Solution to Worksheet 1: Chapter 2

Fred Azizi

2023-02-07

For questions 1-5, indicate the level of measurement being used in the given scenario: Nominal, ordinal, interval, or ratio.

1. A local animal shelter keeps track of the breeds of dogs that come in.

   Answer: Nominal.

2. A local animal shelter keeps track of the weight of dogs that come in.

   Answer: Ratio.

3. The number of miles joggers run per week.

   Answer: Ratio.

4. The starting salaries of graduates of MBA programs.

   Answer: Ratio.

5. Before leaving a restaurant, customers are provided a feedback form. For each question, determine the level of measurement:

   a. What is the approximate distance between this restaurant (in miles) & your residence?

      Answer: Ratio.

   b. Have you ever eaten at this restaurant before?

      Answer: Nominal.

   c. On how many occasions have you eaten at this restaurant before?

      Answer: Ratio.

   d. Which of the following attributes of this restaurant do you find most attractive: service, prices, quality of the food, or the menu?

      Answer: Nominal.

   e. What is your overall rating of this restaurant: excellent, good, fair or poor?

      Answer: Ordinal.

6. Fifteen students answered the question of how many siblings they have. The answers were:

$$1, 1, 2, 0, 3, 2, 1, 4, 2, 3, 1, 0, 0, 1, 2.$$

Construct a simple frequency distribution which includes frequencies, relative frequencies, percent frequencies and cumulative relative frequencies.

Answer:

This data has a *discrete* structure. Therefore, our classes should be 0, 1, 2, 3, 4. Frequency column is filled by frequency of each class. Relative frequency is filled by dividing the frequency of each class by the total number of data (in this case 15). Percent frequencies are constructed by multiplying each relative frequency by 100. For cumulative relative frequencies, first cell would be the relative frequency of that cell. The rest would be filled as relative frequency of that cell+ cumulative relative frequency of the previous cell. Last cell must be 1.

Classes (Number of siblings)	Frequency	Relative frequency	%frequency	CRF
0	3	0.2	20	0.2
1	5	0.333	33.3	0.533
2	4	0.267	26.7	0.8
3	2	0.133	13.3	0.933
4	1	0.067	6.7	1
Total	15	1	100	

7. A paper by Appleton, French, and Vanderpump (1996, American Statistician) goes over a survey of 1314 English woman in 1972-74 were their smoking habits along with other information were recorded and then there was a follow up 20 years later to see how many survived. They summarized the result in a cross-classification table.

   a. During the publication process, a part of table got lost. Fill the table with the given information.

      Answer:

Smoker	Dead	Alive	Total
Yes	139	443	582
No	230	502	732
Total	369	945	1314

   b. What percentage of smokers survived? What is your conclusion?

      Answer:

      Percentage of smoking survivor: $\frac{443}{582} \approx .76$. Percentage of non-smokers who stayed alive during the study: $\frac{502}{732} \approx 0.69$. Naievly, we may conclude that smokers seems to survive more comparing to non-smokers!!

   c. Now look at the following table which shows the results of the study broken down by age group, does your conclusion change?

      Answer:

Age group	$18-24$		$25-34$		$35-44$		$45-54$		$55-54$	
Smoker	Y	N	Y	N	Y	N	Y	N	Y	N
Dead	2	1	3	5	11	7	27	12	51	40
Alive	53	61	121	152	95	114	103	66	64	81
Mortality Percentage	0.036	0.016	0.024	0.032	0.103	0.058	0.208	0.154	0.44	0.331

      Mortality is higher among smokers with a small exception of 25-34 age range group. The previous table was biased toward younger people. This is a classic example of a phenomena called **Simpson Paradox**.