

STAT 145 Practice for Part 2 of the Unit 1 Exam: Technology Required; One-sided cheat-sheet allowed; Scrap paper allowed. If rounding is needed, please leave 3 decimal places.

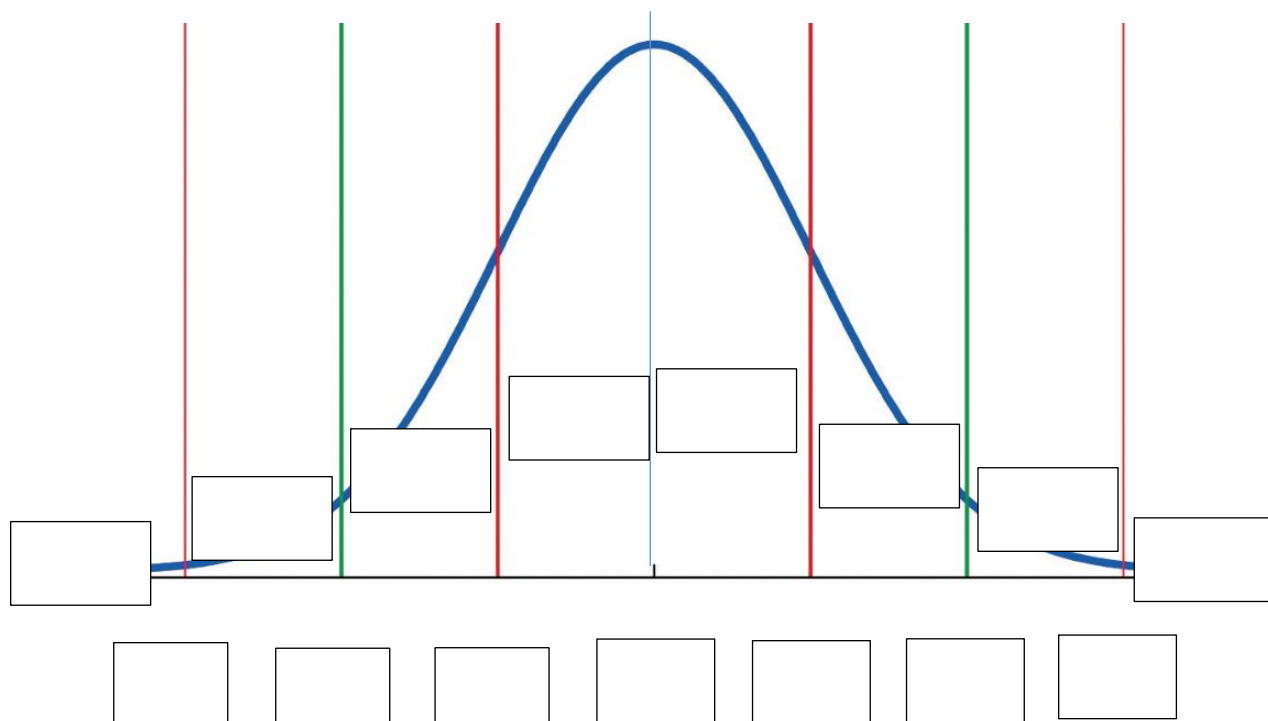
Part I of your exam will be a 10-question MyCourses quiz. An OPTIONAL 30-question MyCourses quiz to practice for Exam 1 is available to you.

Once submitted, a Word document becomes available in MyCourses→Content along with an Excel file with any data. You will type into this document or create a new document with your answers and technology output. These practice problems will give you practice with the format of your exam. Text boxes are provided for you to type into. If that becomes challenging, delete the textbox and type your answer in the space provided.

Problem #1

Owners of a minor league baseball team believe that a *Normal Model* is useful in projecting the number of fans who will attend home games. They use a mean of 7500 fans and a standard deviation of 700 fans.

- A. Complete a picture of the empirical rule for this data – label the values and percentages.
[On your exam, if you would rather draw this out and insert an image, that would be fine]



Answer the following questions:

- B. What percentage of home games do you expect to have between 6100 and 7500 fans in attendance?

- C. What number of fans represents the 16th percentile?

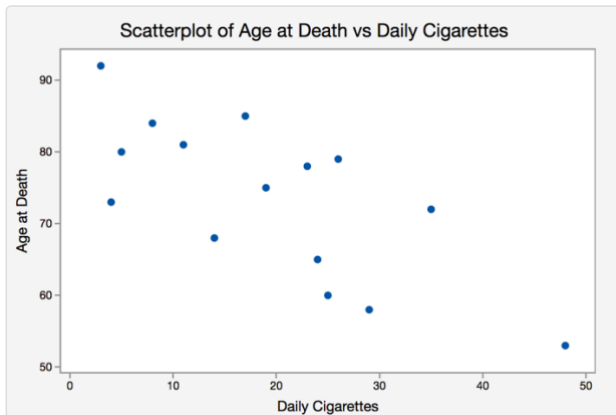
- D. Eighty-four percent of home games are expected to have an attendance less than...

E. In August, there was a home game with 10,400 fans in attendance. Calculate the Z-score for this number of fans. Show your work. Round to the nearest hundredth.

F. Are 10,400 fans considered unusual for this sample? Explain using your results from part E.

Problem #2

A random sample of cigarette smokers were studied and, upon their death, their age (in years) and number of daily cigarettes smoked were recorded. The data are found in the file: 4 Week 4 Review for Exam.xlsx and the sheet is labeled **Smokers**.



Daily Cigarettes	Age at Death
5	80
23	78
25	60
48	53
17	85
8	84
4	73
26	79
11	81
19	75
14	68
35	72
29	58
3	92
24	65

A. Which variable is the predictor variable?

B. Calculate the correlation value for these data:

Interpret the correlation value, *in the context of the problem*. (Use a complete sentence)

C. Provide the equation of the regression line for this data.

D. Interpret the slope of this line, *in the context of the problem*. (Use a complete sentence)

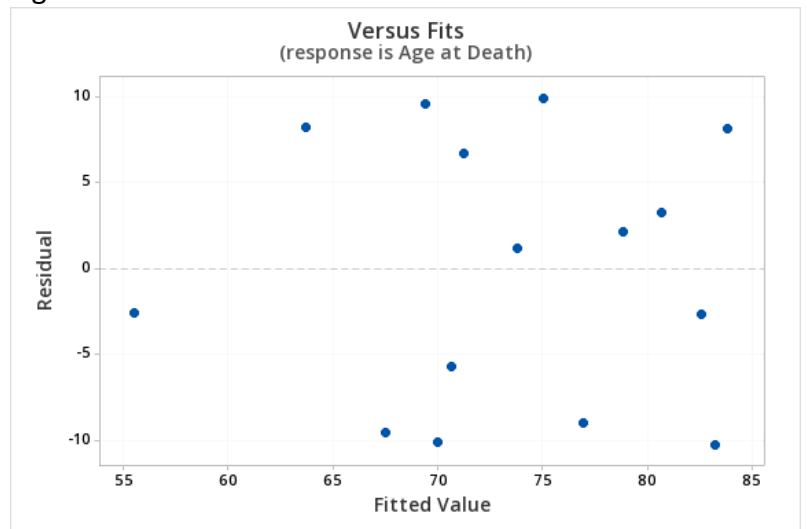
E. Calculate the value of r^2 for these data:

Interpret the value of r^2 , *in the context of the problem*. (Use a complete sentence)

- F. If the number of cigarettes smoked daily is 14, what do we predict is their age at death? Show work or output.

- G. Calculate the residual for when 14 cigarettes are smoked daily. Show your work.

- H. A residual plot is provided for this linear regression model. Can we assume a linear model is a valid model for these data? Explain.



Problem #3

Scientists have studied how long it takes a person to recall a specific memory. The data is recall times (seconds) for a random sample of subjects when asked to recall a **pleasant memory**. The data can be found in the sheet labeled: **Pleasant Memory**.



0.95	1.17	1.22	1.42	1.63	1.98	2.12	2.32	2.38	2.43
2.93	3.03	3.15	3.22	3.42	4.63	4.70	4.88	6.17	

- A. Using the recall times for pleasant memories, calculate the following and fill in the blank column.

Mean =	
Stand. Dev =	
Min	
Q1	
Median	
Q3	
Max	
IQR	

- B. Calculate the lower and upper fences or paste the output.

- C. Interpret the lower and upper fences:

- D. Does this data set have any outliers? Explain how you know.

E. Create a boxplot using technology and paste it here:

F. Describe the shape of the distribution for **pleasant** memories and explain.

G. Indicate the recall time that is typical for **pleasant** memories and explain why you chose this value.

Problem #4

The Scholastic Aptitude Test (SAT) is taken by many high school students. The Math SAT scores for a sample of local high school students are approximately bell-shaped and the summary statistics provided below.

Mean =	531.0
Stand. Dev =	38.0
Min	341
Q1	505
Median	533
Q3	555
Max	800

A. Use a complete sentence and the context of the problem to interpret the sample mean.

B. Explain why the empirical rule intervals and percentages can be applied to these data.

You do not have to show the empirical rule intervals and percentages for these data but will need them to answer the following questions:

C. About 68% of the students scored between which two values on the Math SAT?

D. Only 2.5% of students scored above _____ on the Math SAT.

E. Find and interpret the 84th percentile of scores in the context of the problem.

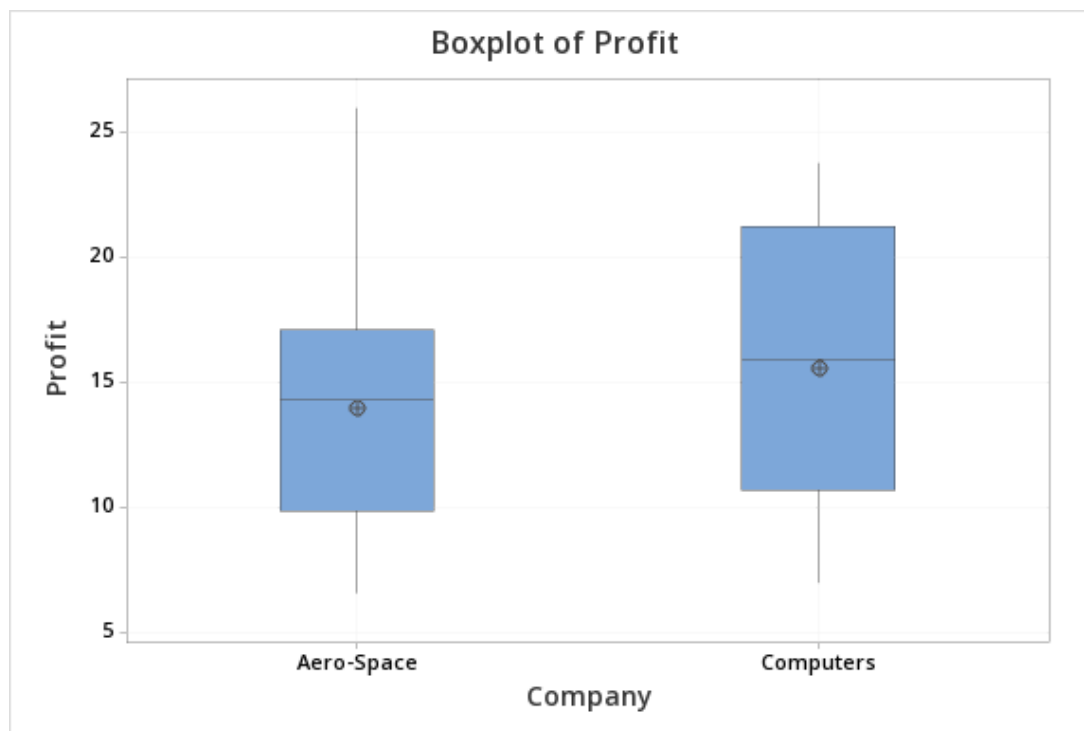
- F. Your cousin earned a Math SAT score of 528. Find the z-score for her Math SAT score. Round to the nearest hundredth.

- G. Interpret this Z-score in the context of the problem.

- H. Based on the Z-score, how did your cousin do, compared to the students in the sample?

Problem #5

A random sample of annual profits per employee (in thousands of dollars) has been collected from two types of companies: Computer and Aero-space. Comparative box plots were produced and can be seen below.



- A. An analyst has determined that both companies profit distributions are somewhat symmetric. Explain why this is a reasonable statement.

- B. Based on the comparative box plots, which company has the higher profits per employee, on average? Explain your decision.

- C. Your colleague claims that the Aero-space companies profit data has much more variability than the computer data. Using a reliable measure of spread, explain why this is not true.