

Solution for HW2 of Stat 13

Problem 1

a)

(1 point)

Backpack Code	PM	MM	BL	DS	CM	MR	MP	RP	JC	CB	EW	EA
Ave. Volume	33.7	41.1	45.8	37.5	36.9	39.8	34.1	37.5	22.4	29	34.7	30

b)

Backpack code	Price(\$)	Volume	Number of Books	Ave. Volume
MR	35	2000	49	40
CM	35	1800	49	37
JC	35	1600	73	22
PM	38	1600	47	34
MM	40	1800	44	41
EW	40	1700	49	35
EA	40	1600	50	30
MP	45	1700	49	34
DS	48	1900	50	38
BL	50	2200	48	46
RP	50	1500	40	38
CB	95	1100	38	29
Average	46	1700	49	35

The price of the backpack doesn't have a closed relationship with the total volume of backpacks or number of books they can hold or the average volume each book has for different backpacks. One backpack has an extreme capacity of holding 73 books.

(1 point for rounding the data, especially round volume to only two effective digits.

1 point for arranging the observations into rows.

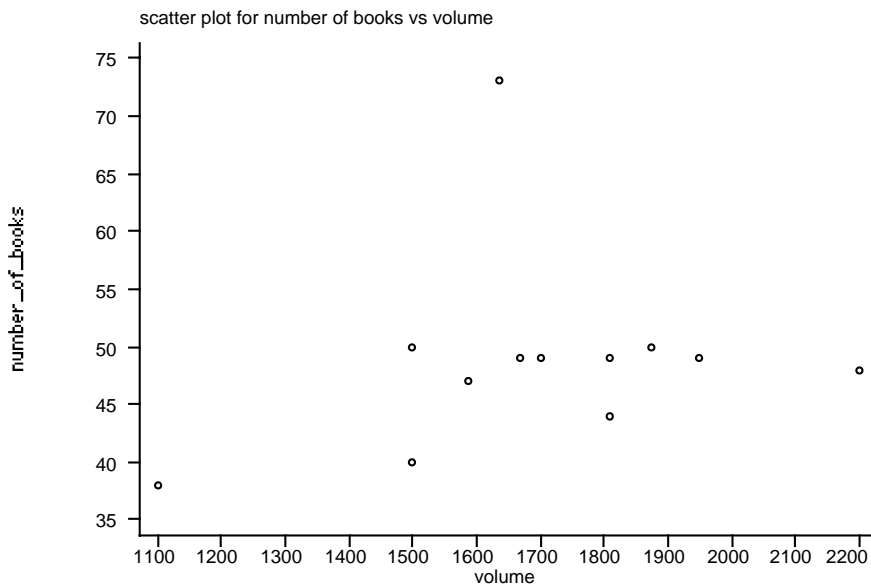
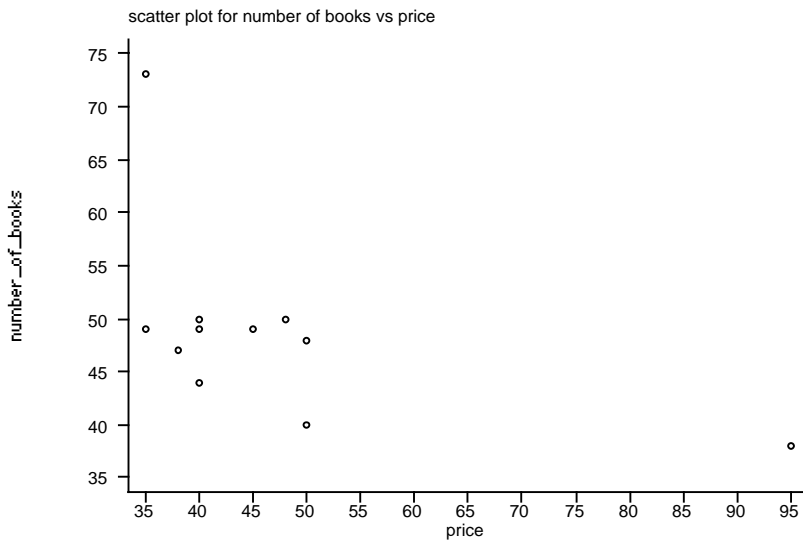
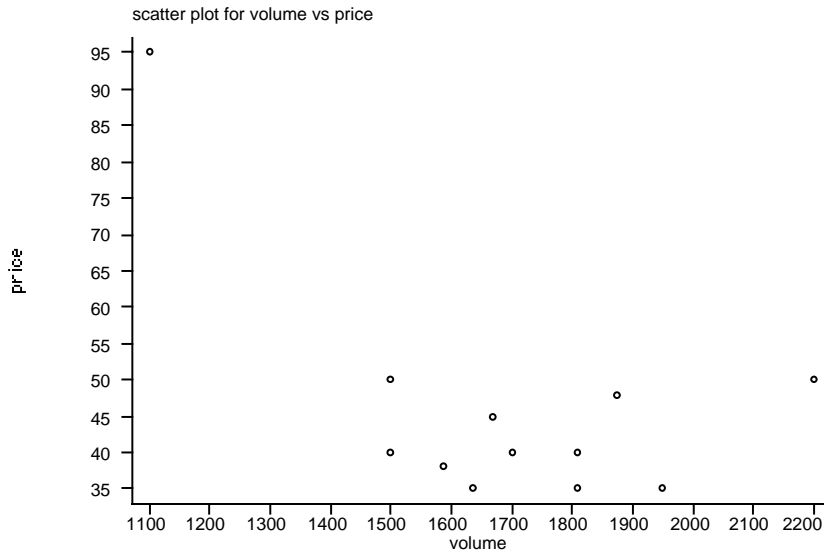
1 point for order the observations by price, and for those with same price by a secondary variable.

1 point for add column average, not row average.

1 point for a short graph of summary)

c)

(2 points for each plot. For a plot, 1 point for the right thing to plot and 1 point for appropriate titles)



d) (4 points)

The data show us that there is no clear relationship between price and volume or number of books or ave. volume. There is an outlier for price, i.e. CB. Its price far exceeds the common range of other ones. A typical price for a bag is 40, with minimum price being 35, and maximum price being 50 if we don't take the outlier into account. On average, high capacity of holding books implies higher volume. Again, there is no clear relationship between ave. volume per book and number of books. But the ave. volume per book increases as the total volume of the bag increases in general. For total volume, there is one outlier: 1100 for CB. And for the number of books, there are two outliers: 38 for CB and 73 for JC. For the ave. volume, 22 for JC should be a suspected outlier. After considering this, we can see that for CB, it is especially expensive, and also especially low volume and low capacity of holding books, but its efficiency is very good: ranked No. 2. Another outlier should be JC, with an extremely high capacity. (Suspected data error?).

Another factor which may affect customer's decision may be the number of auxiliary small bags of the backpack, because we may need some individual small bags to hold small items.

As a recommendation, I would suggest buying MR, because it has the highest volume with the backpacks with lowest price, and its capacity of holding books are also very good (if not considering the suspected outlier, it ranks No.2).

Problem 2

Outlier: some extreme cases with value(s) well away from the main body of the data. (2 points)

Skewed Distribution: the distribution is not symmetric: the observations for one end are larger than that of the other end. (2 points)

Sampling and non-sampling error: sampling error is the error because of taking only part of population into account, which is inevitable for random sampling. Non-sampling error is the error because of bad method of sampling. We can decrease or even eliminate non-sampling error by applying better method. (4points)

Examples for three basic ways: (1 point for each example)

Survey:

1. we are going to take the opinion of all the UCLA freshmen about LA Lakers. We go to the administration department of the university and get the whole list of freshmen, then we randomly choose 100 people from the list and interview them either by phone or personally.

2. we are interested at the customer opinion of all the customers of Verizon about the service of Verizon, We randomly choose 500 telephone number from the telephone number list of the company and interview them by phone.

Experiments:

1. we are interested at whether extra Vitamin C consumption has some effect on the growth of the mice. We randomly choose 30 baby mice, divided into 2 groups. For one group, give the mice an tablet of VC everyday, but no tablet for the mice in the other group. Measure their weight and after 3 month, measure weight again to compare the difference.

2. we are interested at whether a new fertilizer are effective or not. We divide the whole experiment area into small blocks, and randomly assign that kind of fertilizer to half of the blocks, then compare the output.

Observational study:

1. we are interested at whether aspirin has effect on decreasing the hazard of lung cancer. We observe some people, and record at what extent they take aspirin: seldom, light, median, heavy. Then check their death rate w.r.t. lung cancer.

2. we are interested at whether sun light will increase the hazard of skin cancer. We observe some people, and record how often they are under sunlight. And we also record whether they at last develop skin cancer or not.