Introduction to Statistics for Business and Economics Workbook 2

(Version 1.0)

Peter Dalley

Table of Contents

A Note to Instructors

I hope you find this workbook useful, I just want to point out three key features:

This book is totally free to you and your students. Feel free to copy it or post it to your course website and feel free to share it with colleagues.

Although I am widely distributing a PDF file, I have gone to great effort to make a fully editable Word version of this document. Please contact me if you'd like to have a copy of the Word version. You can edit any of these problems to better fit in your class or simply copy and paste an entire problem into an assignment or test, with the attribution "Source: statisticsworkbook.com", or "Adapted from: statisticsworkbook.com".

Every problem in this workbook has a video walkthrough available at http://statisticsworkbook.com. I suspect the true value in this book lies in the video walkthroughs, as it will be useful for homework and particularly useful for "flipping the classroom".

Please let me know if you would like to see additional question-types or topics included in the future. I intend to add to this book frequently based on your input. Also, any feedback you can provide (particularly student feedback) would be greatly appreciated.

Please note, you do not have my permission to use this for a commercial purpose, nor do you have permission to recreate the videos found at http://statisticsworkbook.com. Send me an email if you have any questions about use or attribution.

Thanks for checking out this workbook, and I hope you'll have a look at the companion website: http://statisticsworkbook.com!

Peter Dalley

profdalley@gmail.com

Module 9: Hypothesis Testing – Single Population Mean and Proportion

9-1A Single population mean, one-tail test, σ known

A local craft brewery claims the amount of beer in its bottles is 12oz (355ml). It knows that making false claims on its labels would result in serious penalties if it overstated the true volume. Every Monday morning, a sample of 30 bottles is taken to test the accuracy of their filling machines. Over the past few years of weekly sampling, they have calculated the standard deviation of the population to be $\sigma = 1.6$oz. This week, the sample resulted in a mean filling volume of $\bar{x} = 11.2$oz. Are they at risk of facing any penalties? Use $\alpha = 0.05$.

a) Formulate the null and alternative hypothesis. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

9-1B Single population mean, one tail test, σ known

An instructor at your university has produced a series of problems and accompanying video walkthroughs in hopes of improving his students' understanding of course content. Having taught the course many times, he calculates the overall average to be 71% and he determines that the population standard deviation is $\sigma = 0.12$ (or 12 percentage points). At the end of the following semester his class of 40 students, who had access to the video walkthroughs, had an average grade of $\bar{x} = 74.6\%$. Using a level of significance of $\alpha = 0.05$, test to determine whether or not this shows an improvement over the historical average.

a) Formulate the null and alternative hypothesis. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

9-1C Single population mean, two-tail test, σ known

A lumber yard produces batches of 2x4s in lengths of 8 and 12 feet (96 and 144 inches, respectively). Each batch contains 50 2x4s. Because the lumber is being used primarily for framing walls in home construction, it is imperative that the lengths be accurate. If a 2x4 is too short, or too long, it will cause delays in construction as adjustments will have to be made. The lumber yard knows the standard deviation of 8 foot lengths is $\sigma_8 = 0.82$ inches and for 12 foot lengths is $\sigma_{12} = 1.05$ inches. On the first business day of each month, one batch of each length is taken out of production and measured in order to test the accuracy of their cuts. The average length of their 8 foot 2x4s is $\bar{x}_8 = 95.8$ inches and the average length of their 12 foot 2x4s is $\bar{x}_{12} = 144.3$ inches. Using a level of significance of $\alpha = 0.05$, test to determine whether or not the two lengths of lumber are being cut accurately.

a) Formulate the null and alternative hypotheses. Justify your formulation.
b) Calculate the test statistics for both tests.
c) Use the p-value approach to draw your conclusions.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.
f) Verify your findings using the confidence interval approach.

9-1D Single population mean, two-tail test, σ known

A local farmer produces hay for nearby cattle ranchers. The hay is rolled into 50lb (22Kg) bails and are sold by quantity. Therefore, when a rancher buys 100 bales of hay, they can expect to receive 5,000lbs of hay. To ensure the cattle ranchers are getting what they expect, the farmer periodically samples batches of 30 hay bales to test that they are averaging 50lbs. The most recent batch provided a sample weight of $\bar{x}$ = 49.1lbs. The population standard deviation is known to be σ = 3.37. Using a level of significance of α = 0.05 test to determine whether the farmer is producing what the ranchers are expecting.

a) Formulate the null and alternative hypotheses. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

9-1E Probability of Type Two Error and the Power of the Test

Copied from problem 9-1A:

A local craft brewery claims the amount of beer in its bottles is 12oz (355ml). It knows that making false claims on its labels would result in serious penalties if it overstated the true volume. Every Monday morning, a sample of 30 bottles is taken to test the accuracy of their filling machines. Over the past few years of weekly sampling, they have calculated the standard deviation of the population to be $\sigma = 1.6$oz. This week, the sample resulted in a mean filling volume of $\bar{x} = 11.2$oz. Are they at risk of facing any penalties? Use $\alpha = 0.05$.

Formulate the null and alternative hypotheses. Justify your formulation.

a) Calculate the test statistic.
b) Use the p-value approach to draw your conclusion.
c) Verify your conclusion using the critical value approach.
d) Interpret your conclusion.
e) If the actual population mean is $\mu_a = 11$oz, what is the probability of committing a Type II error?
f) If the manager states that she is willing to risk a $\beta = 0.10$ probability of not rejecting the null if the average volume within 0.5oz of specification, how large should the sample size be?

9-2A Single population mean, one-tail test, σ unknown

Red light cameras are often an effective deterrent to reducing the number of people who run red lights at intersections. However, they can be expensive to install and maintain. For these reasons, they are only installed at intersections where there tends to be the most accidents caused by drivers running red lights. Let's assume the threshold number of accidents is 10 per year; any more than this warrants a camera. At one location, researchers found the average number of accidents over a ten-year period was 10.6. Based on this data, they immediately submitted a recommendation to install a red light camera at this intersection. After all, 10.6 is greater than 10.

a) Was this the correct decision?
b) What is the proper approach to making this decision? Show each step.
   (Hint: Assume you have access to their data and were able to calculate the sample standard deviation to be s = 1.19)

9-2B Single population mean, two-tail test, σ unknown

A local fire department has a goal of responding to house fires in 10 minutes or less. In order to determine whether they are achieving their goal, samples of 40 response times are tested every week. The most recent sample resulted in a mean response time of $\overline{x}$ = 10.6 minutes with a sample standard deviation of s = 2.1.

a) Formulate the appropriate null and alternative hypothesis. Justify your formulation.
b) Calculate your test statistic and discuss your results. Show all your work.

9-2C Single population mean, two-tail test, σ unknown

It is common among some universities to target a specific average grade in certain courses. At the end of each semester, instructors faced with this constraint are required to determine whether their class average is statistically different from the targeted average for the course. If they are either above or below the target, adjustments must be made. Let us assume that in a particular course, instructors are expected to have an average grade of 70%. This semester, the class average was 66% with 53 students. The sample standard deviation is $s = 0.13$, or 13 percentage points. Use $\alpha = 0.05$.

a) Formulate the appropriate null and alternative hypothesis. Justify your formulation.
b) Calculate your test statistic and discuss your results. Show all your work.
c) Produce a 95% confidence interval estimate consistent with this test.

9-3A Single population proportion, one-tail test

Election years always bring us reports full of statistics on whose winning in the latest polls. One recent pollster argued that the republican candidate has support from more than half of registered voters. After digging a little deeper into the article, you find more details on the pollster's findings. In a footnote at the bottom of the page, you find the following information: Out of a sample of 200 registered voters, 110 stated that they support the Republican candidate.

a) On what information is the pollster's statement based? Is this fair and accurate?
b) Formulate the appropriate null and alternative hypothesis for a proper test.
c) Calculate your test statistic.
d) State and interpret your conclusion.

9-3B Single population proportion, two-tail test

It has been argued that drunk drivers cause 50 percent of accidents on the nation's highways. In order to test this claim, you obtain the following information: out of the last 100 accidents in your state, you find that 45 of them were caused by drunk driving.

a) Formulate the appropriate null and alternative hypothesis. Justify your formulation.
b) Calculate your test statistic.
c) State your conclusion and interpret its meaning.

Module 10: Hypothesis Testing – Two population means and proportions

10-1A Two population means, one-tail test, σ known

A friend of yours has claimed that Ford owners are, on average, faster drivers and Honda owners. Although you may not disagree with him, being the young statistician that you are, you decide to gather some data and perform a test. You set up a radar on the highway and begin collecting data. After one week, you've found the average speed of 42 Hondas was 61.3mph and the average speed of 53 Fords was 63.6mph. Assume that we know the population standard deviation of the Hondas to be $\sigma_H = 5.2$ and the Fords to be $\sigma_F = 5.4$.

a) Formulate the null and alternative hypothesis. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

10-1B Two population means, One tail test, σ known

Many students tend to choose courses based on which professor gives the easiest marks. One student seems very certain that Prof Dell is much easier than Prof Balley. In order to test this claim, you talk with other students who have taken courses with each of them. After talking with 43 of Prof Dell's past students, you find the average grade was 69%. You talk with 57 of Prof Balley's past students and calculate their average grade to be 66%. Assume that we know the population standard deviation of Prof Dell's grades to be σ_D= 0.12 and Prof Balley's grades to be σ_B = 0.14.

a) Formulate a hypothesis to test to test the student's claim. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

10-1C Two population means, One tail test, σ known

Imagine the following results come from two independent random samples taken from two populations.

	Sample A	Sample B
Count	35	30
Mean (Yards)	235	228
Standard deviation	5.8	3.9

a) Formulate a hypothesis to test to determine that the mean of population A is 5 yards greater than the mean of population B. Use α 0.03.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

10-1D Two population means, Two tail test, σ known

Anybody who has had siblings, knows how competitive they can become. Imagine, two brothers play catch with a baseball. It may start of as innocent play, but it won't be long before it becomes competitive with one bragging about their ability to throw further than the other. In order to settle the argument, Dad comes out to take measurements. After each brother throws the ball 50 times, Dad calculates Kid A, had an average distance of 44 feet and Kid B had an average distance of 46 feet. As one would expect, Kid B begin bragging as soon as he hears this news. Dad suggests that on average, the two are throwing the ball equal distance. Assume that we know the population standard deviation of Kid A's throws to be $\sigma_A = 7.3$ and Kid B's to be $\sigma_B = 8.2$.

a) Formulate a hypothesis to test to test the Dad's claim. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.
f) Confirm your findings with a confidence interval estimate.

10-1E Two population means, Two tail test, σ known

Imagine the following results come from two independent random samples taken from two populations.

	Sample A	Sample B
Count	40	40
Mean (lbs)	128	115
Standard deviation	4.8	5.7

a) Formulate a hypothesis to test to determine that the difference between the two means is 10 pounds. Use α 0.03.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.
f) Confirm your results using a confidence interval estimate.

10-2A Two population means, One tail test, σ unknown

A friend once told you that Golden Retrievers are a much faster breed of dog than Border Collie. As a dog lover, you become interested in determining whether or not the data would support such a claim. Assume you manage to gather 24 Golden Retrievers and 32 Border Collie for a massive 100-meter dog race! After the race, you gather all their times. You find the average time for the Golden Retriever to 10.5 seconds and for the Border Collies, 11.3 seconds. You calculate the sample standard deviations to be $s = 1.3$ and $s = 0.8$ seconds for the Retrievers and Border Collies, respectively.

a) Formulate a hypothesis to test your friend's claim. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

10-2B Two population means, Two tail test, σ unknown

University classes are becoming increasingly diversified, with students moving from all parts of the planet to study in difference countries. Imagine your statistics instructor gives you the following assignment:

Measure the heights of the students in your classes, and sort them by continent of origin. Perform a hypothesis test to determine if the average height of students from North America is different than the average height of students from Europe.

As the good student that you are, you awkwardly go around asking all of your classmates in your classes how tall they are. You obtain the following data:

	European	**North American**
Count	83	109
Mean (inches)	71	69
Standard deviation	6.2	4.7

f) Formulate the appropriate hypothesis test. Justify your formulation.
g) Calculate the test statistic.
h) Use the p-value approach to draw your conclusion.
i) Verify your conclusion using the critical value approach.
j) Interpret your conclusion.

10-3A Two population means, One tail test, Matched Sample

As a bilingual country, there are benefits to speaking both official languages in Canada. This is especially true for employees of the Canadian Public Service. In order to promote bilingualism, public servants have the opportunity to take courses in their second language at no cost to themselves. As a taxpayer funded training program, it's important to verify that it is effective. In order to test this, students are given an entry exam when they begin language training and an exit exam once they are finished. The difference between their grades on these two exams are used to determine if they have improved in their language proficiency. The following data is from the most recent data collection:

Student	**Entrance Exam**	**Exit Exam**	**Difference (Exit minus Entrance)**
1	25	34	9
2	45	52	7
3	68	67	-1
4	53	57	4
5	57	62	5
6	40	46	6

k) Formulate the appropriate hypothesis test. Justify your formulation.
l) Calculate the test statistic.
m) Use the p-value approach to draw your conclusion.
n) Verify your conclusion using the critical value approach.
o) Interpret your conclusion.

10-3B Two population means, Two-tail test, Matched Sample

Retail gasoline outlets frequently advertise the benefits of their fuel additives in maintaining a clean and smooth running engine. However, there is some disagreement on whether or not it affects fuel efficiency. In order to determine if the additive affect fuel efficiency, one chain of gas stations measured fuel efficiency of 5 cars without the additive, then again with the additive. The following table contains the data they collected:

Car	Without Additive	With additive	Difference (with minus without)
1	21	24	3
2	16	15	-1
3	13	16	3
4	19	17	-2
5	24	26	2

p) Formulate the appropriate hypothesis test. Justify your formulation.
q) Calculate the test statistic.
r) Use the p-value approach to draw your conclusion.
s) Interpret your conclusion.

10-4A Two population proportion, One tail test

Tourism plays an important role in the local economies of many small towns. Because tourism can be difficult industry to define, due to the broad spectrum of activities involved, of hotel occupancy rates are frequently used as a rough measure of the industry's overall performance. The following table provides hotel occupancy data for your home town, during the month of August for two consecutive years.

	This Year	**Last Year**
Occupied Rooms	1498	1365
Total Rooms	1600	1500

t) Formulate a test to determine if occupancy rates of increased.
u) Calculate the test statistic.
v) Draw your conclusion at the α=0.05 level of significance.
w) Interpret your conclusion.

10-4B Two population proportion, two tail test

Many employers are beginning to recognize the importance of maintaining a happy and healthy workforce. Some employers even provide their workers with arcades, child care facilities, libraries and all kinds of sporting activities such as tennis courts, pickle ball course and more. However, some people argue that these resources are not enjoyed equally by men and women. Therefore, it is important to ensure that the resources being offered are benefitting both genders equally. In order to ensure this to be true, one employer periodically samples its employees and asks whether or not they are satisfied with their current working environment. The following table provides the data collected.

	Women	**Men**
Satisfied respondents	51	49
Total number of employees	57	61

x) Formulate a test to determine both genders are equally satisfied with their workplace environment.
y) Draw your conclusion at the $\alpha=0.05$ level of significance.
z) Interpret your conclusion.

Module 11: Hypothesis Testing – Variances

11-1A Hypothesis Testing: Single population variance.

A local craft brewery claims the amount of beer in its bottles is 12oz (355ml). It knows that making false claims on its labels would result in serious penalties if it overstated the true volume. Every Monday morning, a sample of 30 bottles is taken to test the accuracy of their filling machines. Over the past few years of weekly sampling, they have calculated the standard deviation of the population to be $\sigma = 1.6$oz. They would like this to be a maximum level of variation in their filing process so, periodically, this is tested as well. This week, the sample resulted in a sample standard deviation of $s = 2.01$. Perform a test to determine if they are exceeding their maximum desired variance. Use $\alpha = 0.05$.

a) Formulate the null and alternative hypothesis. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

11-1B Hypothesis Testing: Single population variance.

A local farmer produces hay for nearby cattle ranchers. The hay is rolled into 50lb (22Kg) bails and are sold by quantity. Therefore, when a rancher buys 100 bales of hay, they can expect to receive 5,000lbs of hay. To ensure the cattle ranchers are getting what they expect, the farmer periodically samples batches of 30 hay bales to test that they are averaging 50lbs. The most recent batch provided a sample weight of $\overline{x}$ = 49.1lbs. It is important that the bails be relatively consistent in size as well. The historical population standard deviation is known to be $\sigma = 3.37$ and is considered to be an acceptable maximum. However, in this recent sample, the standard deviation was found to be 3.97. Using a level of significance of $\alpha = 0.05$ test to determine whether the variance in the size of hay bales is within the acceptable range.

a) Formulate the null and alternative hypotheses. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

11-1C Hypothesis Testing: Single population variance.

A lumberyard produces batches of 2x4s in lengths of 8 and 12 feet (96 and 144 inches, respectively). Each batch contains 50 2x4s. Because the lumber is used primarily for framing walls in home construction, it is imperative that the lengths be accurate. If a 2x4 is too short, or too long, it will cause delays in construction, as adjustments will have to be made. In an attempt to reduce the variation in cutting lengths, the lumber yard as taken steps to increase the precision of its saws. The lumber yard knows the current standard deviation of 8 foot lengths is $\sigma_8 = 0.82$ inches and for 12 foot lengths is $\sigma_{12} = 1.05$ inches. On the first business day of each month, one batch of each length is used to test the accuracy of their cuts. The standard deviation of their 8 foot 2x4s is $s = 0.62$ inches and the standard deviation of their 12 foot 2x4s is $s = 0.84$ inches. Using a level of significance of $\alpha = 0.01$, test to determine whether they have succeeded at reducing the standard deviation of the two lengths of lumber.

a) Formulate the null and alternative hypotheses. Justify your formulation.
b) Calculate the test statistics for both tests.
c) Use the p-value approach to draw your conclusions.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.

11-1D Hypothesis Testing: Single population variance.

An instructor at your university has produced a series of problems and accompanying video walkthroughs in hopes of improving his students' understanding of course content. Having taught the course many times, he calculates the overall average to be 71% and he determines that the population standard deviation is $\sigma = 0.12$ (or 12 percentage points). At the end of the following semester his class of 40 students, who had access to the video walkthroughs, had an average grade of $\bar{x} = 74.6\%$. He also calculated the sample standard deviation to be $s = 0.15$. Using a level of significance of $\alpha = 0.05$, test to determine whether or not the assumed population standard deviation of $\sigma = 0.12$ is correct.

a) Formulate the null and alternative hypothesis. Justify your formulation.
b) Calculate the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Verify your conclusion using the critical value approach.
e) Interpret your conclusion.
f) Develop a 95% confidence interval for the population standard deviation.

11-1E Hypothesis Testing: Single population variance.

A sample of 25 observations has a standard deviation of s = 1.9. Test the following hypothesis:

$$H_o: \sigma^2 = 7.9$$
$$H_a: \sigma^2 \neq 7.9$$

a) Use the p-value approach to draw your conclusion.
b) Develop a 95% confidence interval for the population variance.

11-2A Hypothesis Testing: Two population variance.

One of the challenges in writing in-class exams is to ensure that they can be completed within the allotted time. Even though the average completion time is within requirements, the variance can often be a problem. Some students might finish in ten minutes, while others run out of time. In an attempt to reduce the variance of completion times, a new computerized method of testing has been implemented. In order to determine if the new method was successful at reducing the variance of completion times, a sample 30 students were asked to write the exam using the old method and a sample of 35 students wrote the exam using the new method. The standard deviation of completion times using the old method was s = 9.2 minutes and the standard deviation of completion times using the new method was s = 6.7 minutes. Test to determine if the new method resulted in a lower variance of completion time.

a) Develop the null and alternative hypothesis.
b) Draw your conclusion using the p-value and the critical value approaches.

11-2B Hypothesis Testing: Two population variance.

Modern office buildings often have hi-tech centralized heating and cooling systems. These can greatly improve energy efficiency if used properly. A problem can arise if, in response to being too cold, one person decides to bring their own private space heater for their office. By heating their own space they may cause the thermostats to think the building is warmer than it actually is. The result is that the heating system fails to turn on in one area because the other area is being heated by a different source. The result is that others become cold causing them to each bring in their own space heaters and the problem continues. In order to try to remedy this problem, measurements are taken in a variety of locations across two different sections of the building. In the first section, the average temperature in 15 locations was 20.8 degrees Celsius with a standard deviation of 2.8. In the second section, the average temperature across 21 locations was 18 with a standard deviation of 1.9.

a) Develop a test to determine if there is a difference in variance between the two sections.
b) Draw your conclusions using the P-value approach and critical value approach.
c) Develop a test to determine if there is a difference in average temperature between the two sections.
d) Draw your conclusions using the p-value approach.

11-2C Hypothesis Testing: Two population variance.

Copied from exercise 10-2A:

A friend once told you that Golden Retrievers are a much faster breed of dog than Border Collie. As a dog lover, you become interested in determining whether or not the data would support such a claim. Assume you manage to gather 24 Golden Retrievers and 32 Border Collie for a massive 100-meter dog race! After the race, you gather all their times. You find the average time for the Golden Retriever to 10.5 seconds and for the Border Collies, 11.3 seconds. You calculate the sample standard deviations to be s = 1.3 and s = 0.8 seconds for the Retrievers and Border Collies, respectively.

a) Develop a test to determine if the assumption of unequal variances was appropriate.
b) Draw your conclusions at the 0.05 level of significance.

11-2D Hypothesis Testing: Two population variance.

Copied from exercise 10-2B:

University classes are becoming increasingly diversified, with students moving from all parts of the planet to study in difference countries. Imagine your statistics instructor gives you the following assignment:

Measure the heights of the students in your classes, and sort them by continent of origin. Perform a hypothesis test to determine if the average height of students from North America is different than the average height of students from Europe.

As the good student that you are, you awkwardly go around asking all of your classmates in your classes how tall they are. You obtain the following data:

	European	North American
Count	83	109
Mean (inches)	71	69
Standard deviation	6.2	4.7

a) Develop a test to determine if the assumption of unequal variances was appropriate.
b) Draw your conclusions at the 0.05 level of significance.

Module 12: Multiple Proportions, Independence and Goodness of Fit

12-1A Testing Equality Across Multiple Population Proportions

As part of an undergraduate research project, you decide to determine whether pet owners are satisfied with their choice of pet. In order to gather sample data, you develop a survey to ask respondents what type of pet they have (Dog, Cat, Other) and whether they are likely to adopt the same pet upon its death, or a different one. The following table contains the observed frequencies:

		Type of Pet			
		Cat	**Dog**	**Other**	**Total**
Likely to readopt	**Yes**	34	52	24	110
	No	56	40	42	138
	Total	90	92	66	248

a) Formulate the null and alternative hypothesis.
b) Compute the Expected frequencies.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.
e) If appropriate, use the Marascuillo procedure to determine where any difference exist.

12-1B Testing Equality Across Multiple Population Proportions

After the most recent election, you decide to determine if there was a difference in the proportion of voters who changed their voting intentions at the last minute. Some voters choose early on in the campaign who they will vote for and stick with it, while others may change their minds when new information becomes available. This might shed some light on which voters are more susceptible to information released close to election day. In order to gather data, you produce a survey that asks each respondent which party they voted for and if this was a result of a change of intentions within the 4 weeks prior to Election Day. The following table provides the observed frequencies:

		Political Party			
		Republican	**Democratic**	**Other**	**Total**
Changed voting intention	**Yes**	112	82	83	277
	No	98	79	49	226
	Total	210	161	132	503

a) Formulate the null and alternative hypothesis.
b) Compute the Expected frequencies.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.
e) If appropriate, use the Marascuillo procedure to determine where any differences exist.

12-1C Testing Equality Across Multiple Population Proportions

A local car dealership is interested in determining customer satisfaction and brand loyalty. The sample owners of three different brands of vehicles and ask whether or not they are likely to be the same brand when they purchase a new car, or shop around. The following table contains the observed frequencies:

		Brand of vehicle			
		Ford	**GMC**	**Chevy**	**Total**
Likely to repurchase	**Yes**	95	78	103	276
	No	61	89	115	265
	Total	156	167	218	541

a) Formulate the null and alternative hypothesis.
b) Compute the Expected frequencies.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.
e) If appropriate, use the Marascuillo procedure to determine where any difference exist.

12-2A Tests of Independence

The local animal shelter is interested in knowing if people decision to adopt a pet or purchase from a breeder is independent of the type of pet. Knowing which pets are more likely to be adopted will help them manage their inventories. The following table provides the observed frequencies:

	Type of Pet			
	Cat	Dog	Other	Total
Adopt	62	59	51	173
Purchase	35	68	63	166
Total	97	127	114	339

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-2B Tests of Independence

Is political affiliation independent of gender? The following table provides the observed frequencies from a recent survey of students at your college:

	Political Party			
	Republican	**Democratic**	**Other**	**Total**
Male	48	36	15	99
Female	43	39	16	98
Total	91	75	31	197

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-2C Tests of Independence

A local car dealership has reason to believe that whether a customer is married or single will influence the brand of vehicle they choose to buy. The following table contains the observed frequencies collected using a survey.

	Brand of vehicle			
	Ford	**GMC**	**Chevy**	**Total**
Married	90	74	131	295
Single	70	91	101	262
Total	160	165	232	557

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and the test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-3A Multinomial Goodness of Fit Test

Is there a preference in the type of pet people own? The following table provides the observed frequencies:

	Type of Pet				
	Cat	**Dog**	**Bird**	**Other**	**Total**
Number of Pets owned	97	120	114	100	431

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-3B Multinomial Goodness of Fit Test

Is there a preferred political party among your classmates? The following table provides the observed frequencies from a recent survey of students at your college:

	Political Party			
	Republican	**Democratic**	**Other**	**Total**
Total	89	98	42	229

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-3C Normal Distribution Goodness of Fit Test

A simple random sample of 30 grades from a principles of microeconomics course are listed below. They have been sorted from largest to smallest for convenience. The mean grade is 0.61 with a standard deviation 0f 0.17.

20%	54%	69%
30%	56%	70%
38%	59%	71%
40%	60%	72%
46%	60%	78%
49%	63%	79%
49%	65%	82%
50%	66%	82%
54%	67%	90%
54%	69%	94%

a) Formulate the null and alternative hypothesis.
b) Compute the expected frequencies and test statistic.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

12-3D Normal Distribution Goodness of Fit Test

A local painting company that employs students is doing some analysis on the completion times of its employees. One part of the analysis is to determine if the completion times are normally distributed or not. The data below consists of the number of minutes it took each of 30 employees to finish painting a small bedroom. The average completion time was 73.93 minutes with a standard deviation of 15.41. The data has been sorted from largest to smallest for convenience.

44.17	67.09	80.11
49.23	68.05	81.20
52.46	69.59	83.52
57.00	70.97	84.36
58.27	71.00	86.20
59.31	71.24	95.81
62.40	74.96	98.71
63.33	76.66	99.50
66.29	77.97	99.73
67.06	79.38	102.30

a) Formulate the null and alternative hypothesis.
b) Compute the actual and expected frequencies.
c) Use the p-value approach to draw your conclusion.
d) Interpret your conclusion.

Module 13: Analysis of Variance

13-1A Single Factor Analysis of Variance (ANOVA) – Completely Randomized Design

WhiteTooth Inc. is developing an additive for its line of toothpastes that is designed to whiten teeth in as little time as possible. It currently has two variations of the additive, Type A, and B, but only wishes to produce and market one. In order determine the effectiveness of these new additives, a focus group consisting of 15 people is organized. Five are given type A, five are given type B and five are the control group and are given a placebo. Each person is asked to use the toothpaste and record the time of days it takes before their teeth achieve a predetermined shade of white. The following table contains the data collected:

Observation	**Type A**	**Type B**	**Control Group (Placebo)**
1	5	4	9
2	5	5	7
3	7	6	8
4	4	5	9
5	6	8	7
Standard Deviation	**1.1**	**1.5**	**1.0**
Sample Mean	**5.4**	**5.6**	**8**

a) Test to determine whether or not there is a difference between the two types of additives and the control group.
b) Perform a Fisher's LSD test if necessary.

13-1B Single Factor Analysis of Variance (ANOVA) – Observational Study

Students in different college majors are always complaining (or bragging) about how difficult their field of study is relative to another. You decide that perhaps you could use the number of hours spent studying as a proxy for the level of difficulty. The more hours spent studying, the more difficult the subject matter must be. You survey students across 3 fields of study and ask them how many hours per day then spend studying outside of class time. The following table contains the data collected:

Observation	Accounting	Physics	Sociology
1	2.5	3.2	3.1
2	3.6	4.3	1.8
3	4.3	1.4	2.7
4	3.1	2.3	3.6
5	3.4	2.9	4.1
6		2.5	2.9
7			3.2
Standard Deviation	**0.66**	**0.97**	**0.72**
Sample Mean	**3.4**	**2.8**	**3.1**

a) Test to determine whether there is a difference in the average number of hours spent studying between the three college majors.
b) Perform a Fisher's LSD test if necessary.

13-1C Single Factor Analysis of Variance (ANOVA) – Completely Randomized Design

A new type of glass is being developed to use in areas at risk of earthquakes. Three types of glass have been developed, but the company only wishes to manufacture one. In order to test the strength of the glass, window panes of identical sizes were placed in a machine designed to shake the glass in a manner that simulates the stress it would have to endure in an earthquake. In the most severe earthquakes, the shaking can last as long as 5 minutes. The amount of time before each window pane shattered was recorded. The following table contains the data collected:

Observation	Type I	Type II	Type III
1	5.1	5.3	5.3
2	4.8	5.6	5.9
3	4.9	4.8	6.1
4	5.3	5.3	6.3
5	5.9	5.8	5.8
6		5.7	5.4
7			5.9
Sample Variance	**0.19**	**0.13**	**0.13**
Sample Mean	**5.2**	**5.4**	**5.8**

a) Test to determine whether or not there is a difference between the two types of additives and the control group.
b) Perform a Fisher's LSD test if necessary.

13-2A Single Factor Analysis of Variance (ANOVA) – Randomized Block Design

Everything Co. frequently relies on courier services to deliver sensitive documents between its regional offices. There are three courier companies available, each one offers loyalty discounts giving the incentive for customers to choose one courier and stick with it. You decide to perform a test to determine if there is a difference in delivery times between the three courier options you have. You send three packages of equal size to each of five regional offices. Each package is sent through one of the three couriers. The following table contains the delivery times, in hours:

Regional Office	Option A	Option B	Option C	Block Mean
1	28.3	23.7	28.5	**26.8**
2	49.7	46.8	46.3	**47.6**
3	20.8	21.3	22.5	**21.5**
4	35.8	30.6	34.6	**33.7**
5	22.4	19.2	20.8	**20.8**
Treatment Mean	**31.4**	**28.3**	**30.5**	

a) Test to determine whether or not there is a difference between the three courier services. (Hint: SST = 1511.52)

13-2B Single Factor Analysis of Variance (ANOVA) – Randomized Block Design

Canine Munchies Inc. is developing a new brand of dog food designed specifically for less active dogs. They have developed two types of food all with a lower fat, higher protein blend of ingredients in order to minimize weight gain; a common problem among its target market of inactive dogs. In order to determine if there is a difference between the two new brands of food as compared to the dogs' regular diet, a group of five dogs were each fed the three types of food for a period of 30 days. Two of the three were the new brands, and the third was the dogs' original diet. The data below shows the difference in the dogs' weight between the first and 30th day on the diet. A positive number indicates a weight gain; a negative number indicates a weight loss.

Dog	Type 1	Type 2	Original Diet	Block Mean
1	-1	3	2	**1.33**
2	2	0	2	**1.33**
3	-1	3	-1	**0.33**
4	1	4	0	**1.67**
5	-4	-1	-1	**-2.00**
Treatment Mean	**-0.6**	**1.8**	**0.4**	

a) Test to determine whether or not there is a difference between the two types of dog food and their original diet in terms of their impact on the dogs' weight. (Hint: SST = 63.73)

13-3A Two Factor with Replication Analysis of Variance (ANOVA) – Factorial Design.

The local animal rescue shelter is interested in knowing if there is a significant difference in the number of animal adoptions between its three largest shelters during its busiest weekend of the year. It also is interested in knowing if there is a significant difference between the number of cats and dogs adopted as well. The following data shows the number of each animal adopted at each of its three shelters on each day of the long weekend.

	A	B	C	Treatment Means
	2	2	1	
Dogs	4	1	2	**2.22**
	3	1	4	
Interaction Means	**3.00**	**1.33**	**2.33**	
	4	2	4	
Cats	5	2	3	**3.22**
	4	3	2	
Interaction Means	**4.33**	**2.33**	**3.00**	**Grand mean**
Treatment Means	**3.66**	**1.83**	**2.67**	**2.72**

a) Test to determine whether or not there is a difference between the average number of adoptions by location, by animal, type and interaction. (SST = 25.6)

13-3BTwo Factor with Replication Analysis of Variance (ANOVA) – Factorial Design.

A designer of commercial retail space is doing a study to determine which method of managing line ups at the till works best. Method A involves many smaller line up at individual tills. Method B involves one large line up being served by multiple tills. The table below contains the wait times, in minutes, of customers in three different retail settings using both of the two proposed methods.

	Grocery	**Electronics**	**Toys**	**Treatment Means**
	5	8	2	
Method 1	6	6	1	**4.44**
	5	5	2	
Interaction Means	**5.33**	**6.33**	**1.67**	
	3	5	3	
Method 2	2	3	4	**3.22**
	2	2	5	
Interaction Means	**2.33**	**3.33**	**4.00**	**Grand mean**
Treatment Means	**3.83**	**4.83**	**2.83**	**3.83**

a) Test to determine whether or not there is a difference in the average number of minutes waiting by queue method, retail setting and for interaction. (SST = 60.5)

Module 14: Simple Linear Regression

14-1A Simple Linear Regression

There's a strong belief that student performance is directly linked to the amount of time the student spends studying. In order to test this claim, you gather the following data:

Observation	Grade	Hours of study
1	24	2.2
2	36	3.9
3	68	5.3
4	55	3.7
5	81	5.1
Mean	**52.8**	**4.0**

f) Fill in the blanks in the table below. Interpret your results.

g) Use the estimated regression equation to develop a confidence interval estimate for the average grade for somebody who studies 5 hours.

Regression Statistics	
Multiple R	
R Square	
Adj R Square	0.74
Standard Error	
Observations	

ANOVA

	df	*SS*	*MS*	*F*	*p-value*
Regression		1723.56			
Error			139.75		
Total					

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept		19.85				
Hours of study		4.74				

14-1B Simple Linear Regression

Use the following information to estimate the corresponding demand curve:

Observation	Quantity	Price
1	3	61
2	5	57
3	8	57
4	10	52
5	12	53
Mean	**7.6**	**56**

h) Fill in the blanks in the table below. Interpret your results.

i) Use the estimated regression equation to develop a prediction interval estimate for the quantity demanded at a price of 56.

Regression Statistics	
Multiple R	
R Square	0.83
Adj R Square	0.78
Standard Error	
Observations	

ANOVA

	df	*SS*	*MS*	*F*	*p-value*
Regression					
Error		8.89			
Total					

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept			4.43			
Price		0.24				

14-1C Simple Linear Regression

My parents once told me that the longer they're married, the happier they were. Let's test this claim. You find the following date:

Observation	Years Married	Happiness Index
1	15	80
2	10	60
3	36	75
4	44	90
5	54	95
Mean	**31.8**	**80**

j) Fill in the blanks in the table below. Interpret your results.

k) Use the estimated regression equation to develop a confidence interval estimate for the average happiness for somebody who is married for 20 years.

Regression Statistics	
Multiple R	0.84
R Square	
Adj R Square	0.62
Standard Error	
Observations	

ANOVA

	df	*SS*	*MS*	*F*	*p-value*
Regression					
Error					
Total		750			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept		8.11			34.67	86.27
Years Married		0.23	2.73			

Module 15: Multiple Regression

15-1A Multiple Linear Regression

The following estimated regression equation states that quantity sold of a good, is a function of its own price (P_x), the price of a related good (P_y), advertising expenditures (A) and average household income (M):

$$E(Q_d) = \beta_0 + \beta_1 P_x + \beta_2 P_y + \beta_3 A + \beta_4 M$$

Prices are measured in dollars, while advertising and income are measured in thousands of dollars. The following table provides the estimated regression results:

Regression Statistics	
Multiple R	0.92
R Square	0.85
Adjusted R Square	0.81
Standard Error	86.14
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	671325.08	167831.27	22.62	0.00
Residual	16	118710.43	7419.40		
Total	20	790035.51			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	686.63	169.67	4.05	0.00	326.94	1046.32
own price	-11.64	2.31	-5.04	0.00	-16.54	-6.74
related price	-12.22	2.72	-4.50	0.00	-17.98	-6.46
advertising	6.37	2.12	3.00	0.01	1.88	10.87
average income	6.58	3.77	1.74	0.10	-1.42	14.59

a. Write the estimated regression equation.
b. Interpret each of the estimated coefficients and the corresponding interval estimates.
c. Interpret the R-squared.
d. Interpret the results of the tests for individual parameter significance and overall model significance.
e. Estimate the expected quantity sold, when the prices of the two goods, x and y are $25 and $15, respectively, advertising expenditures are $30,000 and average income is $45,000.

15-1B Multiple Linear Regression - Multicollinearity

With a belief that an individual's salary can be predicted by their age and experience, the following regression equation is estimated.

$$E(Salary) = \beta_0 + \beta_1(Exp) + \beta_2(Age)$$

Salary is measured in thousands of dollars, while experience and age are measured in years.

Regression Statistics	
Multiple R	0.64
R Square	0.41
Adjusted R Square	0.34
Standard Error	10.99
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1423.06	711.53	5.89	0.01
Residual	17	2052.29	120.72		
Total	19	3475.35			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-27.10	47.04	-0.58	0.57	-126.35	72.15
Experience	-2.24	2.26	-0.99	0.34	-7.01	2.54
Age	4.55	2.38	1.91	0.07	-0.47	9.56

a. Write the estimated regression equation.
b. Interpret the R-squared.
c. Interpret the p-values and confidence interval estimates.
d. Discuss the results of the tests on individual parameters and the model.

15-1C Multiple Linear Regression

The following estimated regression equation states that wheat yield (pounds), is a function of it's the average monthly rainfall in inches (R), the density of seed dispersion (S) in seeds per square inch, average daily temperature, degrees Fahrenheit (T) and an index measuring the quality of fertilizer (F):

$$E(Y) = \beta_0 + \beta_1 R + \beta_2 S + \beta_3 T + \beta_4 F$$

The following table provides the estimated regression results:

Regression Statistics	
Multiple R	0.83
R Square	0.68
Adjusted R Square	0.60
Standard Error	115.36
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	458409	114602.26	8.61	0.00
Residual	16	212912.9	13307.06		
Total	20	671321.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1064.68	219.70	4.85	0.00	598.95	1530.42
rainfall	11.35	3.15	3.61	0.00	4.68	18.02
seed density	7.90	3.09	2.56	0.02	1.35	14.45
temperature	5.36	4.55	1.18	0.26	-4.29	15.00
fertilizer quality	13.76	4.32	3.18	0.01	4.60	22.93

a. Write the estimated regression equation.
b. Interpret the R-square.
c. Interpret the coefficients and corresponding confidence interval estimates
d. Interpret the p-values for the tests on individual parameter significance and overall model significance.
e. Estimate the expected wheat yield when the monthly rainfall is 2 inches, the seed is distributed at 20 seeds per square inch, the daily temperature is 45 and the fertilizer quality index is 65.

15-2A Multiple Linear Regression – Dummy Variables

In problem 15-1B, we developed an estimated regression equation that demonstrated the problem of multicollinearity. This revised model excludes years of experience, which was found to be highly correlated with age. We have now added to the model two dummy variables to estimate the effect of educational attainment on salary. The new model is as follows:

$$E(Salary) = \beta_0 + \beta_1(AGE) + \beta_2(MA) + \beta_3(PhD)$$

Salary is measured in thousands of dollars; age is measured in years. MA equals one if the individual has a Master's degree, zero otherwise. PhD equal one if the individual has a doctorate, and is zero otherwise. The estimate regression output is below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.94
R Square	0.88
Adjusted R Square	0.85
Standard Error	10.22
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	11977.32	3992.44	38.25	0.00
Residual	16	1670.01	104.38		
Total	19	13647.33			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	30.57	17.78	1.72	0.10	-7.12	68.27
Age	4.78	0.59	8.06	0.00	3.52	6.03
Master	11.08	5.70	1.94	0.07	-1.00	23.17
Doctorate	31.12	5.78	5.39	0.00	18.88	43.37

a. Write the estimated regression equation.
b. Interpret the R-square.
c. Interpret the coefficients and corresponding confidence interval estimates
d. Interpret the p-values for the tests on individual parameter significance and overall model significance.

15-2B Multiple Linear Regression – Dummy Variables

In problem 15-1A, we estimated a demand equation and found income to be statistically insignificant. The following regression equation removes income from the model, but now includes a dummy variable for gender, with the view that men and women have difference consumption habits. The model now states that quantity demanded, is a function of its own price (P_x), the price of a related good (P_y), advertising expenditures (A) and the gender of the consumer.

$$E(Q_d) = \beta_0 + \beta_1 P_x + \beta_2 P_y + \beta_3 A + \beta_4 G$$

Prices are measured in dollars, while advertising is measure in thousands of dollars. The dummy variable takes the value zero for men and one for women. The following table provides the estimated regression results:

Regression Statistics	
Multiple R	0.84
R Square	0.71
Adjusted R Square	0.64
Standard Error	110.26
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	480027.63	120006.91	9.87	0.00
Residual	16	194517.03	12157.31		
Total	20	674544.67			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	765.29	139.43	5.49	0.00	469.71	1060.87
own price	-8.83	2.55	-3.46	0.00	-14.24	-3.42
related price	-11.13	3.93	-2.83	0.01	-19.45	-2.80
advertising	5.81	1.85	3.15	0.01	1.90	9.73
Gender	221.21	51.07	4.33	0.00	112.94	329.48

a. Write the estimated regression equation.
b. Interpret the R-square.
c. Interpret the coefficients and corresponding confidence interval estimates
d. Interpret the p-values for the tests on individual parameter significance and overall model significance.

15-2C Multiple Linear Regression – ANOVA

Copied from problem 13-1B: Refer to problem 13-1B to see the original data set.

Students in different college majors are always complaining (or bragging) about how difficult their field of study is relative to another. You decide that perhaps you could use the number of hours spent studying as a proxy for the level of difficulty. The more hours spent studying, the more difficult the subject matter must be. You survey students across 3 fields of study and ask them how many hours per day then spend studying outside of class time. We will use regression analysis to estimate the follow regression equation:

$$E(Hours) = \beta_0 + \beta_1(Phy) + \beta_2(Soci)$$

We have defined Accounting to be the base case, with *Phy* identifying students who major in physics, and *Soci* to identify sociology students. The following table provide the estimated regression results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.31
R Square	0.10
Adjusted R Square	-0.02
Standard Error	0.80
Observations	18

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.034	0.51	0.80	0.47
Residual	15	9.60	0.64		
Total	17	10.63			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.38	0.36	9.45	0.00	2.62	4.14
Phys	-0.61	0.48	-1.27	0.22	-1.65	0.42
Soci	-0.32	0.47	-0.69	0.50	-1.32	0.68

a. Write the estimated regression equation.
b. Interpret the R-square.
c. Interpret the coefficients and corresponding confidence interval estimates
d. Interpret the p-values for the tests on individual parameter significance and overall model significance.

15-2D Multiple Linear Regression – ANOVA

Copied from problem 13-1C: Refer to problem 13-1C to see the original data set.

A new type of glass is being developed to use in areas at risk of earthquakes. Three types of glass have been developed, but the company only wishes to manufacture one. In order to test the strength of the glass, window panes of identical sizes were placed in a machine designed to shake the glass in a manner that simulates the stress it would have to endure in an earthquake. In the most severe earthquakes, the shaking can last as long as 5 minutes. The amount of time before each window pane shattered was recorded. The following table contains the data collected:

$$E(Min) = \beta_0 + \beta_1(T2) + \beta_2(T3)$$

We have defined Type 1 glass to be the base case, with *T2* identifying Type 2 glass, and *T3* identifying Type 3 glass. The following table provide the estimated regression results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.59
R Square	0.35
Adjusted R Square	0.26
Standard Error	0.38
Observations	18

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.18	0.59	4.03	0.04
Residual	15	2.20	0.15		
Total	17	3.38			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.20	0.17	30.38	0.00	4.84	5.56
Type 2	0.22	0.23	0.93	0.36	-0.28	0.71
Type 3	0.61	0.22	2.74	0.02	0.14	1.09

a. Write the estimated regression equation.
b. Interpret the R-square.
c. Interpret the coefficients and corresponding confidence interval estimates
d. Interpret the p-values for the tests on individual parameter significance and overall model significance.