Data Science Laboratory 1

Lab 5 Problem Sheet Solutions

1. Classify the following attributes as binary, discrete or continuous. Further, classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

   (a) Age in years.

   Discrete, quantitative, ratio.

   (b) Time in terms of AM or PM.

   Binary, qualitative, ordinal.

   (c) Brightness as measured by a light meter.

   Continuous, quantitative, ratio.

   (d) Brightness as measured by people's judgements.

   Discrete, qualitative, ordinal.

   (e) Angles as measured in degrees between 0 and 360.

   Continuous, quantitative, ratio.

   (f) Bronze, Silver and Gold medals as awarded at the Olympics.

   Discrete, qualitative, ordinal.

   (g) Height above sea level.

   Continuous, quantitative, interval/ratio (depends on whether sea level is regarded as an arbitrary origin)

   (h) Number of patients in a hospital.

   Discrete, quantitative, ratio.

   (i) ISBN numbers for books.

   Discrete, qualitative, nominal (ISBN numbers do have order information, though)

   (j) Ability to pass light in terms of the following values: opaque, translucent, transparent.

   Discrete, qualitative, ordinal.

   (k) Military rank.

   Discrete, qualitative, ordinal.

   (l) Distance from the centre of campus.

   Continuous, quantitative, interval/ratio (depends)

(m) Density of a substance in grams per cubic centimetre.

Discrete, quantitative, ratio.

2. Compute the mean, median and mode for the following statistics:

Name of Dam	Height (in ft)
Oroville Dam	756
Hoover Dam	726
Glen Canyon Dam	710
Don Pedro Dam	568
Hungry Horse Dam	564
Round Butte Dam	440
Pine Flat Lake Dam	440

Then, compute the IQR, AAD, MAD, variance and standard deviation.

Mean $= \frac{1}{7}(756 + 726 + 710 + 568 + 564 + 440 + 440) = \frac{4204}{7} = 600.57$.

To compute the median we must first order the list in numerical order: 756, 726, 710, 568, 564, 440, 440. There is an odd number so the median is 568.

Since the number 440 occurs more often than any of the other numbers on this list, the mode is 440.

To compute the variance, we compute $\frac{1}{6}\sum_{i=1}^{7}(x_i - 600.57)^2 = 17638.29$. It follows that the standard deviation is 132.81.

Next, for AAD and MAD we compute $\frac{1}{7}\sum_{i=1}^{y}|x_i - 600.57|$ and $median(\{|x_1 - 600.57|, \ldots, |x_7 - 600.57|\})$, respectively. Doing so, we find $AAD = 111.51$ and $MAD = 128$.

$Q_1 = 440$, $Q_2 = 568$, $Q_3 = 726$ and so $IQR = 726 - 440 = 286$.

3. Find the median of the following list:

$$42, 48, 24, 25, 42, 28, 31, 33, 51, 57, 68, 33, 75, 36, 79, 85, 79$$

Then compute the IQR, variance and standard deviation.

- $mean = 49.18$
- $median = 42$
- $variance = 434.15$
- $standard\,deviation = 20.84$
- $Q_1 = 32$, $Q_2 = 42$, $Q_3 = 71.5$, $IQR = 39.5$.