

HW1: Descriptive Statistics—Percentages and Central Tendency

1. A vaccine to prevent the flu was provided free of charge in a small community in a two-shot sequence over a 2-week period. Some people received both shots, some took only the first shot, others received neither. A survey of 1,000 residents taken the following spring provided these data:

Treatment	Got the Flu (# of persons)	Did Not Get the Flu (# of persons)
No Vaccine	29	273
One Shot	9	91
Two Shots	14	586

- (1) Calculate descriptive statistics to describe the relationship between “Treatment” variable and whether the individual got the flu or not. (6 pts)
- (2) Describe the relationship between Treatment and whether an individual got the flu or not. (2 pts)

Ans: (1)

Treatment	Got the Flu (# of persons)	Did Not Get the Flu (# of persons)	Total	Got the Flu (Percentage, %)	Did Not Get the Flu (Percentage, %)
No Vaccine	29	273	302	$29/(29+273)*100\%=9.6\%$	$273/(29+273)*100\%=90.4\%$
One Shot	9	91	100	$9/(9+91)*100\%=9.0\%$	$91/(9+91)*100\%=91.0\%$
Two Shots	14	586	600	$14/(14+586)*100\%=2.3\%$	$586/(14+586)*100\%=97.7\%$

- (2) As shown in the table above, with treatments of no vaccine and one shot, the percentages of people who got the flu versus who did not get the flu are similar (9.6% vs 91.0%). This indicates that there is only a slight difference between these two treatments. Instead, people who received two shots have a lower percentage of getting the flu. It decreases from 9.6% to 2.3% for those with no vaccine vs. two shots, and from 9.0% to 2.3% for those with one shot vs. two shots. Therefore, receiving a two-shot vaccine may have some effect on protecting from getting the flu.

2. The following table shows the number of children in a sample of 30 randomly selected families in Nantou.

Number of Children in the Family	Number of Families
0	9
1	7
2	5
3	4
4	3
5	2

- (1) Calculate the mean number of children per family in this sample. (1 pt)
- (2) What is the median number of children per family? (1 pt)
- (3) What is the mode for the above distribution? (1 pt)
- (4) What is the statistical unit here? (1 pt)
- (5) What is the variable here? (1 pt)

Ans: (1) Mean = $(0 \times 9 + 1 \times 7 + 2 \times 5 + 3 \times 4 + 4 \times 3 + 5 \times 2) / 30 = 1.7$ children/family

(2) Md position = $(30+1)/2 = 15.5$

Number of Children in the Family	Number of Families (f)	cf
0 ← Modal class	9	9
1 ← median class	7	16
2	5	21
3	4	25
4	3	28
5	2	30

Median= 1 child/family

(3) The highest frequency of “number of families” is 9, and the corresponding category is 0. → Mode = 0 child/family

(4) The statistical units in this question are 30 families. (Also accept the answer of “A statistical unit in this question is a family”).

****Note to students:** Statistical units are data items used for their unique identification to provide information for which statistics are ultimately compiled. In this context, “30” represents the sample size, and “family” describes the statistical unit (also the analytical unit) of the entity for which the required statistics are compiled.

(5) Number of children in the family