

## Statistics

### Essays

A. The computer manufacturing industry in a country produces basically two classes of computers. An annual price survey of the industry includes two models of each class (see table below). *Model 1a* was discontinued after 2002, and it was replaced in the survey sample by a slightly upgraded version, called *Model 1a+*. The results of the survey from 2000 to 2002 are as follows:

<i>Price (US\$) of surveyed computer models</i>			
	2000	2001	2002
<b>Computers of class 1</b>			
Model 1a	1,500	1,450	
Model 1a+			1,350
Model 1b	1,200	1,180	1,180
<b>Computers of class 2</b>			
Model 2a	425	400	410
Model 2b	540	550	550
<i>Estimated number of computers sold (in thousands)</i>			
	2000	2001	2002
<b>Computers of class 1</b>			
Model 1a	20	35	
Model 1a+			40
Model 1b	10	15	15
<b>Computers of class 2</b>			
Model 2a	360	350	370
Model 2b	140	140	130

<i>Characteristics of computer models of class 1 included in the survey</i>			
	Model 1a	Model 1a+	Model 1b
Hard Disk	40GB	80GB	80GB
Mflops*	201	201	185

\* Millions of floating point operations per second

- 1) The chained Laspeyres and Paasche price index number formula are (where  $p$  stands for price and  $q$  for quantity):

$$P_{t-1,t}(q_{t-1}) = \frac{\sum p_t q_{t-1}}{\sum p_{t-1} q_{t-1}} \quad \text{Laspeyres} \qquad P_{t-1,t}(q_t) = \frac{\sum p_t q_t}{\sum p_{t-1} q_t} \quad \text{Paasche}$$

- (a) Calculate the chained Laspeyres and Paasche price indices for each computer class (using always the previous period as base year).

- (b) Explain assumptions you make to obtain your results, especially regarding the replacement of *Model 1a* with *Model 1a+*.
- 2) The national statistical agency also collects information on production of each computer class on an ongoing basis:

<i>Production shares</i>			
	2000	2001	2002
Computers of class 1	0.15	0.25	0.23
Computers of class 2	0.85	0.75	0.77

- (a) Obtain the aggregate chained Laspeyres and Paasche price indices for computers (i.e. general price indices for the whole computer industry, setting the previous period as the base year for the current price index).
- (b) According to each of the two chained index formulas, how much has the general level of prices for computers increased in 2002 as compared to 2000?
- (c) How can any difference between the results obtained using Laspeyres and Paasche formulas be explained?
- 3) Suppose you have to interpret a time series of production or sales numbers of computers. Why would a corresponding time series of price indexes be useful? Please explain.

B. You have been asked to construct a questionnaire for a survey on housing. The survey will be administered door-to-door by interviewers. Your colleagues remind you that respondents often fail to understand the questions the researcher is trying to ask.

- 1) What are five common causes of this failure to communicate to the respondent (limit your answer to five causes)? What can you do to avoid making these mistakes?

In designing the sample for the housing survey you are asked to make sure it is representative of the population. One of your colleagues suggests that you use clusters in your sample design whilst another suggests that you should stratify your sample.

- 2) What effect does clustering and stratification for this sample have on, for instance, the standard error? Contrast these methods of sampling with simple random sampling.

### Questions

1. Suppose

$$f_{Y|X}(Y|X) = \begin{cases} 1 & X < Y < X + 1 \\ 0 & \text{otherwise} \end{cases}$$

and that  $X$  has the uniform (0,1) distribution

- (a) Find  $E(Y)$
- (b) Find  $f_{X|Y}(X|Y)$
- (c) Find  $P(X + Y < 1)$
2. Define the mean, mode and median and show by diagram how they are normally related in a distribution which is skewed to the right.
3. A bowl contains four lottery tickets with the numbers 111, 221, 212, and 122. One ticket is drawn at random from the bowl and  $A_i$  is the event that 2 is in the  $i$ th place,  $i = 1, 2, 3$ .

Determine whether  $A_1$ ,  $A_2$ , and  $A_3$  are independent.

4. Describe two ways of measuring growth rates of time series and comment on the advantages and disadvantages of each.

5. Distinguish clearly between systematic errors and random errors in a sample inquiry.

6. In a college class, there are 80 men and 40 women. There are 36 smokers and 84 non-smokers. What is the likelihood that any student selected at random is a male smoker under the following conditions:

a. If the percentage of smokers among men is the same as the percentage of smokers among women (i.e. no relationship between sex and smoking)?

b. If the percentage of smokers among men is twice the percentage of smokers among women (i.e. men are twice as likely to smoke as women)?

7. A national service requires all applicants to undergo a drug screening examination. In a batch of 25 applicants one fails the test. The applicant claims never to have used any drugs. It is known that the equipment used in the test has been shown empirically to be able to detect the presence of drugs 99% of the time when the drugs are present. The manufacturer also admits that 5% of the time the equipment detects drugs even when they are not there. It is also known that at least 90% of those who enlisted in the armed forces of the country have not used drugs. What, if anything, can be said about the claims of the applicant?

8. The table below shows the number of "excellent" students (receiving A scores in their exam) and the number of "mediocre" students (receiving C scores in their exam) according to three different teachers, Mr. X, Mr. Y and Mr. Z, that have been asked to mark the same exam.

Test the hypothesis that the number of "mediocre" students is the same for each teacher (Hint: test the hypothesis that being "excellent" or "mediocre" is independent by the teacher who marked the exam).

	Mr. X	Mr. Y	Mr. Z	Total
Excellent (A)	50	47	56	153
Mediocre (C)	5	14	8	27
Total	55	61	64	180

9. Of the 50 people on a bus,  $1/5$  opposes the smoking ban on the bus. If 15 of these passengers are selected at random, without replacement, (a) what is the probability that exactly 5 of them oppose the smoking ban and (b) what is the expected number of passengers in the sample that opposes the smoking ban?

#### List of topics

Suggested reading includes basic textbooks or manuals in the field of statistics covering the following topics:

- Probability Theory
- Descriptive Statistics
- Inferential Statistics
- Demographic and Social Statistics
- Economic Statistics