

Part I. (10 points)

1. (Lind et. al.) A student takes a survey of heights of 450 college women and divides them into 10 classes. Her first two class midpoints are 62.5" and 65.5". There are 91 women in the first class and 110 women in the second class. (5)

- What is the (width of the) class interval (w) ?
- What is the lower limit of the third class?
- What is the relative frequency of the second class?
- What is the cumulative relative frequency of the second class?
- If this distribution is skewed to the left, about what percent of the data is above the median?

Solution: a) Since $65.5 - 62.5 = 3$, the interval must be 3.

b) The lower limit must be 67. If the class interval is 3, the class must extend by 1.5 on either side of the midpoint. The layout for the first three classes is given in the table below. $n = 450$

Class	Midpoint	f	$f_{rel} = \frac{f}{n}$	F	$F_{rel} = \frac{F}{n}$
61 - 64	62.5	91	.2022	91	.2022
64 - 67	65.5	110	.2444	201	.4467
67 - 70	68.5	?	?	?	?

- The relative frequency is .275
- The cumulative relative frequency is .5025, which might be found as (about) the sum of .2022 and .2444.
- the median is defined as a point with 50% of the data above or below it.

2. Indicate whether the following are: Nominal Data, Ordinal Data, Interval Data, Continuous Ratio Data or Discrete Ratio Data. (3)

- A recent cartoon suggested that the numbers on pro Football jerseys be replaced by their salaries. What kind of data would the numbers be before the change? **Ans:** Nominal.
- What kind of data would the numbers usually be considered after the change? **Ans:** Continuous ratio.
- What kind of data is a team's score at half time? **Ans:** Discrete ratio.

3. All my family doctor's patient files are coded as follows: FS (Adult females who currently smoke); FN (Adult females who do not currently smoke); MS (Adult males who currently smoke) and MN (Adult males who do not currently smoke). Are these categories mutually exclusive and collectively exhaustive? If you say 'no' to either characteristic, explain. (2) **Ans:** The categories are mutually exclusive (for instance no person who is coded FS will be in FN, MS or MN), but not collectively exhaustive because there is no reason to assume that all the family doctor's patients are adults.

Part II. Compute an appropriate answer, showing your work (15 Points maximum - if you do more than 15 points, only your right answers will be counted.):

1) A sample of pipe outside diameters gives a mean of 16.0 inches and a standard deviation of 0.2 inches.

a) If the median diameter is 15.9 inches and the mode is unknown

(i) What is the maximum fraction of the pipes that could have a diameter above 16.4

inches? (1) **Ans:** Get a z-score. $k = z = \frac{x - \bar{x}}{s} = \frac{16.4 - 16}{0.2} = 2$. We know from the

Chebyshev rule that the fraction in the tail above $\mu + k\sigma$ is less than $\frac{1}{k^2}$. Since

$k = 2$, this fraction is below $\frac{1}{4} = 25\%$.

(ii) Between what two diameters must at least 15/16 of the pipe diameters lie? (1) **Ans:**

If only $\frac{1}{16} = \frac{1}{k^2}$ are outside the interval, we must have $k^2 = 16$ or $k = 4$. Thus the interval is $\mu \pm k\sigma = 16.0 \pm 4(0.2) = 16.0 \pm 0.8$ or 15.2 to 16.8.

(iii) Is this distribution skewed to the left or the right, or is it symmetrical? (1) **Ans:**

Since the median is below the mean, it should be skewed to the right.

b) If, instead, the median diameter is 16.0 inches and the mode is also 16.0 inches, between what two diameters must almost all the pipe diameters lie? (1) **Ans:** Since $16 \pm 3(0.2) = 16 \pm 0.6$ is 3

standard deviations from the mean, the Empirical Rule (which applies because the mean, median and mode are equal) says that there will be almost none outside 15.4 to 16.6.

c) What is the coefficient of variation for this sample of pipes? (1) **Ans:**

$$C = \frac{s}{\bar{x}} = \frac{0.4}{16} = 0.025.$$

2) The newest computer in the headquarters of a firm that you are liquidating is five months old and the oldest is 99 months old.

a) If the (absolute) frequencies are to be presented in a line graph in seven classes, what intervals would you use? Explain your reasoning using an appropriate formula and use it to fill in the table below.

(3)

Ans: $\frac{99 - 5}{7} = 13.43$ so use 14. This is only a suggestion. Any number, like 15, somewhat

above 13.57 will work, as long as you cover the range. Two possibilities are shown below. You should have shown only one.

Class	From	to	From	to
A	3	16.9	0	14.9
B	17	30.9	15	29.9
C	31	44.9	30	44.9
D	45	58.9	45	59.9
E	59	72.9	60	74.9
F	73	86.9	75	89.9
G	87	100.9	90	104.9

b) What is the name of the type of graph that you are drawing (Is it a histogram?) and what would the x and y coordinates be of the last point on the line that you draw to represent the frequencies? (2) **Ans:** The graph is a frequency polygon and we must create an empty class to end it. For the first classification above, the interval is 14 and the midpoint of the last class on the table is 94, so the last point is $(x = 108, y = 0)$. For the second classification above, the interval is 15 and the midpoint of the last class on the table is 97.5, so the last point is $(x = 112.5, y = 0)$.

