Stat 13: Homework 2

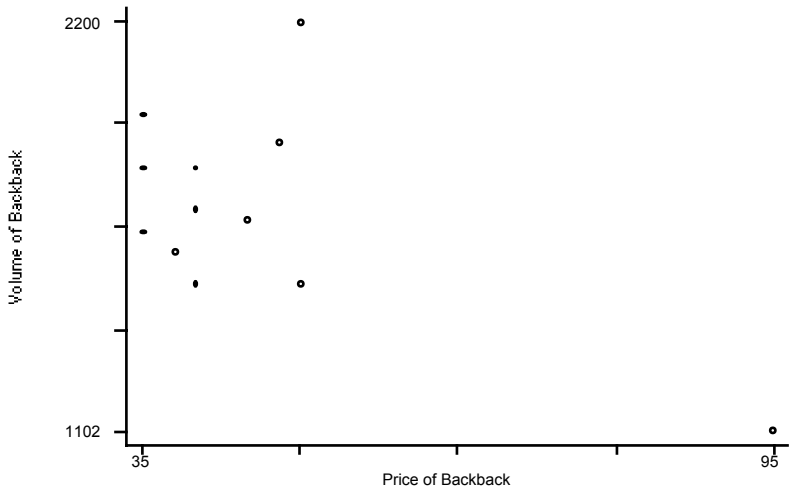
Chart: Should be in this order or reverse order based on price. Avg Volume should contain at most one decimal place of accuracy (the whole list is only here for your knowledge when grading so youÕll know if they only made a simple rounding error.)

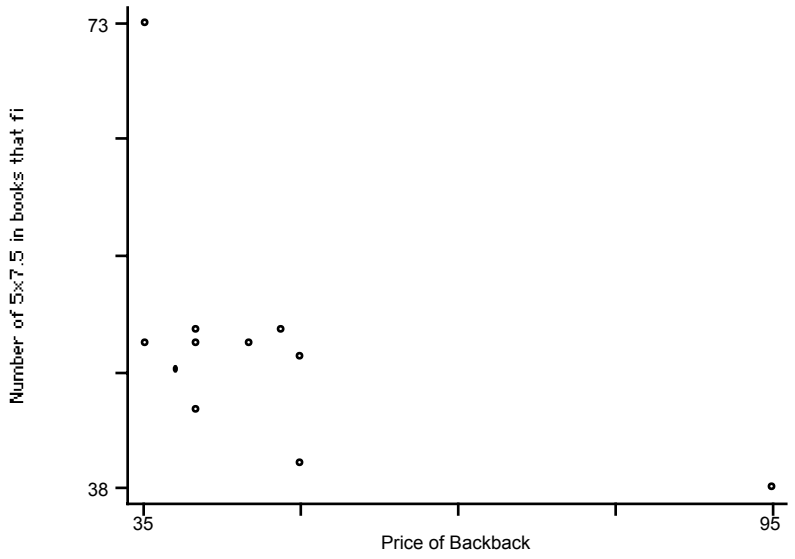
Scoring the Chart: 15 points possible. Take off 5 points if the avg vol per book data is missing or completely inaccurate, take off 3 points if itÕs accurate but not rounded at all. Take off 5 points if not sorted by price (either ascending or descending is fine). Be aware that some people will also round the volume column, either rounding or not rounding is fine-donÕt worry about it). So basically: 5 points for having the last column rounded either to the ones or tenths place (only 2 if present but not rounded), 5 points for sorting by price, and 5 points for simply having a chart with all 12 observations.

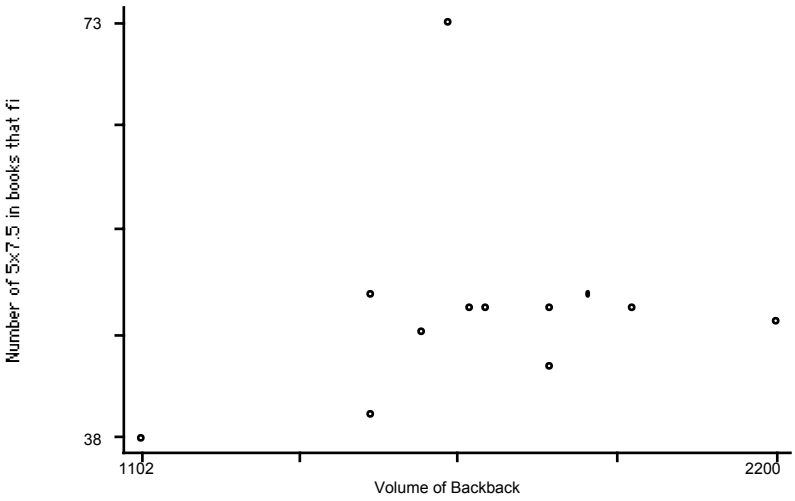
	backpack	price	volume	book	avgvol
1.	JC	35	1635	73	22.39726
2.	CM	35	1810	49	36.93877
3.	MR	35	1950	49	39.79592
4.	PM	38	1586	47	33.74468
5.	EW	40	1700	49	34.6938
6.	EA	40	1500	50	30
7.	MM	40	1810	44	41.13636
8.	MP	45	1670	49	34.08163
9.	DS	48	1874	50	37.48
10.	RP	50	1500	40	37.5
11.	BL	50	2200	48	45.83333
12.	CB	95	1102	38	29

Note: I had a student tell me that Prof Dinov said having data read across in rows rather than columns as shown above is also acceptable so donÕt take off for that.

Graphs:







For the 150-250 word report on the data look for ideas similar to the following:

1 In the first graph we see that price and volume are not as closely linked as we might have suspected. The CB backpack has a very high price (95) and relatively low volume (1102)--(specificially identifying this point as the CB backpack isnÕt necessary but observing that it is interesting should be done).

2 Again in the Num Books v Price graph we see that the CB backpack stands out and there isnÕt a clear pattern or strong relationship in the other points comparing price and number of books that fit.

3 WhatÕs interesting in the third graph is that number of books doesnÕt seem to increase as volume increases as we might expect. Noticing that the JC backpack holds significantly more books than the others and that a backback with 2200 cubic inches of volume holds about as many books as a backpack with only 1586 cubic inches should also be noted.

4 Recommendations for which backpack to buy might mention JC, because it holds the largest number of books and is only $35. MR is also a reasonable recommendation since it has a volume of 1950 cubic inches and costs $35. Consider any recommendation that is well supported.

5 Comments on important variables missing very open to interpretation. Some ideas include number of compartments in the backpack, some kind of measurement about strap cushioning, or simply the height/width/depth measurements of the main compartment. Clearly, there are many acceptable ideas here. Even number of colors the backpack is offered in might be of interest to some people.

Scoring: 5 points for each of the three graphs (make sure axes are labeled, but donÕt be concerned with which variable is on the x-axis and which is on the y-axis), 7 points for backpack recommendation (make sure there is some explaination here), 7 points for missing variable ideas (at least one idea should be given), and 7 points for each of points 1-3 above concerning the graphs. If one of these three ideas is missing but others are present, use your best judgement. This question is obviously very open to interpretation. Maximum: 50 points on graphs and interpretations.

Question 2: Each worth 5 points

Outlier: A point that differs significantly from a majority of the data. Outliers should be investigated further to see why they are so different. Was it a result of recording error or equipment failure? Did the unusual result actually occur. Legitimate points cannot be removed from a dataset for convenience.

Skewed distribution: A histogram of the data shows a shape that is not symmetrical around a mean. Either more than half of the data values fall above, or more than half

the data values fall below the mean. In cases where a distribution is skewed, a median gives a better picture of the center of the data.

Sampling Errors: Òerrors that arise from the decision to take a sample rather than trying to survey the whole populationÓ (a census). Note: this is a line taken directly from the text book. If they have this answer word for word- only give them 3 points of the 5 possible. The question specifically asks them to discuss in their own words.

Non-Sampling Errors:Óerrors that would be present even in a censusÓ - again, only 3 of 5 points if word for word, this is directly from the book.

2 survey examples(5 points total, 3 if only one example given): Examples of a survey should include mention of the population being sampled, the way the sample was collected, and what the study was trying to find out.

2 designed experiment examples(5 points total, 3 if 1): Examples should make clear what the experimentÕs subject is (a person, a board of wood...), what the treatment is, what the response is, and a brief description of how the experiment is organized (is there blocking, blinding, how many subjects are being studied... something, just about anything)

2 observational study examples(5 points total, 3 if 1): Examples should include the subject, treatment, response, and some explaination of how the data was collected (was it from a computer database already in existence, were people collected and studied for pain tolerance (or some other quality) like in the first homework.

The main difference between an observational study and an experiment is whether or not the experimenter has control over which subject receives which treatment. It should be clear in their examples that this difference is understood.

.
.
.
.