Midterm 1 - STAT 301 Section 501/502 1
Spring 2012

Name:

UIN:

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 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.

4. You have 50 minutes to work on this exam. There are 15 multiple choice questions. If you cannot do one question move on to the next.

5. You may use a calculator but not a phone during 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.
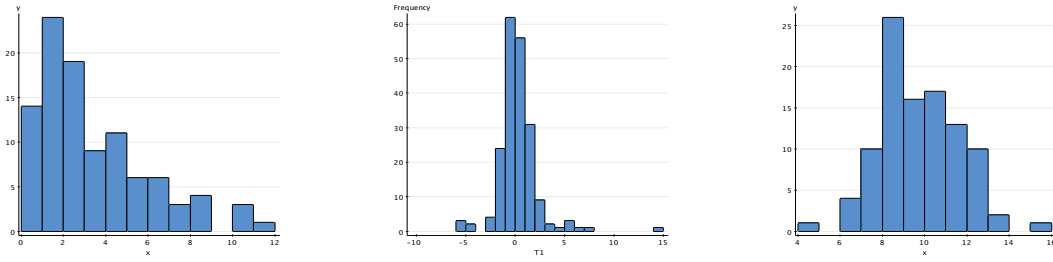
8. Good Luck!!!

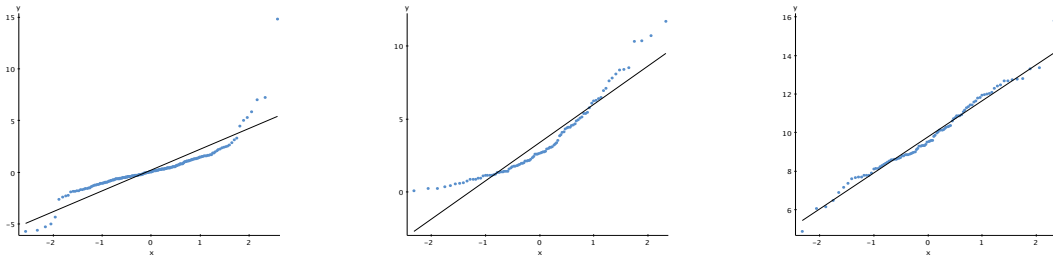Figure 1: First plot I, second II third is III

Figure 2: First plot (a), second (b) third is (c)

(1) Match the histograms (where means of the plots are 3.4, 0.2 and 10 respectively and the standard deviations are 2.6, 2 aand 2 respectively) with the following QQplots

   (A) (Ib), (IIa) and (IIIc)

   (B (Ic), (IIb) and (IIIa)

   (C) (Ia), (IIc) and (IIIb)

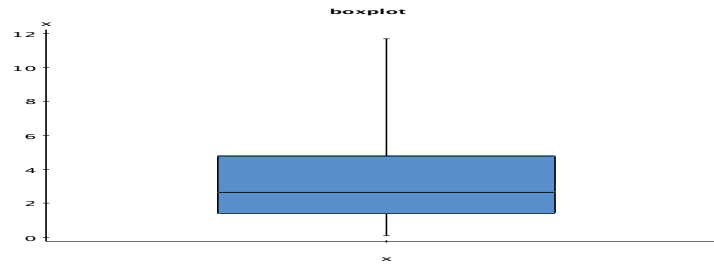   (D) (Ia), (IIc) and (IIIc)

   (E) (Ic), (IIa) and (IIIb)

Figure 3:

(2) Which of the following statements best explains the boxplot in figure 3.

    (A) The distribution is definitely skewed to the left.

    (B) The distribution is definitely skewed to the right.

    (C) It is likely that the distribution is symmetric.

    (D) It suggests that the distribution is skewed to the left.

    (E) It suggests that the distribution is skewed to the right

(3) A small shop pays 3 shop assistants $15K$ dollars, one junior manager $30K$ and the senior manager $60K$.

Which statement is correct?

    (A) The median is 15K, the first quartile is 15K and the third quartile is 60K.

    (B) The mean is 27K, the median is 15K and the standard deviation is 19.6.

    (C) The mean is 27K, the median is 15K, but it is not possible to calculate the standard deviation.

    (D) The mean is 15K, the median is 27K and the standard deviation is 19.6.

    (E) The mean is 15K, the median is 27K and the variance is 19.6.

(4) An online company that sells animal feed to customers over the whole of the US, is trying to determine the effectiveness of the feed that it sells farmers. It does this by analysing questionnaires that are sent to back them. There are a few ways it can get these opinions. Offer a 5% rebate to all customers who return the questionnaire and then use the questionnaires that are completed and returned.

   (A) If the feed is effective, then using this method we may see the effect.

   (B) This method may give a biased result, we will probably not see an effect.

   (C) This method may give a biased result, though it is not clear which way.

(5) A farmer is comparing the weights of two horses each from two breeds. Both breeds follow approximately a normal distribution, where Breed 1 has mean 1000 and standard deviation 100 and Breed 2 has mean 1050 and standard deviation 120. Horse Noddy from breed 1 has weight 970 pounds and horse Dobby from breed 2 has weight 1014 pounds. Taking into account that the horses are from different breeds, which statement is correct.

   (A) Taking into account breed the relative weight of Nobby is more than Doddy.

   (B) Taking into account breed the relative weight of Dobby is more than Noddy.

   (C) The relative weights of Dobby and Noddy are about the same.

   (D) None of the above.

Q(6-9) It is important to monitor a foal in the first few days after birth. In particular it is important that the mare gives him essential colostrum in the first few days (in order that she protects him with antibodies). In order to do this, the foal must feed frequently from his/her mother (on average a newborn foal feeds 4 times in an hour). However, if the foal feeds too frequently, then this suggest that the mother is not providing enough colostrum and the foal may need veterinary assistance.

The distribution of feed times (time between the feeds) of newborn horses approximately follow a normal distribution, but for each breed of horses the distributions vary slightly:

Breed 1: $N(15, 3)$ (mean 15 minutes, standard deviation 3)

Breed 2: $N(15, 2)$ (mean 15 minutes, standard deviation 2)

Breed 3: $N(16, 2.5)$ (mean 16 minutes, standard deviation 2.5)

(6) Vets give a more rigourous examination to newborn foals whose feeding times fall below 13 minutes. This means

   (A) On average, foals from Breed 2 will be examined the most often.

   (B) On average, foals from Breed 3 will be examined the most often.

   (C) On average more foals from Breed 2 will be examined than Breed 3. But overall foals from Breed 1 will be examined most often.

   (D) On average the proportion of foals from all breeds will be the same.

   (E) On average, foals from Breed 1 will be examined more often than foals from Breed 2. But overall foals from Breed 3 will be examined most often.

(7) Using a blanket threshold for all breeds is not very effective in identifying the foals at risk. Instead, it was decided that a 10% percentile should be used. The foals falling in the 10% percentile for their corresponding breeds should be used as the cut off point for further intervention. Using this information what should the intervention times for each of the breeds be?

   (A) Breed 1: $15 - 0.54 \times 3$, Breed 2: $15 - 0.54 \times 2$, Breed 3: $16 - 0.54 \times 2.5$.

   (B) Breed 1: $15 - 1.28 \times 3$, Breed 2: $15 - 1.28 \times 2$, Breed 3: $16 - 1.28 \times 2.5$.

   (C) Breed 1: $15 + 0.54 \times 3$, Breed 2: $15 + 0.54 \times 2$, Breed 3: $16 + .54 \times 2.5$.

   (D) Breed 1: $15 + 1.28 \times 3$, Breed 2: $15 + 1.28 \times 2$, Breed 3: $16 + 1.28 \times 2.5$.

   (E) Actually, it is okay to use a blanket threshold as long as it is $15 - 1.28 \times 2$

(8) Basing a referral (to a vet) on just one measurement turns out to be costly. Therefore it was decided to take the average time over 4 feeds. The foals whose average was falling in the 10% percentile for their corresponding breeds should be used as the cut off point for further intervention. What is the 10% percentile for <u>Breed 1</u> in this case?

(A) $15 - 1.28 \times 3$.

(B) $15 - 0.54 \times \frac{3}{2}$.

(C) $15 - 1.28 \times \frac{3}{2}$.

(D) $15 - 1.28 \times \frac{3}{4}$.

(E) $15 - 0.54 \times 3$.

(9) What is the chance that the time between feeds of a foal from <u>Breed 2</u> should be greater than 15.5 minutes (if you like algebra you can think of this as P(Breed 2 feed time is > 15.5))?

(A) $1 - P(Z < (15.5 - 15)/2) = 0.40$

(B) $1 - P(Z < (15 - 15.5)/2) = 0.56$

(C) $1 - P(Z < (15 - 15.5)) = 0.31$

(D) $P(Z < (15.5 - 15)) = 0.69$

(E) None of the above.

Q(10-13) The time between feeds for a newborn lamb is also important (for a similar reason to those described in the foal question described above). However, the distribution of the times between feeds tends to be right skewed with mean 5 hours and standard deviation 1.5.

(10) Where will the mean and medians for the distribution of feed times for lambs be located?

(A) The mean and median are close.

(B) The mean will be 1.5 times the median.

(C) The mean will be to the left of the median.

(D) It is unclear whether the mean will be on the left or right of the median.

(E) The mean will be to the right of the median.

Q(11-13) In order to determine whether the lamb should be referred for veterinary assistance the first 25 times between feeds is noted and their average is taken.

(11) What is the standard deviation of the average?

   (A) 1.5/25
   (B) 1.5/5
   (C) 5/5
   (D) $1.5/\sqrt{5}$
   (E) 1.5

(12) If the average time between feeds is less than the 10th percentile of the distribution of the average, then the lamb is refered to a vet. Which statement is correct?

   (A) Lambs with an average time between feeds less than 0.10×5 hours will be referred to a vet.

   (B) Lambs with an average time between feeds less than 5 hours will be referred to a vet.

   (C) None of these lambs will be referred to a vet.

   (D) The distribution is unknown and without any additional information we cannot calculate the percentile.

   (E) The sample size is probably large enough for us to use the normal approximation. By using this approximation the 10th percentile is approximately 4.6.

(13) A lamb's average feed time is 4.3 hours (average over 25 feeds). Based on your answer above, should she see a vet?

   (A) By using the z-transform we see that $\frac{4.3-5}{0.3} = -2.33$ is very close to the 10 percentile, and it is not clear whether we should refer the lamb.

   (B) The percentile this lamb corresponds to is very small, so we should refer the lamb to the vet.

   (C) We cannot say anything without additional information.

   (D) 4.3 is not a very far below the mean, so the lamb probably does not need to see the vet.

   (E) Since $\frac{4.3-5}{1.5} = -.467$ is greater than -1.28, the lamb does not need to see the vet.

(14) When sampling from a highly skewed parent (population) distribution, while looking at samples of sizes 20 and 40, which of the following is TRUE about the sampling distribution of the sample mean?

    (A) The sampling distribution will look closer to normal for size 40 than for size 20.

    (B) The mean of the sampling mean will be closer to the parent distribution mean for size 40 than for size 20.

    (C) The standard deviation of the sample mean will larger for size 40 than for size 20.

    (D) All of the above are true.

    (E) Only two of the above are true.

(15) Which of the following statements is FALSE?

    (A) In a right-skewed distribution, the right tail is longer than the left tail.

    (B) For a unimodal symmetric distribution without outliers, mean and median are about the same.

    (C) If all the values in a data-set are 1, then the standard deviation is 0.

    (D) In a left-skewed distribution, the mean is larger than the median.

    (E) If all points are on a straight line in a normal quantile plot, we could say that the data is normally distributed.