Midterm 1 - STAT 301
Spring 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 60 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) Sidney observes the data

$$1, 1.5, 2, 3, 3.5, 3.8, 6, 6.5, 7, 8.$$

The sample mean and standard deviation of this sample is 4.23 and 2.47 respectively. What happens to the mean, standard deviation and quartiles of the sample if the last observation 8 is replaced with a much larger number?

(A) The mean and standard deviation stay about the same, but the quartiles and median change.

(B) Only the mean will change.

(C) The first quartile and median stay the same, but the third quartile, mean and standard deviation become larger.

(D) The quartiles and median stay the same, but the mean and most likely the standard deviation will become much larger.

(E) Unless we know the actual new number it is impossible to say what will happen to the values.
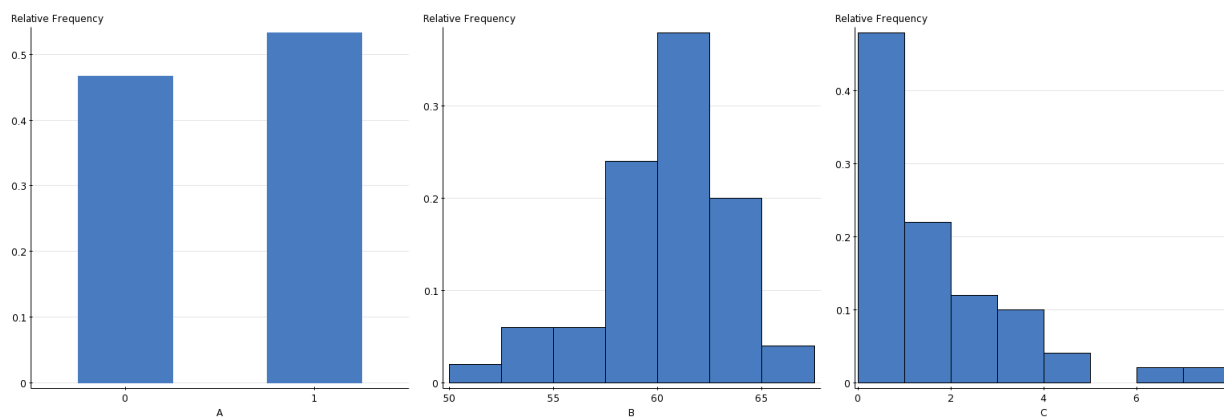
(2) Identify the histogram to the application.

Figure 1: The far left is (A), middle is (B) and right is (C)

(A) A = gender, B = number children in a household, C=height in inches.

(B) A = human heights in feet, B=number of children in a household and C = study times.

(C) A = number of children in a household, B=study times and C = human weights in inches.

(D) A = gender, B=human heights in inches and C = study times in hours.

(E) Could be any of the above.

(3) Baby Einstein is a company that produces videos which they claim makes children more intelligent. To prove their claim, they found some parents who owned Baby Enstein videos and parents who did not own the product. They compared the IQ of these children. They found that children who watched Baby Einstein videos tended to have slightly higher IQs then children who did not watch Baby Einstein. Which statement most accurately describes the conclusion of this study?

(A) This proves in a statistical sense that Baby Einstein's educational tools do improve the IQ of children.

(B) There may be lurking variables in the way that the experient has been designed. It is not possible to determine whether it was the videos that had an effect or whether parents who played such videos to their children also spent more time with thir children. No conclusions can be drawn.

(C) In order to check whether there is a real difference we need to test for statistical significance.

(D) Since the difference is so close it is not possible to determine whether there is a real difference or not.

(E) Baby Einstein should select a larger sample size using the formula margin of error $= 1.96 \times \sigma/\sqrt{n}$.

(4) In a survey, each student interviewed is asked the following four of questions (a) What newspaper do you read? (b) How tall are you? (c) How many siblings do you have? (d) What number bus do you take to school?

What type of variables are these?

(A) (a) = categorical variable, (b) = numerical, continuous variable (c) = numerical discrete variable, (d) = categorical variable.

(B) (a) = categorical variable, (b) = numerical, discrete variable (c) = numerical discrete variable, (d) = numerical discrete variable.

(C) (a) = numerical variable, (b) = numerical, discrete variable (c) = numerical discrete variable, (d) = numerical discrete variable.

(D) (a) = categorical variable, (b) = categorical, continuous variable (c) = categorical discrete variable, (d) = numerical discrete variable.

(E) These variables are very ambigious, we need a larger sample size.

(5) In this question we consider the distribution of heights of a certain population. The mean and standard deviation of the heights of this population is $\mu = 67$ inches and $\sigma = 4$ inches respectively. A random sample of 10 people are taken from this population and the sample mean is denoted as $\bar{x}$. The sample mean is transformed as follows $Y = \frac{\bar{x}-67}{4}$. What is the mean and standard deviation of this new variable.

(A) The mean is zero and the standard deviation is one.

(B) The mean is one and the standard deviation is $4/\sqrt{10}$.

(C) The mean is zero and the standard deviation is $1/\sqrt{10}$.

(D) The mean is 67 and the standard deviation is four.

(E) None of the above.

(6) In this question we compare the weights of Aksai Black Pied pigs and Angeln Saddleback pigs.

The distribution of weights of Aksai Black Pied pigs is approximately normal with mean 600 pounds and standard deviation 50 pounds.

The distribution of Angeln Saddleback Pigs is approximately normal with mean 750 pounds and standard deviation 40 pounds.

A farmer has a Aksai Black Pied pig called Ernie and a Angeln Saddleback Pigs called Bert.

Ernie the Aksai Black Pied weights 680 pounds, whereas **Bert the Angeln Saddleback** weights 700 pounds. Relative <u>to their breed</u>, which pigs weights the most.

(A) Ernie is in the 94.5% percentile, whereas Bert is in the 89.4% percentile. There is not much is dfference between their relative weights, but Ernie's relative weight is slightly higher than Bert's relative weight.

(B) Ernie weighs 680 pounds and Bert weighs 700 pounds so Bert is heavier.

(C) Ernie is in the 94.5% percentile, whereas Bert is in the 10.56% percentile. Thus Ernie would be considered a heavy Aksai Black Pied, but Bert would be considered a very light Angeln Saddleback.

(D) Ernie is in the 6.5% percentile, whereas Bert is in the 10.56% percentile. Thus Ernie relative weight is slightly less than Bert's relative weight.

(E) None of the above.

(7) HEB wants to see whether it's promotions have increased sales in it's products. It is known that the mean weekly sales of ketchup is 400 bottles, the standard deviation is 80. The number of products sold in a week is approximately normally distributed.

HEB introduce a promotion on the ketchup. They want to see whether this promotion has increased sales. During the promotion over a period of 10 weeks the mean weekly sales (sample mean evaluated over 10 weeks) is 420 bottles. Which statement is correct?

(A) Assuming normality, the 95% confidence interval for the mean number of sales after the promotion is $[263, 567]$. As this interval is so wide, we cannot say anything about the mean number of sales.

(B) Assuming normality, 95% of the weekly sales will lie in interval $[370, 470]$. As most of this interval is bigger than 400, it is clear that the promotion has increased sales.

(C) HEB needs to decrease the sample size, there is too much variability in the current sample.

(D) By increasing the sample size the population becomes more normally distributed which allows to identify the mean with a greater degree of precision.

(E) Assuming normality the 95% confidence interval for the mean number of sales after the promotion is $[370, 470]$. The mean number of bottles sold before the promotion (400) is inside this interval, it is not possible to say whether the promotion had an effect or not.

(8) Which statement or statements are correct.

(A) If the sample size is quite large, the distribution of the heights becomes more normal. This is called the central limit theorem.

(B) With 95% confidence, the sample mean lies in a 95% confidence interval.

(C) A 95% confidence interval interval is based on the principle that if it were possible to construct, 100 such confidence intervals, then about 95 of these would contain the true population mean.

(D) Two of the above is true.

(E) None of the above is true.

(9) Gestational diabetes in diagnosed in a patient if at one hour after consuming a sugary drink their blood sugar level is over 140. However, the blood sugar level of a patient is unknown, and a diagnosis is made on a sample or a sample mean being over 140. The blood level will vary sample to sample, but it generally believed the amount of glucose in a blood sample follows a normal distribution with standard deviation 10.

Kate's mean blood sugar level is 143 (thus, unknown to her doctor she actual has gestational diabetes), and each blood sample follows a normal distribution with mean 143 (the mean level) and standard deviation 10. What is the probability that her gestational diabetes will be not be diagnosed?

(A) The probability she will not be diagnosed based on just one sample is 61.8%, however if the average of 6 samples were taken the probability of not being diagnosed will increase to 76.9%.

(B) The probability she will not be diagnosed based on just one sample is 38.2%, however if the average of 6 samples were taken the probability of not being diagnosed will reduce to 23.12%.

(C) The probability she will not be diagnosed based on just one sample is 61.8%, however if the average of 6 samples were taken the probability of not being diagnosed will reduce to 23.12%.

(D) There is too much variability in the sample, but if we increase the sample size then the standard error of the distribution will decrease.

(E) None of the above.

(10) Let us return to the above question above on gestational diabetes.

Which statement or statements are true.

(A) The probability of a misdiagnoses decreases as the number of blood samples increases.

(B) The probability of a misdiagnoses decreases the further the mean glucose level of a patient is from 140.

(C) The probability of a misdiagnoses decreases the closer the mean glucose level is to 140.

(D) Two of the above.

(E) None of the above.

(11) A pharmaceutical company wants to test the effectiveness of one of the new drugs it has developed. They go to the local hospital to collect their data. Which sampling method will give the most reliable answer?

(A) Advertise for patient volunteers to take the drug, then find another group of patients that have not taken the drug. Monitor both groups over a period of time.

(B) Ask a doctor to decide which patients should be given the drug and which patients should be given the placebo and monitor them over a period of time.

(C) A person outside the hospital, who does not know any of the patients, randomly assigns some patients to the drug and other patients to the placebo. Both groups are monitored over a period of time.

(D) Two of the above is true.

(E) None of the above is true.

(12) The population of island A is 0.5 million and the population of island B is 5 million.

Demographers want to estimate the proportion of people over the age of 65 on both islands. They take a simple random sample of 1000 people from island A and simple random sample of 3000 people from island B. Which statement is correct.

(A) If you make the calculation you will find that 1000 is a 0.2% of population island A, whereas 3000 is only 0.06% of island B.

Because, the sample taken from island A is a larger proportion of the population than the sample taken from island A, this suggests that the sample from island is a lot more accurate than the sample from island B.

(B) The sample taken from island B is three times larger than the sample from island A, thus the standard error of the estimator from island B is likely to be smaller than the standard error from island A (assuming the standard deviations in both populations are the same).

This means that the estimator from island B is likely to be a more reliable estimator of the proportion of elderly people than the estimator from island A.

(C) Since both sample are relatively small compared to the size of the population, neither of the estimators will be reliable.

(D) Two of the above is true.

(E) Neither of the above is true.

(13) A mini van which carries 8 people can carry the maximum load of 1500 pounds. It is known that the weight of humans is roughly normal distributed with mean 170 pounds and standard deviation 25 pounds. What is the chance the mini van, when carrying 8 people, will exceed the load maximum?

(A) Assuming normality of the average weight of 8 people, the probability of exceeding 1500 pounds is approximately 24.1%.

(B) Assuming normality of the average weight of 8 people, the probability of exceeding 1500 pounds is so small that it can't be found on the z-tables.

(C) Assuming normality of the average weight of 8 people, the probability of exceeding 1500 pounds is approximately 2.3%.

(D) Since the distribution of the sample mean is unknown we need to use the t-distribution with 7 degrees of freedom, therefore the probability is approximately 5%.

(E) None of the above.

(14) I take a simple random sample, calculate the sample mean and construct the 95% confidence interval for the mean $[10, 14]$. Which statement is correct.

(A) I need the sample mean and standard deviation in order to determine the sample size.

(B) In order that the margin of error is 1, I need to half my sample size.

(C) In order that the margin of error is 1, I need to double my sample size.

(D) In order that the margin of error is 1, I need to quadruple my sample size.

(E) None of the above.

(15) Below there are two plots. One plot is the histogram of a population and the other is the histogram of the sample mean (of size four) taken from that population. You are asked to match the plot with the correct explanation.
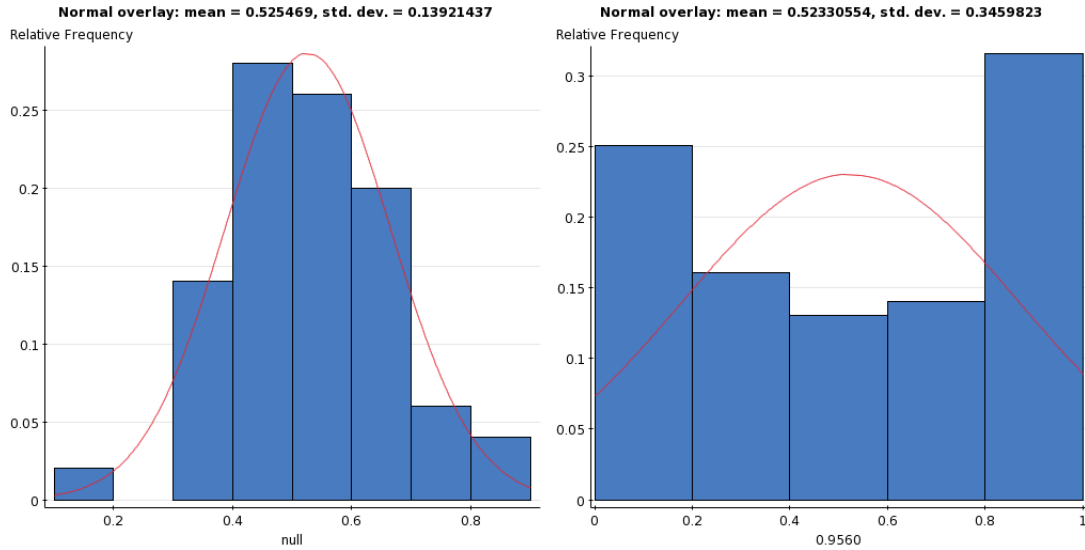

Figure 2: The far plot is (A), the right plot is (B)

(A) We need to use that that the standard error increases with sample size to answer this question, however the standard errors are unknown.

(B) B corresponds to the histogram of the population and A corresponds to the histogram of sample mean, where the samples are taken from the distribution in B. We observe that as the sample size increases that standard deviation decreases and the histogram 'looks' more normal.

(C) A corresponds to the histogram of the population and B is the histogram of sample mean, where the samples are taken from the distribution in A. Observe that that standard deviation has increased as we increase the sample size. This is because the sample mean is more variable than the original population. Of course, in reality we only one sample mean and we have to draw inference based on that. These plots help us to do this.

(D) It is not clear which histogram corresponds to what, this is because there aren't any proper labels on the axis.

(E) The density of a distribution is where the area below the curve gives the chance of an event. The normal distribution is a density and not a histogram, therefore it is not possible to tell whether A or B corresponds to the distribution of the population or sample mean.