RG Cox

ASU ECN221-Business Statistics Practice Homework

November 14, 2016

Nitaka Mountain

Below are problems that you can work for practice. Working these problems does not directly impact your grade; they are not worth any points in the class. I created them for you so that you have an additional opportunity to practice before you do the homework which does count towards your grade.

Contents

1

Practice Homework 1

These problems cover topics that are prerequisites for the course. Complete these before doing Homework 1. The answers are at the end.

1.1 Combinations

Combinations is a topic covered in MAT211. If you have forgotten how to compute the number of combinations or if you took a class other than MAT211 to satisfy the prerequisite then you should review this material. The basic formula to use in these problems is

$$_nC_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

1. Suppose I have a test bank with 50 questions. I am going to give a quiz with 10 questions. How many distinct combinations of questions are possible?
2. If the tutoring center is going to hire 5 tutors and 12 people apply for those openings how many different combinations of hired tutors are possible?

1.2 Mean, median, mode

3. What is the average or mean for the following data: 1, 2, 5, 9, 10, 20, 21?
4. What is the median for the following data: 1, 2, 5, 9, 10, 20, 21?
5. What is the mode for the following data: 17, 22, 5, 5, 5, 9, 10, 20, 20, 21?

1.3 Variance and Standard Deviation

Recall that the standard deviation is the square root of the variance.
6. If the standard deviation is 17 what is the variance?
7. If the variance is 26.4, what is the standard deviation?

1.4 Probabilities

Use this table for the remaining questions.

<div align="center">survey data</div>

Political Party	Num. Surveyed	Num. Supporting Gun Control Laws
Democratic	11	8
Republican	14	1
Other	9	5

8. How many people were surveyed?
9. How many Republicans were surveyed?
10. What is the probability that a person does not support gun control laws?
11. What is the probability that a democrat supports gun control laws?

KEY

1. 50 choose 10 is 50!/(10!40!)=10,272,278,170.
2. 12 choose 5 is 12!/(5!7!)=792. Keep in mind that 5!=5*4*3*2*1.
3. 1+2+5+9+10+20+21=68. Then we take the sum and divide by "n" which is 7, 69/7=9.714 rounding to the third decimal place.
4. The median leaves at least half of the observations at or below and at least half at or above. With 7 observations this will be the middle value which is 9. Check the book for the slightly different case when there are an even number of observations unless you remember how to do this from MAT211.
5. The mode is the value which appears most frequently. In this case the value 5 appears three times, which is more than any other value, so 5 is the mode.
6. Take 17 and square it to get which is 289.
7. Take the square root of 26.4 which is approximately 5.138.
8. 11+14+9=34 people surveyed.
9. 14 republicans, which you see by looking in the row that says "Republican."
10. There are 8+1+5=14 people supporting gun control laws so that there are 34-14=20 people not supporting gun control laws. Then we take 20 out of 34 to get 20/34=.588 as the probability of not supporting gun control laws (the value is approximate since it was rounded at the third decimal).
11. In this case we are focusing only on democrats so we look only in that row and see 8 out of 11 support the laws so we get 8/11=.727 as the rounded probability.

Practice Homework 2

Here are some practice problems for you to work before doing homework 2. The answers are on the last page(s) of this section. This homework primarily requires you to know some definitions in addition to being able to work the types of problems below. For definitions you must simply consult lecture notes and the text book.

2.1 Cross Tabulation Tables

This table shows grades in the prerequisites for ECN221 and whether or not the student has taken a previous course in statistics for a sample of ENC221 students.

Previous statistics class	MAT 211 Grade			
	A	B	C	total
No	109	129	75	313
Yes	57	81	49	187
Total	166	210	124	500

1. How many students received an A in MAT211?
2. How many students did not previously take a statistics class and got a C in MAT211?
3. What percentage of students received a B in MAT211?
4. What percentage of students have already had a statistics class?
5. What percentage of students that have already taken a statistics class got a C in MAT211?

2.2 Reading a stem and leaf display

Warning: not all stem and leaf displays are identical. They have the same basic structure but depending on the software program used they have slightly different appearances.

Stem (10)	Leaf (1)
5	58
6	35568
7	111224677
8	00005699
9	11226

6. What is the largest value in the display above?
7. What is the smallest value in the display above?
8. What is the mode in the data displayed above?

2.3 Frequency Tables and Cumulative Frequency Tables

		MAT 211 Grade	
MAT211 Grade	frequency	relative frequency	percent frequency
E	22		
D	42		
C	124		
B	210		
A	166		

		MAT 211 Grade	
MAT211 Grade	Cumulative freq.	Cumulative relative freq.	Cum. percent freq.
E	22	.039	3.9
D			
C			
B			
A			

9. Complete the frequency table.
10. Complete the cumulative frequency table.

KEY

1. Read from the "Total" row and the "A" column to find 166.
2. Read from the "No" row and the "C" column to find 75.
3. 210/500=.42 which translates to 42%.
4. 187/500=.374 which translates to 37.4%.
5. This questions restricts attention to the students that "have already taken a statistics class" so we are only looking at the "Yes" row of the table. Then 49/187=.262 which translates to 26.2%.
6. 96. 9 is the largest value in the "stem" which is the 10s place. For the stem of 9 the largest leaf is 6. That gives 96.
7. 55.
8. 80. There are four instances of 80 and no other value appears as frequently.
9. The completed table should look as follows:

MAT211 Grade	MAT 211 Grade		
	frequency	relative frequency	percent frequency
E	22	.039	3.9
D	42	.074	7.4
C	124	.22	22
B	210	.372	37.2
A	166	.294	29.4

First you must find the total by adding all the values in the "frequency" column, 166+210+124+42+22=564. Then take each value in the "frequency" column and divide it by the total of 564 to find the relative frequency. Finally, multiply those values by 100 to find the percent frequency.

10. The completed table is below. You need to use the relative frequency table to help you complete the cumulative frequency table.

MAT211 Grade	MAT 211 Grade		
	Cumulative freq.	Cumulative relative freq.	Cum. percent freq.
E	22	.039	3.9
D	64	.113	11.3
C	188	.333	33.3
B	398	.706	70.6
A	564	1	100

3

Practice Homework 3

Here are some practice problems for you to work before doing homework 3. The answers are on the last page(s) of this section. This homework primarily requires you to know some definitions in addition to being able to work the types of problems below. For definitions you must simply consult lecture notes and the text book.

3.1 Mean, variance, and percentiles with or without excel

From the course webpage (ECN221 page) get the spreadsheet with practice homework data and go to the "cotton prices" data set. In column B starting on row 12 are the prices of cotton, given in cents per pound, in the US from 1941 through 1951. I pulled these data from the St. Louis Federal Reserve Bank website.

If you have forgotten how to find averages and variances using excel then please review the videos available on Bb.

1. What was the average price of cotton from 1941-1951 (the average in the data set)?
2. Using the same cotton prices data what is the variance of the price?
3. Using the same data what is the standard deviation?
4. Using the same data what is the median price?
5. Using the same data what is the 25th percentile of the price?
6. Using the same data what is the 75th percentile of the price?
7. Using the same data what is the interquartile range?
8. Using the following values find the mean and the standard deviation: 8, 10, 15, 16, 23.
9. Suppose you have the following values 11, 1, 2, 102, 3, 21, 133, 23, 72, 100, 25, 105, 200. Find the interquartile range.

3.2 Correlation

10. Find the correlation when the standard deviation of x is 9 and the standard deviation of y is 13 and the covariance is -40
11. Find the correlation when the standard deviation of x is 10 and the standard deviation of y is 7 and the covariance is 23.

3.3 z-scores

12. Find the z value or z score for x=7 when the mean is 10 and the standard deviation is 2.4.
13. Find the z value or z score for x=12 when the mean is 10 and the standard deviation is 2.4.

KEY

1. Using the "average" function find 50.3.
2. Using the "var.s" function find 214.44.
3. Take the square root of 214.44 to find 14.64.
4. There are 132 observations. Sort them. You can do this in excel under the "DATA" tools using "SORT" as shown here:
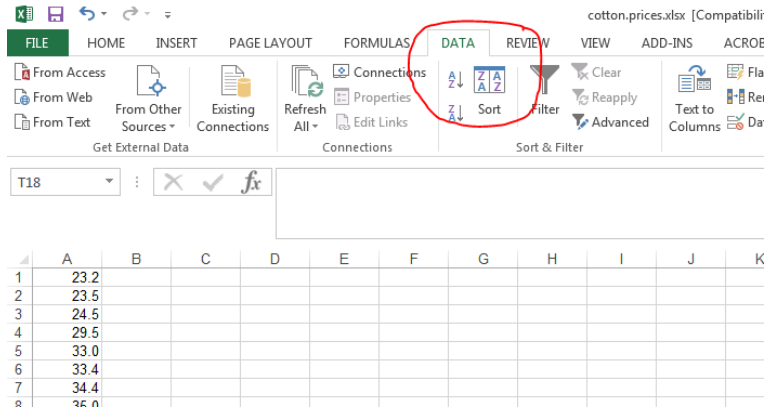
   To find the median you will need to find a value where 66 values are at or below and 66 values are at or above. The 66th and 67th values are 47.3 and 50.0. Taking the average of those two we get 48.7. This is the median.
5. Take .25*132=33 so we need the average of the 33rd and 34th values. (36.8+37.2)/2=37.
6. Take .75*132=99 so we need the average of the 99th and 100th values. (61.2+61.4)/2=61.3.
   Caution: if there had been 134 observations we would have had .75*134=100.5 and we would have simply taken the 101st value without averaging two numbers.
7. IQR=61.3-37=24.3. The numbers 61.3 and 37 came from the previous questions where you found the 25th and 75th percentiles.
8. The mean is (8+10+15+16+23)/5=72/5=14.4. The standard deviation rounds to about 5.857. You could find this using many calculators which will have a variance or standard deviation function or by hand filling out a table such as this:

x	Average	(x-14.4)	$(x-14.4)^2$
8	14.4	-6.4	40.96
10	14.4	-4.4	19.36
15	14.4	0.6	0.36
16	14.4	1.6	2.56
23	14.4	8.6	73.96
		total	137.2
		divide by n-1	34.3
		take square root	5.8566

9. First sort the observations lowest to highest: 1, 2, 3, 11, 21, 23, 25, 72, 100, 102, 105, 133, 200. There are 13 observations so n=13. Find the 25th percentile by taking (.25)(n)=(.25)(13)=3.25. Since this is not an integer take the next integer up or 4th value which is 11. Check the book for what to do if it is an integer or look back at some previous problems where that was the case. Then for the 75th percentile find (.75)(13)=9.75 so go to the 10th value which is 102. Now take 102-11=91 so IQR=91.

10. The correlation is as follows:
    (-40)/(9x13)=-40/117=-.3419.

11. The correlation is as follows:
    23/(10x7)=23/70=.3286.

12. z=(7-10)/2.4=-3/2.4=-1.25.

13. z=(12-10)/2.4=2/2.4=.8333.

4

Practice Homework 4

Here are some practice problems for you to work before doing homework 5. The answers are on the last page.

    You should read the material in teh textbook and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the textbook. Also, for homework 4 you might practice with the "sum" function in excel. There will be columns of data in an excel spreadsheet and you will need to sum up the values.

4.1 Combinations

You have already seen problems dealing with combinations but here are a couple more.

1. I am going on a backpacking trip and will be eating breakfast on the trail for 5 days. I can choose the following for breakfast and I won't take any meal for more than one day: maple and brown sugar oatmeal; apple and cinnamon oatmeal; blueberry oatmeal; cheerios; eggs; pancakes. How many different combinations of breakfasts are there?
2. Suppose I have the same food options as above but I am only going to be on the trail for 4 breakfasts. How many different combinations are there now?

4.2 Calculating Probabilities from a Table

Consider the following table summarizing a data set. In some of the problems you will start with a larger set of data and may want to summarize it into a smaller table such as this one here.

	graduate student	undergraduate	total
resident	7	147	154
non-resident	19	32	51
total	26	179	205

3. What is the probability that a student in this data set is a non-resident?
4. What is the probability that a student is a resident?
5. What is the probability that a student is an undergraduate?
6. What is the probability that a student is a non-resident undergraduate?
7. What is the probability that a student is a resident graduate student?
8. What is the probability that a student is a graduate student or a non-resident?
9. What is the probability that a student is a resident given that they are an undergraduate?
10. What is the probability that a student is a graduate student given that they are a non-resident?

KEY

1. There are 6 different breakfast options so this is a 6 choose 5 problem. 6!/(1!5!)=6.
2. Now we have a 6 choose 4 problem. 6!/(2!4!)=15.
   First thing I would do is convert the table to a joint probability table; divide every entry on the table by the total 205. However, that is not the only way to get the answers.

	graduate student	undergraduate	total
resident	0.034	0.717	0.751
non-resident	0.093	0.156	0.249
total	0.127	0.873	1.000

3. What is the probability that a student in this data set is a non-resident? 0.249
4. What is the probability that a student is a resident? 0.751
5. What is the probability that a student is an undergraduate? 0.873

	graduate student	undergraduate	total
resident	0.034	0.717	0.751
non-resident	0.093	0.156	0.249
total	0.127	0.873	1.000

6. What is the probability that a student is a non-resident undergraduate? .156
7. What is the probability that a student is a resident graduate student? .034
8. What is the probability that a student is a graduate student or a non-resident? .127+.249-.093=0.283 take the P(grad student)+P(non-res)-P(grad student and non-res).
9. What is the probability that a student is a resident given that they are an undergraduate? =P(resident and undergrad)/P(undergraduate)=.717/.873=.821.
10. What is the probability that a student is a graduate student given that they are a non-resident? =P(grad and non-resident)/P(non-resident)=.093/.249=.373.

5

Practice Homework 5

Here are some practice problems for you to work before doing homework 5. The answers are on the last page.

You should read the material in the textbook and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the textbook. Also, this homework requires you to read an article about a company that used its knowledge of discrete probabilities to come up with an advertising campaign. You will find answers to some questions by reading the article. In the homework you will also see that using excel will save you a lot of time on some problems. Watch the Bb video "Finding Binomial Probabilities Using Excel" for some help on this.

5.1 Binomial probabilities with or without excel

1. For a random variable that follows the binomial distribution with probability of success=.35 and n=17 trials, what is the expected value or mean?
2. For a random variable that follows the binomial distribution with probability of success=.53 and n=22 trials, what is the expected value or mean?
3. Using the binomial pmf when p=.28 and n=30, what is P(X=10)?
4. Using the binomial pmf when p=.6 and n=20, what is P(X=5)?
5. Using the binomial pmf when p=.28 and n=30, what is P(X>=10), that is at least 10? Note, this is a cumulative probability problem and you might want to use excel or some other software for help unless your calculator can handle this type of problem.
6. Using the binomial pmf when p=.6 and n=20, what is P(<X=5), that is not more than 5? Note, this is a cumulative probability problem and you might want to use excel or some other software for help unless your calculator can handle this type of problem.
7. If P(X<10)=.37, and P(X=10)=.03, what is P(X>10)?

5.2 Poisson probabilities with or without excel

When using excel follow the same steps as for the binomial distribution except you will type poisson instead of binom.

8. Using the Poisson pmf when the mean is 10, what is P(X=15)?
9. Using the Poisson pmf when the mean is 4, what is P(X=8)?
10. Using the Poisson pmf when the mean is 10, what is P(X>=15)?
11. Using the Poisson pmf when the mean is 4, what is P(X<8)?

5.3 Math review

12. True or false, exp(a)=e^a.
13. What is $e^{1.5}$? (Find the number.)

KEY

1. E(X)=p*n=(.35)(17)=5.95.
2. E(X)=p*n=(.53)(22)=11.66.
3. P(X=x)=n!/(x!(n-x)!)px(1-p)(n-x)=30!/(10!20!)(.28)10(.72)20=.1247.
4. P(X=x)=n!/(x!(n-x)!)px(1-p)(n-x)=20!/(5!15!)(.6)5(.4)15=.00129.
5. .319, you can sum the probabilities that the outcome is 10, 11, .30 or use excel to make the calculation.
6. .0016, you can sum the probabilities that the outcome is 0, 1, .5 or use excel to make the calculation.
7. Probabilities must sum to 1 so 1-(.37+.03)=.6.
8. The formula is $\mu^x e^{-\mu}/x!$ or if we are using Donnelly's notation for the mean we could write it $\lambda^x e^{-\lambda}/x!$ however the key point to remember is that the parameter you are working with is the mean. Typically we use μ to denote the mean but some people prefer to rewrite this with λ. Here the values for the mean is 10 and for x it is 15 so we have

$$\frac{\mu^x e^{-\mu}}{x!} = \frac{10^{15} e^{-10}}{15!} = .0347.$$

9. The formula is $\mu^x e^{-\mu}/x!$. Here the values for the mean is 4 and for x it is 8 so we have

$$\frac{\mu^x e^{-\mu}}{x!} = \frac{4^8 e^{-4}}{8!} = .0298.$$

10. You can add up the probabilities that X=0, 1, 2, ., 14 and then subtract that from 1 or use excel to find a value of .0835.
11. .9489. Add up the probabilities that X is 0,1,.,7.
12. True. Those are two different ways of writing the same thing.
13. $e^{1.5} = 4.48168907033806$.

6

Practice Homework 6

Here are some practice problems for you to work before doing homework 6. The answers are in the key that follows.

   You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

6.1 Uniform Distribution

1. What is the mean for a random variable that is uniformly distributed with minimum 80 and maximum 100?
2. Suppose you have a random variable that is uniformly distributed on the interval [5,13]. What is the expected value for the random variable?
3. What is the formula for the variance of a random variable that follows a uniform distribution?
4. Using the formula that you just found for the variance find the specific variance for a uniformly distributed random variable where the minimum is 15 and the maximum is 21.
5. A random variable is uniformly distributed on [50,75]. What is the probability of drawing an observation with a value less than 60, i.e. $P(X < 60)$?
6. For the previous problem now find $P(X > 60)$.

6.2 Exponential Distribution

   Recall the notation that is sometimes used where the parameter μ is converted to a parameter $\lambda = \frac{1}{\mu}$
7. What is the formula for the cdf for the exponential distribution?
8. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X>16)?

9. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X>21)?

10. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X<12)?

6.3 Standard Normal or z Distribution

11. Suppose that X is a random variable that follows a normal distribution with mean 50 and standard deviation of 10. What is the z value of an observation of 62?

12. Suppose that X is a random variable that follows a normal distribution with mean 50 and standard deviation of 10. What is the z value of an observation of 48?

13. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value greater than 1.45?

14. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value less than 1.45?

15. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value greater than -1.45?

16. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value less than -1.07?

Note: understanding z values and the associated probabilities is extremely important in this class. If you do not understand the exponential distribution you will lose some points and it will be tough to get an A or B but if you do not understand the standard normal distribution it will be tough to get a C. If the topic is challenging for you then please spend extra time with a tutor, classmate or TA or come and visit with me in order to get the necessary help. We both want you to pass the class.

KEY

1. What is the mean for a random variable that is uniformly distributed with minimum 80 and maximum 100? (100+80)/2=90
2. Suppose you have a random variable that is uniformly distributed on the interval [5,13]. What is the expected value for the random variable? (5+13)/2=9
3. What is the formula for the variance of a random variable that follows a uniform distribution? $(\max - \min)^2/12$.
4. Using the formula that you just found for the variance find the specific variance for a uniformly distributed random variable where the minimum is 15 and the maximum is 21. (21-15)2/12 = 36/12 = 3.
5. A random variable is uniformly distributed on [50,75]. What is the probability of drawing an observation with a value less than 60, i.e. P(X<60)? (60-50)/25 = 10/25 = .4.
6. For the previous problem now find P(X>60). 1-.4=.6.
7. What is the formula for the cdf for the exponential distribution? $1 - exp(-x/\mu) = 1 - exp(-x\lambda)$ when using the notation $\lambda = 1/\mu$.
8. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X>16)? Plug in to the formula you found that P(X<=16)=.656 so that P(X>16) is .344.
9. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X>21)? .2466
10. Suppose that X follows the exponential distribution with $\mu = 15$. What is P(X<12)? .5507
    Standard Normal or z Distribution
11. Suppose that X is a random variable that follows a normal distribution with mean 50 and standard deviation of 10. What is the z value of an observation of 62? (62-50)/10=12/10=1.2.
12. Suppose that X is a random variable that follows a normal distribution with mean 50 and standard deviation of 10. What is the z value of an observation of 48? (48-50)/10=-2/10=-.2.
13. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value greater than 1.45? Look at the z table and find 1.45 and you will find a value of .9265 so then take 1-.9265 to get .0735.
14. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value less than 1.45? Similar to the problem before but now because we are looking for the "less than" part we do not subtract from 1, instead we just use the .9265.
15. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value greater than -1.45? Here you should notice that the z distribution is symmetric around 0 so that the answer is .9265. If this doesn't initially make sense then draw

the picture of the z distribution and look at the area to the left of 1.45 and to the right of -1.45. Your eyeballs should tell you that they are about the same or that they are exactly the same because if the distribution is drawn perfectly (as is done with computer software) then they are exactly the same.

16. Suppose that z is a random variable that follows the standard normal distribution. What is the probability of observing a z value less than -1.07? Looking at the z table we see that P(x<=1.07)=.8577. Again, because the distribution is symmetric around 0 this means that P(x>-1.07)=.8577.

7

Practice Homework 7

Here are some practice problems for you to work before doing homework 7. The answers are in the key that follows.

You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

You may also want to review some of the videos on finding average and variance in excel. These are on the "Videos" tab in Bb in case you need a refresher.

7.1 Standard Error

1. Suppose that you have a data set of starting salaries and you find an average starting salary of $47,222 with a standard deviation of $6,528 for a 22 observations. What is the standard error?
2. Suppose that instead of 22 observations you had 33 observations in the previous problem but the average and standard deviation stayed the same. Would the standard error be larger or smaller? What is the new standard error?
3. If the standard deviation is 10 for a sample of 17, what is the standard error?
4. For proportions the formula for standard error is a little different. Suppose you have a sample proportion of .3 and a sample of size 20. What is the standard error in this case?
5. Suppose that 40 people are surveyed and asked if they think the minimum wage should be increased. 23 people say yes and 17 people no. What is the standard error?

7.2 Standard Normal or z Distribution (again)

6. Suppose that you have a sample of 40 textbook prices and $\bar{x} = \$200$. Suppose also that the standard deviation is $50. If the true population mean is $\mu = \$220$ what is the z score for the sample that you found?

7. Using the values in the problem above, what is the probability that you would have found an average price less than $200 if the true mean is $220?

8. Suppose that true average price of a beer at professional sporting events in the US is $6. You attend 33 events and find an average price of $\bar{x} = \$6.40$ and the standard deviation is $0.46. Find the z value for your sample.

9. Using the values in the problem above, what is the probability that you would have found an average price greater than $6.40 if the true mean is $6?

KEY

1. Suppose that you have a data set of starting salaries and you find an average starting salary of $47,222 with a standard deviation of $6,528 for a 22 observations. What is the standard error? $6528/\sqrt{22}$=$6528/4.69=$1391.77 but typically you would omit the dollar sign.

2. Suppose that instead of 22 observations you had 33 observations in the previous problem but the average and standard deviation stayed the same. Would the standard error be larger or smaller? What is the new standard error? Because n has increased the standard error will decrease given that nothing else has changed in this example.
$6528/\sqrt{33}$=$6528/5.74456264653803=$1136.38 but typically you would omit the dollar sign.

3. If the standard deviation is 10 for a sample of 17, what is the standard error?
$$10/\sqrt{17} = 10/4.12310562561766 = 2.425.$$

4. For proportions the formula for standard error is a little different. Suppose you have a sample proportion of .3 and a sample of size 20. What is the standard error in this case? The formula is $\sqrt{((p(1-p))/n)}$= $\sqrt{((.3(1-.3))/20)}$= $\sqrt{(.21/20)}$ =.1025.

5. Suppose that 40 people are surveyed and asked if they think the minimum wage should be increased. 23 people say yes and 17 people no. What is the standard error? The sample proportion is p=23/40=.575. The formula is:

$$\sqrt{((p(1-p))/n)} = \sqrt{((.575(.425))/40)} = \sqrt{(.2444/40)} = .0782.$$

6. Suppose that you have a sample of 40 textbook prices and $\bar{x}$=$200. Suppose also that the standard deviation is $50. If the true population mean is μ=$220 what is the z score for the sample that you found? Use the formula for the z value

$$(200 - 220)/(50/\sqrt{40}) = -20/(50/6.325) = -20/7.906 = -2.5298.$$

7. Using the values in the problem above, what is the probability that you would have found an average price less than $200 if the true mean is $220? Use the z table conclude that it is approximately .0057.

8. Suppose that true average price of a beer at professional sporting events in the US is $6. You attend 33 events and find an average price of $\bar{x}$=$6.40 and the standard deviation is $0.46. Find the z value for your sample. Use the formula for the z value

$$(6.4 - 6)/(.46/\sqrt{33}) = .4/(0.080075721739621) = 4.99527186655481.$$

9. Using the values in the problem above, what is the probability that you would have found an average price greater than $6.40 if the true mean

is \$6? Use the z table conclude that it is approximately 0. Keep in mind when comparing this problem to 7 than here you had "greater" in the question and in 7 you had "less."

8

Practice Homework 8

Here are some practice problems for you to work before doing homework 8. The answers are in the key that follows.

   You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

   You may also want practice reading the z and t tables; knowing how to read those tables will make a difference in your grade.

8.1 Standard Error and Degrees of Freedom

1. Suppose that $\sigma = 20$ and that n=20. What is the standard error?
2. Suppose that the standard deviation is not known and so you will use a t distribution when solving the problem. Suppose n=24. In this case how many degrees of freedom are there? (The context is constructing a confidence interval for one mean).

8.2 Confidence Intervals

3. Suppose that alpha is .1 ($\alpha = .1$). What is the confidence coefficient?
4. Suppose you are constructing 99% confidence intervals for the mean. What is the probability that a confidence interval constructed in that way will not contain the true mean?
5. Construct a 90% confidence interval for μ when $\bar{x} = \$64$, $\sigma = 17$ and n=38?
6. Construct a 95% confidence interval for μ when $\bar{x} = \$41$, $\sigma = 13$ and n=31?
7. True or False. When the population standard deviation is not known we have to find a t value in order to construct a confidence interval.

NOTE/WARNING: In the problems above like 5 and 6 the standard deviation is known as opposed to being estimated. In this case we construct the margin of error using a z value but when the standard deviation is estimated, i.e. we know the *sample* standard deviation (s instead of σ) then we construct the margin of error using a t value.

KEY

1. Suppose that $= 20$ and that n=20. What is the standard error? 4.47
2. Suppose that the standard deviation is not known and so you will use a t distribution when solving the problem. Suppose n=24. In this case how many degrees of freedom are there? (The context is constructing a confidence interval for one mean). n-1 is the formula so 24-1=23.
3. Suppose that alpha is .1. What is the confidence coefficient? 1-c.c. is the equation so we have 1-.1=.9.
4. Suppose you are constructing 99% confidence intervals for the mean. What is the probability that a confidence interval constructed in that way will not contain the true mean? This means that c.c.=.99 and the probability that a confidence interval constructed in this fashion will not contain the true mean is .01.
5. Construct a 90% confidence interval for μ when $\bar{x} = \$64$, $\sigma = 17$ and n=38? The standard error is 2.7578 and the margin of error is 4.537 so that the CI is (59.463, 68.537).
6. Construct a 95% confidence interval for μ when $\bar{x} = \$41$, $\sigma = 13$ and n=31? The standard error is 2.335 and the margin of error is 4.576 so that the CI is (36.424, 45.576).
7. True or False. When the population standard deviation is not known we have to find a t value in order to construct a confidence interval. True. We will work with examples where the standard deviation is not known and we need to find a t value when we get to hypothesis testing.

Practice Homework 9

Here are some practice problems for you to work before doing homework 9. The answers are in the key that follows.

You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

You may also want practice reading the z and t tables; knowing how to read those tables will make a difference in your grade.

9.1 Hypothesis Tests for Means, Test Statistics, Critical Values

1. True or False. The critical value for a one tail test is the same as for a two tail test.
2. True or False. The critical value for a right tail test is the same as for a left tail test.
3. Suppose you have the null hypothesis $H_0 : \mu = 50$ and $\bar{x} = 54$ and s=20 and n=36. Test the null hypothesis at the .05 level of significance, report your test statistic and the critical value. Draw a picture showing the rejection region and why you failed to reject or rejected H_0.
4. Suppose you have the null hypothesis $H_0 : \mu \geq 120$ and $\bar{x} = 129$ and s=26 and n=63. Test the null hypothesis at the .05 level of significance, report your test statistic and the critical value. Draw a picture showing the rejection region and why you failed to reject or rejected H_0.
5. When Mike Leach coached at Texas Tech his players had "fat little girlfriends." (Yes, I think it is funny that the girlfriends were neither "big fat" nor "skinny little" but then perhaps Coach Leach and his players wanted to make me laugh.) Suppose that Mike Leach was arguing that the girlfriends of his football players had an average BMI of at least 28 (the CDC categorizes people with a BMI of 28 as overweight). In other words suppose his null hypothesis is $H_0 : \mu \geq 28$. If you found the average BMI of

20 of the girlfriends was 22.4 with a standard deviation of 1.9 would you reject or fail to reject the null hypothesis at the .1 level of significance? What is the test statistic and what is the critical value? Draw a picture showing the rejection region and your test statistic.
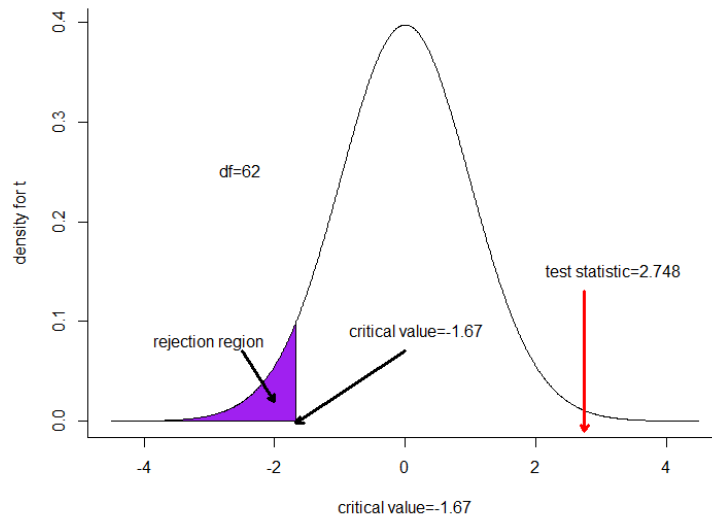
9.2 Hypothesis Tests for Proportions

6. Suppose that alpha is .1. Suppose $H_0 : p = .3$ (the population proportion is .3) but that you find a sample proportion of .34 from a sample of 150. What is your conclusion?

KEY

1. True or False. The critical value for a one tail test is the same as for a two tail test. False
2. True or False. The critical value for a right tail test is the same as for a left tail test. False, the critical value for a right tail test will be positive and for a left tail test it will be negative.
3. Suppose you have the null hypothesis $H_0 : \mu = 50$ and $\bar{x} = 54$ and s=20 and n=36. Test the null hypothesis at the .05 level of significance, report your test statistic and the critical value. Draw a picture showing the rejection region and why you failed to reject or rejected H_0. The test statistic is 1.2 and the critical values are -2.03 and 2.03 so I fail to reject.
4. Suppose you have the null hypothesis $H_0 : \mu \geq 120$ and $\bar{x} = 129$ and s=26 and n=63. Test the null hypothesis at the .05 level of significance, report your test statistic and the critical value. Draw a picture showing the rejection region and why you failed to reject or rejected H_0. This is a one tail test and it is a left tail test. In this case you can tell that you will fail to reject just by noticing that 129>120 but let's get the critical value and test statistics anyway. The test statistic is 2.748 and the critical value is -1.67 (you can approximate this from the t table or you could go to excel to get a more precise value). In this case 2.748>-1.67 so I fail to reject.

5. When Mike Leach coached at Texas Tech his players had "fat little girl-friends." (Yes, I think it is funny that the girlfriends were neither "big fat"

nor "skinny little" but then perhaps Coach Leach and his players wanted to make me laugh.) Suppose that Mike Leach was arguing that the girl-friends of his football players had an average BMI of at least 28 (the CDC categorizes people with a BMI of 28 as overweight). In other words suppose his null hypothesis is $H_0 : \mu \geq 28$. If you found the average BMI of 20 of the girlfriends was 22.4 with a standard deviation of 1.9 would you reject or fail to reject the null hypothesis at the .1 level of significance? What is the test statistic and what is the critical value? Draw a picture showing the rejection region and your test statistic.

The test statistic is -13.18 and the critical value is -1.328 (there are 19 degrees of freedom). The test statistic is "far" to the left of the critical value in this left tail test so that you reject the null hypothesis
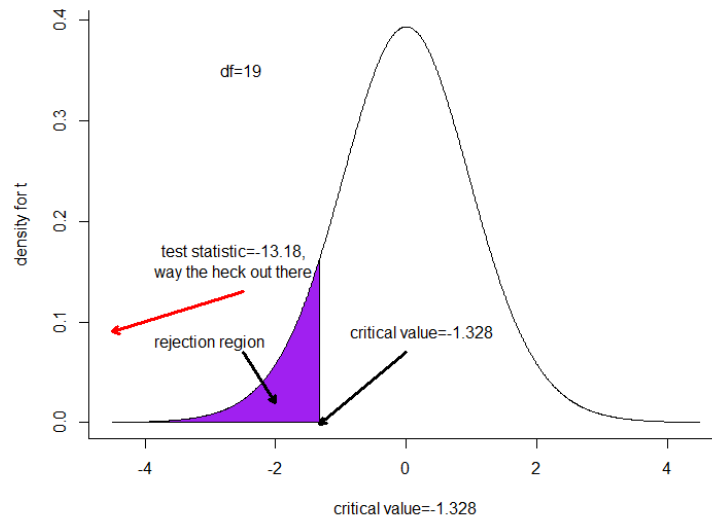
6. Suppose that alpha is .1. Suppose $H_0 : p = .3$ (the population proportion is .3) but that you find a sample proportion of .34 from a sample of 150. What is your conclusion? The critical value for the two tailed test is -1.645 and 1.645 while the test statistic is 1.069 so that you fail to reject the null hypothesis. There is not statistical evidence that the true proportion is different from .3.

10

Practice Homework 10

Here are some practice problems for you to work before doing homework 10. The answers are in the key that follows.

    You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

10.1 Hypothesis Tests for Difference in Two Means

1. Use a .05 level of significance. Test the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$. Suppose that the sample means are 10 and 12 and that the sample standard deviations are 5 and 6 and that the sample sizes are 41 and 45. Note that there are 83 degrees of freedom. Report the standard error and the test statistic and your conclusion.

2. Use a .01 level of significance. Suppose that group 1 is juniors and group 2 is sophomores. The college of business wants to decide whether they should encourage students to wait and take statistics as juniors instead of allowing sophomores to take the class. The dean wants to test the null hypothesis $H_0 : \mu_1 - \mu_2 \leq 0$ (this null hypothesis states that sophomores do at least as well as juniors but if it rejected then she can conclude there is evidence that juniors actually do better). Suppose that the sample means for test scores are 83.2 and 84.5 and that the sample standard deviations are 16.1 and 15.5 and that the sample sizes are 141 and 122. Note that there are 258 degrees of freedom. Report the standard error and the test statistic and your conclusion.

3. Use a .1 level of significance. Test the null hypothesis $H_0 : \mu_1 - \mu_2 \geq 5$. Suppose that the sample means are 46 and 47 and that the sample standard deviations are 3.5 and 4 and that the sample sizes are 30 each. Note that there are 56 degrees of freedom. Report the standard error and the test statistic and your conclusion. Also, what was the point estimate of the difference?

10.2 Hypothesis Tests for Difference in Means for Paired or Matched Samples

4. Set alpha at .05 ($\alpha = .05$). Suppose you have a matched sample and $\bar{d} = 10$ while the sample standard deviation $s_d = 27$ for 20 pairs. Test the hypothesis $H_0 : \mu_d = 0$. What is your conclusion? Report the standard error, test statistic, and critical value.

5. Set alpha at .05. Suppose you have a matched sample and $\bar{d} = 100$ while the sample standard deviation $s_d = 72$ for 10 pairs. Test the hypothesis $H_0 : \mu_d = 80$. What is your conclusion? Report the standard error, test statistic, and critical value and degrees of freedom.

KEY

1. Use a .05 level of significance. Test the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$. Suppose that the sample means are 10 and 12 and that the sample standard deviations are 5 and 6 and that the sample sizes are 41 and 45. Note that there are 83 degrees of freedom. Report the standard error and the test statistic and your conclusion. The standard error is 1.18733 and the test statistic is -1.6699 so that we fail to reject the null hypothesis since the critical values are + and - 1.989.

2. Use a .01 level of significance. Suppose that group 1 is juniors and group 2 is sophomores. The college of business wants to decide whether they should encourage students to wait and take statistics as juniors instead of allowing sophomores to take the class. The dean wants to test the null hypothesis $H_0 : \mu_1 - \mu_2 \leq 0$ (this null hypothesis states that sophomores do at least as well as juniors but if it rejected then she can conclude there is evidence that juniors actually do better). Suppose that the sample means for test scores are 83.2 and 84.5 and that the sample standard deviations are 16.1 and 15.5 and that the sample sizes are 141 and 122. Note that there are 258 degrees of freedom. Report the standard error and the test statistic and your conclusion. The standard error is 1.951 and the test statistic is -.666. This is a one tail (right tail) test and the critical value is 2.341. Because -.666<2.341 you fail to reject the null hypothesis. There is not any statistical evidence that juniors do better in the class.

3. Use a .1 level of significance. Test the null hypothesis $H_0 : \mu_1 - \mu_2 \geq 5$. Suppose that the sample means are 46 and 47 and that the sample standard deviations are 3.5 and 4 and that the sample sizes are 30 each. Note that there are 56 degrees of freedom. Report the standard error and the test statistic and your conclusion. Also, what was the point estimate of the difference? The standard error is .97 and the test statistic is -6.18 and the critical value is -1.297. Since -6.18<-1.297 (this is a left tail test) you reject the null hypothesis. The point estimate of the difference is 46-47=-1.

4. Set alpha at .05 ($\alpha = .05$). Suppose you have a matched sample and $\bar{d} = 10$ while the sample standard deviation $s_d = 27$ for 20 pairs. Test the hypothesis $H_0 : \mu_d = 0$. What is your conclusion? Report the standard error, test statistic, and critical value. The standard error is 6.037 and the test statistic is 1.656 and the critical values are - and + 2.093 so you fail to reject.

5. Set alpha at .05. Suppose you have a matched sample and $\bar{d} = 100$ while the sample standard deviation $s_d = 72$ for 10 pairs. Test the hypothesis $H_0 : \mu_d = 80$. What is your conclusion? Report the standard error, test statistic, and critical value and degrees of freedom. The standard error is 22.768 and the critical values for the two tailed test are -2.262 and 2.262 while the test statistic is .878 so that you fail to reject the null hypothesis. There 9 degrees of freedom for this t test.

11

Practice Homework 11

Here are some practice problems for you to work before doing homework 11. The answers are in the key that follows.

You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

Excel has a special function for conducting ANOVA. There are PCs available on campus in CPCOM and other locations. I recommend using one of those computers and installing the data analysis toolpak and using the ANOVA feature available there. I have posted videos on Bb showing how to install the data analysis toolpak and on how to use the ANOVA feature. If you choose not to use one of the computers on campus and the ANOVA feature then you will have to create the appropriate formulas in excel on your own. The choice is yours.

For the actual homework the dataset has a lot more observations than in the practice homework. If you are using the ANOVA feature in excel this is a meaningless difference and if you typed in the formulas in excel then this should only mean that you scroll over extra cells.

11.1 ANOVA and F values

1. From the course webpage (ECN221 page) open the practice homework file and go to the data set tab with the Houston Astros data set which has attendance at weekend games for the Houston Astros (a baseball team).
2. What is the average attendance on Fridays? Saturdays? Sundays?
3. What is the overall average?
4. Suppose you are conducting ANOVA to test the hypothesis that mean attendance is the same for all three days. What was the MSE (this is also sometimes called MSW, provide a number)?
5. What is the total sum of squares?

6. What is MSTR or MSB? Note that some software packages will call this mean squares between or MSB while others will call it mean squares treatment or MSTR but they are the same thing
7. How is the test statistic distributed?
8. How many degrees of freedom are there?
9. What is the value of the test statistic?
10. Suppose alpha=.05. What is the critical F value when there are 4 numerator and 7 denominator degrees of freedom.
11. Suppose you see the following output. What is SSTR (aslo called SSB)? What is MSE (also called MSW)? What is the total sum of squares? What is the value of the test statistic? What is the p-value? If alpha=.05 do you reject or fail to reject the null hypothesis that all means are equal? How many treatments were used? How many observations were there?

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	230913250.8	3	76971084	6.225924	0.002244	2.946685
Within Groups	346163922.2	28	12362997			
Total	577077173	31				

12. Complete the table below.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Treatments		4	24.69286		0.014078	2.795539
Error	144.2286		6.270807			
Total		27				

KEY

1. From the course webpage (ECN221 page) open the practice homework file and go to the data set tab with the Houston Astros data set which has attendance at weekend games for the Houston Astros (a baseball team).
2. What is the average attendance on Fridays? Saturdays? Sundays? 28,155.7, 28,843.6, and 23,874.9.
3. What is the overall average? 26, 958.
4. Suppose you are conducting ANOVA to test the hypothesis that mean attendance is the same for all three days. What was the MSE (provide a number)? 16994837.
5. What is the total sum of squares? 407376741
6. What is MSTR or MSB? Note that some software packages will call this mean squares between or MSB while others will call it mean squares treatment or MSTR but they are the same thing. 50734841.
7. How is the test statistic distributed? The test statistic follows the F distribution.
8. How many degrees of freedom are there? 2 numerator df and 18 denominator df.
9. What is the value of the test statistic? 2.985 which you get from dividing the MSTR by MSE (same as MSB/MSW).
10. Suppose alpha=.05. What is the critical F value when there are 4 numerator and 7 denominator degrees of freedom. Consult the F table and find a value of 4.12.
11. Suppose you see the following output. What is SSTR (aslo called SSB)? What is MSE (also called MSW)? What is the total sum of squares? What is the value of the test statistic? What is the p-value? If alpha=.05 do you reject or fail to reject the null hypothesis that all means are equal? How many treatments were used? How many observations were there?

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	230913250.8	3	76971084	6.225924	0.002244	2.946685
Within Groups	346163922.2	28	12362997			
Total	577077173	31				

SSTR or SSB=230913250.8 and MSE or MSW =12362997 and TSS=577077173 and F test statistic=6.2259 and p-value=0.002244. Reject the null hypothesis that the means are equal at the .05 (or 5%) level of significance. There were 4 treatments (on the "between groups" row see that there are 3 degrees of freedom and this is the number of groups minus 1). 32 total observations.

12. Complete the table below.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Treatments	98.77143	4	24.69286	3.937748	0.014078	2.795539
Erro	144.2286	23	6.270807			
Total	243	27				

12

Practice Homework 12

Here are some practice problems for you to work before doing homework 12. The answers are in the key that follows.

You should read the relevant chapter and review the lecture notes. Some questions deal with definitions and there are no practice definitions. For those you must simply consult lecture notes and the text book.

Excel has a special function for conducting regression analysis which is available on PCs. There are PCs available on campus in CPCOM and other locations. I recommend using one of those computers and installing the data analysis toolpak and using the regression feature available there. I have posted videos on Bb showing how to install the data analysis toolpak and on how to use the regression feature. If you choose not to use a computer that has these features (which are availableon campus) then you will have to create the appropriate formulas in excel on your own. The choice is yours.

Get the practice data spreadsheet and use the "education spending" data to answer the questions. You can click on the headers to see the data and use more observations if you are curious. In any case here are 7 observations showing the public high school graduation rates and the per student spending for each state. Run a regression using these data and use per student spending as the independent variable and graduation rate as the dependent variable.

12.1 Regression and p values

1. Again, use the course webpage (ECN221 page) to open the practice homework file and go to the data set tab with the "education spending" data set. What is R^2?
2. What is the estimated variance of the error term?
3. What is the estimated standard deviation of the error term?
4. What is the estimated intercept?
5. Is the intercept statistically significantly different from 0 at the .05 level of significance?

6. What is the estimated slope?
7. What is the standard error for the estimated slope?
8. What is the test statistic for the null hypothesis that the slope is 0, i.e. $H_0 : \beta_1 = 0$?
9. Do you reject or fail to reject the null hypothesis?
10. Suppose instead that the null hypothesis is that the slope is .0001. This would mean that a \$1,000 increase in spending increases the graduation rate by a tenth of a percent. In this case what is the test statistic and do you reject or fail to reject at the .05 level of significance?
11. What is the p-value reported for the test that the intercept is 0?
12. Again using the graduation rates data, what is the predicted graduation rate for a state with a per pupil spending of \$10,000?
13. What is the 95% confidence interval for $E(y^*|x=10,000)$? This is the confidence interval for the mean value of the graduation rate when the per pupil expenditure is \$10,000; note that you found the point estimate in the previous problem. Now find the 95% prediction interval, again for x=10,000.
14. If alpha=.05 and p-value=.049 do you reject or fail to reject the null hypothesis?
15. If alpha is .1 and the p-value is .11 do you reject or fail to reject the null hypothesis?
16. Draw graphs of "well behaved" error terms and of error terms that violate the assumptions made in this segment of the class and book.

KEY

Regression Statistics	
Multiple R	0.129591165
R Square	0.01679387
Adjusted R Square	-0.179847356
Standard Error	5.644101222
Observations	7

ANOVA

	df	SS	MS	F	Significance F
Regression	1	2.72060696	2.720607	0.085404	0.781841874
Residual	5	159.279393	31.85588		
Total	6	162			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	77.99567219	7.154354531	10.90185	0.000113	59.60481839	96.386526
spending	-0.000183887	0.000629235	-0.29224	0.781842	-0.001801386	0.00143361

1. $R^2 = .01679$
2. This is the MSE which is 31.8558.
3. This is 5.6441.
4. The estimated intercept is 77.99567
5. Yes, look at the p-value which has a value of .000113<.05 or you can look at the t-stat and compare it to the appropriate t value to get the same answer.
6. The estimated slope is found on the row that says "spending" and the value is -.000183887.
7. Look in the standard error column and find .000629.
8. Look in the t-stat column and find -.292.
9. Fail to reject because -.292>-2.571 which is the appropriate critical value OR you can look at the p-value and notice that .7818>.05 which also implies that you will fail to reject the null hypothesis.
10. Here the test statistic is (-.00018389-.0001)/.0006292=-.451. Then you fail to reject because -.451>-2.571 and -.451<2.571. Keep in mind that this is a two tail test but that you have already seen one-tail tests in previous homework and you should be prepared to see one-tail tests in the future.
11. Look in the p-value column and find .000113.
12. 77.995672-.0001839*10000=76.157 or something very close depending on how you rounded.
13. Start with the point estimate of 76.157 and also recall our critical t value in previous problems of + and - 2.571. We need the standard deviation of $\hat{y}^*$. This is found (check the formula in the book and notes) by taking the standard error of the estimate (5.644) and multiplying it times the square root of

$$(1/7 + (10000 - 10852.71)^2/80457055) = .389736.$$

The standard deviation of $\hat{y}^*$=2.199712. That was .389736*5.6441. I know that is a lot of work so have excel do it for you. You get 76.157 +/- 2.571*2.199712=76.157 +/- 5.655=[70.501, 81.812].

14. If alpha=.05 and p-value=.049 do you reject or fail to reject the null hypothesis? Reject.

15. If alpha is .1 and the p-value is .11 do you reject or fail to reject the null hypothesis? Fail to reject.

16. Draw graphs of "well behaved" error terms and of error terms that violate the assumptions made in this segment of the class and book. Compare your graphs to those in the notes and in the book.