PART I OF THE EXAM MULTIPLE CHOICE QUESTIONS (40 points, 2 points each). You may write on the exam but please MARK YOUR ANSWERS IN THE SCANTRON and turn in the scantron before proceeding to answer PART 2

There are five possible responses to each of the following multiple choice questions. There is only one "BEST" answer. Be sure to read all possible choices before selecting your answer. You may mark on this examination. You may use a calculator but a calculator manual may not be used.

Tables for the Normal, Student's t, and Binomial distributions will be provided. You will have 3 hours to complete the entire exam. You can take a 10 minute break when you have finished parts 1 and 2 before working on part 3. This is a closed book exam. A formula sheet will be provided.

1. Which of these questions from the Spring 2016 MATH1530 class survey produced variables that are quantitative?

   i. What is your age?

   ii. Do you favor or oppose daily prayer in the classroom? (Favor or Oppose)

   iii. How much do you believe minimum wage should be? (Enter in $0.00 format.)

   iv. What is your favorite way of spending an evening? (Staying at home, Watching Television, Resting/Relaxing, Reading, Visiting with friends, Other)

   v. What is your religious identification? (Christian Religion, Non-Christian Religion, None)

   (A) i, iii (B) ii, iv, v (C) i (D) iii (E) i, ii, iii, iv, v

2. The Spring 2016 MATH1530 class survey asked "How many different people do you text on a normal day?" The table below represents the responses of 702 students.

Number of different people texted	0	1	2	3	4	5	6	7	8	9 or more
Number of Students	13	32	75	129	119	142	51	31	22	88

   Estimate the median number of different people that the 702 students texted on a normal day.

   (A) 351 (B) $(119 + 142)/2 = 130.5$ (C) 5 (D) 4.5 (E) 4

   Use the following for the next question. The table below represents the responses of 708 students to the MATH1530 survey question "What is your favorite way of spending an evening?"

Evening Time	Visiting with Friends	Resting/Relaxing	Other	Staying at Home	Watching TV	Reading
Count	259	166	97	78	61	46

3. What can you conclude about this distribution?

   (A) It is somewhat symmetric with 'Other' as the center of the distribution.

   (B) It is left-skewed with a spread from 'Visiting with Friends' to 'Reading.'

   (C) It is right-skewed with 'Other' being the median.

   (D) 'Visiting with Friends' is the most popular way of spending an evening and the least popular time spent is 'Reading.'

   (E) Min = 'Reading', Q1 = 'Watching TV' to 'Reading', Median = 'Other', Q3 = 'Visiting with Friends' to 'Resting/Relaxing', Max = 'Visiting with Friends'

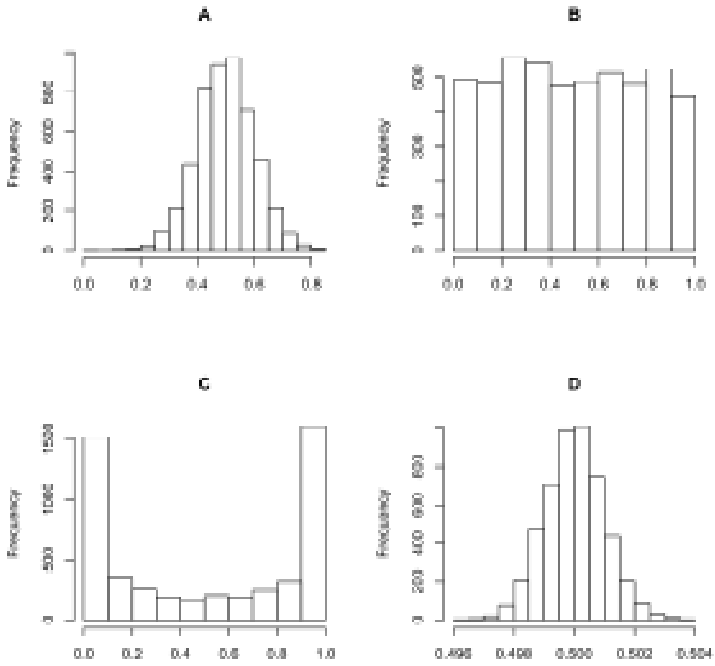
4. The scores students earned on an easy test range from 0 to 100. There are many scores ranging from 85 to 100, some scores ranging from 70 to 80, and a few scores below 70. The distribution of test scores will be

   (A) skewed to the left. (C) roughly symmetric.

   (B) skewed to the right. (D) roughly normal.

   (E) correlated to age of the student.

5. Which of the following histograms has the largest standard deviation? Each histogram ranges from 0 to 1 and is fairly symmetric around 0.5.



(A) Histogram A (B) Histogram B (C) Histogram C (D) Histogram D (E) No way to tell.

6. The use of _____ is the most effective way of establishing causality between two variables.

(A) a scatterplot
(B) a positive correlation
(C) a two sample t-test
(D) an observational study
(E) a randomized comparative experiment

7. Ann Landers asked her readers if they would choose to have children again. An incredible 70% of respondents said that if they lived their lives over again, they would not have children. A similar poll in Good Housekeeping the same year resulted in 95% of people saying they would have children. You can conclude that

(A) these polls used voluntary response, so their results tell us little about the population of all adults.
(B) about (70+95)/2 = 82.5% of adults would not have children if they lived their lives over again.
(C) more parents still need to respond on the question, as a larger sample is required to reduce bias.
(D) both polls reveal that most people would not have children again.
(E) the Good Housekeeping poll is more reliable than the Ann Landers poll since 95% is larger than 70%.

8. A very comprehensive study of an expensive new diet pill was conducted. After looking at 100,000 patients for a year they find that in almost every case everyone that took the diet pill lost exactly one pound. There was little variance, everyone lost that one pound, so there was a statistically significant loss of weight. The explanation is

(A) the results are of practical significance.
(B) that the sample size is large and it isn't surprising to find the results to be statistically significant.
(C) that mean weight loss of 1 pound is large compared to the mean weight loss of most diets.
(D) new diets typically have less variability than standard diets, and so small differences can appear to be statistically significant.
(E) all of the above.

9. Suppose a farmer wishes to work out the average milk yield of each cow type in his herd, which consists of Ayrshire, Friesian, Galloway, and Jersey cows. He could divide up his herd into the four sub-groups and take random samples from each subgroup. This is an example of

   (A) a simple random sample. (C) a stratified random sample.

   (B) a multistage random sample. (D) an observational sample.

   (E) an experiment.

10. There is a very promising new cancer drug that only has a very small amount to test for further study. In an early clinical trial, the research found that there was a dramatic reduction in cancer growth rate on three subjects. Although the cancer growth rate dramatically decreased, the results are not statistically significant. The explanation is

    (A) the placebo effect is present, which limits statistical significance.

    (B) the calculation was in error. The researchers forgot to include the P-value.

    (C) that although the cancer growth rate has decreased, the test statistic was large in magnitude.

    (D) that the sample size is small and it isn't surprising to find the results not statistically significant.

    (E) the results are not of practical significance.

Use the following for the next 2 questions. The MATH1530 survey asked students "What political party do you identify with?" and " Do you favor or oppose daily prayer in the classroom?" The distribution of counts is shown in the table

Political ID	**Favor**	**Oppose**	**Total**
Republican	200	32	232
Democrat	72	94	166
Independent	105	74	179
Other	80	51	131
Total	457	251	708

11. What percent of the students surveyed favor daily prayer in the classroom?

    (A) 86% (B) 65% (C) 44% (D) 33% (E) 28%

12. Given that a student favors daily prayer in the classroom, what percent are Republicans?

    (A) 86% (B) 65% (C) 44% (D) 33% (E) 28%

Use the following for the next question. The Pick 4 games in many state lotteries announce a four-digit winning number each day. Each of the 10,000 possible numbers 0000 to 9999 has the same chance of winning. You win if your choice matches the winning digits. Suppose your choice is 0011.

13. What is the probability that the winning number matches your number exactly?

    (A) $11/10000$ (B) $4/10000$ (C) $1/10000$ (D) $1/9999$ (E) 0

14. If you flip a fair coin and roll a fair 6-sided die, what is the probability that you will flip a heads and roll a 5?

    (A) 3 (B) $2/3$ (C) $1/3$ (D) $1/12$ (E) 0

15. Which of the following is an example of a matched pairs design?

    (A) A teacher compares the scores of students using a computer-based method of instruction with the scores of other students using a traditional method.

    (B) A teacher compares the pre-test and post-test scores of students.

    (C) A teacher compares the scores of students in her class on a standardized test with the national average score.

    (D) A teacher compares her class's average score on a standardized test with the national average score.

    (E) None of the above.

16. Several factors are involved in the creation of a confidence interval. Among them are the sample size, the level of confidence, and the margin of error. Which statement is true?

    (A) For a specified confidence level, larger samples provide smaller margins of error.

    (B) For a given sample size, higher confidence means a smaller margin of error.

    (C) For a fixed margin of error, smaller samples provide greater confidence.

    (D) For a specified confidence level, halving the margin of error means halving the sample size.

    (E) All of the above.

17. The null and alternative statistical hypotheses are: $H_0 : \mu = 25$ and $H_a : \mu > 25$. A random sample of size 40 is selected, the sample mean is 32, We use software to perform test, and the p-value was found to be 0.005. Which is the correct interpretation of that p-value ?

    (A) When the mean of a sample of size 40 is 32, the probability that the population mean is 25 is only 0.005.

    (B) When the mean of the population is 25, the probability of obtaining a sample mean 32 or higher in a random sample of size 40 is only 0.005.

    (C) The chance that the null hypothesis is true is 0.005.

    (D) The chance that the alternative hypothesis is true is 0.005.

    (E) The probability of rejecting H_0 when it is true is 0.005.

18. The results from a t-confidence interval (and the t-test) may NOT be reliable when

    (A) the sample size is very large.

    (B) we do not know the standard deviation of the population.

    (C) the distribution of the variable is normal.

    (D) the sample size is small, the distribution of the variable is very skewed, the sample data has extreme outliers on one side.

    (E) the t-confidence interval and the t-test ARE ALWAYS reliable, when we calculate a t and a p-value we are fine.

19. In a given situation, the consequences of Type I error are extremely serious. Which of these would be the best strategy?

    (A) You choose a p-value smaller than α.

    (B) There is nothing you can do.

    (C) Fix the α at a very small value and work with a large enough sample

    (D) Look at the p-value first and then go back and make α larger than the p-value.

    (E) Look at the p-value first and then go back and make α smaller than the p-value.

20. What is the p-value?

A. The probability that the null hypothesis is true.

B. The proportion of people in the population who have a certain trait

C. The probability that the null hypothesis is false

D. The probability of obtaining a value for the statistic of interest based on the data collected in a survey or experiment, or a more extreme value when the null hypothesis is TRUE

E. The probability of obtaining a value for the statistic of interest based on the data collected in a survey or experiment, or a more extreme value when the null hypothesis is FALSE

MATH 1530 COMPETENCE EXAM FOR GRADUATE TEACHING ASSISTANTS – April 2016
PART B (Computers should not be used for this part, of course calculators can be used)
40 points in this part

1. (9 points,3 each part)
   A large group of male runners walk on a treadmill for 6 minutes. Their heart rates in beats per minutes (BPM) at the end vary from runner to runner according to the Normal distribution with mean 110 BPM and standard deviation 15 BPM.
   a) What percent of runners have heart rates above 130? (Show your work including a sketch of the distribution and shade in the area you are trying to calculate)

   b) For a medical study, we want to select runners who are in the lower 20% of the heart rates. What would be the maximum heart rate that somebody could have in order to qualify for this study? (Show your work including a sketch of the distribution and mark the value you are trying to find)

   c) If we were to take all the possible samples of size 5 from that population and record their sample means. What would be the distribution of all those sample means? **Why?** (show your work and include a sketch of the distribution)

2. (12 points) This question is about opinions about the minimum wage

a) (2 points) One of the candidates to the presidency of the USA is proposing that the minimum wage should be $15. We want to know if the average minimum wage based on the opinions of the population of ETSU students coincides with the proposal of that candidate. Some of us think that maybe ETSU students are not so ambitious. Using μ to represent the mean quantity that ETSU students think the minimum wage should be, write the null and alternative hypotheses.

b) (3points) A survey was conducted among ETSU students, a random sample of 694 students was selected, and one of the questions asked was 'How much do you believe minimum wage should be?' The sample mean and the standard deviation of the answers given by the students were $9.42 and $2.242, respectively.
Calculate the test statistic to test the null hypothesis you wrote in part a). Based on that value, how do you feel about the null hypothesis? Why? Show your work including a sketch.

c)(3 points) Based on the mean and standard deviation calculated from the sample, calculate AND interpret a 95% two-sided confidence interval for the population mean. Show your work.

d)(4 points) Do we really need such a large sample to calculate the confidence interval? Next time we do such a survey we would like to save some work and money so we want to select a smaller sample. How large does the sample size need to be in order to do the estimation with a margin of error of 0.5 with 95% confidence? Assume that the standard deviation in the population is $2.5. (Show your work.)

3) (10 points)

The population of the United States is aging, though less rapidly than in other developed countries. Here are the percents (sorted) of residents aged 65 and older in the 50 states, according to the 2000 census.

5.7	8.5	9.6	9.7	9.9	10.6	11.0	11.2	11.2	11.3	11.3	11.6	11.7	11.7	12.0	12.0	12.1	12.1	12.1	12.1
12.3	12.4	12.4	12.5	12.7	12.8	12.9	13.0	13.0	13.0	13.1	13.2	13.2	13.3	13.3	13.3	13.4	13.5	13.5	13.6
13.8	14.0	14.3	14.4	14.5	14.7	14.9	15.3	15.6	17.6										

a) (3points) Construct a stem and leaf display.

b) (3 points) Sketch the boxplot. (Please be as precise as possible when doing the sketch.)

c) (2points) Are there any outliers?

d) (2points) Do you think the distribution is unimodal or multimodal?

4)(9 points, 3 each) This question is about favoring or opposing daily prayers in the classroom in the school system.

a) Assume that in a certain region of the country only 30% of the people are in favor of daily prayers in the classroom. What is the probability that 8 or more are in favor of daily prayers from a random sample of 15 individuals? (Show your work or explain your answer.)

b) In a random sample of 708 ETSU students, 457 answered that they agree with daily prayers in the classroom. Calculate a 95% confidence interval for the proportion of all ETSU students who agree with daily prayers in the classroom. (Show your work.)

c) Next time we do such a survey, what should our sample size be if we want to have a margin of error of only 2% and still be 95% confident?

PLEASE TURN IN YOUR ANSWERS FOR PART 2 BEFORE WORKING THE NEXT PART USING THE COMPUTER

PART 3 OF THE EXAM (20 points)

PLEASE, TURN ON THE COMPUTER NOW. You are expected to use statistical software (Minitab or R) to solve this part. You will prepare a Word document with your analysis inserting the computer output you get including graphics, interpretations answers to the questions etc. and email it to lewiscn2@etsu.edu & seier@etsu.edu
You will work with the data set 'kneesurgery'. There are two versions of the data file, one in Minitab format and one is a text file (kneesurgery.mtw and kneesurgery.txt). You are welcome to use any of them. The data file contains information for 200 patients who underwent knee surgery. Half of them needed a blood transfusion during or right after surgery and the other half did not. The variables are:

- Trans : Needed transfusion (1 = 'yes', 0= 'no')
- Age : age of patient
- BMI : Body mass index
- preopHb : Pre-operative hemoglobin
- CVCAD : Cardio vascular artery disease (1 = 'yes', 0= 'no')
- Cirsysdis : circulatory system disease (1 = 'yes', 0= 'no')

The purpose of this analysis is to explore what variables are associated to the need for a transfusion during or right after surgery. We will organize the analysis by asking specific questions. Please insert the plots and numerical output in your report

1. We want to know how old are the people that we are studying. Obtain an appropriate plot for the variable age and comment on the shape of the distribution. Calculate the five-number-summary for age. Are there any outliers in the data set (for age)? Justify your answer. Interpret the value of the median.
2. We want to know if the quantitative variables in the data set are correlated. Produce appropriate plots and calculate the necessary statistics to answer this question.
3. Is there a difference in mean preoperative hemoglobin between those who needed transfusion and those who did not? Obtain a plot and perform a test of hypothesis to answer the question. Interpret the results.
4. Produce two plots to explore if there is any difference in median age and median BMI between those who need transfusion and those who don't.
5. Do you think that needing transfusion is independent of having or not Cardio vascular artery disease? Do you think that needing transfusion is independent of having or not circulatory system disease? Perform appropriate tests and state your conclusions.

After answering each question, write a short paragraph summarizing your findings and answering the main research question.