THE OPEN UNIVERSITY OF SRI LANKA
B.Sc./ B. Ed. DEGREE, CONTINUING EDUCATION PROGRAMME
OPEN BOOK TEST 2006/2007
PSU 1182/PSE 3182/PSZ 4182 – BIO STATISTICS

DURATION: ONE AND HALF HOURS

Date: 01 – 08 – 2006	Time: 3.30pm – 5.00pm

ANSWER ALL QUESTIONS.

Non-programmable calculators are permitted.

1. A manufacturer producing mixed fruit drink is interested in finding out the mixing proportions of the ingredients yielding a fruit drink that is most appealing to the majority of the consumers. The fruits used as ingredients are pineapple, passion fruit and banana. The other ingredients are sugar and citric acid. The manufacturer is particularly interested in comparing the attitudes of consumers regarding the taste when pineapple, passion fruit and banana are mixed in the proportions 1:2:1 and 1:2:2. For both mixtures, the manufacturer wishes to add 2 teaspoons of citric acid per two litres of mixed fruit juice. However, the manufacturer suspects that the amount of sugar that need to be added may depend on the mixing proportions. He wishes to study three sugar levels that are .1mg, 2mg and 3mg per two litres of mixed fruit juice.

   The manufacturer has resources to collect information from 200 adults and 100 children who are possible future consumers of the product.

   Suppose the manufacturer wishes to employ you as a statistician to help in designing his study.

   i) Suggest a suitable design for this study. If you use the random number table, clearly describe how you read the values.

   ii) In relation to this study, explain the following terms.
   a) interaction
   b) random variation

2. State whether the following statements are true or false. In each case, if the statement is false, explain why it is false.

   i) In cluster sampling, each experimental unit in the population has a chance of being included in the sample.

   ii) In stratified sampling, each experimental unit in the population has the same chance of being included in the sample.

   iii) Non sampling errors can always be reduced by increasing the sample size.

   iv) Missing values in data sets always lead to non-response bias.

   v) Quota sampling can produce non-representative samples.

3. A researcher is interested in finding out whether exposure to UV rays can cause memory loss in humans. He selected 50 construction workers who are mostly working in the sun through out the day and another 50 clerks who rarely get exposure to sunlight during their work. The researcher carried out the experiment as follows. Each individual was showed how a simple task is carried out and was asked to repeat what they had observed. The researcher then recorded whether the person was able to correctly repeat the task he had observed or whether he failed to repeat. The researcher then intended to compare whether the proportions in each group who completed the task correctly are the same for both groups construction workers and clerks.

   i) Describe the population and state whether the population is finite or infinite.

   ii) Classify the data collected in this experiment as qualitative or quantitative.

   iii) Classify the data collected in this experiment as nominal, ordinal, interval or ratio.

   iv) State whether the research study is observational or experimental. Give reasons for your answer.

   v) Describe the following terms in relation to this study.
      a) bias
      b) confounding
      c) replicate

2006/2007
OBT

3

PSU 1182/ PSE 3182/ PSZ 4182 Bio Statistics
Open Book Test 2006/2007
Solutions

1. The researcher is interested in comparing two fruit drinks produced from pineapple, passion fruit and banana with mixing proportions 1:2:1 and 1:2:2. For both fruit drinks, the amount of citric acid is kept at 2 tea spoons per two liters of the drink. However, he is interested in studying three sugar levels 1mg, 2mg and 3mg. Hence the factors of interest are the fruit proportions having two levels (1:2:1 and 1:2:2) and the amount of sugar having three levels.

   The researcher has resources to collect information from 200 adults and 100 children. We need to address two things, the selection of individuals for the study and how we should run the experiment with the selected individuals.

   The population of adults and population of children are both large and the individuals are geographically quite dispersed. Hence, a non-probability sampling procedure such as convenience sampling can be used to select the individuals for the study.

   Note: In the solutions to this question, some had proposed to use a systematic sampling procedure for selecting individuals for the study. If the researcher is planning to collect information from adults attending chosen super markets and from some selected schools depending on the convenience of the researcher, the systematic sampling procedure appears suitable.

   Now we need to discuss how to allocate the treatments among the selected individuals. In other words, we need to clearly describe which fruit mixture need to be given to each individual and the amount of sugar to be added.

   Since manufacturer is planning to use the same amount of citric acid, the preferred amount of citric acid is not a factor of interest for this study. Hence the researcher is interested only in two factors. One factor is the fruit proportion having two levels 1:2:1 and 1:2:2. The other factor of interest is the amount of sugar having three levels 1mg, 2mg and 3mg.

   The adults are likely to be different from children with respect to the taste preferences. It is therefore advisable to study the preferences for each fruit drink separately with adults and with children.

   The number of treatment combinations is 2x3, which is 6. If we label the two proportions as P_1 and P_2 and the three sugar levels as S_1, S_2 and S_3, the six combinations can e labeled as $P_1S_1, P_1S_2, P_1S_3, P_2S_1, P_2S_2$ and P_2S_3.

Six different drinks are to be compared. Therefore, it is not advisable to use the same individuals to try all six combinations. Instead, we divide the adults as well as children randomly into six groups each so that each group will get only one drink for tasting.

Since there are 200 adults, each combination can be tried with roughly 200/6, or roughly 33 adults and the remaining two adults can be allocated as extra for any two desired combinations. This gives the following allocation of adults for each treatment combination.

$$P_1S_1 = 34, P_1S_2 = 33, P_1S_3 = 33$$
$$P_2S_1 = 34, P_2S_2 = 33, P_2S_3 = 33.$$

Note: Here we argued that it is not much practical to get each adult to try all six combinations. However, if each adult was to try all six combinations we will have 6x200 responses as opposed to 200 responses that we get in the proposed design.

Likewise each child can be asked to taste one combination and the children can be allocated as 100/6, or roughly 16 per each treatment combination. With 16 replicates for each combination we will have 16x6 or 96 children and the balance four can be assigned to any four treatment combinations. This gives the following allocation of children for each treatment combination.

$$P_1S_1 = 17, P_1S_2 = 17, P_1S_3 = 16$$
$$P_2S_1 = 17, P_2S_2 = 17, P_2S_3 = 16.$$

The following procedure can be used to randomly allocate adults. A similar procedure can be used to allocate children for the different treatment combinations.

- Label the adults as $1, 2, \cdots, 200$.
- Select 34 random numbers from 1 to 200 to decide the labels for the combination P_1S_1. Since 200 is a three digit number, we read three digit values from the random number table.
- In reading numbers, we ignore the values greater than 200 and also values that are already chosen.
- We can start reading values from any row and any column. Starting from the first row and the first three columns we select the adults labeled 135,198,60,169,8,50,126,77,160,81,182,83,160,41,21,61,38,64,161,19,126 etc. Here we have continued with row 1 and columns 2nd, 3rd and 4th once we have reached row 40 with columns 1st, 2nd and 3rd.
- When we read values for the second combination say for P_1S_2, we repeat the same but in addition ignoring all the values that are already chosen for the previous combination.

For each chosen individual, we can offer the respective fruit drink and ask them to indicate their preference on a pre-determined scale such as 1 – strongly dislike, 2 – somewhat dislike, 3 – just right, 4 – somewhat tasty, and 5 – very tasty.

ii) a) Interaction
If the preference for fruit proportion depends on the amount if sugar we add, we say that there is an interaction between the two factors, the proportion of fruits and the amount of sugar.

b) Random variation
Random variation is the uncertainty in the observations that we can not explain from the study. For instance, the differences in responses of individuals receiving the same fruit drink can not be explained from the design used in this study. Hence, we refer to such variations as random variations.

2. i) True

ii) False
This statement is not always true. For instance, in stratified sampling, if we make the chance of being selected to depend on the stratum size, this statement will be false in instances of different strata sixes.

iii) False
For instance, if there is an error in the scale we use to measure the yield that will cause a non-sampling error that does not reduce with increasing sample size.

iv) False
If the values are missing in random, the resulting estimate is not biased simply because some observations are missing.

v) True

3. i) The researcher is interested in studying about all humans. Hence, the population consists of all humans including who are not even born at the time of data collection. Hence, the population is infinite.

ii) The data collected are simply information as to whether the person is able to repeat the task correctly or not. This only deals with quality and hence the data are qualitative.

iii) The data collected in this experiment is nominal.

iv) The researcher has no control over the factors he is studying. Hence the study is observational.

v) a) bias

Clerks and construction workers are usually subjected to different stress levels that may cause memory loss. If the clerks chosen for the study are all subjected to less stress levels a comparison simply using such experimental units can lead to a bias.

b) Confounding

Confounding is the mixing up of effects of factors. Here the experimenter can not separate the effects of stress levels on memory loss from the effect of exposure to the sunlight. We refer to this as confounding.

c) Replicate

Replicates are different observations collected under the same experimental conditions. Here, observations collected on memory loss of clerks are replicates. Similarly observations collected on construction workers are also replicates.