Homework Solutions – Chapters 8-9

Homework 8.1

1. $H_o: \mu_M - \mu_W = 0 \;\; vs. \;\; H_a: \mu_M - \mu_W \neq 0$

2.

(a) $H_o: \mu_{Evening} - \mu_{Morning} = 0 \;\; vs. \;\; H_a: \mu_E - \mu_M > 0$ or $H_o: \mu_M - \mu_E = 0 \;\; vs. \;\; H_a: \mu_M - \mu_E < 0$

(b) $H_o: \mu_{Evening} - \mu_{Morning} = 5 \;\; vs. \;\; H_a: \mu_E - \mu_M > 5$

3.

(a) $H_o: \mu_{NoPainPlus} - \mu_{QuickReliev} = 0 \;\; vs. \;\; H_a: \mu_N - \mu_Q > 0$

(b) $H_o: \mu_{NoPainPlus} - \mu_{QuickReliev} = 4 \;\; vs. \;\; H_a: \mu_N - \mu_Q > 4$

Homework 8.2

1.

Problem Type	CI, 2 Large samples, Z test
90% Confidence Interval	$[0.66198, 20.538] \approx [0.7, 20.5]$
Interpretation	We are 90% sure that the true mean difference between men and women for aerobic times is between about 42 seconds and 20 minutes. In other words, men spend at least 42 seconds longer than women, and possibly up to 20 minutes more time on average in an aerobic state.

Problem Type	CI, 2 Large samples, Z test
95% Confidence Interval	$[-1.242, 22.442] \approx [-1.2, 22.4]$
Interpretation	We are 95% sure that the true mean difference between men and women for aerobic times is between about -1 minutes and 22 minutes. Thus, it is possible that there is no difference between men and women.

2.

Problem Type	CI, 2 Large samples, Z test
Hypothesis	$H_o: \mu_{Moab} - \mu_{Hilo} = 0 \;\; vs. \;\; H_a: \mu_{Moab} - \mu_{Hilo} \neq 0$
95% Confidence Interval	$[-0.514, 0.79142] \approx [-0.5, 0.8]$
Decision	Since 0 is in the confidence interval, fail to reject the null hypothesis
Conclusion	There is not overwhelming evidence of a difference between the two plants. It is possible that there is no difference.

3. (a)

Problem Type	CI, 2 Large samples, Z test
Hypothesis	$H_o : \mu_{Evening} - \mu_{Morning} \leq 0 \;\; vs. \;\; H_a : \mu_E - \mu_M > 0$
95% Confidence Interval	$[5.9974, 8.6026] \approx [6.0, 8.6]$
Interpretation	We are 95% sure that the true mean difference between evening and morning drive times is between 6 and 8.6 minutes.

(b)

Problem Type	CI, 2 Large samples, Z test
Hypothesis	$H_o : \mu_{Evening} - \mu_{Morning} \leq 0 \;\; vs. \;\; H_a : \mu_E - \mu_M > 0$
95% Confidence Interval	$[5.9974, 8.6026] \approx [6.0, 8.6]$
Decision	Since the mean difference is greater than 0, reject the null hypothesis
Conclusion	There is sufficient evidence to say that the drive time in the evening is longer than in the morning, on average.

(c)

Problem Type	CI, 2 Large samples, Z test
Hypothesis	$H_o : \mu_{Evening} - \mu_{Morning} \leq 5 \;\; vs. \;\; H_a : \mu_E - \mu_M > 5$
95% Confidence Interval	$[5.9974, 8.6026] \approx [6.0, 8.6]$
Decision	Since the mean difference is at least 6 minutes, reject the null hypothesis
Conclusion	There is sufficient evidence to say that the drive time in the evening is at least 5 minutes longer than in the morning, on average.

4.

Yes	No
No	No
No	Yes

5. (a) Reject (b) Reject (c) Fail to reject (d) Fail to reject (e) Reject (f) Reject

Homework 8.3

1.

Problem Type	HT, 2 Large samples, Z test
Hypothesis	$H_o: \mu_{Men} - \mu_{Women} \geq 0 \;\; vs. \;\; H_a: \mu_{Men} - \mu_{Women} < 0$
p-value	$2.52 * 10^{-6} = 0.00000252 \approx 0$
Decision	Since 0 < 0.01, reject the null hypothesis
Conclusion	There is strong evidence that men spend less time in the bathroom, on average than women.

2.

Problem Type	HT, 2 Large samples, Z test
Hypothesis	$H_o: \mu_{Men} - \mu_{Women} \leq 0 \;\; vs. \;\; H_a: \mu_{Men} - \mu_{Women} > 0$
p-value	0.122
Decision	Since 0.122 > 0.05, fail to reject the null hypothesis
Conclusion	There is not strong evidence that there is a difference in the average times men and women spend studying in the library.

3.

Problem Type	HT, 2 Large samples, Z test
Hypothesis	$H_o: \mu_{Investors} - \mu_{YourMoney} \leq 0 \;\; vs. \;\; H_a: \mu_{Investors} - \mu_{YourMoney} > 0$
p-value	0.197
Decision	Since 0.197 > 0.1, fail to reject the null hypothesis
Conclusion	There is not strong evidence that there is a difference in the average times visitors spend at the two websites.

4.

Problem Type	HT, 2 Large samples, Z test
Hypothesis	$H_o: \mu_{Men} - \mu_{Women} = 0 \;\; vs. \;\; H_a: \mu_{Men} - \mu_{Women} \neq 0$
p-value	$1.68 * 10^{-16} \approx 0$
Decision	Since 0 < 0.1, reject the null hypothesis
Conclusion	There is sufficient evidence to say that the mean time men and women spend in college differs.

Homework 8.4

1. (a)

Problem Type	HT, 2 Small samples, T test
Hypothesis	$H_o : \mu_{Valdosta} - \mu_{OutOfState} = 0 \;\; vs. \;\; H_a : \mu_{Valdosta} - \mu_{OutOfState} \neq 0$
Pooled Variance	Since $^{33.71}\!/_{13.26} = 2.54 > 2$, do not use pooled variance
p-value	0.449
Decision	Since 0.449 > 0.05, fail to reject the null hypothesis
Conclusion	There is not strong evidence of a difference between the mean amounts Valdosta residents spend compared to Out of State people. There may be a significant difference but with such small samples and the large variability we were unable to demonstrate that.

(b)

Problem Type	CI, 2 Small samples, T Interval
Pooled Variance	Since $^{33.71}\!/_{13.26} = 2.54 > 2$, do not use pooled variance
95% Confidence Interval	$[-30.01, 14.051] \approx [-\$30, \$14]$
Interpretation	Out of State people may spend as much as \$30 more than Valdosta residents, on average. However, Valdosta residents may spend up to \$14 more than Out of State people. Thus, it is possible that there is no difference since 0 is in the confidence interval.

2. (a)

Problem Type	HT, 2 Small samples, T test
Hypothesis	$H_o : \mu_{Professors} - \mu_{Students} \leq 0 \;\; vs. \;\; H_a : \mu_{Professors} - \mu_{Students} > 0$
Pooled Variance	$s_{Professors} = 3.21,\ s_{Students} = 4.62$. Since $^{4.62}\!/_{3.21} = 1.44 < 2$, use pooled variance
p-value	0.046
Decision	Since 0.046 < 0.1, reject the null hypothesis
Conclusion	There is sufficient evidence to say that the average faculty distance from the university is further than the average student distance from the university.

(b)

Problem Type	CI, 2 Small samples, T Interval
Pooled Variance	use pooled variance (see above)
95% Confidence Interval	$[0.12517, 9.4748] \approx [0.1, 9.5]$
Interpretation	We are 90% sure that faculty live at least 1/10 of a mile further than students and possibly up to 9.5 miles further.

3.

Hypothesis	$H_o: \mu_{Old} - \mu_{New} \leq 0$ vs. $H_a: \mu_{Old} - \mu_{New} > 0$
Pooled Variance	pooled variance was used
p-value	0.003
Decision	Since 0.003 is very small, there is virtually no risk in rejecting the null hypothesis
Conclusion	There is strong evidence that the new route takes less time than the old route.

90% Lower Bound CI	$[0.471 hours, \infty] \approx$ at least 28 minutes
Interpretation	We are 90% sure that the new route, on average, is at least 28 minutes faster than the old route.

Homework 8.5

1. (a) The factor that is being controlled for is the variability in the women's weights.

(b)

Problem Type	HT, Dependent Samples, 1 sample T test
	$Difference = New - Existing$
Hypothesis	$H_o: \mu_d = 0$ vs. $H_a: \mu_d \neq 0$
p-value	0.554
Decision	Since 0.554 > 0.1, fail to reject the null hypothesis
Conclusion	There is not strong evidence of a difference between the two scales.

(c) $[-3.446, 1.8456] \approx [-3.4, 1.8]$. The fact that there could be a true mean difference of 0 leads us to say that there is no statistical difference in womens' weights under the two different scales.

2. (a) The factor that is being controlled for is the variability due to the location of the plots where the tomatoes are planted.

(b)

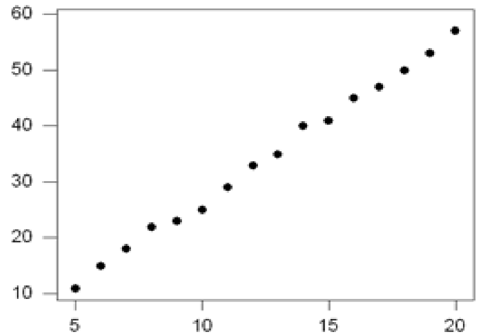
Problem Type	HT, Dependent Samples, 1 sample T test
	$Difference = OppulentOrange - Ruby\,\mathrm{Re}\,d$
Hypothesis	$H_o: \mu_d \leq 0$ vs. $H_a: \mu_d > 0$
p-value	$7.76*10^{-7} \approx 0$
Decision	Since 0 < 0.05, reject the null hypothesis
Conclusion	There is strong evidence that the true mean of the OO variety is larger than the true mean of the RR variety.

(c) $[3.3678, 6.3322] \approx [3.4, 6.3]$. Thus, we see that on average, the OO variety produces 3 to 6 more tomatoes than the RR variety.

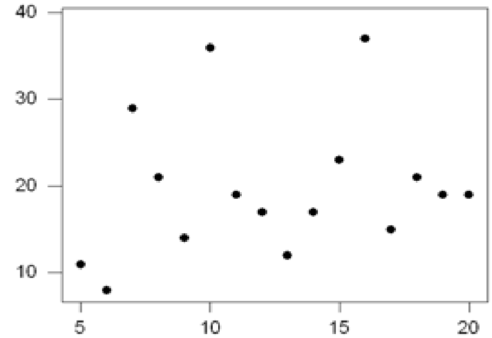
Homework 9.1

1.

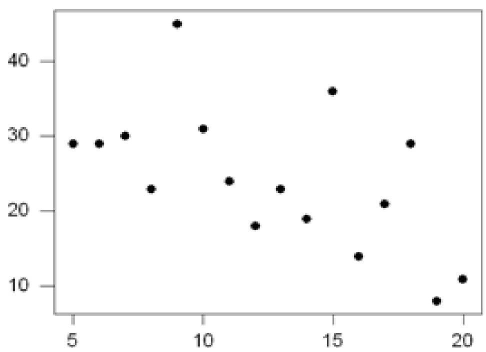
Data Set 1



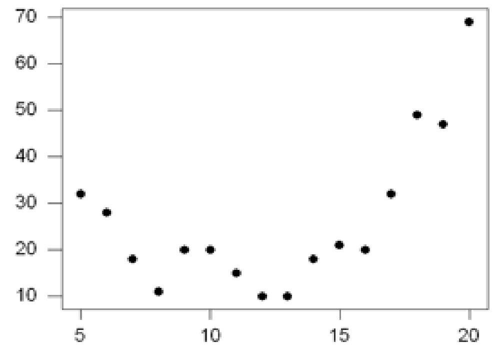
Data Set 2



Data Set 3



Data Set 4



(b)

Data Set 1	Data Set 2
strong, positive, linear correlation	weak positive correlation with some outliers
Data Set 3	Data Set 4
weak, negative correlation, possibly linear	curvature, both negative and positive correlation.

Homework 9.2

1. (a)

Data Set 1	Data Set 2
`Y1 = - 3.32 + 2.99 X`	`Y2 = 16.1 + 0.303 X`
Data Set 3	Data Set 4
`Y3 = 38.6 - 1.14 X`	`Y4 = 1.5 + 1.98 X`

(b) Data Set 1 with an estimated slope of 2.99
(c) Data Set 2 with an estimated slope of 0.303

Homework 9.3

1. (a)

Data Set 1	
Initial Assessment	The scatterplot shows a strong indication that a linear fit is appropriate. OK to proceed with simple linear regression.
Regression Line	`Y1 = - 3.32 + 2.99 X`
Hypothesis	$H_o : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0$
p-value	0.0
Decision	Since $0 < 0.05$, we will reject the null hypothesis
Conclusion	The regression line is statistically significant because we have strong reason to believe that the slope is not zero.
R^2	99.7%
Interpretation	99.7% of the total variability is explained by this regression line, which is an extremely high value.
Final Conclusion	The regression line is significant and it does an excellent job of explaining the data. We would probably use the equation for reasonably accurate predictions.

Data Set 2	
Initial Assessment	There appears to be somewhat of a positive linear correlation in the data. There are several obvious outliers in the scatterplot that deviate from this pattern. We will try linear regression using caution.
Regression Line	`Y2 = 16.1 + 0.303 X`
Hypothesis	$H_o : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0$
p-value	0.514
Decision	Since $0.514 > 0.05$, we fail to reject the null hypothesis
Final Conclusion	The regression line is not statistically significant because the slope can't be distinguished from zero. Thus, we must stop. A linear relationship is not appropriate to model this data. Do not use the regression line.

Data Set 3	
Initial Assessment	The scatterplot shows some negative linear correlation. OK to proceed with simple linear regression, but with a bit of caution.
Regression Line	`Y3 = 38.6 - 1.14 X`
Hypothesis	$H_o : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0$
p-value	0.02
Decision	Since 0.02 < 0.05, we will reject the null hypothesis
Conclusion	The regression line is statistically significant because we have strong reason to believe that the slope is not zero.
R^2	32.9%
Interpretation	32.9% of the total variability is explained by this regression line, which is a very small value. an extremely high value.
Final Conclusion	The regression line is significant but it doesn't do a good job at predicting. Use the regression line with caution.

Data Set 4	
Initial Assessment	The scatterplot shows a clear indication of curvature. Linear regression is not appropriate. We must stop.
Regression Line	`Y4 = 1.5 + 1.98 X - Meaningless`
p-value	0.02 – Meaningless
R^2	32.9% - Meaningless

2. You evaluate the hypothesis: $H_o : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0$ to see if the slope is significantly different than zero. If the p-value is small, then you reject the null hypothesis and conclude that the slope and the regression line are statistically significant.
3. You look at the value of R^2.
4. The first step is to make a scatter plot. Then, you don't want to proceed with linear regression unless a visual assessment reveals that a straight line might describe the general tendency of the data.
5. -0.88
6. 0.5

Homework 9.4

1. Answers will vary

Homework 9.5

1.

Data Set 3	
Initial Assessment	The scatterplot now shows reasonably strong negative linear correlation. OK to proceed with simple linear regression.
Regression Line	`Y3 = 37.6 - 1.32 X`
Hypothesis	$H_o : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0$
p-value	0.0
Decision	Since 0 < 0.05, we will reject the null hypothesis
Conclusion	The regression line is statistically significant because we have strong reason to believe that the slope is not zero.
R^2	77.1%
Interpretation	77.1% of the total variability is explained by this regression line, which is a reasonably high values.
Final Conclusion	The regression line is significant and it does a reasonable job of explaining the data. We would probably use this regression line for predictions.

2. $Y_3 = 37.6 - 1.32(15) = 17.8$
3. 1.32 grams
4. 5 to 20 minutes.