STAT 340

Mid-term Exam 1 Review

Spring 2024

EXAM INSTRUCTIONS

(1) The exam will be on Wednesday, Feb. 21, 11:30am-12:20pm in class.

(2) The exam will cover chapters 1, 2, 3, 4, and 6 (up to slide 25).

(3) You need to bring a calculator for the exam. You are not allowed to use your cellphone's calculator.

(4) You may have one 8 ½" by 11" (front and back) sheet of paper with formulas, definitions, or whatever you think it is important.

(5) You must show work for possible partial credit.

(6) All work on this exam must be completely on your own. Cheating will be penalized with a score of zero.

1. Classify the following measurements observed in the study as nominal, ordinal, continuous or discrete.

   a. Risk of experiencing complications in pregnancy (low, medium, high)

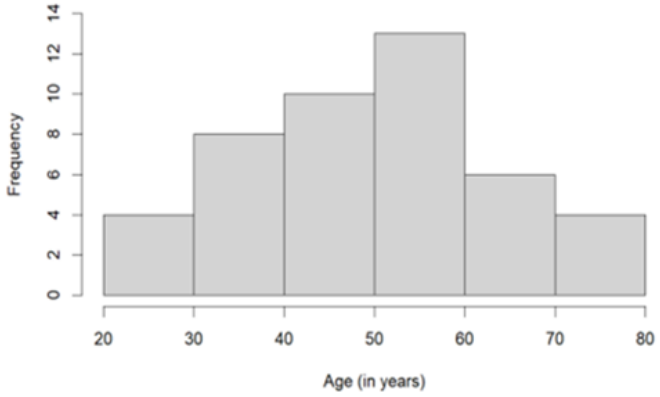
   b. Temperature (in °F)

   c. Blood type (A, B, AB, O)

   d. Number of doctor visits in a month (0, 1, 2, …)

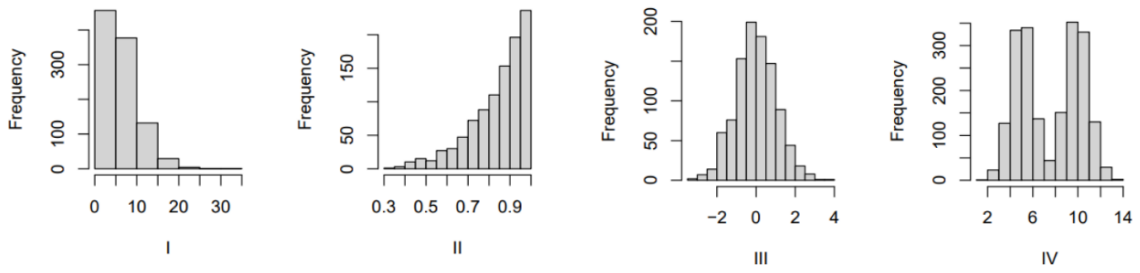
   e. Blood sodium levels (in milliequivalents per liter mEq/L)

2. Here is a histogram of the age of 45 people participated in an online survey, grouped into bins of length 10. The median age must be



   a. between 30 to 40
   b. between 40 to 50
   c. between 50 to 60
   d. between 60 to 70

3. For each histogram, classify it as symmetric, negatively skewed, or positively skewed.



4. Here are the lengths of the 20 Torrey pine needles arranged in increasing order:

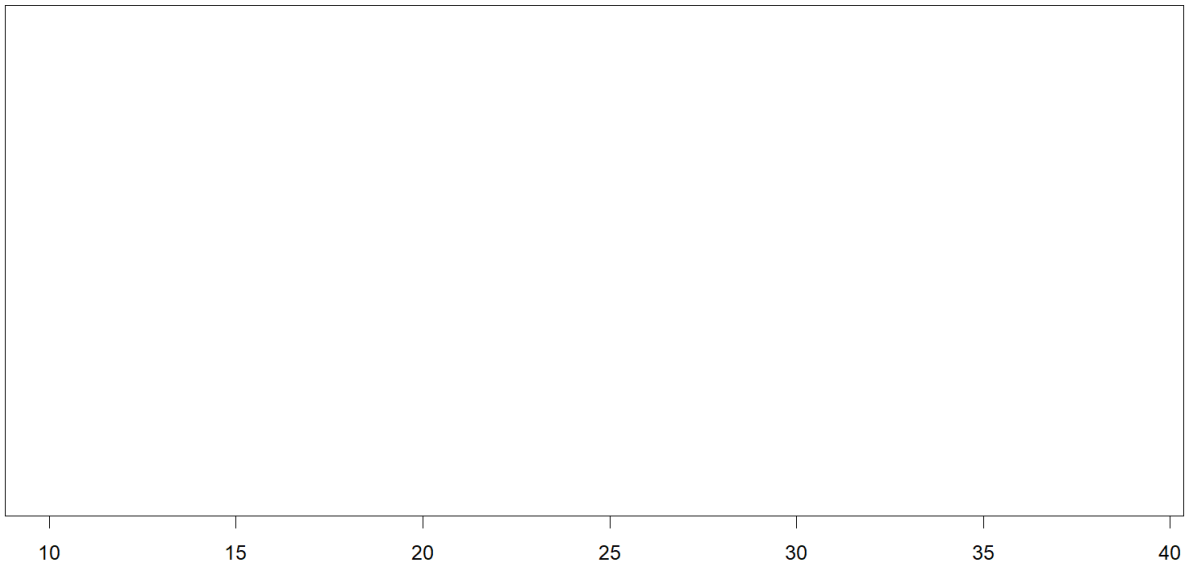
12.6	21.2	21.6	21.7	23.1	23.7	24.2	24.2	25.5	26.6
26.8	28.9	29.0	29.7	29.7	30.2	32.5	33.7	33.7	39.2

a. What is the median for this dataset?

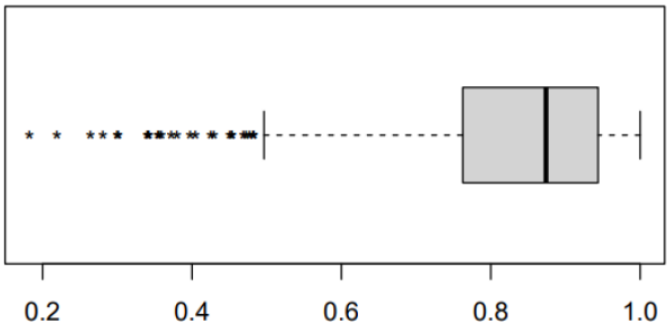
b. Find the first quartile, third quartile and the interquartile range.

c. Does the 1.5 × IQR rule identify any suspected outliers?

d. Draw a horizontal box plot for this data set. You can use the following scale.



5. Based on the box plot below, which of the following do you expect to be true?



   a. Mean > Median
   b. Mean = Median
   c. Mean < Median
   d. Can't know without calculating mean and median

6. The following dataset is a sample containing $n = 7$ observations:

   2, 1, 3, 4, 3, 5, 3

   a. Calculate the sample mean and the sample standard deviation for this dataset.

   b. Provide an interpretation for the value of standard deviation that you obtained in (a).

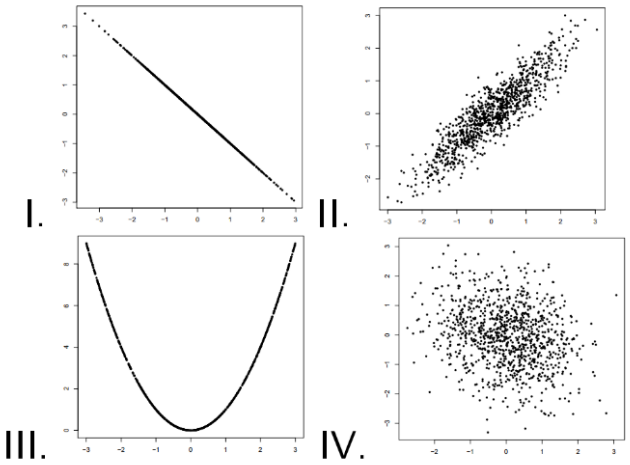
7. What are all the values that a correlation r can possibly take?

   a. $r \geq 0$
   b. $0 \leq r \leq 1$
   c. $-1 \leq r \leq 1$

8. If the points on a scatterplot lie very close to a straight line. The correlation between x and y is close to

   a. -1
   b. 1
   c. either -1 or 1, depending on the direction.

9. For each scatter plot, match the value of correlation coefficient.



   a. $r = -0.2$
   b. $r = -1$
   c. $r = 0$
   d. $r = 0.9$

I	
II	
III	
IV	

10. The toto toucan possesses the largest beak relative to body size of all birds. This exaggerated feature has received various interpretations, such as being a refined adaptation for feeding. However, the large surface area may also be an important mechanism for radiating heat (and hence cooling the bird) as outdoor temperature increases.

    To investigate the relationship between outdoor temperature (x) (in ℃) and percentage heat loss from beak (y), we compute the following quantities.

    - $\bar{x}$ = mean of the values of x = 22.5
    - $\bar{y}$ = mean of the values of y = 47.375
    - s_x = standard deviation of the values of x = 4.761
    - s_y = standard deviation of the values of y = 10.751
    - r = correlation between x and y = 0.914

    a. What is the equation of the least-squares regression line for predicting beak heat loss, as a percentage of total body heat loss from all sources, from temperature? (Give your answers in three decimal places.)

    b. Explain in specific language what the slope and intercept of this line say about the relationship between beak heat loss and temperature.

    c. Use the equation that you found in (a) to predict beak heat loss, as a percentage of total heat loss from all sources, at a temperature of 25℃.

    d. What percentage of the variation in beak heat loss is explained by the straight-line relationship with temperature?

11. List two types of good and two types of bad sampling designs.

12. A researcher is interested in understanding Kansas State University undergraduate student's perceptions of the nutritional quality of campus food in an academic year. It isn't practical to contact all students. He wants to collect data from a sample of students. Classify the following sampling approaches as convenience, voluntary, SRS, stratified, or multistage sampling.

    a. He will prepare a small questionnaire and send it to all the undergraduate KSU students via email and collect data from the responses.

    b. The researcher has prepared a list with all the undergraduate students in KSU and will randomly select 1000 of them to collect data.

    c. He will collect data from undergraduate students who participated in a similar study in the previous academic year.

    d. He will group the students into academic year (freshman, sophomore, junior, senior). Then he will select students from each group with the proportion of overall students selected from a given academic year being equal to the proportion of students in that academic year at the university. The total number of selected students should be 1000.

    e. He will group the undergraduate students into college (10 - agriculture, engineering, arts and science etc.) in the first stage and academic year (4) in the second stage. He will select 20 students from each college-academic year combination.

13. A survey is carried out at Kansas State University to estimate the proportion of all undergraduate students living at home during the current term. Of all the undergraduate students enrolled at the campus, a random sample of 100 was surveyed.

    a. What is the population of interest?

    b. What is the sample?

14. Educational policy researchers randomly selected 400 teachers at random from the National Science Teachers Association database of members and asked them whether or not they believed that evolution should be taught in public schools. They received responses from 252 teachers.

    a. What is the population of interest?

    b. What is the sample?

15. Toxoplasmosis is an infection with a parasite called Toxoplasma gondii. In the USA, roughly 9% of all people who are 12-49 years old have antibodies, suggesting infection. A sample of 100 people were selected for a study to test the effectiveness of a new drug for treating Toxoplasmosis.

    a. What is the population of interest?

    b. What is the sample?