Exercises Sheet 5

Sampling Distributions, Estimation and Hypothesis Testing on One and Two samples

Note: For estimation questions, do not worry about the f.p.c. unless explicitly asked for.
The first few questions are straightforward and designed to reinforce points made in the lecture. You should work on these on your own as far as possible, bringing difficulties only to the tutorial. Needless to say, all questions can and should be attempted on your own anyway, but Tutorial coverage would probably start at Q.6

1. (a) There are three values in a population, (100, 300 and 500). We are interested in investigating the sampling distribution of the mean ($\bar{x}$) for samples of size 2. How many possible distinct random samples are there? Assume sampling **with replacement**. Find the mean of each.

(b) Assuming each sample is equally likely to occur, give probabilities for results in (a), and hence give the *sampling distribution* of $\bar{x}$. On the basis of this, form a hypothesis about the mean of the population.

(c) If we change the sample size to 3 in (a), how many possible random samples are there now? What is the sampling distribution now?

(d) For the random samples generated in (a), what are the sample variances, s^2 ? Give the sampling distribution of s^2.

(e) For the sampling distributions in (b) and (c), find the probability that the mean is within 75, (either above or below), of 300. Which sample size gives more tightly clustered x values; (better precision in statistical terms)?

Note: samples here are, of course, very small – for purposes of illustration only.

2. (a) If X is Normally distributed with $\mu = 200$ and $\sigma = 50$, find $P\{X > 275\}$

(b) What proportion of the X values in (a) are within (i) 1.96 and (ii) 2.58 standard deviations of the mean?

(c) Use the following sampling distribution of ($\bar{x}$) to determine the proportion of $\bar{x}$ values within 37.5 units of 162.5, (include the boundaries of this interval).

$\bar{x}$	100	125	150	175	200	225
$P\{\bar{x}\}$	0.10	0.15	0.25	0.25	0.15	0.10

3. (a) Find the mean of the sampling distribution of the means in Question 1, part(b). Is this a *biased* or *unbiased estimate* of the population mean?

(b) Find the standard error (S.E.) of the sampling distribution of the means in the Question 1. How does this compare with the standard deviation of the population in that problem?

(c) As the sample size, n, increases do you expect the resulting $\bar{x}$ values to be closer to or further from the population mean μ ? In other words, what happens to the S.E.?

4. (a) A researcher selects a random sample of 400 substance weights from the population of weights on a long-running experiment. The population mean is 485 g. and its S.D. is 80 g. Find $P\{\bar{x} > 500\}$, i.e. prob. that a mean from the sampling distribution of the means is greater than 500.

(b) If you are told that the underlying observations (X = scores) are Normally distributed and a weight recorded is selected at random, what is the probability that it is greater than 500? (i.e. looking now for $P\{X > 500\}$.

(c) Comparing your answers to (a) and (b), comment.

5. (a) For Question 1, part (a), what difference would sampling *without replacement* make, i.e. from a *finite* population?

(b) What is the S.E. of the mean for this new sampling distribution?
(c) Compare the S.E.'s obtained here and in Question 3, part (b) and comment.

6. (a) A researcher needs estimates of the mean time required to complete certain processes in the lab. in order to ascertain hours of lab. time required. Randomly selecting 40 records on each of three processes, (s)he obtains a mean time of 12 hours for each. From past experience the S.D.'s for the three sections are taken to be 2hrs. , 4hrs and 6 hrs. respectively. Find a 90% confidence interval for the mean time for each section.
(b) Which of the samples in (a) is likely to provide the most accurate information about its population mean? Justify your answer.
(c) How would your confidence intervals be affected if you
(i) increased the sample size? (ii) wanted a higher level of confidence?

7. In 500 yields of a given experiment, percentage of "successes" (i.e. satisfactory outcome) given by
(a) 64% method A
(b) 83% method B
(c) 83% method A + extended heating period
(d) 51% method A + reduced heating period
(e) 49% method C

Find and interpret a 98% confidence interval for the proportion of successes for each estimate. Can we be 98% confident that over 50% of the population are successes for each of the above?

8. (a) A new piece of equipment is suspected of having faulty temperature control. The desired temperature is supposed to be 68 degrees and a random sample of 45 readings are collected. The mean of these is 69.178 with a S.E. of 0.483. Formulate and test a suitable hypothesis.
(b) If a sample of only 20 readings had given the same mean and S.E., how would your calculation have been affected?
(c) If you were only concerned about temperatures being too low, how would your calculation be affected?

9. (a) A lab. manager claims that faulty readings on a piece of equipment are generated about 10% of the time, but a researcher believes it to be higher. A random sample of 400 records contain 60 errors. What evidence do you have at the 0.05 level of significance to support the lab. manager's statement or otherwise?
(b) A further random sample of the same size, revealed 45 errors. What would your conclusions have been on the basis of this sample?
(c) Is there a significant difference between the two results obtained?

10. To determine if a new process will improve productivity, two labs are randomly selected, one to use the new, the other to maintain the current approach. Sample sizes are 36 working days for each lab. The sample mean of work accomplished in a given period in the first sample is 17.83 tasks and the S.D. is 3.12 tasks. The sample mean in the second case is 20.1 tasks, with an S.D. of 5.47 tasks. Formulate and test a suitable hypothesis.
What would change if the samples had been smaller, say 15 ?

11. Eight randomly selected cultures are screened by Method A and also by Method B, which is supposedly better. Scores are assigned for both and are given below. Formulate and test a suitable hypothesis. Comment on any assumptions made.

Culture		1	2	3	4	5	6	7	8
Scores	Method A	25	32	16	17	16	20	25	24
	Method B	30	35	30	17	22	19	29	29

Note: Samples here are small for a t-test. For purpose of illustration only.