

Solutions to Homework 2

Prof. Jordan C. Hanson

July 31, 2020

Exercises, Chapter 2: 74, 75a, 76, 78, 79, 80, 83, 84, 88, 92, 93, (94-99)

- Exercise 74 (a-b): The 51 and 99 seem to be outliers according to the stem plot (see Tab. 1).

Stem	Leaves
0	
1	
2	
3	
4	
5	[1.0]
6	
7	[7.0, 8.0, 6.0, 9.0]
8	[1.0, 6.0, 2.0, 4.0]
9	[9.0]

Table 1: A stemplot for exercise 74. .

- Exercise 75a:

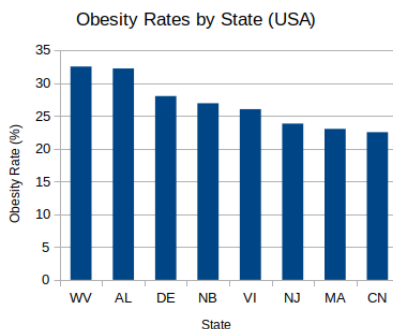


Figure 1: A bar chart of obesity rates in the USA.

- Exercise 76:
 - a: See Tab. 2.
 - b: See Fig. 2.
 - c: The publisher B data has a different bin design. The publisher A data and B data have different N values.
 - d: If the people sampled buy books from all three publishers, and the publishers do not specialize in a particular type of fiction paperback, I would expect Fig. 2 (right) to resemble the others. However, it also has wider bins.
 - e/f: The result is that each plot has the same bin design, and shows the decrease in frequency with increasing number of books purchased. However, the difference for the 0 and 1 bin varies between publisher A and B.
- Exercise 78: See Fig. 3. The relative frequencies are 0.2, 0.36, 0.24, 0.16, and 0.04. The cumulative frequencies are 0.2, 0.56, 0.8, 0.96, and 1.0.

Books	Freq.	Rel. Freq.	Books	Freq.	Rel. Freq.	Books	Freq.	Rel. Freq.
0	10	0.15	0	18	0.15	1	20	0.29
1	12	0.18	1	24	0.20	3	35	0.5
2	16	0.235	2	22	0.20	5	12	0.17
3	12	0.18	3	22	0.18	7	2	0.03
4	8	0.12	4	15	0.13	9	1	0.01
5	6	0.09	5	10	0.08			
6	2	0.03	7	5	0.04			
8	2	0.03	9	1	0.01			

Table 2: (Left) Publisher A, (Middle) Publisher B, (Right) Publisher C.

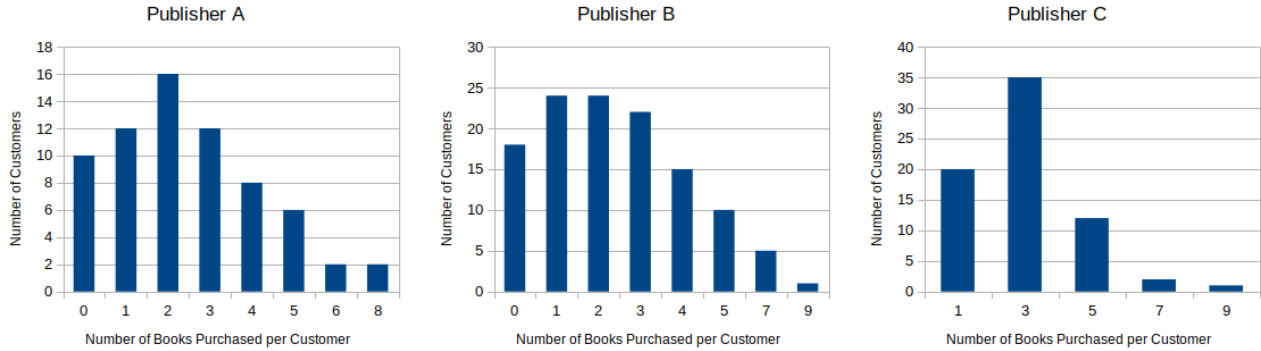


Figure 2: Histograms of books purchased by customers of three different companies.

5. Exercise 79: (a) 41 percent.
6. Exercise 80: (d) Convenience sampling.
7. Exercise 83: (a) 6 percent (b) 45 percent (c) See Fig. 3 (right). (d) 40th is in the 25,000 - 35,000 bin, and 80th is in the 50,000 to 75,000 bin. (e) See Fig. 3 (right).
8. Exercise 84: (a) Q4, 1.0 (b) Q2, 8.0 (c) $Q3 - Q1 = 10$ (d) In the interval 10-13 there is half of the data. In the other interval there is less than half. (e) ii) does not have $1/4$ of the data like the other intervals.
9. Exercise 88: (a) i) Although data 1 has a greater *proportion* above 2 than data 2, we can't know if that's true of the number of total values because the sets could have different N values. ii) Same as (i), we would need to see the individual data points. iii) False: the proportions are equal. (b) Data 2. Seven is at the far edge of the upper quartile, and farther from the median relative to data 1.
10. Exercise 92: 26.5 percent.

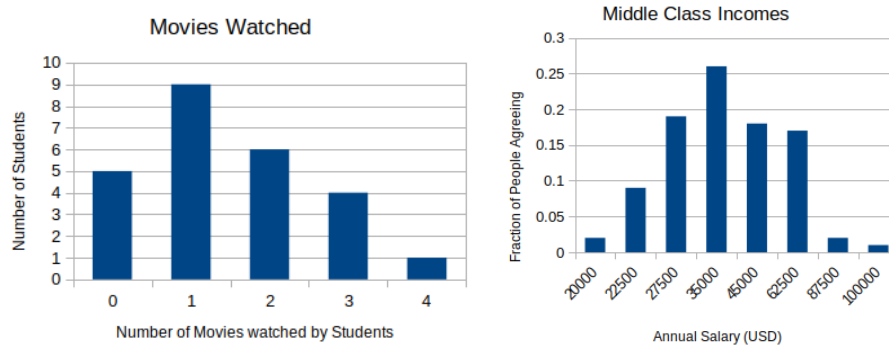


Figure 3: (Left) Movies watched by students. (Right) Annual salaries people associate with the term *middle class*. i) Note the bars do not correspond to constant width. ii) The smallest and largest bins technically do not have fixed size.

11. Exercise 93: (a) This means the upper two quartiles contain a larger fraction of the data. (b) Life expectance could increase, and people could be having fewer children. (c) No, because it depends on the overall population.
12. Exercise 94-99: 94) 6 years, because half of the data is below 1014. 95) i) 1447.5 ii) 528.5 96) 474 97) 50 percent (Q1 to Q3) 98) 919 FTES, the middle 50 percent of the data. 99) 0.03.