm 1.960 \sqrt{\frac{\left(\frac{468}{544}\right)\left(\frac{76}{544}\right)}{544} + \frac{\left(\frac{425}{517}\right)\left(\frac{92}{517}\right)}{517}} = (-0.0060, 0.082)$$

We are 95% confident that James Hardin's true free throw percentage is between 0.60% worse and 8.2% better than Russell Westbook's.

b) Since the interval for the difference in percentage of free throws made includes 0, it is uncertain who the better free throw shooter is.

R6.5. Twins.

H_0: The proportion of preterm twin births in 2000 is the same as the proportion of preterm twin births in 2010. $\left(p_{2000} = p_{2010} \text{ or } p_{2000} - p_{2010} = 0\right)$

H_A: The proportion of preterm twin births in 1990 is the less than the proportion of preterm twin births in 2000. $\left(p_{2000} < p_{2010} \text{ or } p_{2000} - p_{2010} < 0\right)$

Randomization condition: Assume that these births are representative of all twin births.
Independent samples condition: The samples are from different years, so they are unlikely to be related.
Success/Failure condition: $n\hat{p}$ (2000) = 20, $n\hat{q}$ (2000) = 23, $n\hat{p}$ (2010) = 26, and $n\hat{q}$ (2010) = 22 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will perform a two-proportion *z*-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by $SE_{\text{pooled}}\left(\hat{p}_{2000} - \hat{p}_{2010}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2000}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2010}}} = \sqrt{\frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{43} + \frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{48}} \approx 0.1050.$

(You could have chosen to use the unpooled standard error as well. It won't affect your conclusion.)

R6.5. (continued)

The observed difference between the proportions is 0.4651 – 0.5417 = – 0.0766.

$z = \dfrac{(\hat{p}_{2000} - \hat{p}_{2010}) - (0)}{SE(\hat{p}_{2000} - \hat{p}_{2010})} = \dfrac{-0.0766 - 0}{0.1050} \approx -0.73$; Since the P-value = 0.2329 is high, we fail to reject the null hypothesis. There is no evidence of an increase in the proportion of preterm twin births from 2000 to 2010, at least not at this large city hospital.

R6.6. Eclampsia.

a) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 1201, $n\hat{q}$ (mag. sulf.) = 3798, $n\hat{p}$ (placebo) = 228, and $n\hat{q}$ (placebo) = 4765 are all greater than 10, so both samples are large enough.

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

$$\left(\hat{p}_{MS} - \hat{p}_N\right) \pm z^* \sqrt{\frac{\hat{p}_{MS}\hat{q}_{MS}}{n_{MS}} + \frac{\hat{p}_N\hat{q}_N}{n_N}} = \left(\tfrac{1201}{4999} - \tfrac{228}{4993}\right) \pm 1.960 \sqrt{\frac{\left(\frac{1201}{4999}\right)\left(\frac{3798}{4999}\right)}{4999} + \frac{\left(\frac{228}{4993}\right)\left(\frac{4765}{4993}\right)}{4993}} = (0.181,\ 0.208)$$

We are 95% confident that the proportion of pregnant women who will experience side effects while taking magnesium sulfide will be between 18.1% and 20.8% higher than the proportion of women that will experience side effects while not taking magnesium sulfide.

b) H_0: The proportion of pregnant women who will develop eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide.
$(p_{MS} = p_N$ or $p_{MS} - p_N = 0)$
H_A: The proportion of pregnant women who will develop eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide. $(p_{MS} < p_N$ or $p_{MS} - p_N < 0)$

Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 40, $n\hat{q}$ (mag. sulf.) = 4959, $n\hat{p}$ (placebo) = 96, and $n\hat{q}$ (placebo) = 4897 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied (some in part (a)), we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{MS} - \hat{p}_N\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}} = \sqrt{\frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4999} + \frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4993}} \approx 0.002318.$$

(We could also use the unpooled standard error.)

The observed difference between the proportions is 0.00800 – 0.01923 = – 0.01123, which is approximately 4.84 standard errors below the expected difference in proportion of 0.

$z = \dfrac{(\hat{p}_{MS} - \hat{p}_N) - (0)}{SE(\hat{p}_{MS} - \hat{p}_N)} = \dfrac{-0.01123}{0.002318} \approx -4.84$; Since the P-value = 6.4×10^{-7} is very low, we reject the null hypothesis. There is strong evidence that the proportion of pregnant women who develop eclampsia will be lower for women taking magnesium sulfide than for those not taking magnesium sulfide.

R6.7. Eclampsia deaths.

a) H_0: The proportion of pregnant women who die after developing eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide.
$(p_{MS} = p_N \text{ or } p_{MS} - p_N = 0)$
H_A: The proportion of pregnant women who die after developing eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide.
$(p_{MS} < p_N \text{ or } p_{MS} - p_N < 0)$

b) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 11, $n\hat{q}$ (mag. sulf.) = 29,
$n\hat{p}$ (placebo) = 20, and $n\hat{q}$ (placebo) = 76 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}(\hat{p}_{MS} - \hat{p}_N) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}} = \sqrt{\frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{40} + \frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{96}} \approx 0.07895.$$

(We could also use the unpooled standard error.)

c) The observed difference between the proportions is 0.275 – 0.2083 = 0.0667.

$z = \dfrac{(\hat{p}_{MS} - \hat{p}_N) - (0)}{SE(\hat{p}_{MS} - \hat{p}_N)} = \dfrac{0.0667 - 0}{0.07895} \approx 0.84$; Since the P-value = 0.8008 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of women who may die after developing eclampsia is lower for women taking magnesium sulfide than for women who are not taking the drug.

d) There is not sufficient evidence to conclude that magnesium sulfide is effective in preventing death when eclampsia develops.

e) If magnesium sulfide is effective in preventing death when eclampsia develops, then we have made a Type II error.

f) To increase the power of the test to detect a decrease in death rate due to magnesium sulfide, we could increase the sample size or increase the level of significance.

g) Increasing the sample size lowers variation in the sampling distribution, but may be costly. The sample size is already quite large. Increasing the level of significance increases power by increasing the likelihood of rejecting the null hypothesis, but increases the chance of making a Type I error, namely thinking that magnesium sulfide is effective when it is not.

R6.8. Perfect pitch.

a) H_0: The proportion of Asian students with perfect pitch is the same as the proportion of non-Asians with perfect pitch. $(p_A = p_N \text{ or } p_A - p_N = 0)$
H_A: The proportion of Asian students with perfect pitch is the different than the proportion of non-Asians with perfect pitch. $(p_A \neq p_N \text{ or } p_A - p_N \neq 0)$

b) Since P-value < 0.0001, which is very low, we reject the null hypothesis. There is strong evidence of a difference in the proportion of Asians with perfect pitch and the proportion of non-Asians with perfect pitch. There is evidence that Asians are more likely to have perfect pitch.

c) If there is no difference in the proportion of students with perfect pitch, we would expect the observed difference of 25% to be seen simply due to sampling variation in less than 1 out of every 10,000 samples of 2700 students.

R6.8. (continued)

d) The data do not prove anything about genetic differences causing differences in perfect pitch. Asians are merely more likely to have perfect pitch. There may be lurking variables other than genetics that cause the higher rate of perfect pitch.

R6.9. More errors.

a) Since a treatment (the additive) is imposed, this is an experiment.

b) The company is only interested in an increase in fuel economy, so they will perform a one-sided test.

c) The company will make a Type I error if they decide that the additive increases the fuel economy, when it actually makes no difference in the fuel economy.

d) The company will make a Type II error if they decide that the additive does not increase the fuel economy, when it actually increases fuel economy.

e) The additive manufacturer would consider a Type II error more serious. If the test caused the company to conclude that the manufacturer's product didn't work, and it actually did, the manufacturer would lose sales, and the reputation of their product would suffer.

f) Since this was a controlled experiment, the company can conclude that the additive is the reason that the fuel economy has increased. They should be cautious recommending it for all cars. There is evidence that the additive works well for fleet vehicles, which get heavy use. It might not be effective in cars with a different pattern of use than fleet vehicles.

R6.10. Preemies.

a) **Randomization condition:** Assume that these kids are representative of all kids.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$ (preemies) = (242)(0.74) = 179, $n\hat{q}$ (preemies) = (242)(0.26) = 63, $n\hat{p}$ (normal weight) = (233)(0.83) = 193, and $n\hat{q}$ (normal weight) = 40 are all greater than 10, so the samples are both large enough.

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

$$\left(\hat{p}_N - \hat{p}_P\right) \pm z^* \sqrt{\frac{\hat{p}_N \hat{q}_N}{n_N} + \frac{\hat{p}_P \hat{q}_P}{n_P}} = \left(0.83 - 0.74\right) \pm 1.960 \sqrt{\frac{(0.83)(0.17)}{233} + \frac{(0.74)(0.26)}{242}} = \left(0.017,\ 0.163\right)$$

We are 95% confident that between 1.7% and 16.3% more normal birth-weight children graduated from high school than children who were born premature.

b) Since the interval for the difference in percentage of high school graduates is above 0, there is evidence normal birth-weight children graduate from high school at a greater rate than premature children.

c) If preemies do not have a lower high school graduation rate than normal birth-weight children, then we made a Type I error. We rejected the null hypothesis of "no difference" when we shouldn't have.

R6.11. Crawling.

a) H_0: The mean age at which babies begin to crawl is the same whether the babies were born in January or July. $\left(\mu_{Jan} = \mu_{July} \text{ or } \mu_{Jan} - \mu_{July} = 0\right)$

H_A: There is a difference in the mean age at which babies begin to crawl, depending on whether the babies were born in January or July. $\left(\mu_{Jan} \neq \mu_{July} \text{ or } \mu_{Jan} - \mu_{July} \neq 0\right)$

Independent groups assumption: The groups of January and July babies are independent.
Randomization condition: Although not specifically stated, we will assume that the babies are representative of all babies.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are fairly large. The Central Limit Theorem allows us to proceed.

R6.11. (continued)

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 43.68 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

$$SE(\bar{y}_{Jan} - \bar{y}_{July}) = \sqrt{\frac{7.08^2}{32} + \frac{6.91^2}{21}} \approx 1.9596.$$

The observed difference between the mean ages is 29.84 – 33.64 = – 3.8 weeks.

$t = \dfrac{(\bar{y}_{Jan} - \bar{y}_{July}) - (0)}{SE(\bar{y}_{Jan} - \bar{y}_{July})} = \dfrac{-3.8 - 0}{1.9596} \approx -1.94$; Since the P-value = 0.0590 is fairly low, we reject the null hypothesis. There is some evidence that mean age at which babies crawl is different for January and July babies. July babies appear to crawl a bit earlier than July babies, on average. Since the evidence is not strong, we might want to do some more research into this claim.

b) H_0: The mean age at which babies begin to crawl is the same whether the babies were born in April or October. $(\mu_{Apr} = \mu_{Oct}$ or $\mu_{Apr} - \mu_{Oct} = 0)$

H_A: There is a difference in the mean age at which babies begin to crawl, depending on whether the babies were born in April or October. $(\mu_{Apr} \neq \mu_{Oct}$ or $\mu_{Apr} - \mu_{Oct} \neq 0)$

The conditions (with minor variations) were checked in part (a).

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 59.40 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

$$SE(\bar{y}_{Apr} - \bar{y}_{Oct}) = \sqrt{\frac{6.21^2}{26} + \frac{7.29^2}{44}} \approx 1.6404.$$

The observed difference between the mean ages is 31.84 – 33.35 = – 1.51 weeks.

$t = \dfrac{(\bar{y}_{Apr} - \bar{y}_{Oct}) - (0)}{SE(\bar{y}_{Apr} - \bar{y}_{Oct})} = \dfrac{-1.51 - 0}{1.6404} \approx -0.92$; Since the P-value = 0.3610 is high, we fail to reject the null hypothesis. There is no evidence that mean age at which babies crawl is different for April and October babies.

c) These results are not consistent with the researcher's claim. We have slight evidence in one test and no evidence in the other. The researcher will have to do better than this to convince us!

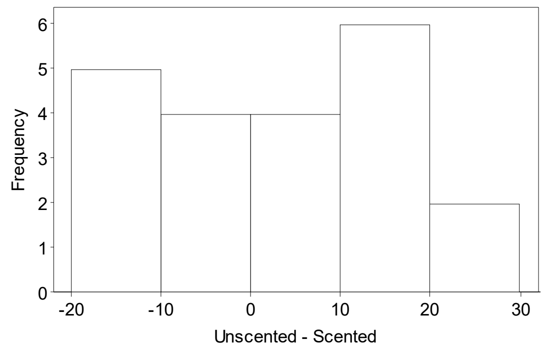
R6.12. Mazes and smells.

H_0: The mean difference in maze times with and without the presence of a floral aroma is zero. $(\mu_d = 0)$

H_A: The mean difference in maze times with and without the presence of a floral aroma (unscented – scented) is greater than zero. $(\mu_d > 0)$

Paired data assumption: Each subject is paired with himself or herself.
Randomization condition: Subjects were randomized with respect to whether they did the scented trial first or second.
Nearly Normal condition: The histogram of differences between unscented and scented scores shows a distribution that could have come from a Normal population, and the sample size is fairly large.

R6.12. (continued)



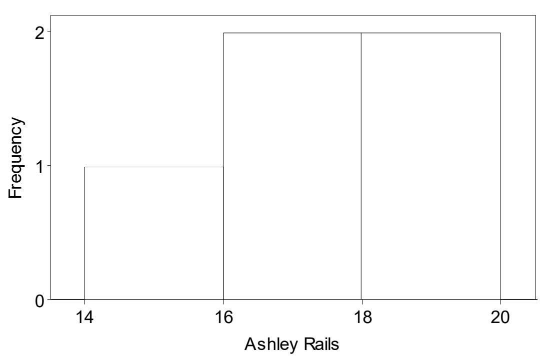
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 21 – 1 = 20 degrees of freedom, $t_{20}\left(0, \frac{13.0087}{\sqrt{21}}\right)$.

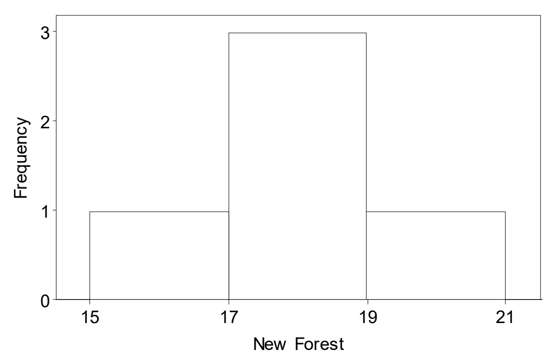
We will use a paired *t*-test, (unscented – scented) with $\bar{d} = 3.85238$ seconds.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} \approx \frac{3.85238 - 0}{13.0087 / \sqrt{21}} \approx 1.36$; Since the P-value = 0.0949 is fairly high, we fail to reject the null hypothesis. There is little evidence that the mean difference in time required to complete the maze is greater than zero. The floral scent didn't appear to cause lower times.

R6.13. Pottery.

Independent groups assumption: The pottery samples are from two different sites.
Randomization condition: It is reasonable to think that the pottery samples are representative of all pottery at that site with respect to aluminum oxide content.
Nearly Normal condition: The histograms of aluminum oxide content 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 7 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{AR} - \bar{y}_{NF}) \pm t_{df}^{*}\sqrt{\frac{s_{AR}^2}{n_{AR}} + \frac{s_{NF}^2}{n_{NF}}} = (17.32 - 18.18) \pm t_7^{*}\sqrt{\frac{1.65892^2}{5} + \frac{1.77539^2}{5}} \approx (-3.37, 1.65)$$

We are 95% confident that the difference in the mean percentage of aluminum oxide content of the pottery at the two sites is between –3.37% and 1.65%. Since 0 is in the interval, there is no evidence that the aluminum oxide content at the two sites is different. It would be reasonable for the archaeologists to think that the same ancient people inhabited the sites.

R6.14. Grant writing.

a) H_0: The proportion of NIH grants accepted is the same for white applicants and black applicants.
$\left(p_{White} = p_{Black} \text{ or } p_{White} - p_{Black} = 0\right)$
H_A: The proportion of NIH grants accepted is the greater for white applicants than for black applicants.
$\left(p_{White} > p_{Black} \text{ or } p_{White} - p_{Black} > 0\right)$

Independence assumption: This is not a random sample of applicants, but is more likely to be all applications. We will assume that the applications are independent of each other.
Independent samples condition: We will assume the groups are independent.
Success/Failure cond.: $n\hat{p}$ (white) = 15,700, $n\hat{q}$ (white) = 42,448, $n\hat{p}$ (black) = 198, and $n\hat{q}$ (black) = 966 are all greater than 10, so the samples are both large enough.

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

$$SE\left(\hat{p}_{White} - \hat{p}_{Black}\right) = \sqrt{\frac{\hat{p}_W\hat{q}_W}{n_{White}} + \frac{\hat{p}_B\hat{q}_B}{n_{Black}}} = \sqrt{\frac{\left(\frac{15,700}{58,148}\right)\left(\frac{42,448}{58,148}\right)}{58,148} + \frac{\left(\frac{198}{1164}\right)\left(\frac{966}{1164}\right)}{1164}} \approx 0.01165483.$$

The observed difference between the proportions is 0.2700 – 0.1701 = 0.0998976.

$$z = \frac{\left(\hat{p}_W - \hat{p}_B\right) - 0}{SE\left(\hat{p}_W - \hat{p}_B\right)} = \frac{0.0998976 - 0}{0.011165483} \approx 8.95;$$ Since the P-value < 0.0001 is very low, we reject the null hypothesis. There is strong evidence that the proportion of white applicants who are receiving NIH grants is higher than the proportion of black applicants receiving grants.

b) This was a retrospective observational study. In order to make the inference, we must assume that the researchers used in the study are representative of all researchers would could possibly apply for funding from NIH.

R6.15. Feeding fish.

a) If there is no difference in the average fish sizes, the chance of observing a difference this large, or larger, just by natural sampling variation is 0.1%.

b) There is evidence that largemouth bass that are fed a natural diet are larger. The researchers would advise people who raise largemouth bass to feed them a natural diet.

c) If the advice is incorrect, the researchers have committed a Type I error.

R6.16. Seat belts 2015.

We have two independent samples, but the sample sizes are very small and it would be hard to generalize the results for New England and the Mountain states based on summary numbers from each state. We also don’t know which states the figures are coming from, and some states are much larger than others in population, geographic area, number of drivers, and so on, yet equally weighted in these data. These values are not appropriate for inference.

If you choose to go ahead with the interval anyway, the interval is (–4.24, 12.08). We are 95% confident that the percentage of seat belt users in the Mountain states is between 4.24 percentage points lower to 12.08 percentage points higher than the percentage of seat belt users in New England. Since the interval contains zero, there is no evidence of a difference in the percentage of seat belt users between the two regions.

R6.17. Age.

a) **Independent groups assumption:** The group of patients with and without cardiac disease are not related in any way.
Randomization condition: Assume that these patients are representative of all people.
Normal population assumption: We don't have the actual data, so we will assume that the population of ages of patients is 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 670 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{Card} - \bar{y}_{None}) \pm t_{df}^{*}\sqrt{\frac{s_{Card}^2}{n_{Card}} + \frac{s_{None}^2}{n_{None}}} = (74.0 - 69.8) \pm t_{670}^{*}\sqrt{\frac{7.9^2}{450} + \frac{8.7^2}{2397}} \approx (3.39,\ 5.01)$$

We are 95% confident that the mean age of patients with cardiac disease is between 3.39 and 5.01 years higher than the mean age of patients without cardiac disease.

b) Older patients are at greater risk for a variety of health problems. If an older patient does not survive a heart attack, the researchers will not know to what extent depression was involved, because there will be a variety of other possible variables influencing the death rate. Additionally, older patients may be more (or less) likely to be depressed than younger ones.

R6.18. Smoking.

a) **Randomization condition:** Assume the patients are representative of all people.
Independent groups assumption: The group of patients with and without cardiac disease are not related in any way.
Success/Failure condition: $n\hat{p}$ (cardiac) = (450)(0.32) = 144, $n\hat{q}$ (cardiac) = (450)(0.68) = 306, $n\hat{p}$ (none) = (2397)(0.237) = 568, and $n\hat{q}$ (none) = (2397)(0.763) = 1829 are all greater than 10, so the samples are both large enough.

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

$$\left(\hat{p}_{Card} - \hat{p}_{None}\right) \pm z^{*}\sqrt{\frac{\hat{p}_{Card}\hat{q}_{Card}}{n_{Card}} + \frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}} = (0.32 - 0.237) \pm 1.960\sqrt{\frac{(0.32)(0.68)}{450} + \frac{(0.237)(0.763)}{2397}}$$
$$= (0.0367,\ 0.1293)$$

We are 95% confident that the proportion of smokers is between 3.67 percentage points and 12.93 percentage points higher for patients with cardiac disease than for patients without cardiac disease.

b) Since the confidence interval does not contain 0, there is evidence that cardiac patients have a higher rate of smokers than the patients without cardiac disease. The two groups are different.

c) Smoking could be a confounding variable. Smokers have a higher risk of other health problems that may be associated with their ability to survive a heart attack.

R6.19. Eating disorders.

a) **Randomization condition:** Hopefully, the students were selected randomly.
10% condition: 150 and 200 are less than 10% of all students.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$ (Muslim) = 46, $n\hat{q}$ (Muslim) = 104, $n\hat{p}$ (Christian) = 34, and $n\hat{q}$ (Christian) = 166 are all greater than 10, so the samples are both large enough.

R6.19. (continued)

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

$$(\hat{p}_M - \hat{p}_C) \pm z^* \sqrt{\frac{\hat{p}_M \hat{q}_M}{n_M} + \frac{\hat{p}_C \hat{q}_C}{n_C}} = \left(\frac{46}{150} - \frac{34}{200}\right) \pm 1.960 \sqrt{\frac{\left(\frac{46}{150}\right)\left(\frac{104}{150}\right)}{150} + \frac{\left(\frac{34}{200}\right)\left(\frac{166}{200}\right)}{200}} = (0.046,\ 0.227)$$

We are 95% confident that the percentage of Muslim students who have an eating disorder is between 4.6 and 22.7 percentage points higher than the percentage of Christian students who have an eating disorder.

b) Although caution in generalizing must be used since the study was restricted to the Spanish city of Ceuta, it appears there is a true difference in the prevalence of eating disorders. We can conclude this because the entire interval is above 0.

R6.20. Cesareans.

H_0: The proportion of births involving cesarean deliveries is the same in Vermont and New Hampshire. $(p_{VT} = p_{NH} \text{ or } p_{VT} - p_{NH} = 0)$

H_A: The proportion of births involving cesarean deliveries is different in Vermont and New Hampshire. $(p_{VT} \neq p_{NH} \text{ or } p_{VT} - p_{NH} \neq 0)$

Random condition: Hospitals were randomly selected.
10% condition: 223 and 186 are both less than 10% of all births in these states.
Independent samples condition: Vermont and New Hampshire are different states!
Success/Failure cond.: $n\hat{p}$ (VT) = (223)(0.166) = 37, $n\hat{q}$ (VT) = (223)(0.834) = 186,
$n\hat{p}$ (NH) = (186)(0.188) = 35, and $n\hat{q}$ (NH) = (186)(0.812) = 151 are all greater than 10, so the samples are both large enough.

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

$$SE_{\text{pooled}}(\hat{p}_{VT} - \hat{p}_{NH}) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{VT}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{NH}}} = \sqrt{\frac{(0.176)(0.824)}{223} + \frac{(0.176)(0.824)}{186}} \approx 0.03782.$$

(We could also use the unpooled estimate of the standard error.)

The observed difference between the proportions is 0.166 – 0.188 = – 0.022.

$$z = \frac{(\hat{p}_{VT} - \hat{p}_{NH}) - (0)}{SE(\hat{p}_{VT} - \hat{p}_{NH})} = \frac{-0.22 - 0}{0.03782} \approx -0.59;$$ Since the P-value = 0.5563 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of cesarean births in Vermont is different from the proportion of cesarean births in New Hampshire.

R6.21. Teach for America.

H_0: The mean score of students with certified teachers is the same as the mean score of students with uncertified teachers. $(\mu_C = \mu_U \text{ or } \mu_C - \mu_U = 0)$

H_A: The mean score of students with certified teachers is greater than as the mean score of students with uncertified teachers. $(\mu_C > \mu_U \text{ or } \mu_C - \mu_U > 0)$

Independent groups assumption: The certified and uncertified teachers are independent groups.
Randomization condition: Assume the students studied were representative of all students.
Nearly Normal condition: We don't have the actual data, so we can't look at the graphical displays, but the sample sizes are large, so we can 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 86 degrees of freedom (from the approximation formula).

R6.21. (continued)

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error

$$SE(\bar{y}_C - \bar{y}_U) = \sqrt{\frac{9.31^2}{44} + \frac{9.43^2}{44}} \approx 1.9977.$$

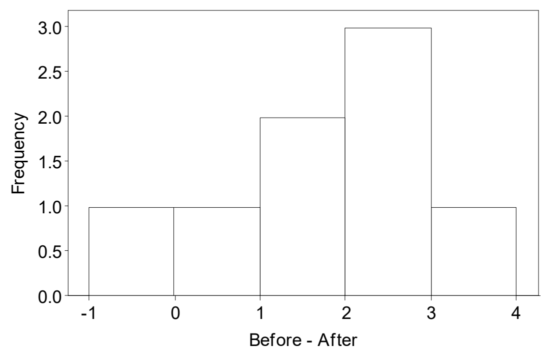
The observed difference between the mean scores is 35.62 – 32.48 = 3.14.

$t = \dfrac{(\bar{y}_C - \bar{y}_U) - (0)}{SE(\bar{y}_C - \bar{y}_U)} = \dfrac{3.14 - 0}{1.9977} \approx 1.57$; Since the P-value = 0.0598 is fairly low, we reject the null hypothesis.

There is some evidence that students with certified teachers had mean scores higher than students with uncertified teachers. However, since the P-value is not extremely low, further investigation is recommended.

R6.22. Legionnaires' disease.

a) **Paired data assumption:** The data are paired by room.
Randomization condition: We will assume that these rooms are representative of all rooms at the hotel.
Nearly Normal condition: The histogram of differences between before and after measurements 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 8 – 1 = 7 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 1.6125 \pm t^*_7\left(\frac{1.23801}{\sqrt{8}}\right) \approx (0.58,\ 2.65)$$

We are 95% confident that the mean difference in the bacteria counts is between 0.58 and 2.65 colonies per cubic foot of air. Since the entire interval is above 0, there is evidence that the new air-conditioning system was effective in reducing average bacteria counts.

b) Answers will vary with each bootstrap attempt.

R6.23. Teach for America, part II.

H_0: The mean score of students with certified teachers is the same as the mean score of students with uncertified teachers. $(\mu_C = \mu_U \text{ or } \mu_C - \mu_U = 0)$

H_A: The mean score of students with certified teachers is different than the mean score of students with uncertified teachers. $(\mu_C \neq \mu_U \text{ or } \mu_C - \mu_U \neq 0)$

Mathematics: Since the P-value = 0.002 is low, we reject the null hypothesis. There is strong evidence that students with certified teachers have different mean math scores than students with uncertified teachers. Students with certified teachers do better.

R6.23. (continued)

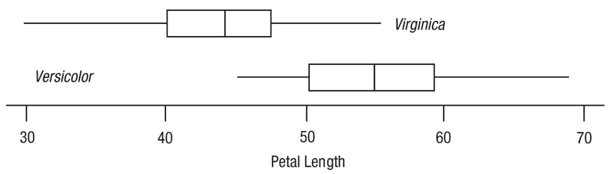
Language: Since the P-value = 0.045 is fairly low, we reject the null hypothesis. There is evidence that students with certified teachers have different mean language scores than students with uncertified teachers. Students with certified teachers do better. However, since the P-value is not extremely low, further investigation is recommended.

R6.24. Fisher's irises.

a) Parallel boxplots of the distributions of petal lengths for the two species of flower are at the right. No units are specified, but millimeters seem like a reasonable guess.

b) The petals of *versicolor* are generally longer than the petals of *virginica.* Both distributions have about the same range, and both distributions are fairly symmetric.

c) **Independent groups assumption:** The two species of flowers are independent.
Randomization condition: It is reasonable to assume that these flowers are representative of their species.
Nearly Normal condition: The boxplots show distributions of petal lengths that are reasonably symmetric with no outliers. Additionally, 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 97.92 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{Ver} - \bar{y}_{Vir}) \pm t^*_{df}\sqrt{\frac{s^2_{Ver}}{n_{Ver}} + \frac{s^2_{Vir}}{n_{Vir}}} = (55.52 - 43.22) \pm t^*_{97.92}\sqrt{\frac{5.519^2}{50} + \frac{5.362^2}{50}} \approx (10.14, 14.46)$$

d) We are 95% confident the mean petal length of *versicolor* irises is between 10.14 and 14.46 millimeters longer than the mean petal length of *virginica* irises.

e) Since the interval is completely above 0, there is strong evidence that the mean petal length of *versicolor* irises is greater than the mean petal length of *virginica* irises.

f) Answers will vary with each bootstrap attempt.

R6.25. Insulin and diet.

a) H_0: People with high dairy consumption have IRS at the same rate as those with low dairy consumption. $\left(p_{High} = p_{Low} \text{ or } p_{High} - p_{Low} = 0\right)$
H_A: People with high dairy consumption have IRS at a different rate than those with low dairy consumption. $\left(p_{High} \neq p_{Low} \text{ or } p_{High} - p_{Low} \neq 0\right)$

Random condition: Assume the people studied are representative of all people.
10% condition: 102 and 190 are both less than 10% of all people.
Independent samples condition: The two groups are not related.
Success/Failure condition: $n\hat{p}$ (high) = 24, $n\hat{q}$ (high) = 78, $n\hat{p}$ (low) = 85, and $n\hat{q}$ (low) = 105 are all greater than 10, so the samples are both large enough.

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

$$SE_{\text{pooled}}\left(\hat{p}_{High} - \hat{p}_{Low}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{High}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Low}}} = \sqrt{\frac{(0.373)(0.627)}{102} + \frac{(0.373)(0.627)}{190}} \approx 0.05936.$$

(We could also use the unpooled standard error.)

R6.25. (continued)

The observed difference between the proportions is 0.2352 – 0.4474 = – 0.2122.

$z = \frac{(\hat{p}_{High} - \hat{p}_{Low}) - (0)}{SE(\hat{p}_{High} - \hat{p}_{Low})} = \frac{-0.2122 - 0}{0.05936} \approx -3.57$; Since the P-value = 0.0004 is very low, we reject the null hypothesis. There is strong evidence that the proportion of people with IRS is different for those who with high dairy consumption compared to those with low dairy consumption. People who consume dairy products more than 35 times per week appear less likely to have IRS than those who consume dairy products fewer than 10 times per week.

b) There is evidence of an association between the low consumption of dairy products and IRS, but that does not prove that dairy consumption influences the development of IRS. This is an observational study, and a controlled experiment is required to prove cause and effect.

R6.26. Cloud seeding.

a) **Independent groups assumption:** The two groups of clouds are independent.
Randomization condition: Researchers randomly assigned clouds to be seeded with silver iodide or not seeded.
Nearly Normal condition: If we deal with only the summary data, we can't look at the distributions, but the means of group are significantly higher than the medians. This is an indication that the distributions are skewed to the right, with possible outliers. The samples sizes of 26 each are fairly large, but we should be careful making conclusions, since there may be outliers. If you look at the data in the provided computer file, you will see highly skewed distributions, which is addressed in a later exercise!

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 33.86 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_S - \bar{y}_U) \pm t^*_{df}\sqrt{\frac{s_S^2}{n_S} + \frac{s_U^2}{n_U}} = (441.985 - 164.588) \pm t^*_{33.86}\sqrt{\frac{650.787^2}{26} + \frac{278.426^2}{26}} \approx (-4.76,\ 559.56)$$

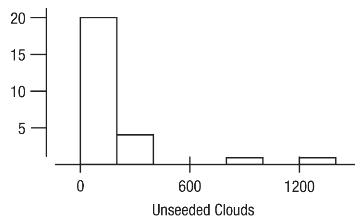
We are 95% confident the mean amount of rainfall produced by seeded clouds is between 4.76 acre-feet less than and 559.56 acre-feet more than the mean amount of rainfall produced by unseeded clouds.

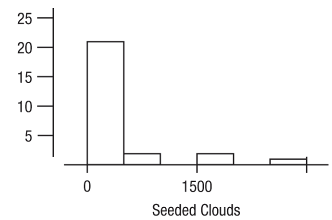
Since the interval contains 0, there is little evidence that the mean rainfall produced by seeded clouds is any different from the mean rainfall produced by unseeded clouds. However, we shouldn't place too much faith in this conclusion. It is based on a procedure that is sensitive to outliers, and there may have been outliers present.

b) Answers will vary. You should find that the bootstrap interval based on shuffling the categories doesn't agree very well with the two-sample *t*-interval constructed in part (a). For example, a bootstrap interval of 1000 randomized differences was (36.38, 548.93). We are 95% confident the mean amount of rainfall produced by seeded clouds is between 36.38 acre-feet and 548.93 acre-feet more than the mean amount of rainfall produced by unseeded clouds. This seems very high. Recall that bootstrapping is sensitive to outliers, just like our *t*-interval. We will explore this in the next exercise.

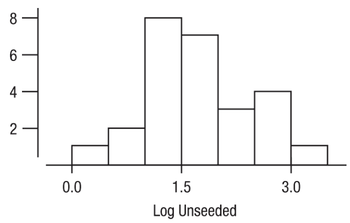
R6.27. Cloud seeding re-expressed.

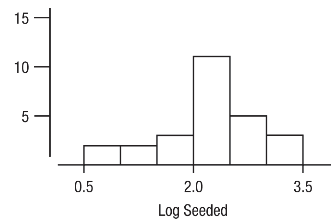
a) Histograms of the cloud seeding data appear below. Both distributions are skewed to the right. Most clouds produced a moderate amount of rain, but some clouds, unseeded and seeded alike, produced much large amounts of rain. The two-sample *t*-interval constructed above is not appropriate, since the Nearly Normal Condition is not met.





b) Re-expressing the cloud seeding data with logarithms makes each distribution much more unimodal and symmetric.





c) Using logarithms to re-express the data allows the Nearly Normal Condition to be 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 49.97 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\overline{\log(y_S)} - \overline{\log(y_U)}) \pm t_{df}^{*}\sqrt{\frac{\left(SD\left(\log(y_S)\right)\right)^2}{n_S} + \frac{\left(SD\left(\log(y_U)\right)\right)^2}{n_U}}$$

$$= (2.229749 - 1.7330111) \pm t_{49.97}^{*}\sqrt{\frac{0.69465993^2}{26} + \frac{0.7130453^2}{26}}$$

$$\approx \left(0.10459974,\ 0.88887595\right)$$

We should certainly let technology do our calculating for us. Since we are using common logs, this is an interval of exponents of 10. We are carrying a large number of decimal places in our calculations to make the formula provide the correct outcomes, but a computer handles this easily.

Lower limit: $10^{0.10459974} \approx 1.27$ Upper Limit: $10^{0.88887595} \approx 7.74$

We are 95% confident the mean amount of rainfall produced by seeded clouds is between 1.27 acre-feet and 7.74 acre-feet more than the mean amount of rainfall produced by unseeded clouds.

As usual, answers will vary for the bootstrap interval. Based on the re-expressed data, our bootstrap interval for the difference in the logarithm of rainfall amount is (0.14346916, 0.87844334).

Lower limit: $10^{0.14346916} \approx 1.39$ Upper Limit: $10^{0.87844334} \approx 7.56$

We are 95% confident the mean amount of rainfall produced by seeded clouds is between 1.39 acre-feet and 7.56 acre-feet more than the mean amount of rainfall produced by unseeded clouds.

The intervals constructed from the re-expressed data are much more consistent with each other than the intervals constructed from the original data.

R6.28. Genetics.

H_0: The proportions of traits are as specified by the ratio 1:3:3:9.
H_A: The proportions of traits are not as specified.

Counted data condition: The data are counts.
Randomization condition: Assume that these students are representative of all people.

Expected cell frequency condition: The expected counts (shown in the table) are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 – 1 = 3 degrees of freedom. We will use a chi-square goodness-of-fit test.

Trait	Observed	Expected	Residual $(Obs - Exp)$	Residual^2 $(Obs - Exp)^2$	Component $(Obs - Exp)^2 / Exp$
Attached, noncurling	10	7.625	2.375	5.6406	0.73975
Attached, curling	22	22.875	– 0.875	0.7656	0.03347
Free, noncurling	31	22.875	8.125	66.0156	2.8859
Free, curling	59	68.625	– 9.625	92.6406	1.35

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp} \approx 5.01,\ \text{df} = 3;\ \text{P-value} = 0.1711$$

Since the P-value is high, we fail to reject the null hypothesis. There is no evidence that the proportions of traits are anything other than 1:3:3:9.

R6.29. Tableware.

a) Since there are 57 degrees of freedom, there were 59 different products included.

b) 84.5% of the variation in retail price is explained by the polishing time.

c) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (59 – 2) = 57 degrees of freedom. We will use a regression slope *t*-interval. For 95% confidence, use $t^*_{57} \approx 2.0025$, or estimate from the table $t^*_{50} \approx 2.009$.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = 2.49244 \pm (2.0025) \times 0.1416 \approx (2.21,\ 2.78)$$

d) We are 95% confident that the average price increases between \$2.21 and \$2.78 for each additional minute of polishing time.

R6.30. Hard water.

a) H_0: There is no linear relationship between calcium concentration in water and mortality rates for males. $(\beta_1 = 0)$

H_A: There is a linear relationship between calcium concentration in water and mortality rates for males. $(\beta_1 \neq 0)$

b) Assuming the conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (61 – 2) = 59 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Mortality} = 1676 - 3.23(Calcium)$, where mortality is measured in deaths per 100,000, and calcium concentration is measured in parts per million.

$t = \dfrac{b_1 - \beta_1}{SE(b_1)} = \dfrac{-3.23 - 0}{0.48} \approx -6.73$; The P-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between calcium concentration and mortality. Towns with higher calcium concentrations tend to have lower mortality rates.

c) For 95% confidence, use $t_{59}^* \approx 2.001$, or estimate from the table $t_{50}^* \approx 2.009$.

$$b_1 \pm t_{n-2}^* \times SE(b_1) = -3.23 \pm (2.001) \times 0.48 \approx (-4.19, -2.27)$$

d) We are 95% confident that the average mortality rate decreases by between 2.27 and 4.19 deaths per 100,000 for each additional part per million of calcium in drinking water.

R6.31. Wealth redistribution 2015.

H_0: Income level and feelings about wealth redistribution are independent.
H_A: There is an association between income level and feelings about wealth distribution.

Counted data condition: The data are counts.
Randomization condition: Although not specifically stated, the Gallup Poll was likely to be random.

	Should Redistribute (*Obs/Exp*)	Should Not (*Obs/Exp*)	No Opinion (*Obs/Exp*)
High Income	426 / 534.33	579 / 460.33	10 / 20.33
Middle Income	558 / 534.33	447 / 460.33	10 / 20.33
Low Income	619 / 534.33	355 / 460.33	41 / 20.33

Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{\text{all cells}} \frac{(Obs - Exp)^2}{Exp} \approx 123.02, \text{ df} = 4; \text{ P-value} < 0.0001$$

Since the P-value is low, we reject the null hypothesis. There is strong evidence of an association between income level and opinion on wealth redistribution. Examination of the components shows that the low-income respondents are more likely to approve of redistribution when compared to the high-income respondents.

R6.32. Wild horses.

a) Since there are 36 degrees of freedom, 38 herds of wild horses were studied.

b) **Straight enough condition:** The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The histogram of residuals is unimodal and symmetric.

c) Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (38 – 2) = 36 degrees of freedom. We will use a regression slope t-interval, with 95% confidence. Use $t_{35}^{*} \approx 2.030$ as an estimate.

$$b_1 \pm t_{n-2}^{*} \times SE(b_1) = 0.153969 \pm (2.030) \times 0.0114 \approx (0.131,\ 0.177)$$

d) We are 95% confident that the mean number of foals in a herd increases by between 0.131 and 0.177 foals for each additional adult horse.

e) The regression equation predicts that herds with 80 adults will have $-1.57835 + 0.153969(80) = 10.73917$ foals. The average size of the herds sampled is 110.237 adult horses. Use $t_{36}^{*} \approx 1.6883$, or use an estimate of $t_{35}^{*} \approx 1.690$, from the table.

$$\hat{y}_\nu \pm t_{n-2}^{*}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n} + s_e^2} = 10.73917 \pm (1.6883)\sqrt{0.0114^2 \cdot (80 - 110.237)^2 + \frac{4.941^2}{38} + 4.941^2}$$
$$\approx (2.26,\ 19.21)$$

We are 90% confident that number of foals in a herd of 80 adult horses will be between 2.26 and 19.21. This prediction interval is too wide to be of much use.

R6.33. Lefties and music.

H_0: The proportion of right-handed people who can match the tone is the same as the proportion of left-handed people who can match the tone. $(p_L = p_R$ or $p_L - p_R = 0)$

H_A: The proportion of right-handed people who can match the tone is different from the proportion of left-handed people who can match the tone. $(p_L \neq p_R$ or $p_L - p_R \neq 0)$

Random condition: Assume that the people tested are representative of all people.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (right) = 38, $n\hat{q}$ (right) = 38, $n\hat{p}$ (left) = 33, and $n\hat{q}$ (left) = 20 are all greater than 10, so the samples are both large enough.

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

$$SE_{\text{pooled}}(\hat{p}_L - \hat{p}_R) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_L} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_R}} = \sqrt{\frac{\left(\frac{71}{129}\right)\left(\frac{58}{129}\right)}{53} + \frac{\left(\frac{71}{129}\right)\left(\frac{58}{129}\right)}{76}} \approx 0.089.$$

The observed difference between the proportions is 0.6226 – 0.5 = 0.1226.

$z = \dfrac{(\hat{p}_L - \hat{p}_R) - (0)}{SE(\hat{p}_L - \hat{p}_R)} = \dfrac{0.1226 - 0}{0.089} \approx 1.38$; Since the P-value = 0.1683 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of people able to match the tone differs between right-handed and left-handed people.

R6.34. AP Statistics scores 2016.

a) H_0: The distribution of AP Statistics scores at Ithaca High School is the same as it is nationally.
H_A: The distribution of AP Statistics scores at Ithaca High School is different than it is nationally.

Counted data condition: The data are counts.
Randomization condition: Assume that this group of students is representative of all years at Ithaca High School.
Expected cell frequency condition: The expected counts (shown in the table) are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 5 – 1 = 4 degrees of freedom. We will use a chi-square goodness-of-fit test.

Score	Observed	Expected	Residual $(Obs - Exp)$	Standardized Residual $(Obs - Exp)/\sqrt{Exp}$	Component $(Obs - Exp)^2/Exp$
5	27	9.591	17.409	5.6214	31.5998
4	18	14.973	3.027	0.7822	0.6120
3	10	17.043	–7.043	–1.7060	2.9105
2	11	10.833	0.167	0.0507	0.0026
1	3	16.56	–13.56	–3.3322	11.1035

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp} \approx 46.23,\ \text{df} = 3;\ \text{P-value} < 0.0001$$

Since the P-value is small, we reject the null hypothesis. There is strong evidence that the distribution of scores at Ithaca High School is different than the national distribution. Students at IHS get fewer scores of 1 than expected, and more scores of 5 than expected.

R6.35. Twins births.

H_0: There is no association between duration of pregnancy and level of prenatal care.
H_A: There is an association between duration of pregnancy and level of prenatal care.

Counted data condition: The data are counts.
Randomization condition: Assume that these pregnancies are representative of all twin births.
Expected cell frequency condition: The expected counts are all greater than 5.

	Preterm (induced or Cesarean) (*Obs / Exp*)	Preterm (without procedures) (*Obs / Exp*)	Term or postterm (*Obs / Exp*)
Intensive	18 /16.676	15 / 15.579	28 / 28.745
Adequate	46 / 42.101	43 / 39.331	65 / 72.568
Inadequate	12 / 17.223	13 / 16.090	38 / 29.687

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp} \approx 6.14,\ \text{df} = 4;\ \text{P-value} \approx 0.1887$$

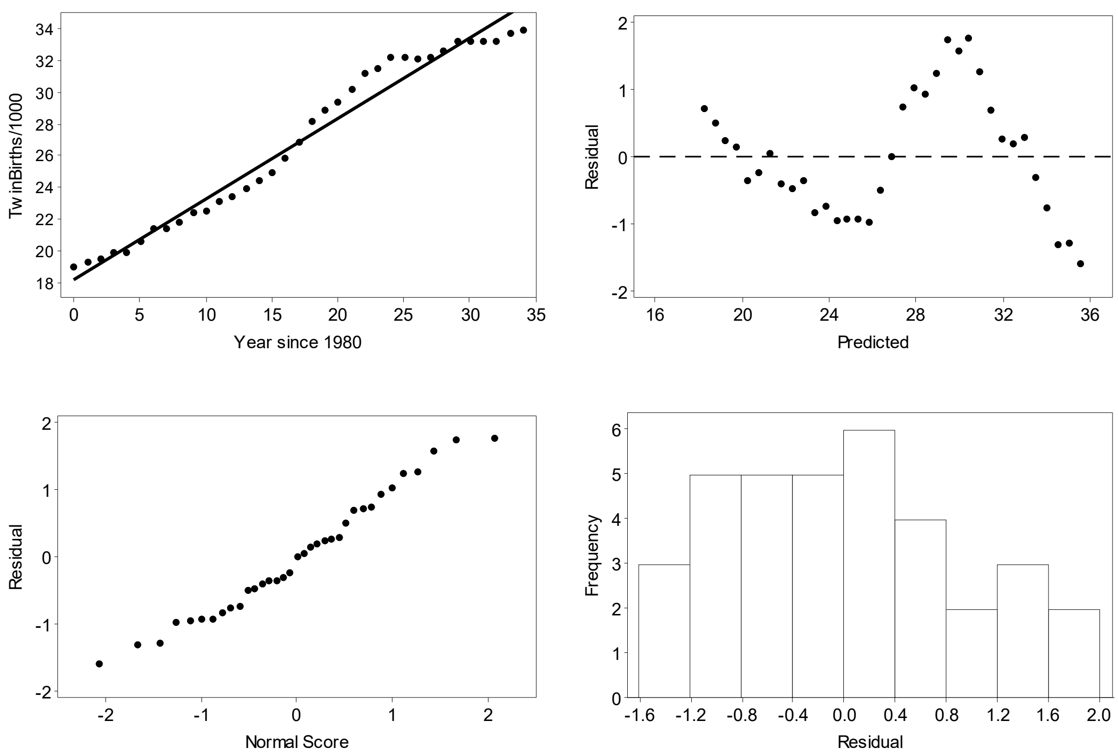
Since the P-value ≈ 0.1887 is high, we fail to reject the null hypothesis. There is no evidence of an association between duration of pregnancy and level of prenatal care in twin births.

R6.36. Twins by year.

a) H_0: There is no linear association between the year and the twin birth rate. $(\beta_1 = 0)$

H_A: There is a linear association between the year and the twin birth rate. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot shows a curved pattern, but wouldn't benefit from re-expression.
Independence assumption: We are examining a relationship over time, so there is reason to be cautious, and we see a strong pattern in the residuals.
Does the plot thicken? condition: The residuals plot shows no obvious trends in the spread.
Nearly Normal condition, Outlier condition: The histogram of residuals is reasonably unimodal and symmetric, and shows no obvious skewness or outliers. The normal probability plot is somewhat curved, so we should be cautious when making claims about the linear association. However, the residuals are all quite small. It should be reasonable to use a linear model, as long as we don't extrapolate beyond the scope of the data.



Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 35 – 2 = 33 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{(TwinBirths/1000)} = 18.2127 + 0.508261(YearSince1980)$

The value of $t = 33.1$. The P-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a positive linear relationship between the twin birth rate and the number of years since 1980. There is evidence that the twin birth rate has been increasing over time.

b) No, simply because you can find an association between the variables does not mean there is a causal relationship. These are most likely two variables that are both increasing, but for very different reasons.

R6.37. Retirement planning.

H_0: The proportion of men who are "a lot behind schedule" in retirement planning is the same as the proportion of women. $(p_M = p_W \text{ or } p_M - p_W = 0)$
H_A: The proportion of men who are "a lot behind schedule" in retirement planning is lower than the proportion of women. $(p_M < p_W \text{ or } p_M - p_W < 0)$

Random condition: Assume the survey was conducted randomly.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (Men) = 267, $n\hat{q}$ (Men) = 455, $n\hat{p}$ (Women) = 301, and $n\hat{q}$ (Women) = 400 are all greater than 10, so the samples are both large enough.

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

$$SE_{\text{pooled}}(\hat{p}_M - \hat{p}_W) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_W}} = \sqrt{\frac{\left(\frac{568}{1423}\right)\left(\frac{855}{1423}\right)}{722} + \frac{\left(\frac{568}{1423}\right)\left(\frac{855}{1423}\right)}{701}} \approx 0.0260.$$

(We could also choose to use the unpooled standard error.)

The observed difference between the proportions is 0.3698 – 0.4294 = 0.0596.

$$z = \frac{(\hat{p}_M - \hat{p}_W) - (0)}{SE(\hat{p}_M - \hat{p}_W)} = \frac{-0.0596}{0.0260} \approx -2.29;$$ Since the P-value = 0.0109 is low, we reject the null hypothesis.

There is evidence that the proportion of women who will say they are "a lot behind schedule" for retirement planning is higher than the percentage of men that would say the same thing.

R6.38. Age and party 2016.

a) There is one sample, classified according to two different variables, so we will perform a chi-square test for independence.

b) H_0: There is no association between age and political party for white voters.
H_A: There is an association between age and political party for white voters.

Counted data condition: The data are counts.
Randomization condition: These data are from a representative phone survey.
Expected cell frequency condition: The expected counts are all greater than 5.

	Leaning Republican (*Obs / Exp*)	Leaning Democrat (*Obs / Exp*)
18 – 29	148 / 189.34	248 /206.66
30 – 49	330 / 347.6	397 / 379.4
50 – 64	375 /351.43	360 / 383.57
65 +	284 / 248.63	236 / 271.37

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{\text{all cells}} \frac{(Obs - Exp)^2}{Exp} \approx 31.68, \text{ df} = 3; \text{ P-value} < 0.0001$$

R6.38. (continued)

c) Since the P-value < 0.0001 is low, we reject the null hypothesis. There is strong evidence of an association between age and political party for white voters.

The table of standardized residuals is useful for the analysis of the differences. Looking for the largest standardized residuals, we can see there are more Democrats and fewer Republicans than we expect in the 18-29-year-old group, with standardized residuals of –3 and 2.88. There are also more Republicans and fewer Democrats among those 65 and older than we expect, with standardized residuals of 2.24 and –2.15.

	Standardized Residuals	
	Leaning Republican	Leaning Democrat
18 – 29	–3	2.88
30 – 49	–0.94	0.9
50 – 64	1.26	–1.2
65 +	2.24	–2.15

R6.39. Eye and hair color.

a) This is an attempt at linear regression. Regression inference is meaningless here, since eye and hair color are categorical variables.

b) This is an analysis based upon a chi-square test for independence.

H_0: Eye color and hair color are independent.
H_A: There is an association between eye color and hair color.

Since we have two categorical variables, this analysis seems appropriate. However, if you check the expected counts, you will find that 4 of them are less than 5. We would have to combine several cells in order to perform the analysis. (Always check the conditions!)

Since the value of chi-square is so high, it is likely that we would find an association between eye and hair color, even after the cells were combined. There are many cells of interest, but some of the most striking differences that would not be affected by cell combination involve people with fair hair. Blonds are likely to have blue eyes, and not likely to have brown eyes. Those with red hair are not likely to have brown eyes. Additionally, those with black hair are much more likely to have brown eyes than blue.

R6.40. Barbershop music.

a) With an R^2 of 90.9%, your friend is right about being able to predict singing scores.

b) H_0: When including the other predictor, performance does not change our ability to predict singing scores. $(\beta_P = 0)$

H_A: When including the other predictor, performance changes our ability to predict singing scores. $(\beta_P \neq 0)$

With $t = 8.13$, and 31 degrees of freedom, the P-value is less than 0.0001, so we reject the null hypothesis and conclude that *Performance* is useful in predicting singing scores.

c) Based on the multiple regression, both *Performance* and *Music* (even with a P-value equal to 0.0766) are useful in predicting singing scores. According to the residuals plot the spread is constant across the predicted values. The histogram of residuals is unimodal and symmetric. Based on the information provided we have met the Similar Variance and Nearly Normal conditions.

R6.41. Cereals and fiber.

a) For the simple regression: The model predicts that calories increase by about 1.26 calories per gram of carbohydrate.

For the multiple regression: After allowing for the effects of *fiber* in these cereals, the model predicts that calories increase at the rate of about 0.83 calories per gram of carbohydrate.

b) After accounting for the effect of *fiber*, the effect of *carbo* is no longer statistically significant.

c) I would like to see a scatterplot showing the relationship of *carbo* and *fiber*. I suspect that they may be highly correlated, which would explain why after accounting for the effects of *fiber*, there is little left for *carbo* to model. It would be good to know the R^2 value between the two variables.

R6.42. Family planning.

H_0: Unplanned pregnancies and education level are independent.
H_A: There is an association between unplanned pregnancies and education level.

Convert the table to counts, planned versus unplanned pregnancies.

	Education Level		
	< 3 Yr HS (*Obs/Exp*)	3 + Yr HS (*Obs/Exp*)	Some College (*Obs/Exp*)
Planned	391/341.12	337/350.93	102/137.95
Unplanned	200/249.88	271/257.07	137/101.05

Counted data condition: The percentages must be converted to counts.
Randomization condition: Assume that these women are representative of all women.
Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 2 degree of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp} \approx 40.72, \text{ df} = 2; \text{ P-value is essentially } 0$$

Since the P-value is low, we reject the null hypothesis. There is strong evidence of an association between unplanned pregnancies and education level. More educated women tend to have fewer unplanned pregnancies.

R6.43. Cereals with bran.

These three cereals are highly influential in the relationship between *carbo* and *fiber*, are creating a strong, negative collinearity between them. Without these points, there is little association between these variables.

R6.44. Old Faithful.

a) H_0: There is no linear relationship between duration of the eruption and interval until the next eruption. $(\beta_1 = 0)$

H_A: There is a linear relationship between duration of the eruption and interval until the next eruption. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The histogram of residuals is unimodal and symmetric.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 222 – 2 = 220 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Interval} = 33.9668 + 10.3582(Duration)$.

The value of $t \approx 27.1$. The P-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between duration and interval. Relatively long eruptions tend to be associated with relatively long intervals until the next eruption.

b) The regression equation predicts that an eruption with duration of 2 minutes will have an interval until the next eruption of $33.9668 + 10.3582(2) = 54.6832$ minutes. ($t_{220}^* \approx 1.9708$)

$$\hat{y}_\nu \pm t_{n-2}^* \sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n}} = 54.6832 \pm (1.9708)\sqrt{0.3822^2 \cdot (2 - 3.57613)^2 + \frac{6.159^2}{222}}$$
$$\approx (53.24,\ 56.12)$$

We are 95% confident that, after a 2-minute eruption, the mean length of time until the next eruption will be between 53.24 and 56.12 minutes.

c) The regression equation predicts that an eruption with duration of 4 minutes will have an interval until the next eruption of $33.9668 + 10.3582(4) = 75.3996$ minutes. ($t_{220}^* \approx 1.9708$)

$$\hat{y}_\nu \pm t_{n-2}^* \sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$$
$$= 75.3996 \pm (1.9708)\sqrt{0.3822^2 \cdot (4 - 3.57613)^2 + \frac{6.159^2}{222} + 6.159^2}$$
$$\approx (63.23,\ 87.57)$$

We are 95%confident that the length of time until the next eruption will be between 63.23 and 87.57 minutes, following a 4-minute eruption.

R6.45. Togetherness.

a) H_0: There is no linear relationship number of meals eaten as a family and grades. $(\beta_1 = 0)$
H_A: There is a linear relationship. $(\beta_1 \neq 0)$

Since the conditions for inference are satisfied (given), the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (142 – 2) = 140 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{GPA} = 2.7288 + 0.1093(Meals / Week)$.

$t = \dfrac{b_1 - \beta_1}{SE(b_1)} = \dfrac{0.1093 - 0}{0.0263} \approx 4.16$; The value of $t \approx 4.16$. The P-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between grades and the number of meals eaten as a family. Students whose families eat together relatively frequently tend to have higher grades than those whose families don't eat together as frequently.

b) This relationship would not be particularly useful for predicting a student's grade point average. $R^2 = 11.0\%$, which means that only 11% of the variation in GPA can be explained by the number of meals eaten together per week.

c) These conclusions are not contradictory. There is strong evidence that the slope is not zero, and that means strong evidence of a linear relationship. This does not mean that the relationship itself is strong, or useful for predictions.

R6.46. Learning math.

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

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

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 459.24 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:

$$SE(\bar{y}_A - \bar{y}_T) = \sqrt{\frac{84.29^2}{231} + \frac{74.68^2}{245}} \approx 7.3158.$$

The observed difference between the mean scores is 560.01 – 549.65 = 10.36.

$t = \dfrac{(\bar{y}_A - \bar{y}_T) - (0)}{SE(\bar{y}_A - \bar{y}_T)} = \dfrac{10.36 - 0}{7.3158} \approx 1.42$; Since the P-value = 0.1574, we fail to reject the null hypothesis.

There is no evidence that the Accelerated Math students have a different mean score on the pretest than the traditional students.

R6.46. (continued)

b) H_0: Accelerated Math students do not show significant improvement in test scores. The mean individual gain for Accelerated Math is zero. $(\mu_d = 0)$

H_A: Accelerated Math students show significant improvement in test scores.
The mean individual gain for Accelerated Math is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The Accelerated Math students had a mean individual gain of $\bar{d} =$ 77.53 points and a standard deviation of 78.01 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's *t*-model, with 231 – 1 = 230 degrees of freedom, $t_{230}\left(0, \frac{78.01}{\sqrt{231}}\right)$. We will perform a paired *t*-test.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{77.53 - 0}{78.01/\sqrt{231}} \approx 15.11$; Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The Accelerated Math students showed significant improvement.

c) H_0: Students taught using traditional methods do not show significant improvement in test scores. The mean individual gain for traditional methods is zero. $(\mu_d = 0)$

H_A: Students taught using traditional methods show significant improvement in test scores. The mean individual gain for traditional methods is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The students taught using traditional methods had a mean individual gain of $\bar{d} =$ 39.11 points and a standard deviation of 66.25 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's *t*-model, with 245 – 1 = 244 degrees of freedom, $t_{244}\left(0, \frac{66.25}{\sqrt{245}}\right)$. We will perform a paired *t*-test.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} \approx \frac{39.11 - 0}{66.25/\sqrt{245}} \approx 9.24$; Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The students taught using traditional methods showed significant improvement.

d) H_0: The mean individual gain of Accelerated Math students is the same as the mean individual gain of traditional students. $(\mu_{dA} = \mu_{dT}$ or $\mu_{dA} - \mu_{dT} = 0)$

H_A: The mean individual gain of Accelerated Math students is greater than the mean individual gain of traditional students. $(\mu_{dA} > \mu_{dT}$ or $\mu_{dA} - \mu_{dT} > 0)$

R6.46. (continued)

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

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 452.10 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

$$SE(\bar{d}_A - \bar{d}_T) = \sqrt{\frac{78.01^2}{231} + \frac{66.25^2}{245}} \approx 6.6527.$$

The observed difference between the mean scores is 77.53 – 39.11 = 38.42.

$t = \dfrac{(\bar{d}_A - \bar{d}_T) - 0}{SE(\bar{d}_A - \bar{d}_T)} = \dfrac{38.42 - 0}{6.6527} \approx 5.78$; Since the P-value is less than 0.0001, we reject the null hypothesis.

There is strong evidence that the Accelerated Math students have an individual gain that is significantly higher than the individual gain of the students taught using traditional methods.

R6.47. Juvenile offenders.

a) **Randomization condition:** We will assume that the youths studied are representative of other youths that might receive this therapy.
10% condition: 125 and 125 are less than 10% of all such youths.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$ (Ind) = 19, $n\hat{q}$ (Ind) = 106, $n\hat{p}$ (MST) = 5, and $n\hat{q}$ (MST) = 120 are all not greater than 10, but with only one equal to 5, the test two-proportion interval should be reliable.

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

$$\left(\hat{p}_I - \hat{p}_M\right) \pm z^* \sqrt{\frac{\hat{p}_I \hat{q}_I}{n_I} + \frac{\hat{p}_M \hat{q}_M}{n_M}} = \left(\tfrac{19}{125} - \tfrac{5}{125}\right) \pm 2.576 \sqrt{\frac{\left(\frac{19}{125}\right)\left(\frac{106}{125}\right)}{125} + \frac{\left(\frac{5}{125}\right)\left(\frac{120}{125}\right)}{125}} = \left(0.0178,\ 0.206\right)$$

We are 99% confident that the percentage of violent felony arrest among juveniles who receive individual therapy is between 1.78 and 20.6 percentage points higher than the percentage of violent felony arrest among juveniles who receive MST.

b) Since the entire interval is above 0, we can conclude that MST is successful in reducing the proportion of juvenile offenders who commit violent felonies. The population of interest is adolescents with mental health problems.

R6.48. Dairy sales.

a) Since the CEO is interested in the association between cottage cheese sales and ice cream sales, the regression analysis is appropriate.

b) There is a moderate, linear, positive association between cottage cheese and ice cream sales. For each additional million pounds of cottage cheese sold, an average of 1.19 million pounds of ice cream are sold.

c) The regression will not help here. A paired *t*-test will tell us whether there is an average difference in sales.

d) There is evidence that the company sells more cottage cheese than ice cream, on average.

R6.48. (continued)

e) In part (a), we are assuming that the relationship is linear, that errors are independent with constant variation, and that the distribution of errors is Normal.

In part (c), we are assuming that the observations are independent and that the distribution of the differences is Normal. This may not be a valid assumption, since the histogram of differences looks bimodal.

f) The equation of the regression line is $\widehat{IceCream} = -26.5306 + 1.19334(CottageCheese)$. In a month in which 82 million pounds of ice cream are sold we expect to sell:

$\widehat{IceCream} = -26.5306 + 1.19334(82) = 71.32$ million pounds of ice cream.

g) Assuming the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (12 – 2) = 10 degrees of freedom. We will use a regression slope t-interval, with 95% confidence.

$$b_1 \pm t_{n-2}^{*} \times SE(b_1) = 1.19334 \pm (2.228) \times 0.4936 \approx (0.09,\ 2.29)$$

h) We are 95% confident that the mean number of pounds of ice cream sold increases by between 0.09 and 2.29 pounds for each additional pound of cottage cheese sold.

R6.49. Diet.

H_0: Cracker type and bloating are independent.
H_A: There is an association between cracker type and bloating.

Counted data condition: The data are counts.
Randomization condition: Assume that these women are representative of all women.
Expected cell frequency condition: The expected counts are all (almost!) greater than 5.

	Bloat	
	Little/None (*Obs / Exp*)	Moderate/Severe (*Obs / Exp*)
Bran	11 / 7.6471	2 / 5.3529
Gum Fiber	4 / 7.6471	9 / 5.3529
Combination	7 / 7.6471	6 / 5.3529
Control	8 / 7.0588	4 / 4.9412

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp} \approx 8.23,\ \text{df} = 3;\ \text{P-value} \approx 0.0414$$

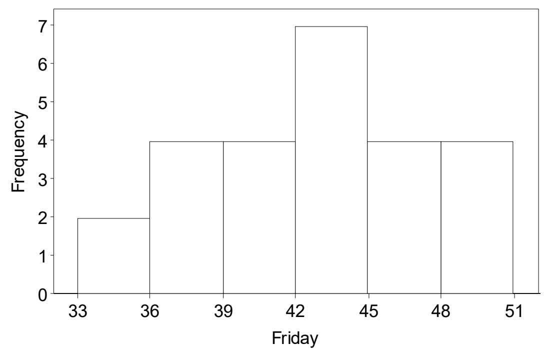
Since the P-value is low, we reject the null hypothesis. There is evidence of an association between cracker type and bloating. The gum fiber crackers had a higher rate of moderate/severe bloating than expected. The company should head back to research and development and address the problem before attempting to market the crackers.

R6.50. Cramming.

a) H_0: The mean score of week-long study group students is the same as the mean score of overnight cramming students. $(\mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0)$

H_A: The mean score of week-long study group students is greater than the mean score of overnight cramming students. $(\mu_1 > \mu_2 \text{ or } \mu_1 - \mu_2 > 0)$

Independent Groups Assumption: Scores of students from different classes should be independent.
Randomization Condition: Assume that the students are assigned to each class in a representative fashion.
Nearly Normal Condition: The histogram of the crammers is unimodal and symmetric. We don't have the actual data for the study group, but the sample size is large enough that it should be safe to proceed.



We are given the following information.

$\bar{y}_1 = 43.2$ $\quad$ $\bar{y}_2 = 42.28$

$s_1 = 3.4$ $\quad$ $s_2 = 4.43020$

$n_1 = 45$ $\quad$ $n_2 = 25$

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 39.94 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

$$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{3.4^2}{45} + \frac{4.43020^2}{25}} \approx 1.02076.$$

The observed difference between the mean scores is 43.2 – 42.28 = 0.92.

$t = \dfrac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE(\bar{y}_1 - \bar{y}_2)} \approx \dfrac{0.92 - 0}{1.02076} \approx 0.90$; Since the P-value = 0.1864 is high, we fail to reject the null hypothesis. There is no evidence that students with a week to study have a higher mean score than students who cram the night before.

b) H_0: The proportion of study group students who will pass is the same as the proportion of crammers who will pass. $(p_1 = p_2 \text{ or } p_1 - p_2 = 0)$

H_A: The proportion of study group students who will pass is different from the proportion of crammers who will pass. $(p_1 \neq p_2 \text{ or } p_1 - p_2 \neq 0)$

R6.50. (continued)

Random condition: Assume students are assigned to classes in a representative fashion.
10% condition: 45 and 25 are both less than 10% of all students.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n_1\hat{p}_1 = 15$, $n_1\hat{q}_1 = 30$, $n_2\hat{p}_2 = 18$, and $n_2\hat{q}_2 = 7$ are not all greater than 10, since only 7 crammers didn't pass. However, if we check the pooled value, $n_2\hat{p}_{\text{pooled}} = (25)(0.471) =$ 11.775. All of the samples are large enough.

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

$$SE_{\text{pooled}}\left(\hat{p}_1 - \hat{p}_2\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_1} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_2}} = \sqrt{\frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{45} + \frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{25}} \approx 0.1245.$$

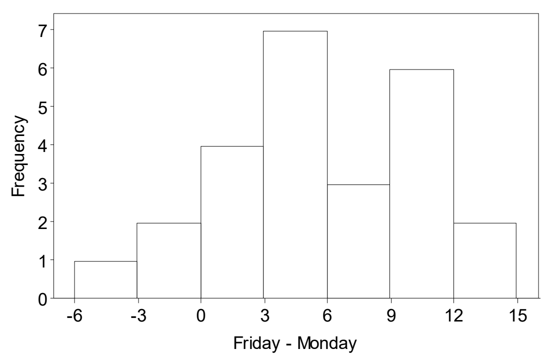
The observed difference between the proportions is 0.3333 – 0.72 = –0.3867.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (0)}{SE(\hat{p}_1 - \hat{p}_2)} = \frac{-0.3867 - 0}{0.1245} \approx -3.11;$$ Since the P-value = 0.0019 is low, we reject the null hypothesis. There is strong evidence to suggest a difference in the proportion of passing grades for study group participants and overnight crammers. The crammers generally did better.

c) H_0: There is no mean difference in the scores of students who cram, after 3 days. $(\mu_d = 0)$

H_A: The scores of students who cram decreases, on average, after 3 days. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Assume that students are assigned to classes in a representative fashion.
Nearly Normal condition: 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 25 – 1 = 24 degrees of freedom, $t_{24}\left(0, \frac{4.8775}{\sqrt{25}}\right)$.

We will use a paired *t*-test, with $\bar{d} = 5.04$.

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} \approx \frac{5.04 - 0}{4.8775/\sqrt{25}} \approx 5.17;$$ Since the P-value is less than 0.0001, we reject the null hypothesis.
There is strong evidence that the mean difference is greater than zero. Students who cram seem to forget a significant amount after 3 days.

R6.50. (continued)

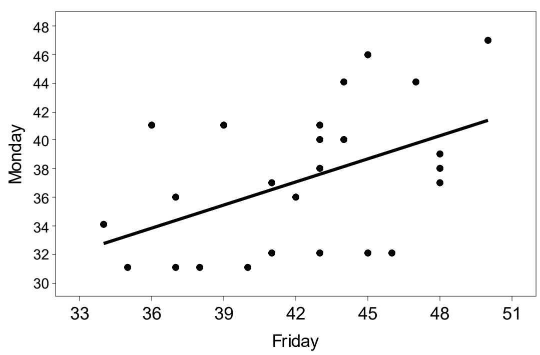
d) $\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 5.04 \pm t_{24}^{*}\left(\frac{4.8775}{\sqrt{25}}\right) \approx (3.03,\ 7.05)$

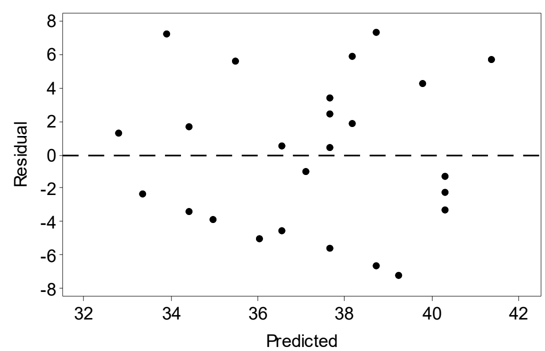
We are 95% confident that students who cram will forget an average of 3.03 to 7.05 words in 3 days.

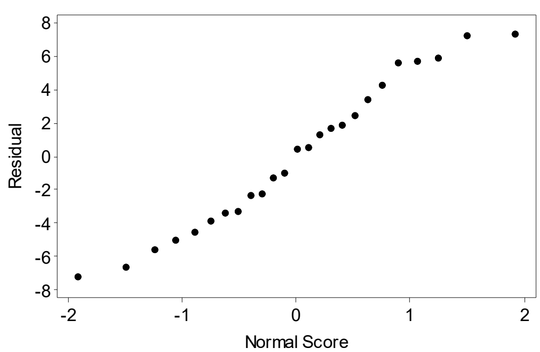
e) H_0: There is no linear relationship between Friday score and Monday score. $(\beta_1 = 0)$

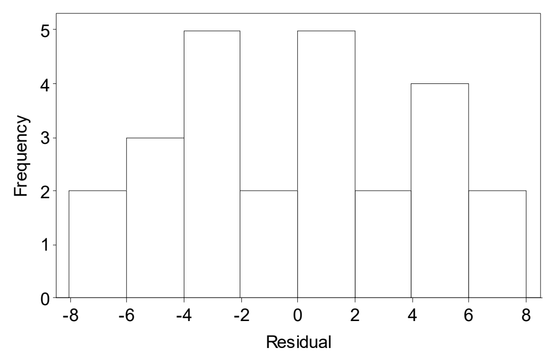
H_A: There is a linear relationship between Friday score and Monday score. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is reasonably straight, and the histogram of the residuals is roughly unimodal and symmetric.









Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (25 – 2) = 23 degrees of freedom. We will use a regression slope t-test.

Dependent variable is: **Monday**
No Selector
R squared = 22.4% R squared (adjusted) = 19.0%
s = 4.518 with 25 - 2 = 23 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	135.159	1	135.159	6.62
Residual	469.401	23	20.4087	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	14.5921	8.847	1.65	0.1127
Friday	0.535666	0.2082	2.57	0.0170

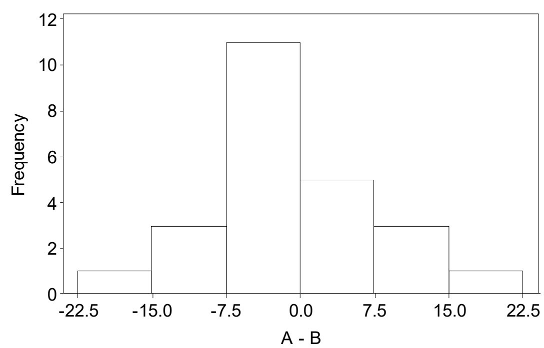
R6.50. (continued)

The equation of the line of best fit for these data points is: $\widehat{Monday} = 14.5921 + 0.535666(Friday)$.

The value of $t \approx 2.57$. The P-value of 0.0170 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between Friday score and Monday score. Students who do better in the first place tend to do better after 3 days. However, since R^2 is only 22.4%, Friday score is not a very good predictor of Monday score.

R6.51. Hearing.

Paired data assumption: The data are paired by subject.
Randomization condition: The order of the tapes was randomized.
Normal population assumption: The histogram of differences between List A and List B is roughly unimodal and symmetric.



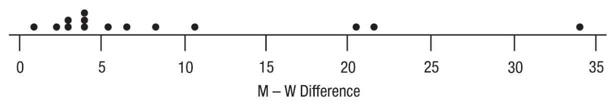
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 24 – 1 = 23 degrees of freedom. We will find a paired t-interval, with 95% confidence.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = -0.\bar{3} \pm t_{23}^{*}\left(\frac{8.12225}{\sqrt{24}}\right) \approx (-3.76, 3.10)$$

We are 95% confident that the mean difference in the number of words a person might misunderstand using these two lists is between –3.76 and 3.10 words. Since 0 is contained in the interval, there is no evidence to suggest that that the two lists are different for the purposes of the hearing test when there is background noise. It is reasonable to think that the two lists are still equivalent.

R6.52. Newspapers.

a) An examination of a graphical display reveals Spain, Portugal, and Italy to be outliers. They are all Mediterranean countries, and all have a significantly higher percentage of men than women reading a newspaper daily.



b) H_0: The mean difference in the percentage of men and women who read a daily newspaper in these countries is zero. $(\mu_d = 0)$

H_A: The mean difference in the percentage of men and women who read a daily newspaper in these countries is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by country.
Randomization condition: Samples in each country were random.
Nearly Normal condition: With three outliers removed, the distribution 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 11 – 1 = 10 degrees of freedom, $t_{10}\left(0, \frac{2.83668}{\sqrt{11}}\right)$.

We will use a paired *t*-test, (Men – Women) with $\bar{d} = 4.75455$.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} \approx \frac{4.75455 - 0}{2.83668/\sqrt{11}} \approx 5.56$; Since the P-value = 0.0001 is very low, we reject the null hypothesis.

There is strong evidence that the mean difference is greater than zero. The percentage of men in these countries who read the paper daily appears to be greater than the percentage of women who do so.