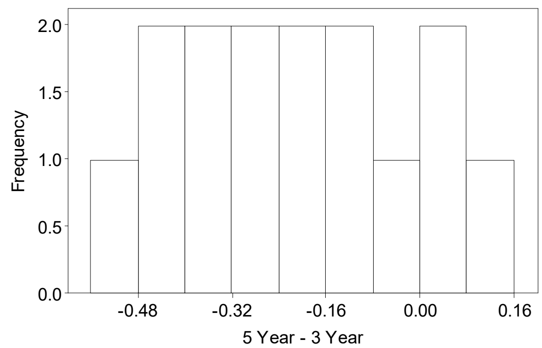
Review of Part VII – Inference When Variables Are Related

R7.1. Mutual fund returns 2017.

a) Paired data assumption: These data are paired by mutual fund.
Randomization condition: Assume that these funds are representative of all mutual funds.
Nearly Normal condition: The histogram of differences shows plausible normality, even if not symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 15 – 1 = 14 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

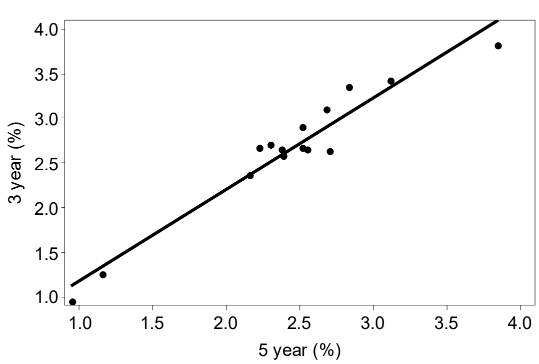
$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = *-0.2153 \pm t^*_{14}\left(\frac{0.187116}{\sqrt{15}}\right) \approx (0.112,\ 0.319)$$

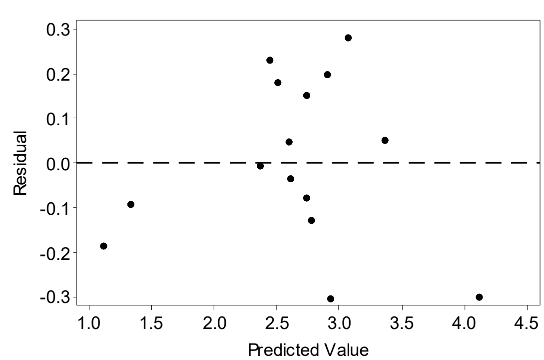
Provided that these mutual funds are representative of all mutual funds, we are 95% confident that, on average, 5-year yields are between 11.2% and 31.9% higher than 3-year yields.

b) H_0: There is no linear relationship between 1-year and 3-year rates of return. $(\beta_1 = 0)$

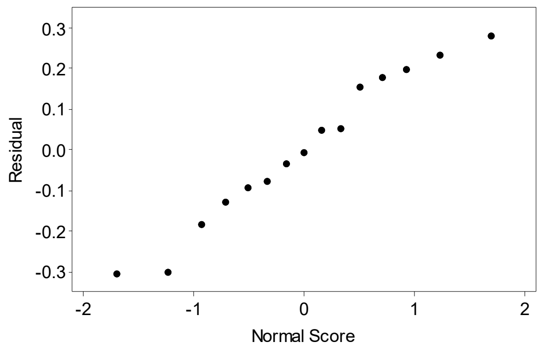
H_A: There is a linear relationship between 1-year and 3-year rates of return. $(\beta_1 \neq 0)$

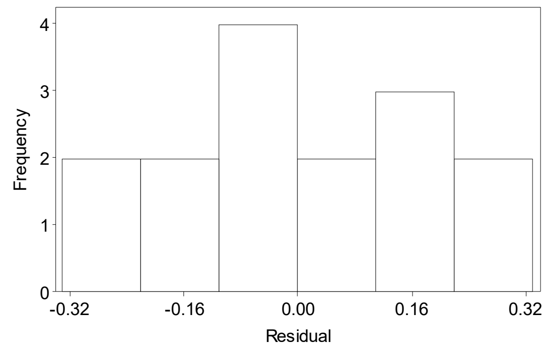
Straight enough condition: The scatterplot is straight enough to try linear regression, though one point seems to be influencing the association.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is straight and the histogram of residuals shows plausible normality, even if not symmetric. With a sample size of 15, it's difficult to determine the true nature of the distribution, so it is probably okay to proceed, but we won't have too much faith in our results.





R7.1. (continued)





Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (13 – 2) = 11 degrees of freedom. We will use a regression slope *t*-test.

The equation of the line of best fit for these data points is $\widehat{3\text{-}year} = 0.134 + 0.1.03(5\text{-}year)$.

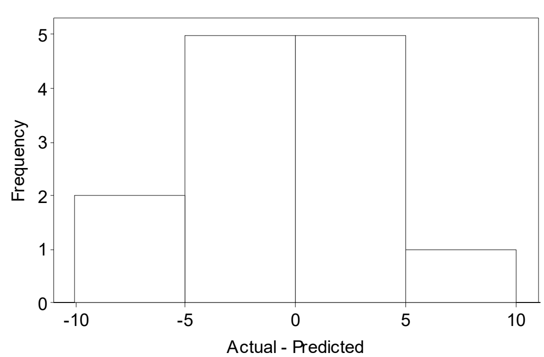
The value of $t \approx 13.95$. The P-value < 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is evidence of a linear relationship between the rates of return for 3-year and 5-year periods. But we don't know if this sample of mutual funds is a random sample (or even representative) of all mutual funds.

R7.2. Polling.

a) H_0: The mean difference in the number of predicted Democrats and the number of actual Democrats is zero. $(\mu_d = 0)$

H_A: The mean difference in the number of predicted Democrats and the number of actual Democrats is different than zero. $(\mu_d \neq 0)$

Paired data assumption: The data are paired by year.
Randomization condition: Assume these predictions are representative of other predictions.
Nearly Normal condition: The histogram of differences between the predicted number of Democrats and the actual number of Democrats is roughly unimodal and symmetric. The year 1958 is an outlier, and was removed.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 13 – 1 = 12 degrees of freedom, $t_{12}\left(0, \frac{3.57878}{\sqrt{13}}\right)$.

R7.2. (continued)

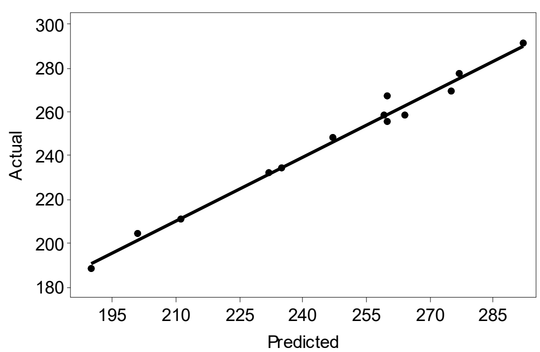
We will use a paired t-test, with $\bar{d} = -0.846$.

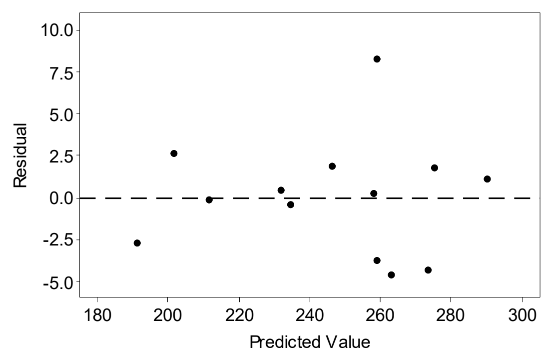
$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{-0.846 - 0}{3.57878 / \sqrt{13}} \approx -0.85;$$ Since the P-value = 0.4106 is high, we fail to reject the null hypothesis. There is no evidence that the mean difference between the actual and predicted number of Democrats was different than 0.

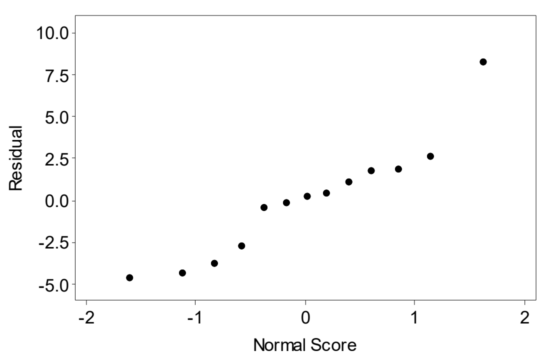
b) H_0: There is no linear relationship between Gallup's predictions and the actual number of Democrats. $(\beta_1 = 0)$

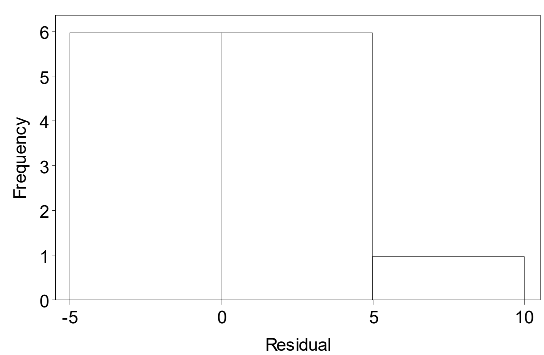
H_A: There is a linear relationship between Gallup's predictions and the actual number of Democrats. $(\beta_1 \neq 0)$

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: After an outlier in 1958 is removed, the Normal probability plot of residuals still isn't very straight. However, the histogram of residuals is roughly unimodal and symmetric. With a sample size of 13, it is probably okay to proceed.









Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (13 – 2) = 11 degrees of freedom. We will use a regression slope t-test.

R7.2. (continued)

The equation of the line of best fit for these data points is $\widehat{Actual} = 6.00180 + 0.972206(Predicted)$.

Dependent variable is: **Actual**
No Selector
R squared = 98.7% R squared (adjusted) = 98.6%
s = 3.628 with 13 - 2 = 11 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	10874.4	1	10874.4	826
Residual	144.805	11	13.1641	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	6.00180	8.395	0.715	0.4895
Predicted	0.972206	0.0338	28.7	≤ 0.0001

The value of $t \approx 28.7$. 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 the number of Democrats predicted by Gallup and the number of Democrats actually in the House of Representatives. Years in which the predicted number was high tend to have high actual numbers also. The high value of $R^2 = 98.7\%$ indicates a very strong model. Gallup polls seem very accurate.

R7.3. Football.

a) He should have performed a one-way ANOVA and an F-test.

b) H_0: The mean distance thrown is the same for each grip. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$
H_A: The mean distances thrown are not all the same.

c) With a P-value equal to 0.0032, we reject the null hypothesis and conclude that the distance thrown is not the same for all grips, on average.

d) **Randomization condition:** The grips used were randomized.
Similar Variance condition: The boxplots should show similar spread.
Nearly Normal condition: The histogram of the residuals should be unimodal and symmetric.

e) He might want to perform a multiple comparison test to see which grip is best.

R7.4. Golf.

a) She should have performed a one-way ANOVA and an F-test.

b) H_0: The mean distance from the target is the same for all four clubs. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$
H_A: The mean distances from the target are not all the same.

c) With a P-value equal to 0.0245, we reject the null hypothesis and conclude that the mean distances from the target are not the same for all four clubs.

d) **Randomization condition:** The clubs used were randomized.
Similar Variance condition: The boxplots should show similar spread.
Nearly Normal condition: The histogram of the residuals should be unimodal and symmetric.

e) She should perform a multiple comparison test to determine which club is best.

R7.5. Horses, a bit less wild.

a) *Sterilized* is an indicator variable.

b) According to the model, there were, on average, 6.4 fewer foals in herds in which some stallions were sterilized, after allowing for the number of adults in the herd.

c) The P-value for *sterilized* is equal to 0.096. That is not significant at the $\alpha = 0.05$ level, but it does seem to indicate some effect.

R7.6. Hard water.

a) *Derby* is an indicator variable.

b) According to the model, the mortality rate north of Derby is, on average, 158.9 deaths per 100,000 higher than south of Derby, after allowing for the linear effects of the water hardness.

c) The regression with the indicator variable appears to be a better regression. The indicator has a highly significant (P-value equal to 0.0001) coefficient, and the R^2 for the overall model has increased from 43% to 56.2%.

R7.7. Lost baggage 2018.

The Counted Data Condition is not met. We cannot use a chi-square test.

R7.8. Bowling.

a) H_0: The mean number of pins knocked down is the same for all three weights. $(\gamma_{Low} = \gamma_{Med} = \gamma_{High})$

H_A: The mean number of pins knocked down are not all the same.

H_0: The mean number of pins knocked down is the same whether walking or standing. $(\tau_S = \tau_W)$

H_A: The mean number of pins knocked down are not the same for the two approaches.

b) The *weight* sum of squares has 3 – 1 = 2 degrees of freedom.
The *approach* sum of squares has 2 – 1 = 1 degrees of freedom.
The error sum of squares has (24 – 1) – 2 – 1 = 20 degrees of freedom.

c) If the interaction plot shows any evidence of not being parallel, she should fit an interaction term, using 2(1) = 2 degrees of freedom.

R7.9. Video racing.

a) H_0: The mean time is the same for all three types of mouse. $(\gamma_{Ergo} = \gamma_{\text{Re}g} = \gamma_{Cord})$

H_A: The mean times are not all the same for the three types of mouse.

H_0: The mean time is the same whether the lights are on or off. $(\tau_{On} = \tau_{Off})$

H_A: The mean times are not the same for lights on and lights off.

b) The *mouse* sum of squares has 3 – 1 = 2 degrees of freedom.
The *light* sum of squares has 2 – 1 = 1 degrees of freedom.
The error sum of squares has (6 – 1) – 2 – 1 = 2 degrees of freedom.

c) No, he should not fit an interaction term. The interaction term would use 2 degrees of freedom. This would not leave any degrees of freedom for the error sum of squares and thus would not allow any interpretation of the factors.

R7.10. Resume fraud.

In order to estimate the true percentage of people have misrepresented their backgrounds to within ± 5%, the company would have to perform about 406 random checks.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.05 = 2.326\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(2.326)^2(0.25)(0.75)}{(0.05)^2}$$

$$n \approx 406 \text{ random checks}$$

R7.11. Paper airplanes.

a) It is reasonable to think that the flight distances are independent of one another. The histogram of flight distances (given) is unimodal and symmetric. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t -model, with 11 – 1 = 10 degrees of freedom. We will use a one-sample t -interval with 95% confidence for the mean flight distance.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 48.3636 \pm t^*_{10}\left(\frac{18.0846}{\sqrt{11}}\right) \approx (36.21,\ 60.51)$$

We are 95% confident that the mean distance the airplane may fly is between 36.21 and 60.51 feet.

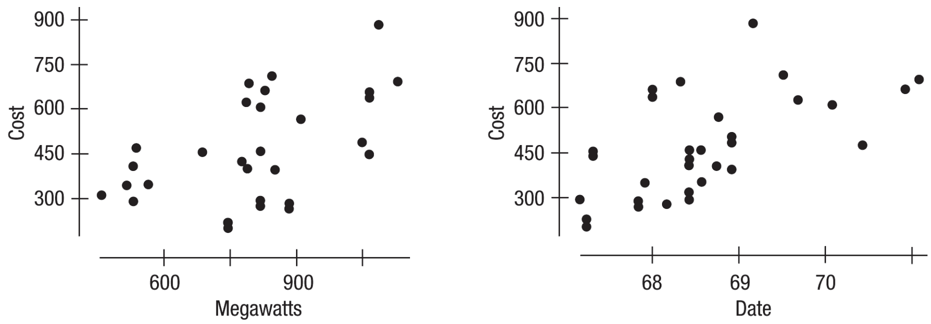
b) Since 40 feet is contained within our 95% confidence interval, it is plausible that the mean distance is 40 feet.

c) A 99% confidence interval would be wider. Intervals with greater confidence are less precise.

d) In order to cut the margin of error in half, she would need a sample size 4 times as large, or 44 flights.

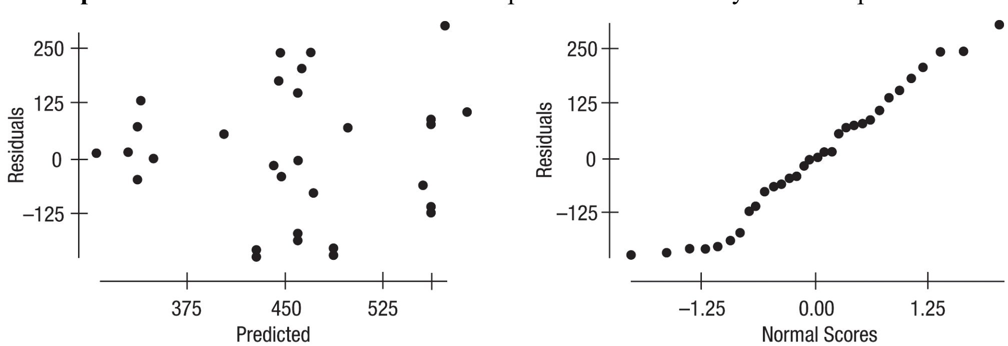
R7.12. Nuclear plants.

a) Both scatterplots are straight enough. Larger plants are more expensive. Plants got more expensive over time.



b) $\widehat{cost} = 111.741 + 0.423831(mwatts)$; According to the linear model, the cost of nuclear plants increased by \$42,383, on average, for each Megawatt of power.

c) **Straight enough condition:** The plot of residuals versus predicted values looks reasonably straight.
Independence assumption: It is reasonable to think of these nuclear power plants as representative of all nuclear power plants.
Nearly Normal condition: The Normal probability plot looks straight enough.
Does the plot thicken? condition: The residuals plot shows reasonably constant spread.

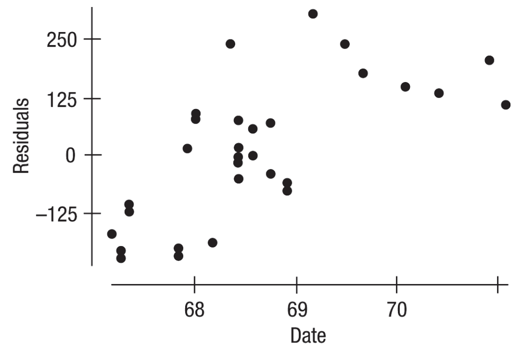


d) H_0: The power plant capacity (*mwatts)* is not linearly associated with the cost of the plant. $(\beta_{Mwatts} = 0)$

H_A: The power plant capacity (*mwatts*) is linearly associated with the cost of the plant. $(\beta_{Mwatts} \neq 0)$

R7.12. (continued)

With $t = 2.93$, and 30 degrees of freedom, the P-value equals 0.0064. Since the P-value is so small, we reject the null hypothesis, *mwatts* is linearly associated with the cost of the plant.

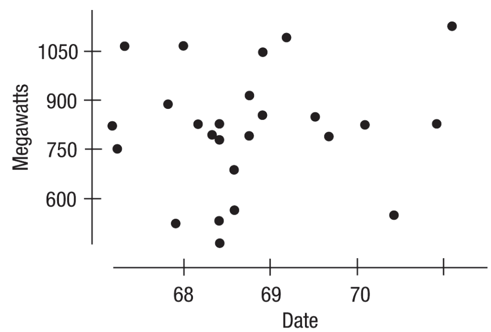


e) From the regression model we find the cost of a 1000 *mwatt* plant is $111.741 + 0.42383(1000) \approx 535.57$, which is $535,570,000.

f) The R^2 means that 22.3% of the variation in *cost* of nuclear plants can be accounted for by the linear relationship between *cost* and the size of the plant, measured in megawatts (*mwatts*). A scatterplot of residuals against *date* shows a strong linear patter, so it looks like *date* could account for more of the variation.

g) The coefficients of *mwatts* changed very little from the simple linear regression to the multiple regression.

h) Because the coefficient changed little when we added *date* to the model, we can expect that *date* and *mwatts* are relatively uncorrelated. In fact, their correlation is 0.02.



R7.13. 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. $(\beta_{Prs} = 0)$

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

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.

R7.14. Sleep.

a) H_0: The mean hours slept is the same for both sexes. $(\gamma_F = \gamma_M)$
H_A: The mean hours slept are not the same for both sexes.

H_0: The mean hours slept is the same for all classes. $(\tau_{Fr} = \tau_{So} = \tau_{Jr} = \tau_{Sr})$
H_A: The mean hours slept are not the same for all classes.

b) Based on the interaction plot, the lines are not parallel, so the Additive Enough condition is not met. The interaction term should be fit.

c) With all three P-values so large, none of the effects appear to be significant.

d) There are a few outliers. Based on the residuals plot, the Similar Variance condition appears to be met, but we do not have a Normal probability plot of residuals. The main concern is that we may not have enough power to detect differences between groups. In particular it may be that upper-class women (juniors and seniors) sleep more than other groups, but we would need more data to tell.

R7.15. Study habits.

a) H_0: The mean hours studied is the same for both sexes. $(\gamma_F = \gamma_M)$
H_A: The mean hours studied not the same for both sexes.

H_0: The mean hours studied is the same for all classes. $(\tau_{Fr} = \tau_{So} = \tau_{Jr} = \tau_{Sr})$
H_A: The mean hours studied are not the same for all classes.

b) Based on the interaction plot, the lines are not parallel, so the Additive Enough condition is not met. The interaction term should be fit.

c) With all three P-values so large, none of the effects appear to be significant.

d) There are a few outliers. Based on the residuals plot, the Similar Variance condition appears to be met, but we do not have a Normal probability plot of residuals. The main concern is that we may not have enough power to detect differences between groups. We would need more data to increase the power.

R7.16. Is Old Faithful getting older?

a) $\widehat{interval} = 35.2463 + 10.4348(duration) - 0.126316(day)$

b) H_0: After allowing for the effects of duration, *interval* does not change over time. $(\beta_{Day} = 0)$

H_A: After allowing for the effects of duration, *interval* does change over time. $(\beta_{Day} \neq 0)$

With a P-value equals 0.0166, we reject the null hypothesis. It appears there is a change over time, after we account for the *duration* of the previous eruption.

c) The coefficient for *day* is not about the two-variable association, but about the relationship between *interval* and *day* after allowing for the linear effects of *duration*.

d) The amount of change is only about –0.126316 (60) = –7.58 seconds per day. This doesn't seem particularly meaningful, although we expect a change of about 46 minutes per year.

R7.17. Traffic fatalities 2013.

a) **Straight enough condition:** The scatterplots of the response versus each predicted value should be reasonably straight. The residuals show a time-related pattern of rising and falling. We should be cautious in interpreting the regression model.
Independence: These data are measured over time. We should plot the data and the residuals against time to look for failures of independence.
Nearly Normal condition: The Normal probability plot should be straight, and the histogram of residuals should be unimodal and symmetric.
Similar Variance condition: The spread of the residuals plot is constant.

R7.17. (continued)

b) $R^2 = 53.0\%$, so 53.0% of the variation in traffic deaths is accounted for by the linear model.

c) $\widehat{deaths} = 689,752 - 324.68(year)$

d) *Deaths* have been declining at the rate of –324.68 per *Year*. (With a standard deviation of the errors estimated to be 3013, this doesn't seem like a very meaningful decrease.)

R7.18. Births.

a) $\widehat{births} = 4422.98 - 15.1898(age) - 1.89830(year)$

b) According to the model, births seem to be declining, on average, at a rate of 1.89 births per 1000 women each year, after allowing for differences across the age of women. According to the model, from 1990 to 1999, birth rate decreased by about 17 births per 1000 women.

c) The scatterplot shows both clumping and a curved relationship. We might want to re-express *births* or add a quadratic term to the model. The clumping is dues to having data for each year of the decade of the 90s for the age bracket of women. It probably does not indicate a failure of the linearity assumption.

R7.19. Typing.

a) H_0: Typing speed is the same whether the gloves are on or off. $(\gamma_{On} = \gamma_{Off})$

H_A: Typing speeds are not the same when the gloves are on and off.

H_0: Typing speed is the same at hot or cold temperatures. $(\tau_H = \tau_C)$

H_A: Typing speeds are not the same at hot and cold temperatures.

b) Based on the boxplots, it appears that both the effects of *temperature* and *gloves* affect typing speed.

c) The interaction plot is not parallel so the additive enough condition is not met. Therefore, an interaction term should be fit.

d) Yes, I think it will be significant because a difference in typing speed due to temperature seems to be significant only when he is not wearing the gloves.

e) The P-values for all three effects: *gloves, temperature* and the interaction term, are all very small, so all three effects are significant.

f) Gloves decreased typing speed at both temperatures. A hot temperature increases typing speed with the gloves off, but has little effect with the gloves on.

g) $s = \sqrt{\dfrac{58.75}{28}} = 1.45$ words per minute. Yes, this seems consistent with the size of the variation shown on the partial boxplots.

h) Tell him to type in a warm room without wearing his gloves.

i) The Normal probability plot of residuals should be straight. The residuals plot shows constant spread across the predicted values. The Similar Variance condition appears to be met. For the levels of the factors that he used, it seems clear that a warm room and not wearing gloves are beneficial for his typing.

R7.20. Typing again.

a) H_0: Typing speed is the same when the television is on or when it is off. $(\gamma_{On} = \gamma_{Off})$

H_A: Typing speeds are not the same when the television is on and off.

H_0: Typing speed is the same when the music is on or when it is off. $(\tau_{On} = \tau_{Off})$

H_A: Typing speeds are not the same when the music is on and off.

R7.20. (continued)

b) Based on the partial boxplots, both effects appear to be small and probably not statistically significant.

c) The interaction plot is not parallel so the additive enough condition is not met. Therefore, an interaction term should be fit.

d) The interaction term may be real. It appears that the effect of television is stronger with the music off than on.

e) All three effects have large P-values. *Music* has a P-value equal to 0.4312, *television* has a P-value equal to 0.2960, and the interaction effect has a P-value equal to 0.5986. None of these effects appear to be significant.

f) None of the effects seem strong. He seems to type just as well with the music and/or the television on.

g) $s=\sqrt{\frac{197.5}{28}}=2.66$ words per minute. This seems consistent with the size of the variation shown in the partial boxplots.

h) Turning the television and/or the music on will not increase his typing speed, nor will it decrease it.

i) The Normal probability plot of residuals should be straight. The residuals plot shows constant spread across the predicted values. The Similar Variance condition appears to be met. For the levels of the factors that he used, it seems that to the level of experimental error present, neither the music nor the television affects his typing speed. If he wanted to see smaller effects, he would have to increase his sample size.

R7.21. NY Marathon.

a) The number of finishers per minute appears to increase by about 0.519037 finishers per minute.

b) There is a definite pattern in the residuals. This is a clear violation of the linearity assumption. While there does seem to be a strong association between time and number of finishers, it is not linear.

R7.22. Learning math.

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

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

Independent groups assumption: Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to 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.

$t=\frac{(\bar{y}_A-\bar{y}_T)-(0)}{SE(\bar{y}_A-\bar{y}_T)}=\frac{10.36-0}{7.3158}\approx 1.42$; 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.

R7.22. (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.

$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{39.11 - 0}{66.25/\sqrt{245}} \approx 9.24$; 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. $(\mu_{dA} = \mu_{dT} \text{ or } \mu_{dA} - \mu_{dT} = 0)$

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

R7.22. (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.

R7.23. Pesticides.

H_0: The percentage of males born to workers at the plant is 51.2%. ($p = 0.512$)
H_A: The percentage of males is less than 51.2%. ($p < 0.512$)

Independence assumption: It is reasonable to think that the births are independent.
Success/Failure Condition: $np = (227)(0.512) = 116$ and $nq = (227)(0.488) = 111$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.512$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.512)(0.488)}{227}} \approx 0.0332$. We can perform a one-proportion *z*-test. The observed proportion of males is $\hat{p} = 0.40$..

$z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \dfrac{0.4 - 0.512}{\sqrt{\frac{(0.512)(0.488)}{227}}} \approx -3.35$; The value of $z \approx -3.35$, meaning that the observed proportion of males is over 3 standard deviations below the expected proportion. The P-value associated with this *z*-score is approximately 0.0004.

With a P-value this low, we reject the null hypothesis. There is strong evidence that the percentage of males born to workers is less than 51.2%. This provides evidence that human exposure to dioxin may result in the birth of more girls.

R7.24. Video pinball.

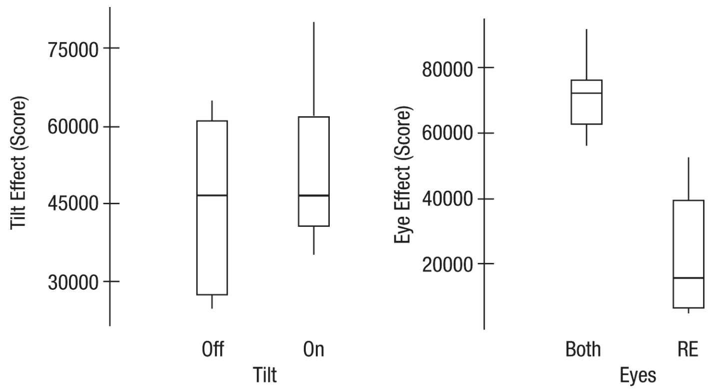
a) H_0: Pinball score is the same whether the tilt is on or off. ($\gamma_{On} = \gamma_{Off}$)

H_A: Pinball scores are different when the tilt is on and when it is off.

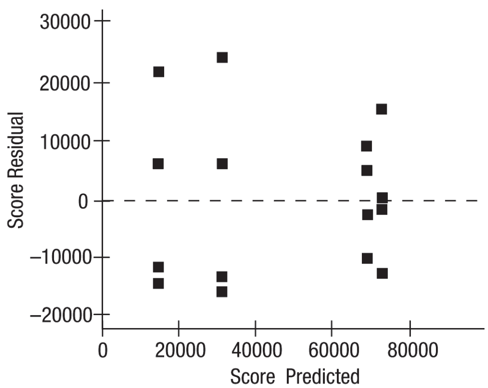
H_0: Pinball score is the same whether both eyes are open or the right is closed. ($\tau_B = \tau_R$)
H_A: Pinball scores are different when both eyes are open and when the right is closed.

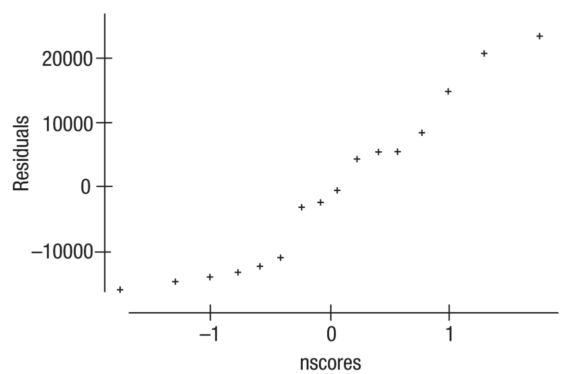
R7.24. (continued)

b) The partial boxplots show that the *eye* effect is strong, but there appears to be little or no effect due to *tilt*.

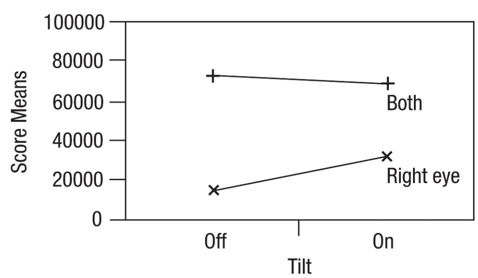


Randomization condition: The experiment was performed in random order.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals is reasonably straight.





There is a possible interaction effect



According to the ANOVA, only eye effect is significant with a P-value less than 0.0001. Both the *tilt* and interaction effects are not significant. We conclude that keeping both eyes open improves score but that using the tilt does not.

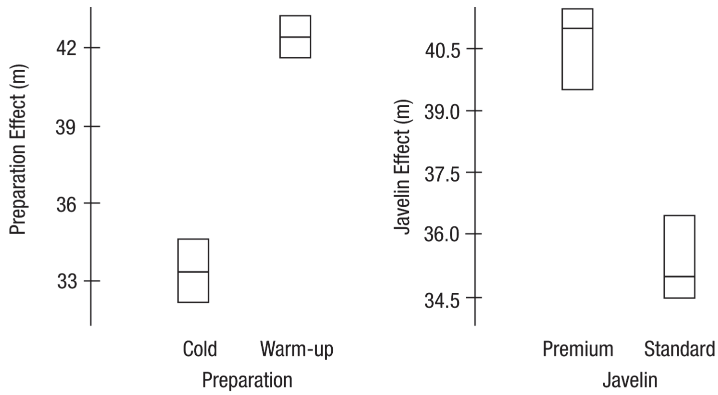
R7.25. Javelin.

a) H_0: Type of javelin has no effect on the distance of her throw. $(\gamma_P = \gamma_S)$
H_A: Type of javelin does have an effect on the distance of her throw.

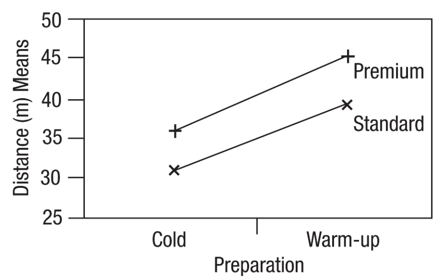
H_0: Preparation has no effect on the distance of her throw. $(\tau_W = \tau_C)$
H_A: Preparation does have an effect on the distance of her throw.

R7.25. (continued)

b) The partial boxplots show that both factors seem to have an effect on the distance of her throw.



The interaction plot is parallel. The additive enough condition is met. No interaction term is needed.



Randomization condition: The experiment was performed in random order.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals should be straight.

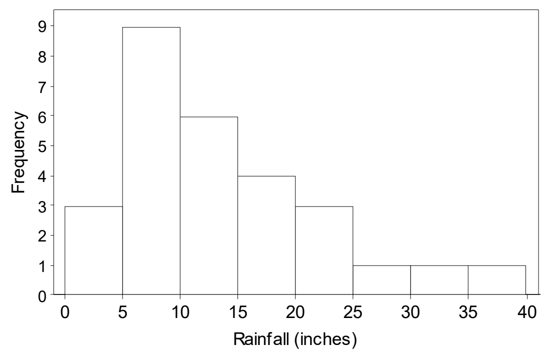
Analysis of Variance For **Distance (m)**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	11552	11552	10502	≤ 0.0001
Prn	1	162	162	147.27	≤ 0.0001
Jvn	1	60.5000	60.5000	55.000	0.0007
Error	5	5.50000	1.10000		
Total	7	228			

With very small P-values, both effects are significant. It appears that depending on the cost, the premium javelin may be worth it — it increases the distance about 5.5 meters, on average. Warming up increases distance about 9 meters, on average. She should always warm up and consider using the premium javelin.

R7.26. LA rainfall 2018.

a) **Independence assumption:** Annual rainfall is independent from year to year.
Nearly Normal condition: The histogram of the rainfall totals is skewed to the right, but the sample is fairly large, so it is safe to proceed.



The mean annual rainfall is 14.0111 inches, with a standard deviation 8.34765 inches. Since the conditions have been satisfied, construct a one-sample t -interval, with 28 – 1 = 27 degrees of freedom, at 90% confidence.

$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 14.0111 \pm t_{27}^{*}\left(\frac{8.340765}{\sqrt{28}}\right) \approx (11.33,\ 16.70)$$

We are 90% confident that the mean annual rainfall in Los Angeles is between 11.33 and 16.70 inches.

b) Start by making an estimate, either using $z^{*} = 1.645$ or $t_{27}^{*} = 1.703$ from above. Either way, your estimate is around 51 years. Make a better estimate using $t_{50}^{*} = 1.676$. You would need about 49 years of data to estimate the annual rainfall in Los Angeles to within 2 inches.

$$ME = t_{50}^{*}\left(\frac{s}{\sqrt{n}}\right)$$

$$2 = 1.676\left(\frac{8.340765}{\sqrt{n}}\right)$$

$$n = \frac{(1.676)^2 (8.340765)^2}{(2)^2}$$

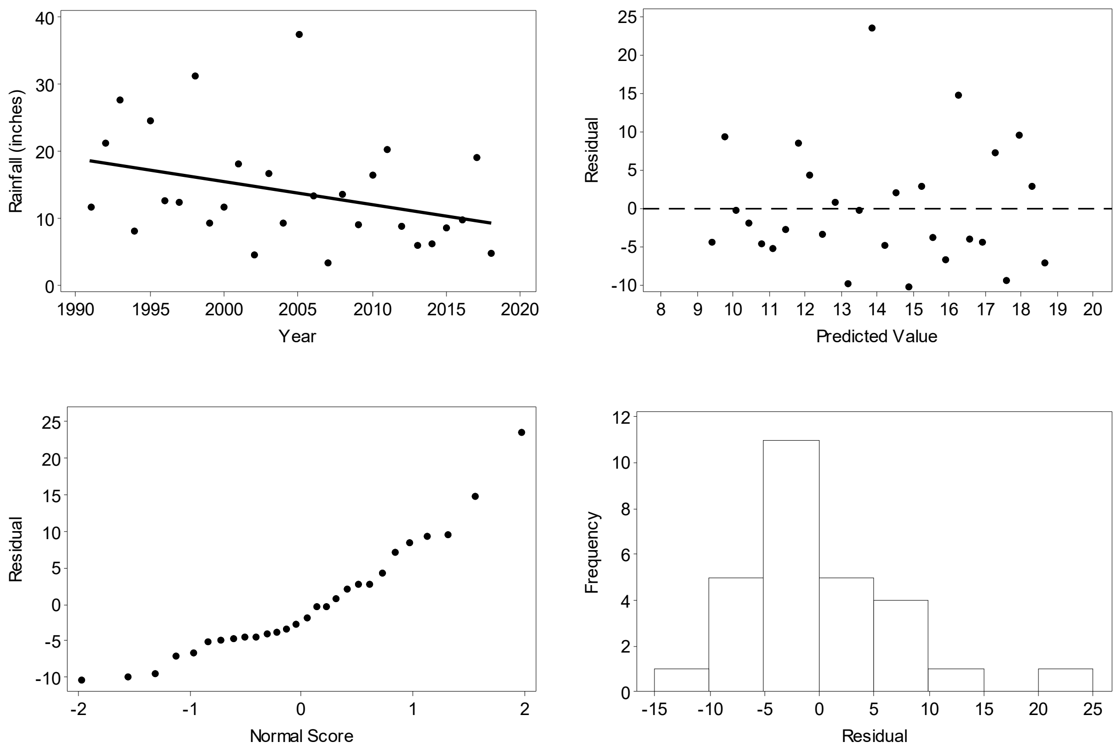
$$n \approx 49 \text{ years}$$

c) H_0: There is no linear relationship between year and annual LA rainfall. $(\beta_1 = 0)$

H_A: There is a linear relationship between year and annual LA rainfall. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot is straight enough to try linear regression, although there is no apparent pattern.
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 (without the outlier) is not straight, but a sample of 28 years is large enough to proceed.

R7.26. (continued)



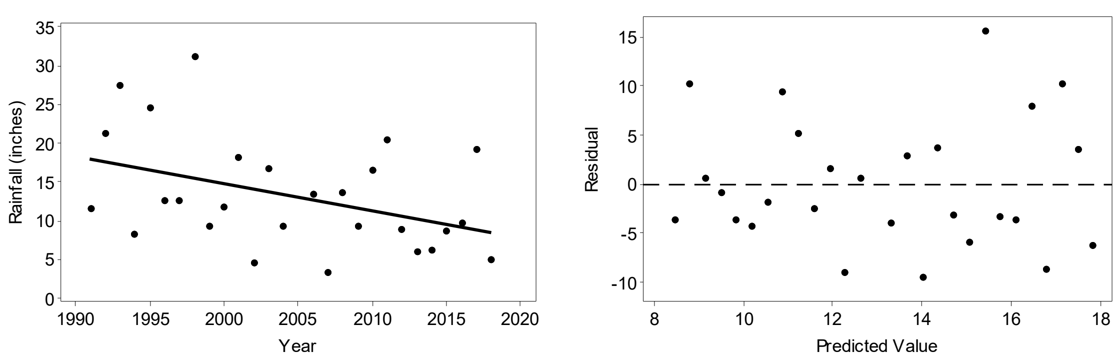
Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (27 – 2) = 25 degrees of freedom. We will use a regression slope *t*-test.

The equation of the line of best fit for these data points is $\widehat{Rain} = 697.543 - 0.3410(Year)$.

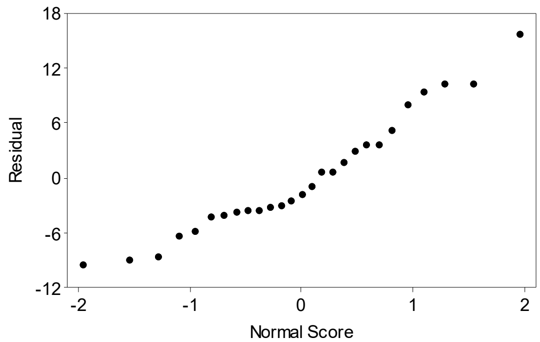
The value of $t \approx -2.24$. The P-value of 0.034 means that the association we see in the data is not likely to occur by chance. We reject the null hypothesis, and conclude that there is evidence of a linear relationship between the annual rainfall in LA and the year.

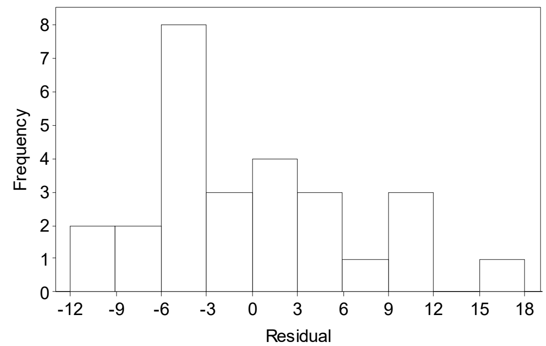
d) 2005 was an extraordinarily wet year, setting some records. Without the outlier, the R^2 increases to 16.7% although the slope remains about the same.

The equation of the line of best fit without 2005 is $\widehat{Rain} = 709.996 - 0.3476(Year)$.



R7.26. (continued)





R7.27. TV and athletics.

a) H_0: The mean hours of television watched is the same for all 3 groups. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean hours of television watched are not the same for all 3 groups.

b) The variance for the *none* group appears to be slightly smaller, and there are outliers in all three groups. We do not have a Normal probability plot of the residuals, but we suspect that the data may not be Normal enough.

c) The ANOVA *F*-test indicates that athletic participation is significant with a P-value equal to 0.0167, we conclude that the number of television hours watched is not the same for all three types of athletic participation.

d) It seems that the differences are evident even when the outliers are removed. The conclusion seems valid.

R7.28. Weight and athletics.

a) H_0: Mean weight is the same for all 3 groups. $(\mu_1 = \mu_2 = \mu_3)$

H_A: Mean weights are not the same for all 3 groups.

b) According to the boxplots the spread appears constant for all three groups. There is one outlier. We do not have a Normal probability plot, but we suspect that the data may be Normal enough.

c) With a P-value equal to 0.0042, we reject the null hypothesis, there is evidence that the mean weights are not the same for all three groups. It may be that more men are involved with athletics, which might explain the weight differences. On the other hand, it may simply be that those who weigh more are more likely to be involved with sports.

d) It seems that differences are evident even when the outlier is removed. It seems that conclusion is valid.

R7.29. Weight loss.

Randomization Condition: The respondents were randomly selected from among the clients of the weight loss clinic.
Nearly Normal Condition: The histogram of the number of pounds lost for each respondent is unimodal and symmetric, with no outliers.

The clients in the sample had a mean weight loss of 9.15 pounds, with a standard deviation of 1.94733 pounds. Since the conditions have been satisfied, construct a one-sample t-interval, with 20 – 1 = 19 degrees of freedom, at 95% confidence.

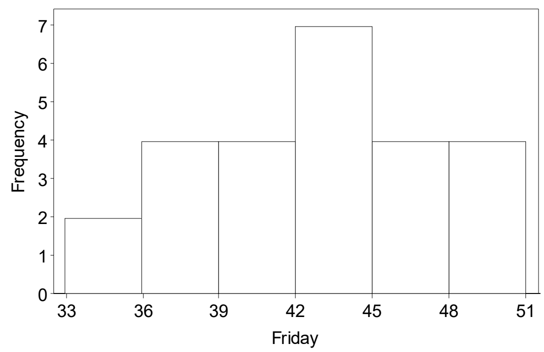
$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 9.15 \pm t_{19}^{*}\left(\frac{1.94733}{\sqrt{20}}\right) \approx (8.24, 10.06)$; We are 95% confident that the mean weight loss experienced by clients of this clinic is between 8.24 and 10.06 pounds. Since 10 pounds is contained within the interval, the claim that the program will allow clients to lose 10 pounds in a month is plausible. Answers may vary, depending on the chosen level of confidence.

R7.30. 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.



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

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

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

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

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

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

$t = \dfrac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE(\bar{y}_1 - \bar{y}_2)} = \dfrac{0.92 - 0}{1.02076} \approx 0.90;$ 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. $(p_1 \neq p_2 \text{ or } p_1 - p_2 \neq 0)$

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.

R7.30. (continued)

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

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

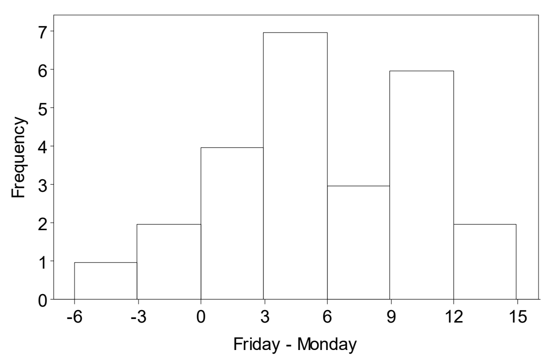
The observed difference between the proportions is 0.3333 – 0.72 = –0.3867.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (0)}{SE(\hat{p}_1 - \hat{p}_2)} = \frac{-0.3867 - 0}{0.1245} \approx -3.11;$$ 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. $\left(\mu_d = 0\right)$

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

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 $\bar{d} = 5.04$.

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{5.04 - 0}{4.8775/\sqrt{25}} \approx 5.17;$$ 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.

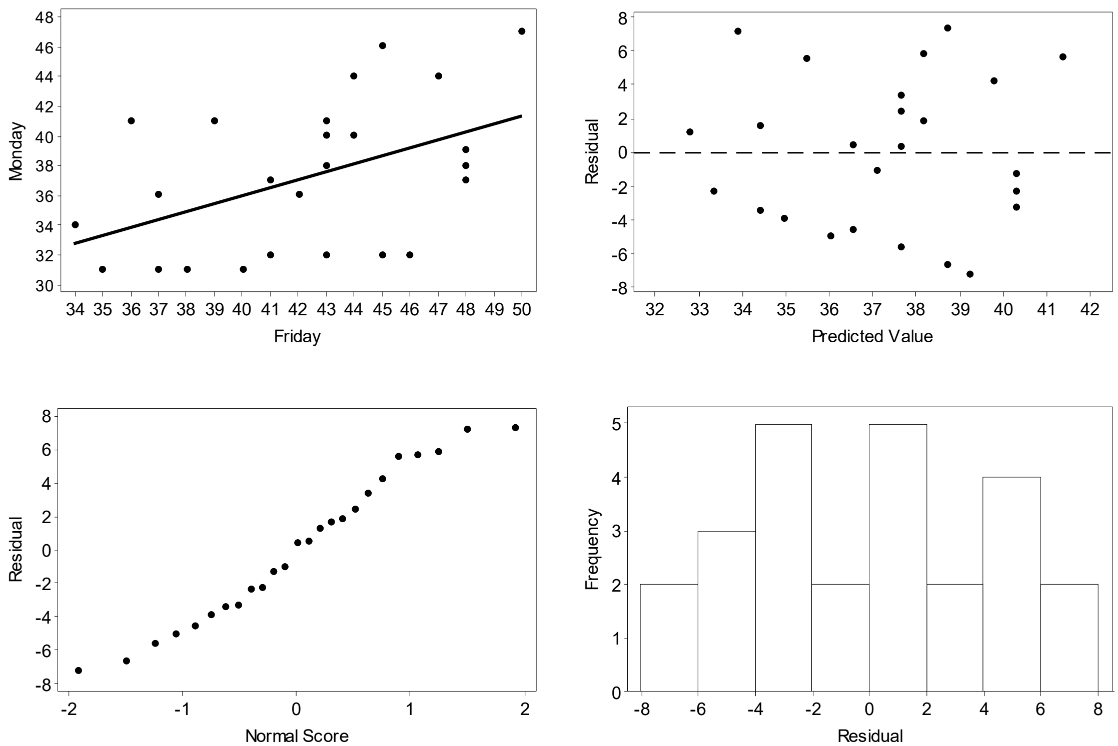
d) $\bar{d} \pm t_{n-1}^*\left(\frac{s_d}{\sqrt{n}}\right) = 5.04 \pm t_{24}^*\left(\frac{4.8775}{\sqrt{25}}\right) \approx (3.03,\ 7.05)$

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. $\left(\beta_1 = 0\right)$

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

R7.30. (continued)



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.

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

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

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.