MM207 Final Project

Name: Eddie S. Jackson

1.

Using the MM207 Student Data Set:

a) What is the correlation between student cumulative GPA and the number of hours spent on school work each week? Be sure to include the computations or StatCrunch output to support your answer.
My answer :
 0.27817234

(from StatCrunch):
Correlation between Q10 What is your cumulative Grade Point Average at Kaplan University? and Q11 How many hours do you spend on school work each week? is:
0.27817234

b) Is the correlation what you expected?
My answer:
No. I expected the correlation to be much higher because the more hours you study should equate to a much higher GPA – in theory that is.

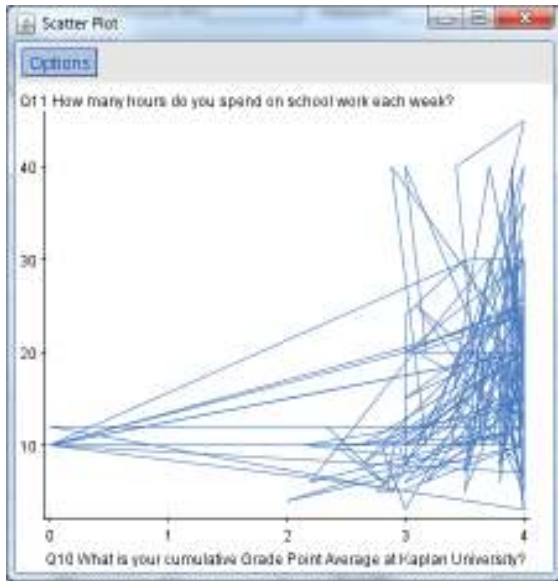**c)** Does the number of hours spent on school work have a causal relationship with the GPA?
My answer:
Yes.
I was going to say no (because of the low correlation above), until I did a scatter plot. This shows that

there definitely is a casual relationship between study time and GPA.

d) What would be the predicted GPA for a student who spends 16 hours per week on school work? Be sure to include the computations or StatCrunch output to support your prediction.
My answer:
3.6

from StatCrunch
Group by: Q11 How many hours do you spend on school work each week?

Q11 How many hours do you spend on school work each week?	Mean	n	Variance
3	3.6666667	3	0.33333334
4	2	1	NaN
5	3.3775	8	0.3129357
6	3.0714285	7	0.42641428
7	3.75	2	0.125
8	3.352	5	0.26252
10	2.9693334	30	1.6706271
11	3.6466668	3	0.14423333

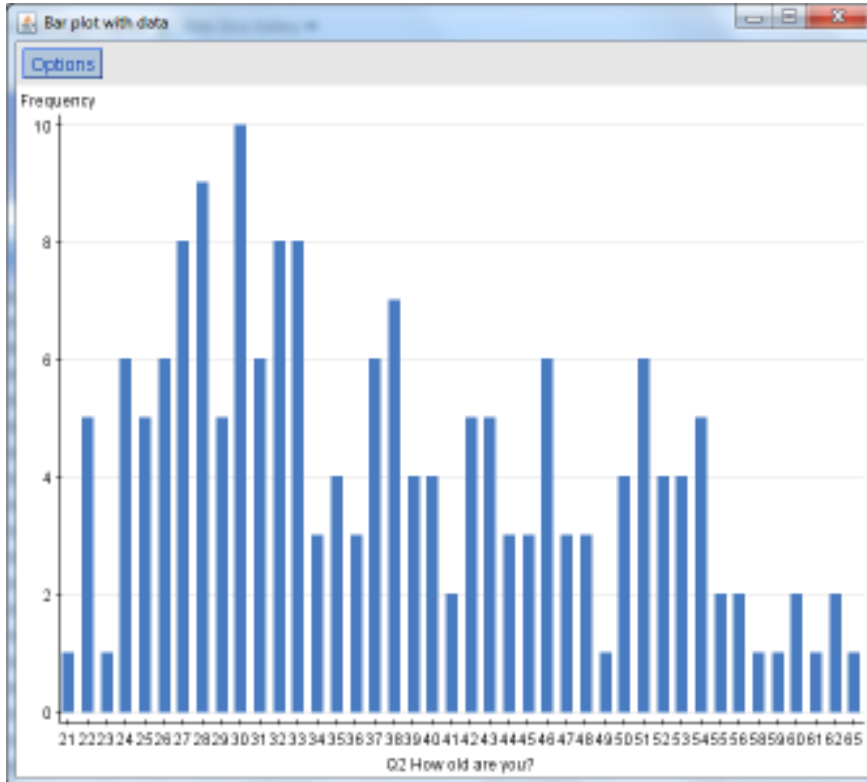| 12	3.290909	11	1.4214091
13	4	2	0
14	3.93	2	0.0098
15	3.7127273	11	0.11040182
16 hours	**3.6**	**3**	**0.07**

2.

Select a continuous variable that you suspect **would not** follow a normal distribution.

a) **My answer:**

my continuous variable is "Age"

b) Create a graph for the variable you have selected to show its distribution.

My answer:

a) Explain why these data might not be normally distributed.
My answer:
These may not be normally distributed due to the fact that people of all ages go to school – you will notice that all values are not tightly gathered around the mean.
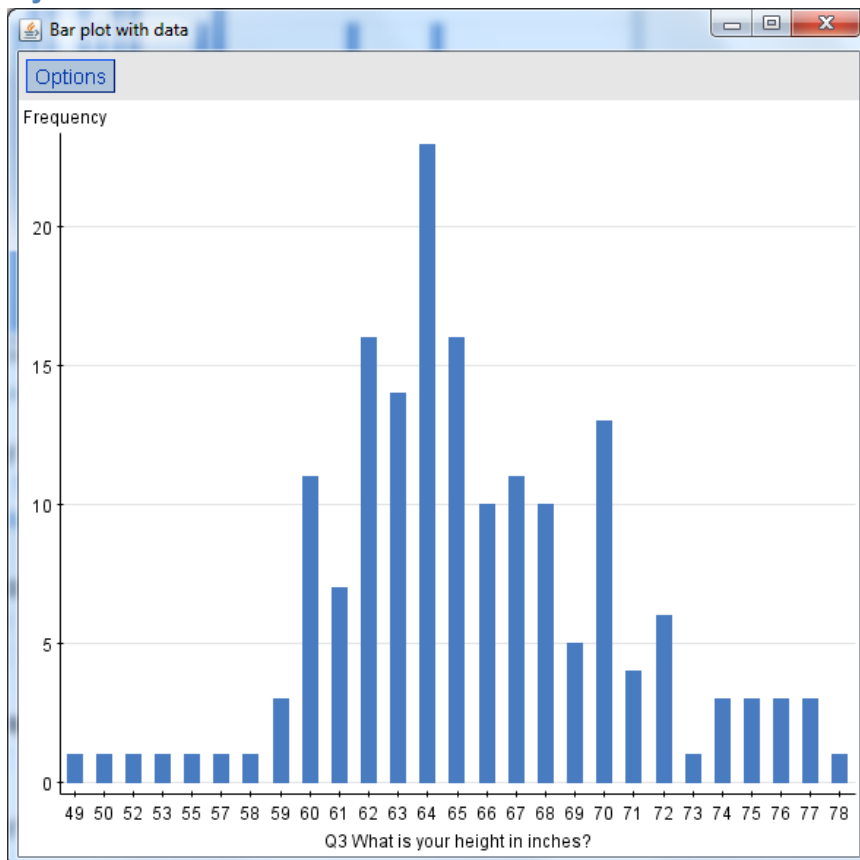
b) Select a second continuous variable that you believe **would** approximate a normal distribution
My answer:
my continuous variable is "Height"

c) Create a graph to show its distribution.
My answer:

d) Explain why these data might be normally distributed.
My answer:
People are different heights of course, however you see an obvious tighter grouping around the mean; suggesting these values are closer to a normal distribution.

3.

Jonathan is a 42 year old male student and Mary is a 37 year old female student thinking about taking this class. Based on their relative position, which student would be farther away from the average age of their gender group based on this sample of MM207 students?
My answer:
Jonathan

from StatCrunch
Summary statistics for Q2 How old are you?:
Group by: Gender

Gender	n	Mean	Variance	Std. Dev.	Std. Err.	Median	Range	Min	Max	Q1	Q3
Female	138	37.746376	104.015495	10.198799	0.86817944	37	44	21	65	30	46
Male	35	38.8	160.28235	12.660267	2.1399755	35	38	24	62	28	51

4.

4. If you were to randomly select a student from the set of students who have completed the survey, what is the probability that you would select a male? Explain your answer.
My answer:
0.2
35 males+138 females+2 no gender listed=175 students total
That makes the probability equal to
35 males/175 total students or 35/175 = **0.2**

Calculations
Calculator says 0.2
Turn that into a percent = 20%

MM207 Final Project

from StatCrunch
Frequency table results for Gender: = 173 count + 2 that did not list gender = 175 students total
Group By: Gender

Results for Gender=Female

Gender	Frequency	Relative Frequency
Female	138	1

Results for Gender=Male

Gender	Frequency	Relative Frequency
Male	35	1

5.

Using the sample of MM207 students:
What is the probability of randomly selecting a person who is conservative and then selecting from that group someone who is a nursing major?
My answer:
For conservative it is: 41 conservative/175 total count or 41/175 or **0.2343** or 23% rounded to the nearest percentage
For a nursing major: 12 conservative-nursing students/175 total count = 12/175 or **0.0686** or 7% rounded to the nearest percent.

Calculator says 0.0685714285714285714285714285714143
from StatCrunch

Frequency table results for Q13 What best describes your political philosophy?: = 170+5 who did not answer = 175 total count

Q13 What best describes your political philosophy?	Frequency
Conservative	41
Liberal	40
Moderate	89

Contingency table results for Q13 What best describes your political philosophy?=Conservative:
Rows: Q13 What best describes your political philosophy?
Columns: Q9 What is your college major?

	Business	IT	Legal Studies	Nursing	Other	Psychology	Total
Conservative	4	1	5	12	4	14	40
Liberal	0	0	0	0	0	0	0
Moderate	0	0	0	0	0	0	0
Total	4	1	5	12	4	14	40

a)   What is the probability of randomly selecting a liberal or a male?
My answer:
0.3886

175 total count of students who took the survey
For a liberal it is: 23.81% or 40/168 (includes males and females)
For a male who is either liberal/moderate/conservative = 35/168 or 20.83%
Minus those that are Male AND Liberal -7
but to actually get the probability, make sure to count all students in the survey 175
So that would be 40+35-7 = 68/175 = **0.3886**

Or the 168 students who answered the question
For a liberal it is: 23.81% or 40/168 (includes males and females)
For a male who is either liberal/moderate/conservative = 35/168 or 20.83%
Minus those that are Male AND Liberal -7

Add those together
So that would be 40+35-7 = 68/168 = **0.40476**190476190476190476190476

Contingency table results:
Rows: Q13 What best describes your political philosophy?
Columns: Gender

Cell format
Count (Row percent) (Column percent)

	Female	Male	Total
Conservative	27 (67.5%) (20.3%)	Add 13 (32.5%) (37.14%)	40 (100.00%) (23.81%)
Liberal	33 (82.5%) (24.81%)	Subtract 7 (17.5%) (20%)	Total liberals 40 (100.00%) (23.81%)
Moderate	73 (82.95%) (54.89%)	Add 15 (17.05%) (42.86%)	88 (100.00%) (52.38%)
Total	133 (79.17%) (100.00%)	35 (20.83%) (100.00%)	168 total (100.00%) (100.00%)

6.

Facebook reports that the average number of Facebook friends worldwide is 175.5 with a standard deviation of 90.57. If you were to take a sample of 25 students, what is the probability that the mean number Facebook friends in the sample will be 190 friends or more?
My answer:
0.2119

My mean is 175.5
My standard deviation is=90.57
Sample=25

So the probability is
Formula to be used: P(X>190)=P((X-mean)/s
sqrt of my sample is = 5
(190-175.5) **/** 90.57/5)
14.5 **/** 18.114
calculator says 0.8004858120790548746825659710721
I check the z-table and I see -.8 under the z of 0.00 correlates to .2119
=0.2119

7. Select a random sample of 30 student responses to question statcruch #6→ "How many credit hours are you taking this term?" Using the information from this sample, and assuming that our data set is a random sample of all Kaplan statistics students, **estimate the average number of credit hours that all**

MM207 Final Project

Kaplan statistics students are taking this term using a 95% level of confidence. Be sure to show the data from your sample and the data to support your estimate.

My answer:
lower 9.725972 upper 11.740694

My mean=10.733333333333333333333333333333
my sample = (12+11+10+18+11+10+10+18+12+10+3+10+12+11+15+4+12+6+10+6+12+10+12+10+11+11+11+11+12+11)
322/30

My Standard deviation=2.8151245

Summary statistics:

Column	n	Std. Dev.
Q6 How many credit hours are you taking this term?	175	2.8151245

My Sample size=30
my sample = (12+11+10+18+11+10+10+18+12+10+3+10+12+11+15+4+12+6+10+6+12+10+12+10+11+11+11+11+12+11)=322

95% confidence interval results:

μ : population mean
Standard deviation = 2.8151245

Mean	n	Sample Mean	Std. Err.	L. Limit	U. Limit
μ	30	10.733334	0.51396906	9.725972	11.740694

Summary statistics:

Column	N	Mean	Variance	Std. Dev.	Std. Err.	Median	Range	Min	Max	Q1	Q3
Q6 How many credit hours are you taking this term?	175	10.748571	7.9249263	2.8151245	0.21280341	11	16	2	18	10	12

8.

Assume that the MM207 Student Data Set is a random sample of all Kaplan students; estimate the proportion of **all Kaplan students who are male** using a 90% level of confidence.
My answer:
lower .150 upper.250

Frequency table results for Gender: so that's 35/175 = 0.2
1-0.2=0.8

sqrt(0.2*(0.8/175) = sqrt(.2*0.004571428571428571428571428571428571143)
=0.030237157840738178177716132289874

my z = 90% or 1.645
margin of error = 1.645*0.030237157840738178177716132289874
=0.049740124648014303101430376168434

lower 0.2-0.049740124648014303101430376168434
=0.15025987535198569689856962383157
upper 0.2+0.049740124648014303101430376168434
=0.24974012464801430310143037616843

lower=.150
upper=.250

to get total students for this calculation

Gender	Frequency
Female	138
Male	35

+ The 2 who didn't answer = **175 students**

9.

Assume you want to estimate with the proportion of students who commute less than 5 miles to work within 2%, what sample size would you need?
My answer:
2128 students are needed for the sample size (based upon the 175 student total count)

<u>**175 total students took the survey**</u>

42+4+3+7+1+1 =58 students travel less than 5 miles to work

58/**175**=0.331428571428571428571428571428571428

1-0.331428571428571428571428571428571428=0.668571428571428571428571428571428571143

(1.96/.02)^2*0.331428571428571428571428571428571428*0.668571428571428571428571428571428571143

=9604 * 0.331428571428571428571428571428571428 = 3183.04

=3183.04 * 0.668571428571428571428571428571428571143

=**2128**.0896

<u>173 answered the question</u>

42+4+3+7+1+1 =58 students travel less than 5 miles to work

58/173=0.335260115606936416184971098
2659

1-0.335260115606936416184971098
2659=0.66473988439306358381502890
17341

(1.96/.02)^2 * 0.66473988439306358381502890
17341 =

= 9604*0.335260115606936416184971098
2659*0.66473988439306358381502890
17341

=**2140**.354839787497076414180226
5361

10. A professor at Kaplan University claims that the average age of all Kaplan students is 36 years old.

Use a 95% confidence interval to test the professor's claim. Is the professor's claim reasonable or not? Explain.

My answer:

YES, it is because the intervals are roughly 36-41…the professor's claim is pretty accurate.

interval are 36.36-41.137

size= 175

mean=37.94857

standard deviation=10.726628

sum=6641

6641/175=**37.94**857142857142857142857142857142857142857 or 37.95

sqrt of 175=13.2287565553229529525080787
68196

st deviation/sqrt of 175

= 10.726628/13.2287565553229529525080787
68196=<u>0.81085685983720420676031901680653</u>

z = 1.96 or 95%

margin of error = 1.96 * 0.81085685983720420676031901680653

= 1.5892794452809202452502252729408

37.94857-1.5892794452809202452502252729408=36.35929055471907975474977472 7059

37.94857+1.5892794452809202452502252729408=41.12712889056184049050045054 5882

From StatCrunch

Summary statistics:

Column	n	Mean	Variance	Std. Dev.	Std. Err.	Median	Range	Min	Max	Q1	Q3	Sum
Q2 How old are you?	175	37.94857	115.060555	10.726628	0.8108569	36	44	21	65	29	46	6641