Midterm Exam for MA5701 – Statistical Methods

Problem 1 (12 points)

Multiple Choice Questions. For each question, only one of the statements is correct. Simply Circle or state the correct statement. You do not need to explain why a statement is correct or incorrect.

(1) (**3 points**) The heights of eight people (in centimeters) are: 160, 170, 165, 180, 181, 174, 158, 171. Which of the following statements is correct:

a. The sample size is 8, and the height is a nominal variable.
b. The sample median is 170 and the sample mean is 169.875.
c. The range (the difference between the largest value and the smallest value) is 23.
d. The height is a continuous variable, and the sample median is 171.
e. None of the above is correct.

(2) (**2 points**) If the interquartile range is zero, you can conclude that:

a. The range must also be zero.
b. The sample mean is also zero.
c. All of the observations have the same value.
d. At least 50% of the observations have the same value.
e. None of the above is correct.

(3) (**2 points**) A sample of 200 test scores for MA5701 produced the following statistics: mean = 85; lower quartile = 80; median = 89; upper quartile = 93. Which of the following statements is correct:

a. Half of the scores are less than 88.
b. The middle 50% of scores are between 85 and 93.
-2 c. One-quarter of the scores are greater than 80.
d. The most common score is 89.
e. None of the above is correct.

(4) (**2 points**) The median is a better measure of central tendency than the mean if:

a. The variable is discrete.
b. The distribution is highly skewed.
c. The variable is continuous.
d. The distribution is not symmetric.

e. None of the above is correct.

(5) **(3 points)** Which one of the statements is correct:

a. If two events A and B are independent, then $Pr(A \text{ or } B) = Pr(A) + Pr(B)$.

b. The probability distribution function of a continuous random variable can have a value greater than 1.

c. For two events A and B, we always have $Pr(A \text{ and } B) = Pr(A) * Pr(B)$.

d. Both (b) and (c) are correct.

e. None of the above is correct.

Problem 2 (14 points)

The data used in this problem is Florida lake data. This data was obtained from the Web site of Florida Lakewatch (http://lakewatch.ifas.ufl.edu/), a volunteer organization coordinated through the University of Florida's Institute of Food and Agricultural Sciences, Fisheries and Aquatic Sciences. Part of the data are shown in the following table.

Lake	County	soil	Winter Month	Winter TP	Winter TN	Winter TC	Winter Secc
ALICE	ALACHUA	CS	JAN	609	810	10	5.5
ALTO	ALACHUA	CS	JAN	18	763	9	2.5
BIVANS	ALACHUA	CS	JAN	290	3753	22.7	0.8

(1) **(4 points)** Determine the types of following variables. Check all that apply.

Variable Name	Notes	Type of Variable
Lake	Name of lake	Qualitative, Nominal, Ordinal Quantitative, Continuous, Discrete
Soil	Name of the dominant soil type	Qualitative, Nominal, Ordinal Quantitative, Continuous, Discrete
Winter TN	Winter Value for total nitrogen, in ug/L	Qualitative, Nominal, Ordinal Quantitative, Continuous, Discrete
Winter Secc	Winter value for Secchi depth, in feet	Qualitative, Nominal, Ordinal Quantitative, Continuous, Discrete

-0.5

(2) **(6 points)** The following is the box plot of Winter TN - Winter Value for total nitrogen, in ug/L. Specify the following part used in the box plot.

a): outlier

b): Maximum observation below upper fence

-0.5 upper inner fence

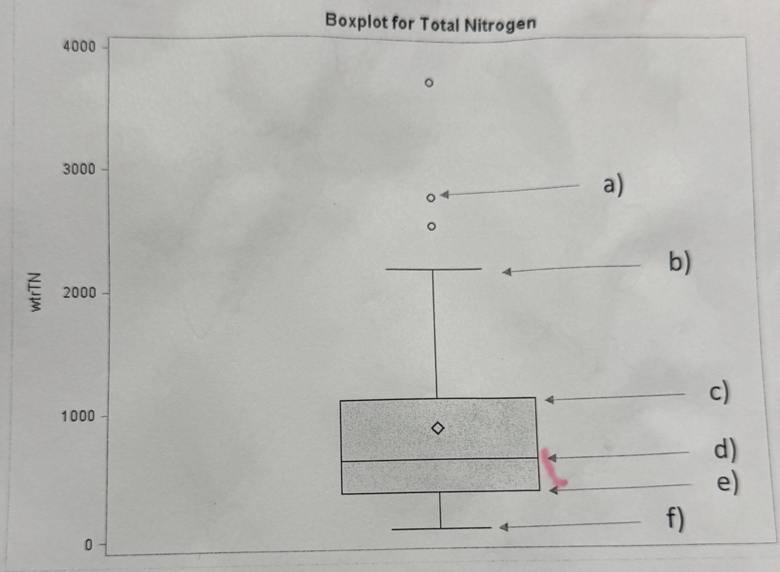
c): 75th percentile (upper quartile)

d): Median

e): 25th percentile (lower quartile)

f): Minimum observation & above ~~the~~ lower ~~inner~~ fence

(3) (**4 points**) Comment on interesting features of the boxplot from Part (2).



Problem 3 (24 points)

The following is a probability distribution of the number of defects on a given contact lens produced in one shift on a production line:

Number of Defects	0	1	2	3	4
Probability	0.45	0.20	0.20	0.10	0.05

(1) **(8 points)** Let A be the event that at most two defects occurred, and B be the event that 2, 3, or 4 defects occurred. Find $\Pr(A)$, $\Pr(B)$, $\Pr(A \text{ and } B)$, and $\Pr(A \text{ or } B)$. Please provide sufficient details about your calculation.

(2) **(6 points)** Let the random variable X be the number of defects on a contact lens randomly selected from lenses produced during the shift, find the mean and variance of X for the shift. Please provide sufficient details about your calculation.

(3) **(6 points)** Assume that the lenses are produced independently. Let Y be the number of defect-free lenses among 10 lenses drawn randomly from the production line during the shift. Write out the probability mass function, the expected value, and the variance of Y.

(4) **(4 points)** Assume that the lenses are produced independently. What is the probability that all five lenses drawn randomly from the production line during the shift are defect-free? Please provide sufficient details about your calculation.

K. Tapond Rao

2

Intersting features of the boxplot from part (2)

There appear to be 3 outliers. The distribution is skewed. The interquartile range lies between above 0 and close to 2000 High variance in total wingers.

problem 3

$Pr(A)$ = at most two defects occured.

$Pr(B)$ = 2, 3 or 4 defects occured.

$$Pr(A) = Pr(A=0) + Pr(A=1) + Pr(A=2)$$

$$= 0.45 + 0.20 + 0.20$$

$$= 0.85$$

$$Pr(B) = Pr(A=2) + Pr(A=3) + Pr(A=4)$$

$$= 0.20 + 0.10 + 0.05$$

$$= 0.35$$

$$Pr(A \cap B) = Pr(A=2) = 0.20$$

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

$$= 0.85 + 0.35 - 0.20$$

$$= 1.00$$

(2)

X - no. of defects → mean = Expected val... $= E(X)$

$E[X] = \sum_{i=0}^{4} p_i x_i$

$= (0)(0.45) + (1)(0.20) + (2)(0.20) + (3)(0.10) + (4)(0.0...)$

$E(X) = 1.1$

$Var(X) = \sum_{i=0}^{4} p_i (x_i - \mu)^2$

$= (0-1.1)^2(0.45) + (1-1.1)^2(0.20) + (2-1.1)^2(0.20) + (3-1.1)^2(0.10) + (4-1.1)^2(0.05)$

$Var(X) =$ 1.49

(3)

Y - number of defect-free level among 10 lens drawn randomly from the production line

$Pr(Y=0)$ = no defect $= 0.45 = p$

$Pr(Y>0) = 1 - Pr(Y=0)$

$= 1 - 0.45$

$= 0.55 = (1-p)$

$Y \sim$ Binomial $(10, 0.45)$

$n = 10$ (number of time drawn)

...lity mass function

$(Y=0) = 0.45 = p$

$pr(Y>0) = 0.55 = (1-p)$

Expected value $E[Y] = np$

$= 10 \times 0.45$

$= 4.5$

Variance $= np(1-p)$

$= 10 \times 0.45(1-0.45)$

$= 4.5(0.55)$

$= 1.36125$

(4) probability that all five lenses drawn randomly are defect-free

$p(Y=0) = 0.45$, $p(Y>0) = 0.55$

zero defe   atleast one defect

—4

all 5 defect free $p(Y=0) \times p(Y=0) \times p(Y=0) \times p(Y=0) \times p(Y=0)$

independent events. $= 0.45 \times \cancel{0.45} \times 0.45 \times 0.45 \times 0.45$

$= 0.\cancel{0.0145}$