Midterm 1 - STAT 301
Fall 2013

Name:

UIN:

Signature:

Version A:

1. Do not open this test until told to do so.

2. This is a closed book examination, However you may use one single-sided sheet of formulas that you have brought with you and the tables. You should have no other printed or written material with you on the exam. But scrap paper is allowed.

3. You have 50 minutes to work on this exam. There are 15 multiple choice questions.

4. On the scantron please state the version of exam that you have.

5. You may use a calculator in the exam.

6. If there is no correct answer or if multiple answers are correct, select the **best** answer.

7. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.

8. Please only give one answer per question (the one that is closest to the solution).

9. Good Luck!!!

(1) Suppose I add 1000 to every value in a data set. Which statement is true?

  (A) The mean, first quartile and standard deviation increase by 1000.
  (B) The mean increases by 1000 but the median stays the same.
  (C) The mean and first quartile increase by 1000 but the standard deviation stays the same.
  (D) Two of the above.
  (E) None of the above.

(2) The standard deviation of a data set is zero. Which statement is true?

  (A) The mean must be zero.
  (B) Each value in the data set must be zero.
  (C) All the values in the data set must be the same.
  (D) Two of the above.
  (E) None of the above.

(3) A television program is doing a survey to find out the average number of pets in a household. They randomly sample customers who go to PetCo (the pet supplies shop). According to national statistics, the average number of pets in a household is 0.8.

Comment on the reliability and accuracy of their estimate:

  (A) Since their sample consists of only customers from a pet supply shop their sample is likely to be be biased towards larger outcomes.
  (B) Since their sample consists of only customers from a pet supply shop their sample is likely to be be biased towards smaller outcomes.
  (C) Since they are randomly sampling from a pet supply shop their sample is likely to be reliable estimate of 0.8.
  (D) Two of the above.
  (E) None of the above.

(4) Machine A and Machine B are packing M&Ms. An inspector wants to know whether both machines pack M&Ms with the same distribution. He takes a random sample of 50 bags of M&Ms from both machines. The histogram and summary statistics are given in figure 2 and the table below. Which statement is correct?

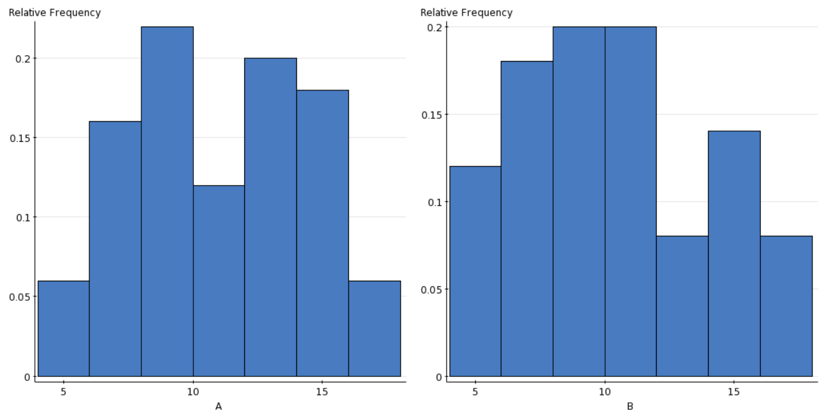


Figure 1: Left: Histogram for sample Machine A. Right: Histogram for sample Machine B.

Machine	sample mean	sample standard deviation
A	10.48	3.36
B	12.92	3.59

(A) The means, standard deviations and histograms are different therefore the number of M&Ms that machine A and B are packing <u>is different</u>.

(B) The means, standard deviations and histograms are different therefore the distribution of the number of M&Ms that machine A and B are packing is <u>probably different</u>.

(C) The means, standard deviations and histograms are similar therefore the number of M&Ms that machine A and B are packing <u>have the same distribution</u>.

(D) There are some differences in the means, standard deviations and histograms, but it is possible that these differences could be due to sample variations. It is not possible to say whether they have the same distribution or not.

(E) Yes it is.

(5) The mean height of a 5 year is 50 inches with standard deviation 5 inches. Below are the distributions of the sample means (based on 200 simple random samples) for two different sample sizes and one red herring (not the sample mean of 5 year olds). Place histogram with the sample size.

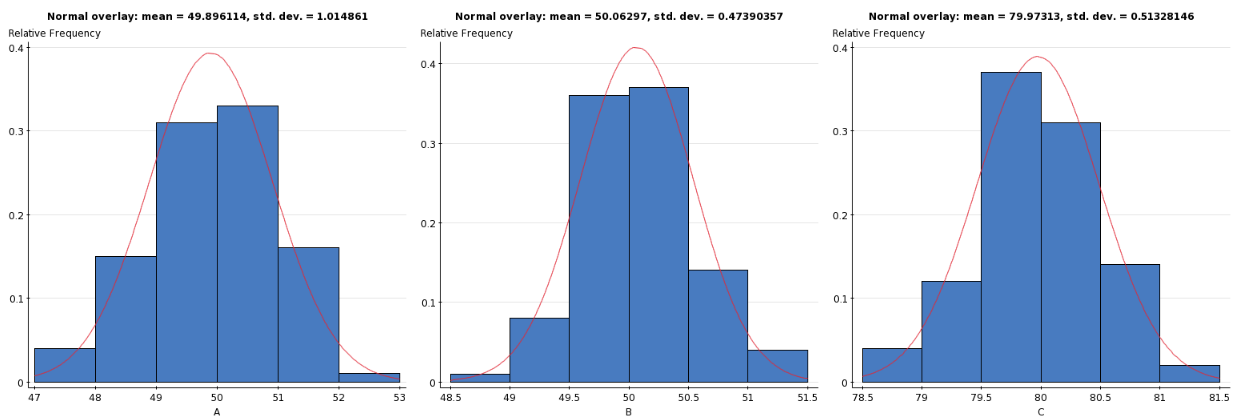


Figure 2:

Plot	sample mean	sample standard deviation
A	49.8	1.01
B	50.06	0.47
C	79.9	0.51

(A) A = red herring, B = sample size 25, C = sample size 100.

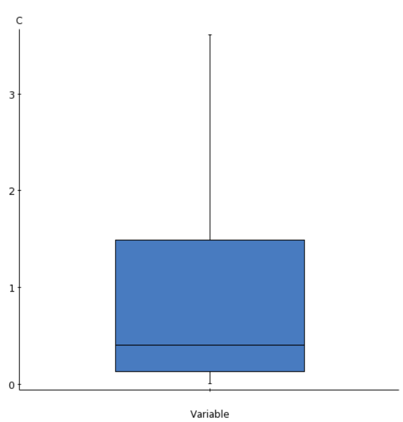
(B) A = sample size 25, B = red herring, C = sample size 100.

(C) A = sample size 25, B= sample size 100, C = red herring.

(D) A = sample size 100, B= sample size 25, C = red herring.

(E) A = red herring, B = sample size 100, C = sample size 25.

(6) Given the boxplot below, what distribution is it likely to have?



(A) A tri-modal distribution.

(B) Left Skewed.

(C) Thick tails.

(D) Right Skewed.

(E) The mean is likely to be less than the median.

(7) A researcher in calf breeding wants to know the mean body temperature of new born calves. The temperature varies from calf to calf. The mean temperature (μ) is unknown, but standard deviation of calf temperatures is known to be $\sigma = 1.2$F.

She takes the temperatures of a random sample of 100 new born calves. The sample mean is $\bar{x} = 98.5$F. She uses this information to construct a 95% confidence interval for the mean temperature.

Suppose she converts her data into Celsius (using the formula: Celsius $= 0.55 \times$Fahrenheit - 17.7). Which statement is true:

(A) The 95% confidence interval for mean temperature in Fahrenheit is $[98.5 \pm 1.96 \times 0.12]$, in Celsius the confidence interval is $[36.47 \pm 1.96 \times 0.066]$.

(B) The 95% confidence interval for mean temperature in Fahrenheit is $[98.5 \pm 1.96 \times 1.2]$, in Celsius the confidence interval is $[36.47 \pm 1.96 \times 17.04]$.

(C) The 95% confidence for mean temperature in Fahrenheit is $[98.5 \pm 1.96 \times 1.2]$. As we do not know the calf temperatures in Celcius we cannot construct the confidence interval in Celsius.

(D) The 95% confidence interval for mean temperature in Fahrenheit is $[98.5 \pm 1.96 \times 0.12]$, in Celsius the confidence interval is $[36.47 \pm 1.96 \times 17.04]$.

(E) None of the above.

(8) The objective of a survey is to find the mean number vehicles in a household. The survey samples a hundred household and evaluates the average (sample mean) of this sample. What can we say about their distribution?

(A) The distribution for the number of vehicles cannot be normally distributed, but the sample mean will be close to normal.

(B) The sample size is large, therefore both the distribution of the number vehicles and the sample mean will be close to normal.

(C) We cannot know the distribution of the number of vehicles or the distribution of the sample mean.

(D) The distribution of the number vehicles is normal.

(E) None of the above.

(9) Female adult heights are normally distributed with mean 65 inches and standard deviation 2 inches. Rebecca is in the top 5% for her height. What is Rebecca's height?

(A) Rebecca is 65.1 inches tall. (B) Rebecca is 64.9 inches tall.

(C) Rebecca is 66.9 inches tall. (D) Rebecca is 61.7 inches tall.

(E) Rebecca is 68.3 inches tall.

(10) The mean height of females is 65 inches with standard deviation is 2 inches. I take a sample of 5 women and evaluate their average height (call it $\bar{x}$). Which statement is true?

(A) The mean of $\bar{x}$ is unknown but the standard error of $\bar{x}$ is $2/5$.

(B) The mean of $\bar{x}$ is 65 inches and standard error of $\bar{x}$ is 2.

(C) The mean of $\bar{x}$ is 65 inches and standard error of $\bar{x}$ is $2/5$.

(D) The mean of $\bar{x}$ is 65 inches and standard error of $\bar{x}$ is $2/\sqrt{5}$.

(E) None of the above.

(11) The heights of adult males is known to be normally distributed with mean 73 inches and standard deviation 3 inches.

The heights of adult females is known to be normally distributed with mean 65 inches and standard deviation 2 inches.

Max is male and 70.5 inches tall. Jane, his sister, is 66.75 inches. Relative to their gender which sibling is taller?

(A) Max is taller than Jane.

(B) Max is in the 0.833 percentile while Jane is in the 0.875th percentile. Max is taller.

(C) We need to know the standard errors to calculate the z-transform, both standard errors are unknown.

(D) Max is in 20th percentile whereas Jane is in the 19th precentile. They are similar.

(E) Max is in the 20th percentile whereas Jane is in the 81st percentile. Mary is taller.

(12) We observe the following variables (I) = Bus number, (II) = number of M&Ms in a bag, (III) = average number of M&Ms in 1000 bags, (IV) = Tennis player rankings. Place variable with type.

(A) I=Numerical Discrete, II= Numerical Discrete, III=Numerical Discrete, IV=Categorical.

(B) I=Categorical, II= Numerical Discrete, III=Numerical Continuous (almost), IV=Numerical Discrete.

(C) I=Numerical Discrete (almost), II= Numerical Discrete, III=Numerical Discrete, IV=Numerical Discrete.

(D) I=Categorical, II= Numerical Continuous, III=Numerical Continuous, IV= Categorical.

(E) I=Categorical, II= Numerical Continuous, III=Numerical Continuous, IV=Numerical Discrete.

(13) The mean number of cans of coke a person drinks at a wedding is 4, with standard deviation is 2.

A catering company is organising a wedding with 100 guests. What is the chance that the hundred guests will consume between 360 to 460 cans of coke?

(A) About 16% (B) About 97.7% (C) About 58%

(D) About 95.4% (E) 100%.

(14) Hypokalemia is diagnosed when a patient's potassium level is below 3.5. Bob has hypokalemia since his mean level is 3.45. His potassium levels vary according to a normal distribution with mean 3.45 and standard deviation 0.4. Evaluate the chance of detection based on the average of four blood samples and suggest how to improve the detection rate.

(A) The chance of detection is about **40%**. To improve detection we need to increase the sample size.

(B) The chance of detection is about **25%**. To improve detection we need to increase the sample size.

(C) The chance of detection is about **75%**. To improve detection we need to decrease the sample size.

(D) Since 3.45 is less than 3.5 there is a **100%** chance his condition will be detected.

(E) The chance of detection is about **60%**. To improve detection we need to increase the sample size.

(15) Which is the most effective method (may be not realistic) for establishing a causality between lactose (dairy) and the incidence of fraternal twins (non-identical twins)?

(A) Compare the rates of fraternal twins amongst vegans (people who don't eat meat/dairy or eggs) and people who consume milk.

(B) Randomly place two groups of women (of child bearing age) into one of two groups (one group given dairy and the other not) and monitor over a few years.

(C) Compare rates of fraternal twins amongst women during the second world war (when dairy and meat was scarce) and now (when dairy is abundant).

(D) Ask a group of randomly selected women to place themselves into either the dairy and non-dairy groups and moniter them over a period of time.

(E) At a baby advise website ask for volunteers to give information about dairy consumption and whether or not they had fraternal twins.