

Statistics 2017

Homework Assignment 1

Due: 1pm, Tuesday 14 February.

Question 1. A swimming coach thinks that a particular technique may improve the times of his swimmers. As an experiment, he offers extra daily practice to work on this technique for swimmers who want to participate. At the end of the season, the coach concludes that the technique is effective, because swimmers who participated in the extra practice have faster average time than those who did not. Explain why this conclusion may not be valid.

Question 2. A school district plans to survey 1000 out of 50,000 parents with enrolled children regarding their preferences on enrolment. A complete alphabetical list of parent names is available. In each of the following, name the sampling method used.

- (a) One of the first 50 names on the complete list is randomly chosen; that person and every 50th person on the list after that person are surveyed.
- (b) The complete list is divided into separate lists by their oldest child's year in school. Random numbers are assigned to names, and each list is ordered by the assigned random numbers. The first 2% of the ordered names in each list are surveyed.
- (c) A random number is assigned to each name using a random number generator. Names are ordered by the random numbers and the first 1000 are surveyed.

Question 3. To determine whether the artificial sweetener aspartame causes headaches, researchers gave capsules containing aspartame or placebo to subjects and observed their responses. After a period which allowed the subjects to rid their bodies of chemicals, those originally given placebo were given aspartame and vice versa. Similar rates of headaches were reported for both groups.

- (a) What are the control and treatment groups in this study?
- (b) This is an example of a *cross-over* design, in which each person is assigned both treatments in random order. What advantages does this design have over a study in which people receive only one treatment?
- (c) If the study noticed, among other things, that one of the treatment groups suffered higher than average rates of stomach upsets, would the researchers be justified in concluding that aspartame can cause stomach upsets?

Question 4. Three different techniques for making bread are tested to determine if one technique is faster. Some dough is divided into 24 equal portions, and each of 4 workers receives 6 portions.

- (a) Identify the treatment factor and the noise factor.

- (b) Construct a randomised block design.
- (c) Suppose the dough is divided into 20 portions instead. Come up with a suitable randomised block design.

Question 5. Body temperatures (in °F) were measured on 30 healthy individuals, resulting in the following data:

97.0, 97.2, 97.3, 97.6, 97.6, 97.7, 97.9, 98.2, 98.2, 98.4,
 98.4, 98.5, 98.6, 98.6, 98.6, 98.6, 98.6, 98.7, 98.8, 98.9,
 99.0, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.5, 99.7, 99.8.

- (a) Calculate the mean, median, standard deviation, and quartiles.
- (b) Construct a box plot. Are there any outliers?

Question 6. The monthly average unleaded gasoline prices per gallon, for the 55 largest cities in the U.S., from Jan 1996 to Dec 1997 are given below.

112.9, 112.4, 116.2, 125.1, 132.3, 129.9, 127.2, 124.0, 123.4, 122.7, 125.0, 126.0,
 126.1, 125.5, 123.5, 123.1, 122.6, 122.9, 120.5, 125.3, 127.7, 124.2, 121.3, 117.7

- (a) Make a time-series plot of the data.
- (b) Calculate the EWMA and the associated MAPE using 3 different parameters: $\alpha = 0.3, 0.4, 0.5$. Which α is the best?

Question 7. We wish to conduct a survey on a sensitive issue, namely to find the proportion x of people who regularly chew gum. n people are randomly chosen. Each is asked to toss a biased coin. If the coins shows H (which occurs with probability q), the person is asked to answer Question A; otherwise s/he is to answer Question B.

Question A: Do you regularly chew gum?

Question B: Is your birthday within 2 weeks of Christmas?

Suppose proportion p of the responses are 'Yes'.

- (a) Estimate x ; state any assumptions.
- (b) Is it possible for your estimate to lie outside the range $[0, 1]$?
- (c) Are there any unhelpful values of q ?

Question 8. A *population* consists of $N = 4$ numbers: 1, 2, 3, 4. Each number has equal chance of being selected.

- (a) Calculate the population variance σ^2 .
- (b) A random sample of size $n = 2$ is selected *with* replacement from the population. Find the sampling distribution for the sample variance (e.g. find all the possible values of the sample variance and the corresponding frequencies).
- (c) Calculate the expected value of the sample variance from (b), and compare it with the population variance obtained in (a).

Question 9. (A twist on the German Tank Problem). Suppose that the enemy has tanks numbered $0, 1, 2, \dots, N$. You observe n of the tank *with* replacement at random and note down their numbers. Using the sample mean of these numbers, find an unbiased estimator for the total number of tanks.