Midterm 1 - STAT 301
Fall 2015

Name:

UIN:

Seat Number:

Signature:

1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You can only bring the cheat sheet that I have posted and the normal z-tables to this exam.
4. NO BASEBALL CAPS can be worn in this test (unless you have a good reason).
5. You have 60 minutes to work on this exam. There are 15 multiple choice questions. If you cannot do one question move on to the next.
6. You may use a calculator in the exam (but not a phone).
7. Partial credit is only given to answers which are correct (not partially correct).
8. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
9. I use sd to denote standard deviation.
10. Good Luck!!!

(1) The IQR (interquartile range) of a data set is zero. What does this tell us about the data set?

(A) At least 50% of the data takes the same value.

(B) All of the data takes the same value.

(C) At least 75% of the data takes the value.

(D) All of the values in the data are zero.

(E) The median is zero.

(2) Everytime an agent takes a call at a call center, the following data is collected.

1	2	3
Type of problem	Number code of Agent	Length of call with customer

What type of variables are the above?

	1	2	3
A	Categorical	Categorical	Numerical continuous
B	Numerical continuous	Categorical	Numerical discrete
C	Numerical discrete	Categorical	Numerical continuous
D	Numerical discrete	Numerical discrete	Numerical continuous
E	Numerical discrete	Categorical	Categorical

(3) Suppose human weights are normally distributed. What proportion of human weights are within **0.5** standard deviations (pay attention to the number of standard deviations) of the mean (calculated to the nearest percentage)?

(A) 64% (B) 38% (C) 68% (D) 34%

(E) It is impossible to say without knowing the mean and standard deviation.

(4-5) We consider the following four data sets: (i) Rolls of a 'fair' six-sided die (Dice) (ii) Maximum monthly temperatures in Antarctica (Max) (iii) Minimum monthly temperatures in Antarctica (Min) (iv) Number of siblings of a student (Siblings). Temperatures are given in Fahrenheit. Their relative frequency histograms are given below.

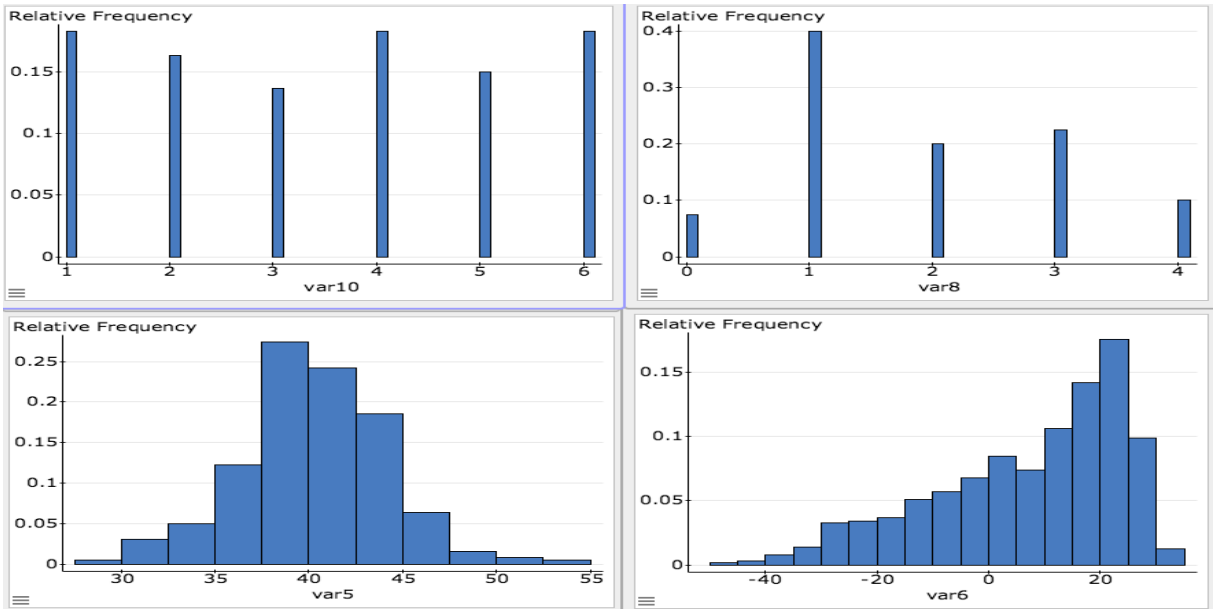


Figure 1: Label: Top Left = (1), Top Right (2), Bottom Left = (3), Bottom Right = (4).

(4) Match the plot to the data set.

	1	2	3	4
(A)	Siblings	Dice	Max	Min
(B)	Dice	Siblings	Max	Min
(C)	Siblings	Dice	Min	Max
(D)	Max	Min	Siblings	Dice
(E)	Min	Siblings	Max	Dice

(5) Match the mean and standard deviation to the data set (using the histogram as a guide).

	Min Temp	Max Temp	Die	Siblings
(A)	Mean = 3, sd = 1.7	Mean = 1.8, sd = 0.9	Mean = 40, sd = 5	Mean = 1.8, sd = 16.5
(B)	Mean = 40, sd = 7	Mean = 6.8, sd = 16.5	Mean = 1.8, sd = 1	Mean = 3, sd = 2.2
(C)	Mean = 6.8, sd = 60	Mean = 40, sd = 30	Mean = 3, sd = 6	Mean = 1.8, sd = 5
(D)	Mean = 6.8, sd = 17	Mean = 40, sd = 4	Mean = 3, sd = 1.7	Mean = 1.8, sd = 0.9
(E)	Mean = 40, sd = 40	Mean = 6.8,sd = 60	Mean = 3, sd = 6	Mean = 1.8, sd = 5

(6)

1. A study of 2050 children in 4th-6th grade in 30 schools recordered their behaviours at lunch. One of the conclusions of the study was that girls tend to discard more food than boys.
2. A magazine article tells the story of an elderly couple (male and female) who took Pilates classes. After taking the classes, the chronic back pain the man was suffering from had gone whereas the chronic back pain the female was suffering from remained. The article concludes than Pilates benefits males more than females.
3. A class of students were randomly assigned to two groups. Group A was asked to do power weights everyday whereas Group B was asked to run one mile everyday. After one month it was reported that the bone strength in Group A was 30% more than the bone strength in Group B.

What type of studies are the above?

	1	2	3
(A)	Biased	Anecdotal	Observational
(B)	Biased	Observational	Confounding
(C)	Observational	Anecdotal	Experimental
(D)	Observational	Experimental	Anecdotal
(E)	Observational	Experimental	Biased

(7) Which statement is **wrong**?

(A) The mean salary of a worker in HEB is \$10 an hour the standard deviation is \$1. HEB announces that it will increase salaries by 2%. The mean salary increases to \$10.20 and the standard deviation is now \$1.02.

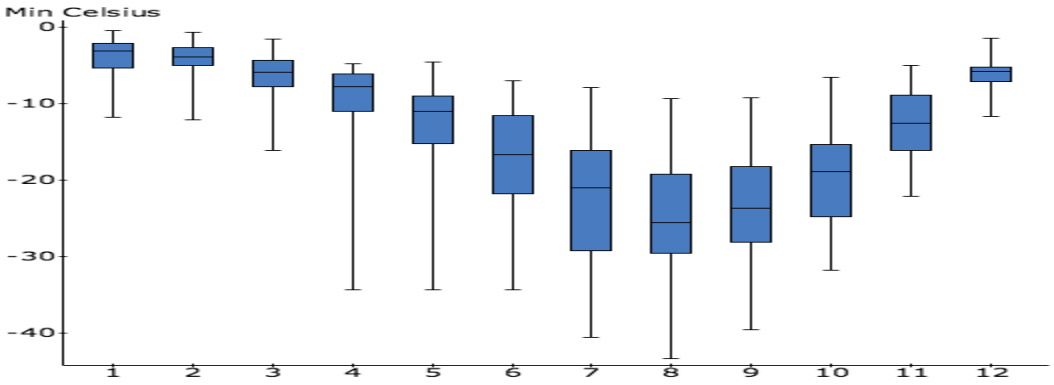
(B) If all the values in a data set are 50, the mean is 50 and the standard deviation is zero.

(C) In a LEFT skewed distribution the mean tends to be LESS than the median.

(D) The mean age of a sample of students is 20.1 years with standard deviation 1 year. In 3 years time the mean age of the students in this sample will increase to 23.1 years and the standard deviation will increase to 3.1 years.

(E) The distribution of the number of siblings a student **is not** normal distribution.

(8) From 1950-2000 the minimum temperatures every month at a station in Antarctica was recorded. A boxplot for the temperatures each months is given below.



Remember Antarctica is in the **Southern hemisphere**, so Month 1 = January (which corresponds to summer) and Month 6 = June (which corresponds to winter).

Based on the plot, which description best describes the data.

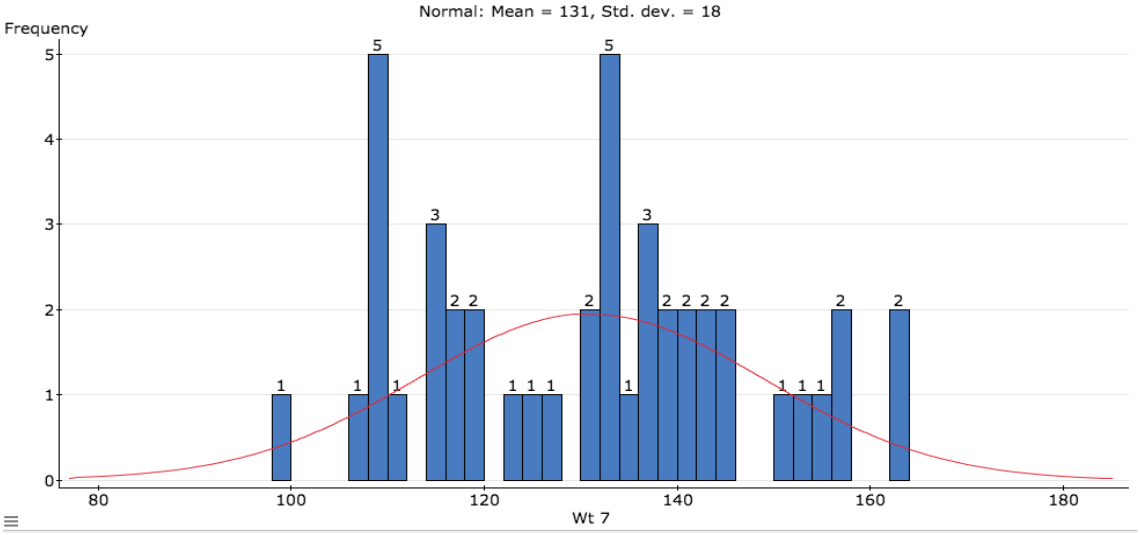
(A) The temperatures tend to be lower in the winter and there is more variability in temperatures over winter.

(B) The temperatures tend to be lower in the winter and there is more variability in temperatures over summer.

(C) For some of the months, the temperatures appear to the LEFT SKEWED.

(D) Answers [A] and [C] (E) Answers [B] and [C].

(9) The weights of **44** calves are followed from week 0 to week 7. The plot below gives the **frequency histogram** for the weights of calves at week 7 (note a bin width of 2 pounds was used). Calculate the proportion of calves (in **this sample of 44** calves) whose weight lies within **two** standard deviations of the mean.



(A) 95% (B) 100% (C) 44% (D) 68% (E) 92%

(10) Sam scores a C in a multiple choice exam. He brags to his friends that he just 'guessed' every answer.

In a multiple choice exam the chance of getting a C by simply guessing every answer is 0.05%. What can we say about Sam's claim (give the answer which most accurately describes the conclusion)?

(A) Sam was definitely guessing.

(B) Since 0.05% is a small probability, it is **impossible** that Sam guessed his way through the exam. Therefore Sam definitely knew the material.

(C) 0.05% is a relatively small probability, therefore it is **unlikely** that Sam guessed all the answers.

(D) We can only answer this question if the data is normally distributed.

(E) [B] and [D].

(11) The mean height of Dutch men is normally distributed with mean 74 inches with standard deviation 5 inches. What proportion of Dutch men are between 71 to 78 inches?

(A) 79% (B) 51% (C) 27% (D) 20% (E) 80%

(12) The mean height of Dutch men is normally distributed with mean 74 inches with standard deviation 5 inches. A Dutch man is the **top** 15th percentile. How tall is he?

(A) 68.18 (B) 74.75 inches (C) 73.25 inches (D) 78.25 inches (E) 79.2 inches.

(13) The height of Dutch men is normally distributed with mean height **74** inches and standard deviation 5 inches, whereas Dutch women are normally distributed with mean height **70** inches and standard deviation 4 inches. In a Dutch family, a brother is **73** inches tall and the sister is **71** inches tall. **Relative to their gender** (compare percentiles), which statement is correct?

(A) The brother is in the 58% percentile whereas the sister is in the 60% percentile. The sister is slightly taller.

(B) The brother is in the 20% percentile whereas the sister is in the 25% percentile. The sister is slightly taller.

(C) The brother is in the 42% percentile whereas the sister is in the 40% percentile. The brother is slightly taller.

(D) The brother is in the 42% percentile whereas the sister is in the 60% percentile. The sister is taller.

(E) The brother is 2 inches taller than his sister, so he is taller.

(14) The mean weight of calves at week 1 is 89 pounds with standard deviation 8.5 pounds. A one week old calf who weighs 70 pounds is presented. How many standard deviations is she from the mean. Given the data what can you say about the calf?

(A) The calf's weight is 2.23 standard deviations to the LEFT of the mean. This is quite far from the bulk of weights for heavy calves, and suggests she is too light for a healthy calf.

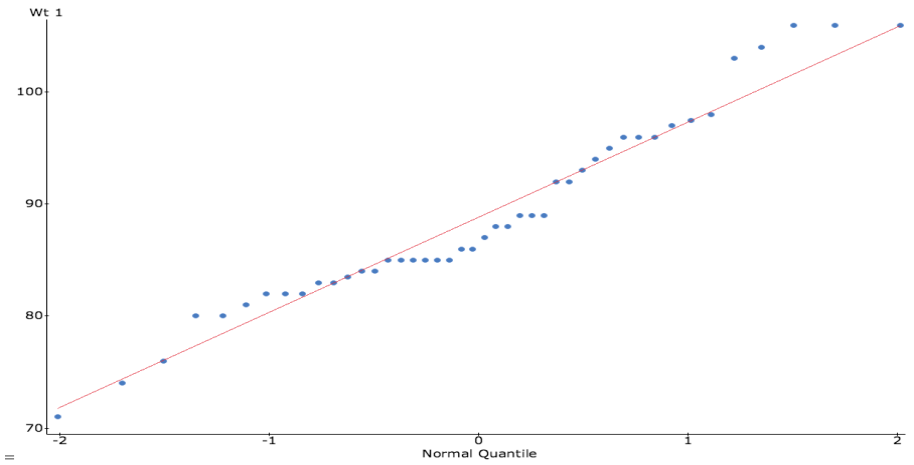
(B) The calf's weight is 2.23 standard deviations to the RIGHT of the mean. This is quite far from the bulk of weights for heavy calves, and suggests she is too heavy for a healthy calf.

(C) The calf's weight is 2.23 standard deviations to the LEFT of the mean. This is quite close to where the majority of healthy calf weights lie, so she must be healthy.

(D) The calf's weight is 19 standard deviations to the LEFT of the mean weight, she must be ill.

(E) The calf's weight is 7.9 standard deviations to the LEFT of the mean weight, she must be well.

(15) The QQplot of week 1 calves is given below. What does this plot suggest?



(A) The plot does not deviate hugely from the line (though it does not fit exactly), suggesting there is a clear **correlation** between weight of calf and age.

(B) The plot does not deviate hugely from the line (though it does not fit exactly), suggesting the data does not deviate hugely from normality.

(C) Due to rounding of the weights, some of the weights in the data set are the same.

(D) Answers [A] and [B] (E) Answers [B] and [C].