Taylor Watanabe (Whittier I.D: 20594796)
20 July 2020
Homework #2
Math 080


74.     a)

| Steam | Leaf |
|---|---|
| 5 | 1 |
| 6 | |
| 7 | 6,7,8,9 |
| 8 | 1,2,4,6 |
| 9 | 9 |

        b)      51 and 99 are outliers because 51 is (-) 25 points from the next closest exam score of 76 and 99 is (+) 13 points from its nearesy exam score of 86. 51 and 99 are separate from all the remaining scores thus outliers.


75.     a) (Used TI-84 calculator) generated {49,47,8,27,21,38,3,18} which correlate to the states, [ West Virgina, Virgina, Delaware, Montana, Maryland, Oregon, Arizona, Kentucky] with subsequent percentages represented in bar graph below of those eight states.
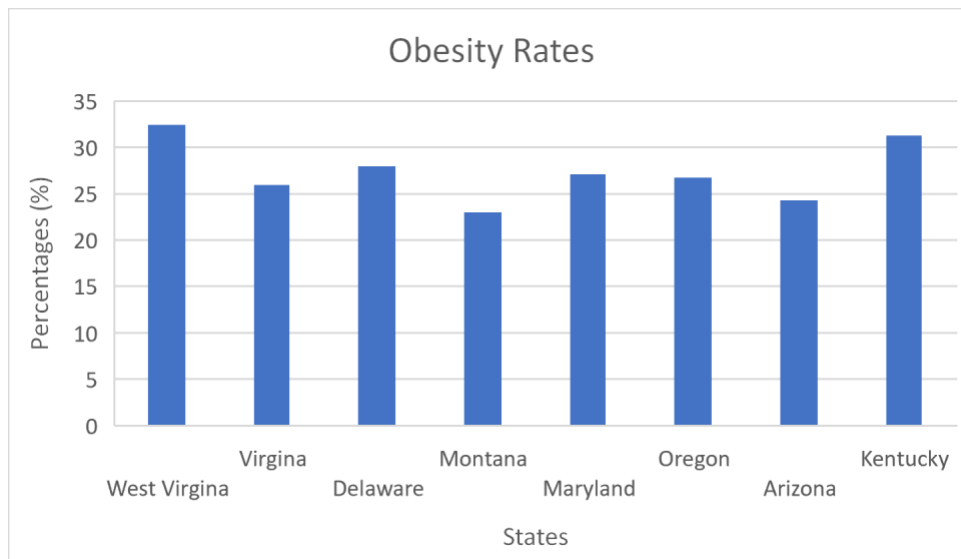


76.     a)

| # of books | Freq. | Rel. Freq. |
| --- | --- | --- |
| 0 | 10 | 0.15 |
| 1 | 12 | 0.18 |
| 2 | 16 | 0.24 |
| 3 | 12 | 0.18 |
| 4 | 8 | 0.12 |
| 5 | 6 | 0.09 |
| 6 | 2 | 0.03 |
| 8 | 2 | 0.03 |

**Table2.62** Publisher A

| # of books | Freq. | Rel. Freq. |
| --- | --- | --- |
| 0 | 18 | 0.15 |
| 1 | 24 | 0.20 |
| 2 | 24 | 0.20 |
| 3 | 22 | 0.18 |
| 4 | 15 | 0.13 |
| 5 | 10 | 0.08 |
| 7 | 5 | 0.04 |
| 9 | 1 | 0.01 |

**Table2.63** Publisher B

| # of books | Freq. | Rel. Freq. |
| --- | --- | --- |
| 0–1 | 20 | 0.29 |
| 2–3 | 35 | 0.50 |
| 4–5 | 12 | 0.17 |
| 6–7 | 2 | 0.03 |
| 8–9 | 1 | 0.01 |

**Table2.64** Publisher C

b)



Publisher A

Publisher B
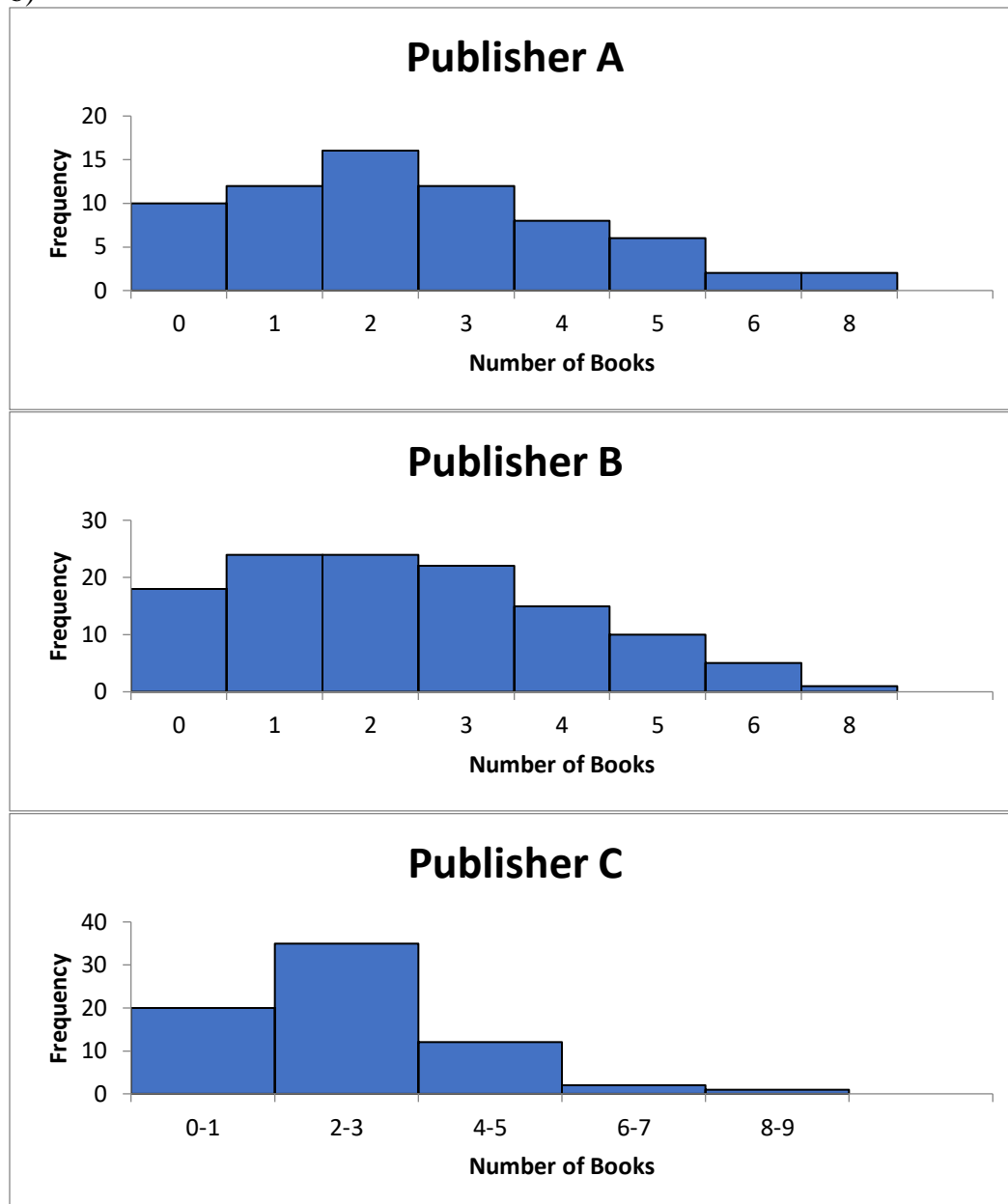
Publisher C

c)      The histograms of publisher's A and B are not identical because of many possible reasons, a couple include different manufactures, different titles, and thus different pricing. Another reason could be the data itself consisting of different populations of customers that differ in reading preference and needs.

d)      No, I wouldn't expect the histogram of publisher C to look like the other two as the bin width is double the size thus changing the number of bins and subsequently the frequency in each bin.
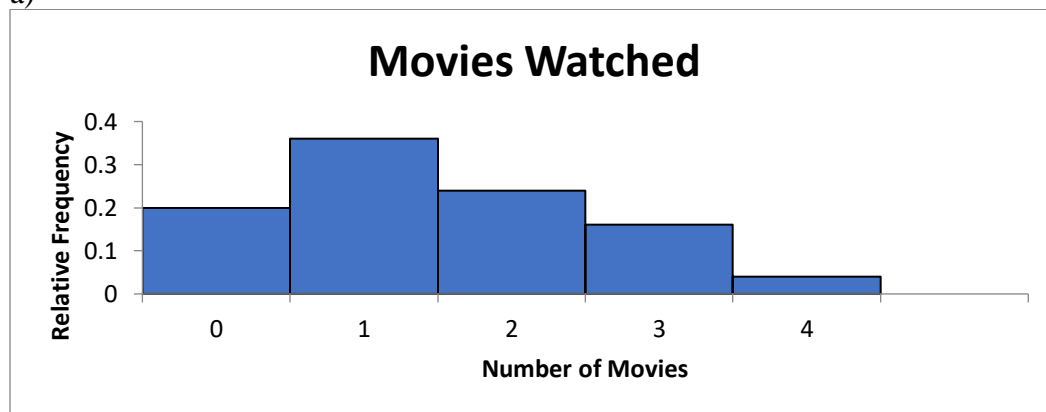
e)

## Publisher A
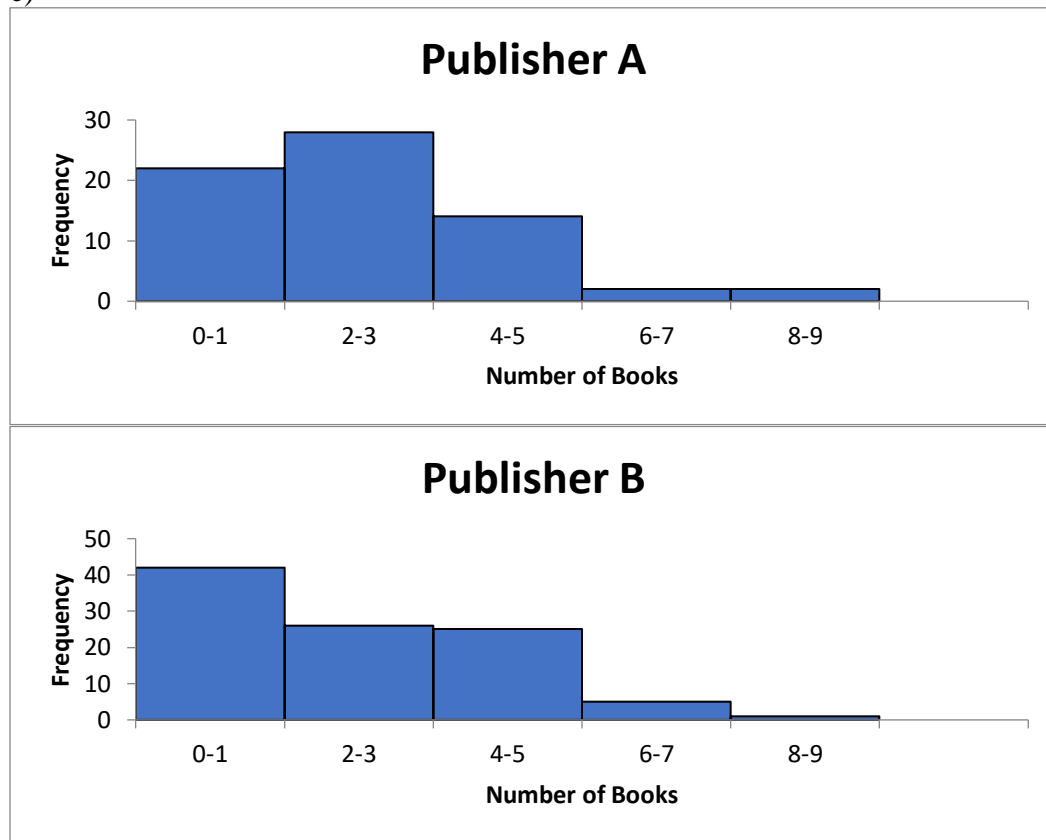


## Publisher B



f)      All three histograms look more alike now as all have the same bin width and all have a common pattern of a decrease in frequency as there's an increase in the number of books with the exception of the bin (0-1) which is seen to be less in publisher C's histogram compared to publisher's A and B.

78.     a)

## Movies Watched

b)

| # of movies | Frequency | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|---|
| 0 | 5 | 0.20 | 0.20 |
| 1 | 9 | 0.36 | 0.56 |
| 2 | 6 | 0.24 | 0.80 |
| 3 | 4 | 0.16 | 0.96 |
| 4 | 1 | 0.04 | 1 |

**Table2.67**

79.    **C. 41**

80.    **D. convenience**

83.    a)    **0.06**

       b)    **0.63**

c)



Histogram-Middle Class Income
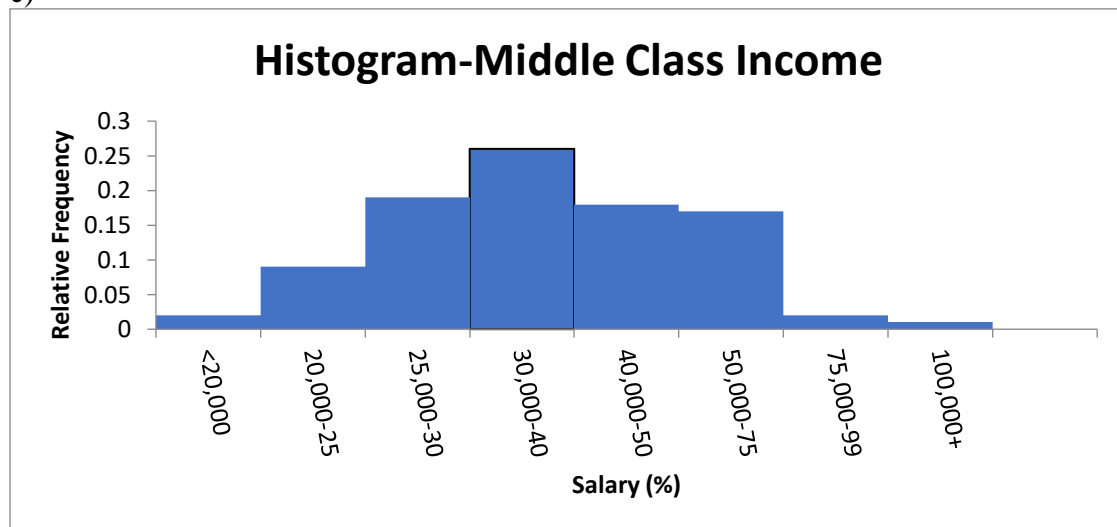
     i.        All bars should not have the same width as the bins, <20,000 and 100,000+ will contain more or less data values.

     ii.       The <20,000 and 100,000+ intervals should be handled by calculating the relative frequencies of the data that falls within those intervals.

d)        The $40_{th}$ percentile = between 30,000 and 40,000; while the $80_{th}$ percentile = between 50,000 and 75,000.

e)



**Middle Class Income**

84.    a)        Smallest spread of data: $Q_3$ to max (spread from 12-13)

b)        Largest spread of data: $Q_1$ to $Q_2$ (spread from 2-10)

c)        IQR $= Q_3 - Q_1 = 12 - 2 = \mathbf{10}$

d)        There's more data in interval; 10-13 as 50% of the data is distributed within the 10-13 interval. It was calculated for interval 2-10 to hold 25% of distribution thus 5-10 would contain less than 25% making interval 10-13 to contain the most data.

e)        ii (interval 2-4). Interval 2-4 is contained within interval 2-10 and thus has the smallest amount of data (<25%).

88.    a)
        i.        There's not enough information given to know the data values in each dataset. So, you don't know if data set 2 has more data values above two like it falsely states.

        ii.        There's not enough information given to know the modes of the datasets as no individual data values are given. So the datasets have the same probability of having the same mode as not having the same mode.

        iii.        Overall, it does seem that there is a larger proportion of data values below four but you are unable to state that that are more below four than there are above four as there is not enough information given to allow for that analysis of comparison.

b) The value of "7" is more likely to be an outlier in dataset 2. By looking at the box plot of dataset 2, it's clearly shown that majority of the data is below 4 and thus makes 7 an outlier as it's separate from the rest of the data.

92.    $\bar{x} = 1328.65/50 = \mathbf{26.75}$

93.    a)    If the medium age rises, it could indicated an overall increase in population age.

       b)    The medium age could rise because the overall birthrates decreased/the number of women wanting a family decrease, or the overall life span of the population is increasing.

       c)    Comparing 1980 data with 1991, it shows that in 1980 the medium age was one year less than the medium age in 1991 possibly indicating that there were a smaller number of children but more information would be needed for analysis.

94.    **6 years** will be expected to have a FTES of 1014 or above as the sample number is 11 (11 years) so the medium of those 11 data points is 6. Thus 6 years would be expected to have a FTES at or below the medium.

95.    a)    **1447.5 FTES**

       b)    **528.5 FTES**

96.    **474 FTES**

97.    The percent of the FTES that were from 528.5 to 1447.5 is **50%.** The medium is 1014 FTES, $Q_1$ equals 528.5 FTES, and $Q_3$ equals 1447.5 FTES, with 25% of the data is below $Q_1$ and 75% of the data below $Q_3$ so the percent of FTES would be 75% - 25% equaling 50%.

98.    **IQR = 919**. The IQR represents where the middle point is within the dataset. IQR = $Q_3$-$Q_1$.

99.    **0.029**