Test-1

level of measurement

1. **(15pts)** Classify each variable as qualitative or quantitative.

ratio ← a. Number of bicycles sold in 1 year by a large sporting goods store. Qn (Discrete)

nominal ← b. Colors of students 'eyes. Ql

ratio ← c. Times it takes to go to school. Qn (Continuous)

ratio ← d. Weights of fish caught in Lake George. Qn (Continuous)

nominal ← e. Marital status of faculty members in a large university. Ql

2. **(15pts)** Classify each variable as discrete or continuous.

ratio ← a. Number of students in the statistics class. Discrete

interval ← b. The temperature in the statistics class. Continuous

ratio ← c. Weights of fish caught in Lake George Lifetime (in hours) of 12 flashlight batteries. Continuous

ratio ← d. Number of books sold each day by a book shop in Phnom Penh. Discrete

ratio ← e. Number of DVDs rented each day by a video store. Discrete

3. **(50pts)** Given the following data (Sample data)

75	61	41	90	91	40	39	69	45	47
32	12	70	38	97	34	94	77	88	91
89	90	43	40	89	85	71	30	25	21

a. Find the mean, variance, standard deviation, mode, median, Q_1, and Q_3 of the data.

b. Construct a frequency table with five classes.

c. Using the grouped data formula, find the mean, variance, standard deviation, mode, median, Q_1, and Q_3 for the table in part (b) and compare it to the results in part (a).

d. Construct a histogram, frequency polygon and cumulative frequency polygon.

4. **(20pts)** If 3 books are picked at random from a shelf containing 5 novels, 3 books of poems, and a dictionary, what is the probability that

(a) the dictionary is selected?

(b) 2 novels and 1 book of poems are selected?

2. (a). The mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 60.466$

. The median $MD = 65$

. The mode $Mode = 40, 89, 90, 91$

} Measures of central tendency (or location)

. The variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 700.9441$

. The standard deviation $s = \sqrt{s^2} = 26.47541$

. Range $R = \max(x_i) - \min(x_i) = 97 - 12 = 85$

} Measures of variation

. Q_1 = value of $\frac{n+1}{4}^{th}$ term $= \frac{x_7 + x_8}{2} = \frac{38+39}{2} = 38.5$

. $Q_2 = MD$ = value of $\frac{n+1}{2}^{th}$ term $= \frac{x_{15} + x_{16}}{2} = \frac{61+69}{2} = 65$

. Q_3 = value of $\frac{n+1}{4} \times 3^{th}$ term $= \frac{x_{23} + x_{24}}{2} = \frac{89+89}{2} = 89$

Q_1 Q_2

12 21 25 30 32 34 38 39 40 40 41 43 45 47 61 69 70 71 75 77 85
88 89 89 90 90 91 91 94 97

Q_3 L_{60}

. Find the 60^{th} percentile

$c = \frac{n \cdot p}{100} = \frac{30 \times 60}{100} = 18$ is a whole number

$\Rightarrow$ the 60^{th} percentile is $x = \frac{x_{18} + x_{19}}{2} = \frac{71+75}{2} = 73$

. Find the percentile corresponding to $x = 45$

$p = \frac{(\text{number of value} < 45) + 0.5}{n} \times 100 = \frac{12 + 0.5}{30} \times 100 = 41.66 \approx 42$

Thus $x = 45$ corresponds to 42^{th} percentile.

. Check for outliers

$IQR = Q_3 - Q_1 = 89 - 38.5 = 50.5$

So we have the interval $[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$

$= [38.5 - (1.5)(50.5), 89 + (1.5)(50.5)] = [-37.25, 164.75]$

Any value that is out of this interval is an outlier.

We see that no value is out of this interval.

Thus the given data does not have any outlier.

(b) Construct a frequency table

We have $k = 5$

- The width of the class interval is $i \geq \frac{\max(x_i) - \min(x_i)}{k}$

$$i \geq \frac{97-12}{5} = 17$$

$$\Rightarrow i = 18$$

- We choose the lower limit for the first class is 12. Then we have frequency distribution as follow:

class	Midpoint	f	cf	cf%
12 up to 30	21	3	3	10%
30 up to 48	39	11	14	46.7%
48 up to 66	57	1	15	50%
66 up to 84	75	5	20	66.7%
84 up to 102	93	10	30	100%
		30		

(c) Find mean, mode, median, variance, sd, Q_1, Q_3:

- Mean $\bar{x} = \frac{\sum x_m \cdot f}{\sum f} = 61.8$

• Modal class is $[30, 48)$

$$\text{Mode} = L + \frac{w(f_m - f_b)}{(f_m - f_b) + (f_m - f_a)}$$

where L = lower limit of modal class

w = width of the class

f_m = frequency of the modal class

f_b = frequency of the class before the modal class

f_a = frequency of the class after the modal class

Here $L = 30, w = 18, f_m = 11, f_b = 3, f_a = 1$

So $\text{Mode} = 30 + \frac{18(11-3)}{(11-3)+(11-1)} = 38$

• Median $MD = L + \frac{w}{f_m}\left[\frac{n}{2} - cfb\right]$

where L = lower limit of the median class

w = class interval

$n = \Sigma f$ total frequency

f_m = frequency for the median class

cfb = cumulative frequency before the median class

From cf, we have the median class is $[48, 66)$, $L = 48$, $w = 18$, $f_m = 1$, $n = 30$, $cfb = 14$

$$\Rightarrow MD = 48 + \frac{18}{1}\left[\frac{30}{2} - 14\right] = 66$$

- Variance $s^2 = \frac{n(\Sigma f \cdot X_m^2) - (\Sigma f \cdot X_m)^2}{n(n-1)}$ = 735.8897

- $S = \sqrt{s^2}$ = 27.12729

- $Q_1 = L + \frac{w}{f_m}\left[\frac{n}{4} - cfb\right]$

From cf%, Q_1 is in the class $[30, 48)$
($\leq 25\%$)

$L = 30$, $w = 18$, $f_m = 11$, $n = 30$, $cfb = 3$

$$\Rightarrow Q_1 = 30 + \frac{18}{11}\left[\frac{30}{4} - 3\right] = 37.36$$

- $Q_2 = MD = 66$

- $Q_3 = L + \frac{w}{f_m}\left[\frac{3n}{4} - cfb\right]$

From cf% ($\leq 75\%$), Q_3 is in the class $[84, 103)$

$L = 84$, $w = 18$, $f_m = 10$, $n = 30$, $cfb = 20$

So $Q_3 = 84 + \frac{18}{10}\left[\frac{3\times 30}{4} - 20\right] = 88.5$

(d) Construct a histogram, frequency polygon and cumulative frequency polygon.