Graded Problem Set # 1 Solutions

1 a.) $\bar{X} = 6.2$ Median $= 6.5$

$s^2 = 17.955$ $s = 4.237$

b.) Mean + median are the same.
Population variance and population standard deviation are different because the population variance is obtained by dividing the sum of squared deviations from the mean by all measurements (N) whereas the sample variance is obtained by dividing the sum of squared mean deviations by the number of measurements in the sample minus 1 $(n-1)$.

$\therefore \sigma^2 = 16.16$ $\sigma = 4.02$

2.)

a.) $\bar{X} \pm s \Rightarrow (19.4, 35.2)$

$\bar{X} \pm 2s \Rightarrow (11.5, 43.1)$

$\bar{X} \pm 3s \Rightarrow (3.6, 51)$

2 b.) The empirical rule holds that data sets with frequency distributions that are mound shaped and symmetric will approximately satisfy the following conditions:

i.) 68% of the measurements will fall within one standard deviation of the mean

ii.) 95% of the measurements will fall within two standard deviations of the mean.

iii.) 99% of the measurements will fall within three standard deviations of the mean.

Yes, the data seem to obey the rule. 22/32 = .6875 of observations fall within $\bar{x} \pm s$. 31/32 = .969 fall in $\bar{x} \pm 2s$

c.) Q1 : $(n+1)\frac{p}{100} = (33)\frac{25}{100} = 8.25 \Rightarrow x_{8.25} = x_8 + .25(x_9 - x_8)$

$= 21.5$

Q3 : $(33)\frac{75}{100} = 24.75 \Rightarrow x_{24.75} = x_{24} + .75(x_{25} - x_{24})$

$= 33$

IQR = Q3 − Q1 = 33 − 21.5 = 11.5

d.) The middle 50% of the measurements fall within the inter-quartile range. The middle 68% of the measurements fall within one standard deviation of the mean when the empirical rule is applicable.

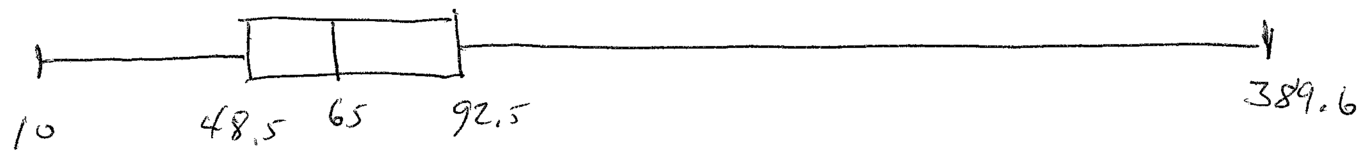
3.) $\Sigma(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$ by Rule #3

$= \Sigma x_i - n\bar{x}$ by Rule #1

$= \Sigma x_i - \cancel{n}\left(\frac{\Sigma x_i}{\cancel{n}}\right)$ by definition of $\bar{x}$

$= \Sigma x_i - \Sigma x_i = 0$ ✓

4.) a.) $x_{min} = \$10,000$ $Q1 = \$48,500$ $Q2 = \$65,000$
$Q3 = \$92,500$ $x_{max} = \$389,599$

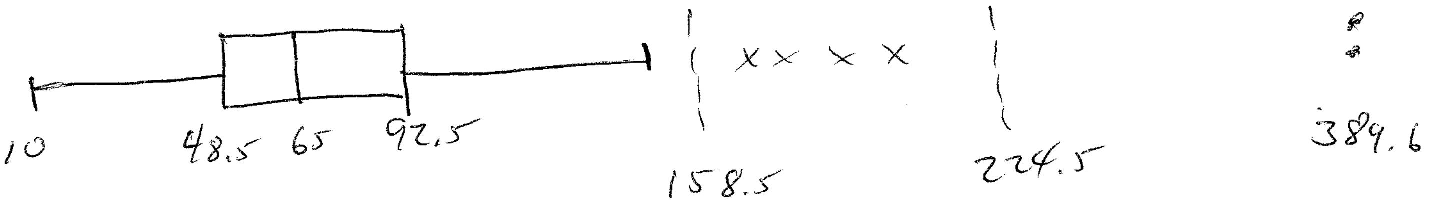


The central tendency or middle of the data is at \$65,000. That is, the typical HR manager earns \$65,000 per year. The inter-quartile range indicates that the middle 50% of the observations fall between \$48,500 and \$92,500. Thus there is considerable variability (dispersion) in HR manager earnings. By virtue of the long right whisker, we see considerable positive skewness in the incomes of the HR managers.

4b.) Inner fences $\left\langle \begin{array}{l} -\$17{,}500 \\ +\$158{,}500 \end{array} \right.$

Outer fences $\left\langle \begin{array}{l} -\$83{,}500 \\ +\$224{,}500 \end{array} \right.$

The lower fences are negative so can be ignored since income is strictly positive



4 unusual observations + 2 outliers.

5a.) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .7$

b.) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = .75$

c.) $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = .5$

d.) $.6 = P(A) \neq P(A \mid B) = .75$, therefore A and B are dependent.

e.) A and B are not mutually exclusive since there is a .3 probability that they are observed together, i.e.,

$P(A \cap B) = .3$.