Handout for the Spotlight Presentation in Statistics Education Research Area

An Investigation of Students' Knowledge Retention of Statistical Concepts Using Problem-Posing Methodology

Carl Lee
Central Michigan University
Mt. Pleasant, MI 48859
EMAIL: Carl.lee@cmich.edu

This presentation proposes to investigate how well students retain the knowledge of statistical concepts one semester after taking the introductory statistics using a problem-posing and interview methodology.

Students who took their introductory statistics course in the Fall of 2004 were the target population of our study. It is often to hear comments from colleagues who teach higher level courses that require introductory statistics as their pre-requisite:

" My students just completed Introductory Statistics last semester, but they do not remember any thing about statistics. They don't even have any clue about very basic statistics such as median, standard deviation, confidence interval".

We design a study to investigate, if any, what statistical concepts students retain and why they can/can't retain basic statistical concepts.

From the target population about 400 students taught by seven instructors, who all have been teaching introductory statistics for many years at Central Michigan University. Students are mostly majoring in College of Business Administration. About 45 percents are mail and 55 percent are female. Their grade levels range from freshman to senior with majority being junior.

The subjects are selected based on the following design. Six students are selected from each instructor's class. Two students are randomly selected from the A and A- level students, two from B level and 2 from C or lower. A total of 42 students are selected and are given a set of five different problem scenarios. The sampling design is a stratified random sampling design. This random sample represents students from different instructors and from different grades. It allows us to make some comparison between the level of problems they pose among different levels of grades (from A to C or lower). Students are then asked to pose any question that is related to the statistical topics covered by the problem scenario. Each student is given the same instruction before posing questions:

"Now, thinking that you are a professor, and you are going to prepare questions to test your students. Here are five different problem scenarios. Each problem scenario covers a variety of statistical concepts and topics you learned from last semester. You may pose as many questions as you can and pose some simple and some difficult and some challenging questions to test your students. You have a total about 25 to 30 minutes to pose your questions. Once you complete your problem posing for one scenario, move on to next".

Most students completed their problem posing within 25 minutes. An interview is then conducted with each student to discuss the type of questions they pose, the purpose of posing each question, and what is the correct answer for each question they posed.

The interview lasted for about 30 to 40 minutes. The interviews were completed recently. We begin to edit the video clips and to record their posed questions. The following are some examples students posed for each problem scenario.

(A) Probability/Conditional Probability
The following is a table that gives the preference of 1000 students for buying brand name shoes:

	Brand Name				
Gender	Nike-Sports	Nike-Nonsport	Adidas-Sports	Adidas-Nonsports	Total
Female	150	180	150	50	500
Male	210	60	100	100	500
Total	360	240	250	150	1000

One student is chosen at random from this group of 1000 students. Pose as many questions as you can to test different concepts related to probability and conditional probability.

Here is an example of questions posed by a "D" student and an "A" student:

A 'D' student posed the following questions:

(1) What brand to you wear?
(2) How does your choice relate to overall?
(3) Is your brand prefered over gender?
(4) How does your brand compare to opposite gender same brand?
(5) Does the gender affect brand or vice versa?

An 'A' student posed the following questions:
(1) Out of the 1000 students, what is the probability that a student is wearing Adidas sports shoes?
(2) What is the probability that a male student is wearing Nike-Nonsport shoes?
(3) What is the probability of finding a female wearing Adidas-nonsport shoes?
(4) What is the probability that if someone is wearing Adidas-sports shoes then the person is male?
(5) What is the probability that if someone is male then he is wearing Adidas-nonsports shoes?

The "D" student did not seem to have any idea about any statistical concepts related to this problem. The student basically posed very descriptive questions about opinions. No question is related to probability concepts.

The "A" students, although was able to pose questions about probability of an event (Q1) , conditional probability (Q4 and Q5), the questions (2) and (3) are confusing. For example, Q2 can be interpreted as probability of intersection:

" probability that the student is a male and wearing Nike-Nonsport"

or a conditional probability

"If the student is a male, find the probability this male wearing Nike-Nonsport shoses".

It is noticed that similar language has been used by some other students as well.
When asking them what statistical concepts they wanted to test when they first read the problem scenario, no student were able to describe in terms of the statistical terminology such as "testing students if they can find the probability of intersection or union of two events".

When carefully examining the questions posed, it becomes clear that students could only pose procedural questions or one-step solution questions. We rarely notice that students could pose questions that are conceptual or require a good understanding of more than one concepts. No student posed any questions related to, for example, mutually exclusive and sample space. Only a few could pose questions related to intersection, union or conditional probability. Even so, many of such questions were confusing similar to the "A" student described above.

We are beginning to analyze the questions and try to make sense out of them. The following is the five scenarios used in this study.

Problem Posing Interview Protocol

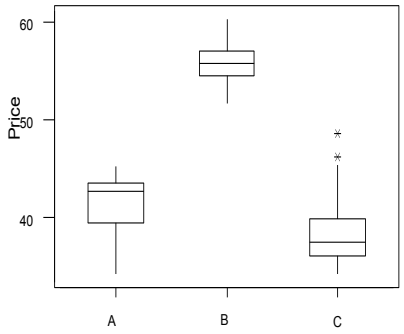
Imagine you are a statistics instructor. You need to prepare for test questions to test if students have learned the knowledge and skills you want them to learn and can challenge the best students in your class.

(A) Probability/Conditional Probability
The following is a table that gives the preference of 1000 students for buying brand name shoes:

	Brand Name				
Gender	Nike-Sports	Nike-Nonsport	Adidas-Sports	Adidas-Nonsports	Total
Female	150	180	150	50	500
Male	210	60	100	100	500
Total	360	240	250	150	1000

One student is chosen at random from this group of 1000 students. Pose as many questions as you can to test different concepts related to probability and conditional probability.

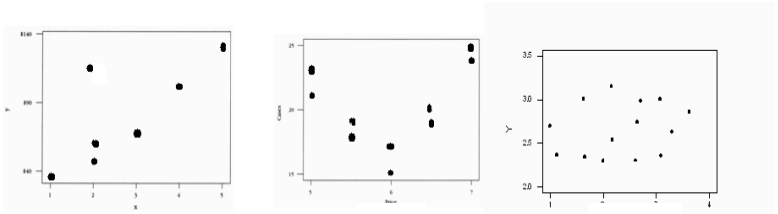
(B) The following summaries are from 40 days of three stocks. In the box plots, they are labeled as A, B and C. In the Descriptive Summary, they are labeled as X, Y and Z. You may pose questions related descriptive statistics, estimation as well as hypothesis testing.
(Note: A,B,C do not necessarily match the order of X,Y,Z).

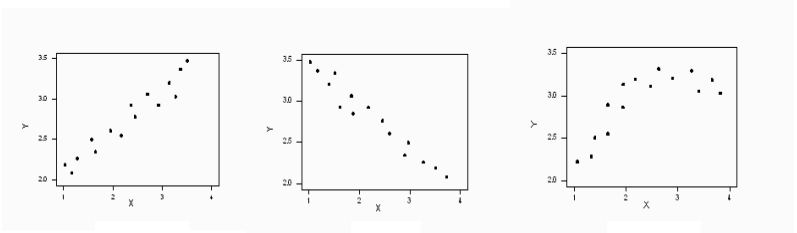


Variable	N	Mean	Median	StDev	Q1	Q3
Stock X	40	41.7	_____	2.86	39.5	43.6
Stock Y	40	55.9	54.8	1.89	54.6	57.2
Stock Z	40	38.4	_____	_____	36.0	39.9

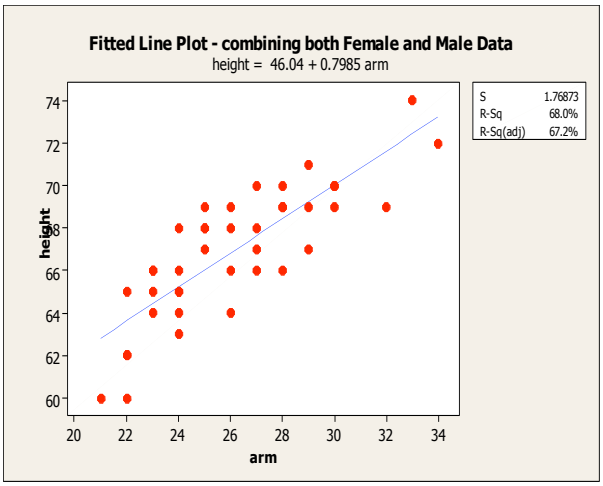
(C) The saving accounts in a large city are generally not normally distributed, instead, they are very skewed to the right, that is many accounts have small savings except a few. Suppose the saving accounts for this city have the average $\mu = \$3000$ and s.d. $\sigma = \$3000$. Bank A randomly takes 5 accounts daily and computes the average of the five for 500 days. Bank B randomly takes 50 accounts daily and computes the average of the 50 for 500 days. Use this information to pose questions related to distribution and sampling distribution of sample mean.

(D) On the right are six patterns of relationships between two variables. Use these plots to pose questions





(E) In a statistics class, students were asked to conduct a project to study if arm length can predict the height. 40 students were randomly selected and their arm lengths and heights were recorded.
The computer output is given below.



Regression Analysis: height versus arm length

The regression equation is:

height = 46.04 + 0.7985 arm

Predictor	Coef	SE Coef	T	P
Constant	46.040	2.338	19.69	0.000
arm	0.79849	0.08884	_____	_____

S = 1.76873 R-Sq = 68.0% R-Sq(adj) = 67.2%