STATISTICS FOR ENGINEERS

Exercise Sheet 4

Hand in solutions to the two starred questions.

1. (a) The sample $x_1, x_2, \ldots, x_n$ has sample mean $\bar{x}$ and sample variance s_x^2. Show that if $y_i = ax_i + b$, where a and b are constants, then the sample mean and variance are $\bar{y} = a\bar{x} + b$ and $s_y^2 = a^2 s_x^2$ respectively.

(b) It is required to find the sample mean and variance of the following set of numbers:

1878905, 1878901, 1878906, 1878904, 1878906, 1878909

Show how you could use the results in part (a), with a = 1 and b = 1878900, to evaluate the sample mean and variance quite easily.

2. The synchronization time, T seconds, of a certain gearbox is the time for the gears to become fully engaged following the gear level being pushed into gear. The probability density function of T is $f(t)=1/\theta$ for $0.03<t<0.03+\theta$ and $f(t)=0$ for other values of t, where $\theta>0$ i.e. T only takes values between these two limits and any value in this range is as likely as any other value (a *Uniform* distribution).

(i) Sketch a graph of $f(t)$ against t and confirm that $f(t)$ is a probability density function (i.e. is always non-negative and integrates to one). Find the mean and variance of T.

(ii) A random sample of observations $T_1, T_2, \ldots, T_n$ have been made at these times and $\overline{T} = \sum_i T_i / n$ is the sample mean. Show that $\hat{\theta} = 2\overline{T} - 0.06$ is an unbiased estimator of θ.

(iii) Suppose that the observed times, in seconds, are 0.054, 0.035, 0.034 and 0.041. Evaluate $\hat{\theta}$. Say, with reasons, whether this estimate appears reasonable.

3*. A soft drinks machine fills cups with volumes that are known to be independently Normally distributed with standard deviation σ = 2.5 ml. A random sample of 6 cups was found to contain the following volumes, in ml, of soft drinks:

224.6, 225.1, 222.7, 224.2, 231.3, 226.0.

(i) Find a 95% confidence interval for the mean volume of soft drink in filled cups.

(ii) Suppose now that it is not known that σ = 2.5 ml. Recalculate the 95% confidence interval for the mean volume.

(iii) If the mean volume is to be estimated to within ±1ml with 95% confidence, approximately how many more cups need to be sampled? [Use the estimate of the standard deviation found in part (ii).]

4*. Two different sound insulation systems, A and B, have been designed for an engine. Eighty randomly selected people have been asked to listen to engines insulated with the two systems and report which they preferred. Denote by X the number in the sample who prefer system A. Show that $X/80$ is an unbiased estimator of p, the proportion of all people who prefer A.

It is found that $X = 46$; find a 95% confidence interval for p. Bearing in mind the values included in your confidence interval, would you say that there is very strong evidence that a majority of people prefer A in general?

Approximately how many more people need to be sampled in order for the 95% confidence interval for p to be of width 0.10 (i.e. $\pm$ 0.05)?

5. A polling company questions N random voters about their support for the Winning Party. Let X= the number of people who say they will support the Winning Party.

i. If the Winning Party has a fraction p of population's vote, calculate the standard deviation of the poll estimate, given by $\hat{p} = X/N$.
ii. Calculate the value of p that maximizes the error on $\hat{p}$, and find the largest range of the 95% confidence interval if N=1500.
iii. [*for enthusiasts*] The voters can be divided into various sub-populations that have significantly different voting intentions. For example 50% of voters are men, 50% are women, but they often have significantly different probabilities p_M and p_W of voting for a particular party. Show that if the selection of sexes is completely random then $\hat{p} = \frac{X_w + X_m}{N_w + N_M}$ is an unbiased estimator of p, where X_W and X_m are the responses for men and women, and N_w and N_M are the number of men and women in the sample.
iv. [*for enthusiasts*] Does it matter if the population can be divided into lots of different subgroups with different intentions if the sampling is random?
v. [*for enthusiasts*] An alternative estimator is $\hat{p}' = \frac{1}{2}\left(\frac{X_W}{N_W} + \frac{X_M}{N_M}\right)$. Show this is also unbiased, and discuss why it might be a better estimator.

Hint for iii,v: you may find it helpful to calculate expectation values over P(X) using $P(X) = \sum_N P(X|N)P(N)$.

6. If you haven't done any of the questions on sheet 1 or 2, do them now!