Stat13 Homework 2

(50 points, student scores will be converted to scores out of 100)

Suggested Solutions

(1) see table. (2 points)

(2) The table: (5 points)
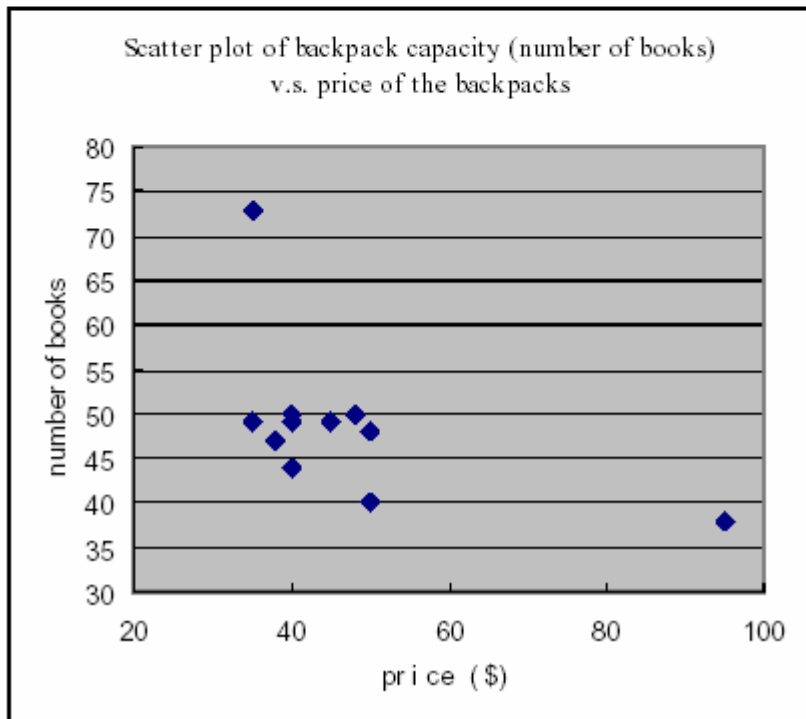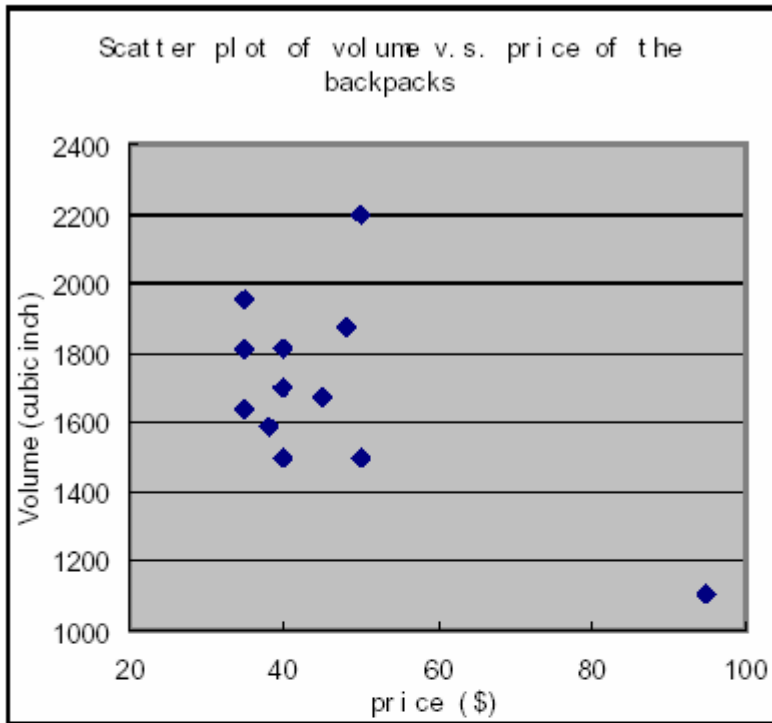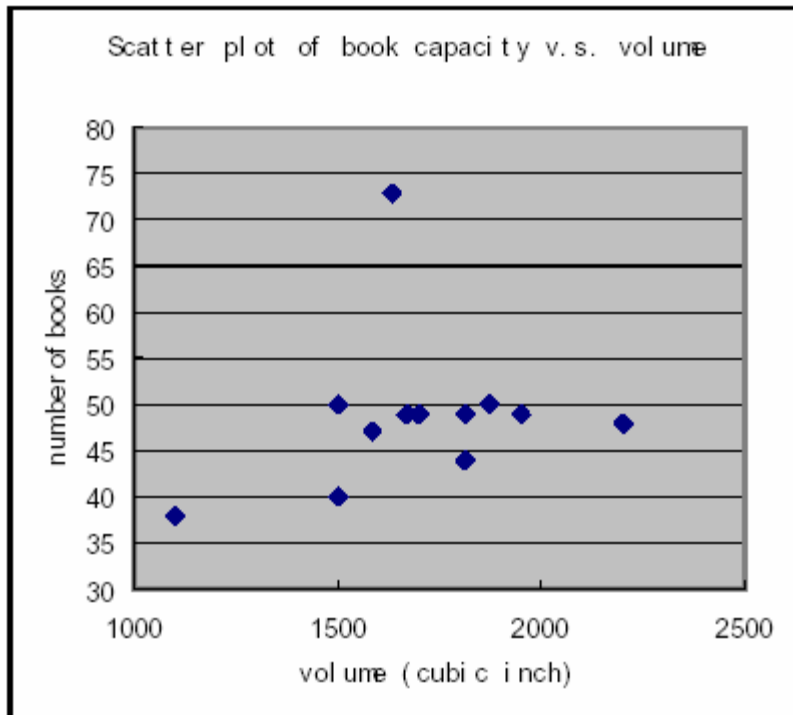
Table 1. Comparison between backpacks.

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

*unit: cubic inch

summary: please refer to (4)

(3) The plots: (15 points)

Scatter plot of volume v.s. price of the backpacks

Scatter plot of backpack capacity (number of books) v.s. price of the backpacks

Scatter plot of book capacity v.s. volume

number of books

80
75
70
65
60
55
50
45
40
35
30

1000   1500   2000   2500

volume (cubic inch)

 (4) summary of data (8 points)
I am not good at writing summaries, here's the key points I want to convey:
Outliers: one price outlier (CB), and one number of books outlier (JC), one volume outlier (CB). The volume outlier is somewhat subjective. BL can also be one.
Trends: There is no obvious correlation between price and volume, or price and number of books; there is some correlation between volume and number of books.
Other data of interest: Since the price is not correlated with either volume or book capacity, there should be other factors influencing the price, for example: material of the backpacks, quality examination data (how durable the backpacks are), value of the brand, where the backpacks are made (made in China means cheaper, generally), decorations on the backpacks, number of chambers in each backpack, fancy or not (may be a score given by some students), features for comfortness … …
A purchase suggestion: Based on current data, JC is the one to suggest: low price, moderate volume and high capacity. Other factors that may contribute to decision-making are already listed in the previous paragraph.

HW_2_2 (Mostly copied from websites) (3 points for each part, 12 points total)
(1) Outlier
Outliers are very atypical observations. It refers to observations in a distribution of data that deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism, and therefore discarding of the observations might be considered.
(2) Skewed distribution
A distribution that is not symmetrical - that is, the majority of the scores are lumped toward one end or the other.
(3) Sampling and non-sampling errors

There is generally some difference between a statistic i.e. the value obtained from a sample and its corresponding parameter of the population. This difference is called **sampling error.** This error can be minimized by increasing the size of the sample and by taking sample of extremely random in nature.

Non-sampling error is the error attributable to all sources other than sampling error. Non-sampling errors arise during the planning, conducting, data processing and final estimation stages of all types of survey.

(4) Two examples of each of the three basic ways of data collection.

Since every student's answer will be different, there is no point making an answer key to this question. Instead, I put the key points of the three types here for checking against the students' answers.

"Survey": To obtain the characteristic of a population, systematically collect and analyze data from **a sample from the population**.

"designed experiment": systematically **create differences** in the treatment factor and examine any resulting changes in the response variable.

"observational study": **observe differences** in the treatment factor and notice any related differences in the response variable.

HW_2_3

(a) (2 points)
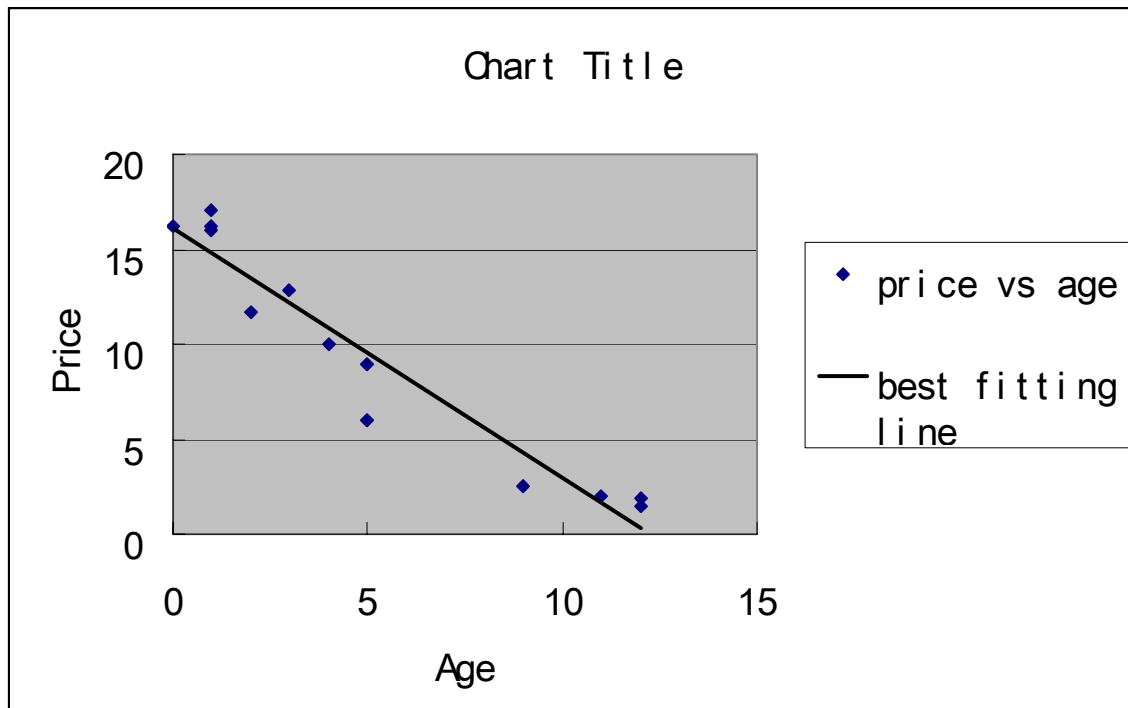
(b) As a general trend: the higher the age, the lower the price.   (2 points)
(c) from left to right, the trend line will go down, i.e., negative slope. (2 points)
(d) see the chart. (2 points)