ECE 315 - INTRODUCTION TO STATISTICS - INVEST 21 SAMPLING

WINTER 2004 **A.P. FELZER**

To do "well" on this investigation you must not only get the right answers but must also do neat, complete and concise writeups that make obvious what each problem is, how you're solving the problem and what your answer is. You also need to include drawings of all circuits as well as appropriate graphs and tables.

In our Investigations on probability we always did the random experiments a *whole bunch of times* - enough times to get "exact" values for the probabilities. In the world of statistics, on the other hand, we only do the random experiments *enough times* to get "good estimates" of the probabilities. The objective of this Investigation is to introduce how to go about getting the samples for making the estimates.

1. The first problem we run into in sampling is making sure our samples are *representative* - are really *random* samples from the **population** - that the samples are not **biased**. A classic example of biased samples is the polling that was done in the 1948 presidential race between Truman and Dewey. The mistake made by the pollsters is that they did their polling by telephone during a time when many poorer people didn't have phones. As a result the pollster's results were skewed towards the more affluent who on the whole tended to be Republicans. What problems do pollsters encounter today when they try to do telephone sampling.

2. Once we have n unbiased samples we can calculate the sample mean just like we did in our Investigations on probabilities as follows

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   Find the sample mean $\overline{X}$ for the samples (1, 4, 2, 3, 1, 2)

3. Show that if each sample is randomly chosen from the population then the expectation of the random variable $\overline{X}$ is the same as that of X as follows

$$E\left[\overline{X}\right] = E[X] = \mu$$

4. In addition to being unbiased we also want our samples to be independent. In particular we want to make use of the fact that if $X_1, X_2, \ldots, X_n$ are independent random variables then

$$Var\left[X_1 + X_2 + \cdots + X_n\right] = Var\left[X_1\right] + Var\left[X_2\right] + \cdots + Var\left[X_n\right]$$

   a. Verify that $Var\left[X_1 + X_2\right] = Var\left[X_1\right] + Var\left[X_2\right]$ for two fair and independent coins with random variable X as follows

$$X = \begin{cases} 1 & Heads \\ 0 & Tails \end{cases}$$

   b. Show that if X_1 and X_2 are any two independent random variables then

$$Var\left[X_1 + X_2\right] = Var\left[X_1\right] + Var\left[X_2\right]$$

5. In the world of probability the variance of a random variable X is as follows

$$Var[X] = E\left[(X-\mu)^2\right]$$

which we can calculate from the results of a whole bunch of random experiments as follows

$$Var[X] = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$$

But in the world of statistics, on the other hand, we can only approximate the variance from the sample values as follows

$$Var[X] \approx s^2 = Average\left[(X_i-\bar{X})^2\right]$$

because we don't have μ - we have to approximate it by the sample mean $\bar{X}$. As a result of this approximation it can be shown that for the average of the sample variances to equal σ^2 as follows

$$Average\left[s^2\right] = \sigma^2$$

we need to divide by $n-1$ instead of n as follows

$$s^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$$

when we calculate s^2. Calculate s^2 for the samples (1, 4, 2, 3, 1, 2)

6. A big question in statistics is how large n has to be in order for $\bar{X}$ to be a good estimate of μ. To do this we make use of the variance of the sample means $\bar{X}$.
   a. Write out in words what we mean by the variance of the sample means
   b. Would you expect a sample mean $\bar{X}$ to be a better estimate of μ if the variance of the sample means is large or is small.

7. The objective of this problem is to calculate the variance of the sample mean of a dice
   a. Sketch $f_X(x)$ of a fair dice
   b. Calculate Var[X]
   c. Repeat the following ten times

      (1) Throw the dice 5 times
      (2) Calculate the average of the 5 throws

   d. Make use of your result in part (c) to estimate the variance of $\bar{X}$
   e. Is the variance of $\bar{X}$ less than, more than or the same as the variance of X. Explain why

8. Does increasing n increase, decrease or not affect the variance of $\bar{X}$. Explain why