Chapter 5

Sampling Distributions

5.1 Introduction

In the chapter1 we concentrated on pure description of the data .1-though we recognised that this might prompt us to ask pertinent questions about the *population* from which the sample was drawn. What exactly does the sample, an often tiny subset, tell us about the population? We can never observe the whole population, even if finite, except at enormous expense so the population mean and variance and indeed any aspect of the population distribution can never be known exactly. We call these unknown population quantities parameters and use Greek letters to denote them: μ ('mu') is the symbol commonly used for the population mean and σ ('sigma') for the population standard deviation. As we have a sample of *n* observations we need to ask the question: Is $\bar{x}$ a 'good' *estimate* of μ? How do we *infer* (find something about) the unknown μ using $\bar{x}$ *?* We know that for almost all samples $\bar{x}$ is not equal to μ but how close is it? Can we answer this question without knowing what μ is? Does $\bar{x}$ get closer to μ as n increases? We need to study the properties of the sample mean as an estimator of the population mean and we achieve this by looking at the values $\bar{x}$ can take over all possible samples: the sampling *distribution.* Of course we can never examine all possible samples but the easy availability of a statistical package like MINITAB enables us to study sampling properties much more readily. We can actually illustrate the theoretical results of this chapter by conducting a *simulation* experiment..

Computing for this Chapter: It is essential that you work through the class exercise and if possible conduct the simulations yourself using the journal file.. Once you have understood what is being demonstrated, namely the results of Theorem 1 below and the Central Limit Theorem, then you should try experimenting with different distributions and sample sizes. You may need to adjust downward the number of samples drawn (10000 in the supplied file) to allow for computation time on, say an IBM PC or equivalent.

Once again we give a set of basic definitions which may differ slightly from the standard texts. They are not intended to be learnt by rote but should help to give you a firm foundation for the proper understanding of statistical concepts..

DEFINITION 1: A parameter is a numerical characteristic of the population of interest. Parameters are usually unknown and we make inferences on them
using the sample data.
(**Examples** : p, the probability of 'success' is a parameter of a Binomial population distribution. The rate of 'failure' A is a parameter of a Poisson distribution and also of an exponential distribution of 'lifetimes' i.e. time to 'failure' in a Poisson process where 'failures' occur randomly in time.)

DEFINITION 2 The population mean p ('mu') is a common parameter. If the population is modeled or described by a p.d.f. (probability density function) $f_X(x)$ for a continuous variable *X* then

$$\mu = E(X) = \int x f_X(x)dx$$

If however *X* is a discrete random variable with probability function (or p.d.f.) $p_X(x)$ = *P[X* = x] then

$$\mu = E[X] = \sum x p_X(x) .$$

Other parameters measuring location can be defined in terms of the c.d.f. $F_X(x) = P[X \le x]$ for example the population median *M* and the upper and lower quartiles Q_1 and Q_2 respectively.

DEFINITION 3: The population variance σ^2('sigma-squared') is a common parameter measuring variability:
$\sigma^2 = Var[X] = E[(X-\mu)^2] = E[X.^2] - \mu^2 = \int x^2 f_X(x)dx - \mu^2$ (for X continuous, and similarly in the discrete case.)

DEFINITION 4 :An estimate is a statistic that we hope will be 'near' to the parameter of interest. (For example, $\bar{x}$ is an estimate of μ.)
An estimator is a rule for calculating an estimate from any sample, usually a random sample.

DEFINITION 5: A random variable (r.v.) is a statistic whose value is determined once the sample data have been observed. Thus an estimator is a r.v. but, in general, a r.v. need not estimate anything.

Use upper case letters for r.v.s and the corresponding lower case letter for the values taken by the r.v.s
If X_i is the r.v. denoting the measurement of variable X on unit i in the sample (i = 1,2,... , *n)* then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

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

are the sample mean and variance respectively, considered as r.v.s.

DEFINITI0N 6 The *sampling* distribution of a r.v. is the collection or distribution of all possible values of the r.v. over all possible samples. The properties of the sampling distributions of these two r.v.s determine how we make inferences on the unknown p (and a^2) from any sample
If the sample is a random sample of size *n* from an infinite population then $X_1, X_2, \ldots, X_n$ are independent r.v.s each with the same distribution (i.e. same p.d.f or probability function) as the population so that

$$E[X_i] = \mu \quad \text{and} \quad Var[X_i] = \sigma^2 \qquad (i = 1,2,\ldots,n)$$

The main result of this Chapter is the following theorem:

Theorem 1 Averaging over all random samples of size n from an arbitrary population with mean μ and variance o^2, the sample mean $\bar{X}$ and sample variance s^2 have the following three properties:

$E[\bar{X}] = \mu$ i.e is an unbiased estimator of μ

$Var[\bar{X}] = \dfrac{\sigma^2}{n}$ i.e. the variability of $\bar{X}$ as an estimator decreases with n.

$E[s^2] = \sigma^2$ i.e.s^2 is an unbiased estimate of σ^2.

Thus s^2/n is used as an unbiased estimate of the variability or variance of $\bar{X}$ as an estimator of μ. It is vital in Statistics to have such an estimate so that inferences using Probability can be made.

Example 3: An infinite population is described by an asymmetrical discrete distribution with just two values: —3 with probability 0.3 and +1 with probability 0.7. Thus we have

$$\mu = E[X] = (-3 \times 0.3) + (1 \times 0.7) = -0.2$$
$$\sigma^2 = Var[X] = E[X^2] - \mu^2 = (-3)^2 \times 0.3 + (1)^2 \times 0.7 - (-0.2)^2 = 3.36$$

These are the values of the (usually unknown) population parameters. Let us now look at all samples of size vs =3. There are infinitely many, but we can tabulate them as follows:

Sample observations	$\bar{x}$	s^2	P(sample)
—3, —3, —3	—3	0	$(0.3)^3 = 0.027$
—3, —3,1	—5/3	32/6	$3(0.3)^2 \times 0.7 = 0.189$
—3,4,1	—1/3	32/6	$3(0.7)^2 \times 0.3 = 0.441$
1,1,1	1	0	$(0.7)^2 = 0.343$

Thus we see that 2 as an estimate of p is2.8 below the 'true 'value in 2.7% of samples, 1.2 above in 34.3% of samples etc. Taking the average over all samples or equivalently the expectation over the sampling distribution, we see that

$$E(\bar{X})) = -3\text{x}0.027 + \frac{-5}{3}\text{x O.189} + \frac{-1}{3}\text{ x O.441} + 1\text{ xO.343} = -0.2$$

exactly, confirming the first result of Theorem 1. Then

$$V[\bar{X}] = E[\bar{X}^2] - (E[\bar{X}])^2$$

$$= (-3).^2\ x\ 0.027 + \left(\frac{-5}{3}\right)^2 \times 0.189 + \left(\frac{-1}{3}\right)^2 \times 0.441 + (1)^2 \times 0.343 - (-0.2)^2$$

$$=1.12 = \sigma^2$$

which confirms the second result. The third result is verified for this example by averaging over all possible values of s^2 thus:

$$E(s^2) = 0 \times 0.027 + \frac{32}{6} \times 0.189 + \frac{32}{6} \times 0.441 + 0 \times 0.343 = 3.36 = \sigma^2$$

Before we prove Theorem 1, recall some basic properties of the means and variances of random variables. You should know from chapter 2 and chapter 3

Let a be a constant and let *X* and *Y* be random variables, then

$P1: E[\alpha X] = \alpha E[X]$

$P2: E[X + Y] = E[X] + E[Y]$

$P3: Var[\alpha X] = \alpha^2 Var[X]$

$\text{P4}: \text{Var}[\text{X} + \text{Y}] = \text{Var}[\text{X}] + \text{Var}[\text{Y}]$ if X and Y are *independent*

We do not define independence formally here: see Chapter 2 . However, the n observations from a random sample are independent random variables (with our definition of random sample, only provided the population is large).

Proof of Theorem 1

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$$

$$E[\bar{X}] = .\frac{1}{n}(E[X_1] + \ldots E[X_n]) \qquad (P1)$$

$$= \frac{1}{n}(\mu + \cdots + \mu)$$

$$= \frac{1}{n}(n\mu)$$

$$= \mu$$

$$Var[\overline{X}] = \frac{1}{n^2} Var[X_1 + \cdots + X_n] \qquad (P_3)$$

$$= \frac{1}{n^2} (Var[X_1] + \cdots + Var[X_n]) \qquad (P_4)$$

$$= \frac{1}{n^2} (\sigma^2 + \cdots + \sigma^2)$$

$$= \frac{n\sigma^2}{n^2} = .\frac{\sigma^2}{n}$$

P4 can be used because $X_1, \cdots, X_2$ are independent (random sample)

To show $E[s^2] = \sigma^2$, we first write

$$(n-1)s^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} X._i^2 - n\overline{X}^2$$

using the definition of s^2 and the fundamental identity (p). Also note that fo any random variable Z, say,

$E[Z^2] = Var[Z] + (E[Z])^2$, by definition of $Var[Z]$.

$$E[\sum_{i=1}^{n} X._i^2] = .\sum_{i=1}^{n} X_i^2 \qquad P2$$

$$= \sum_{i=1}^{n} (Var[X_i] + (E[X_i])^2)$$

$$= \sum_{i=1}^{n} (\sigma^2 + \mu^2)$$

$$= n(\sigma^2 + \mu^2) \qquad \text{and}$$

$$E[n\overline{X}^2] = nE[\overline{X}^2] \qquad (P1)$$

$$= n(Var[\overline{X}] + (E[\overline{X}])^2)$$

$$= n\left(\frac{\sigma.^2}{n} + \mu^2\right)$$

$$= \sigma^2 + n\mu^2$$

Combining these two intermediate results, we have

$$E[(n-1)s^2] = n(\sigma^2 + \mu^2) - (\sigma^2 + n\mu^2) = (n-1)\sigma^2$$

Thus by Pl for n >1,E[s^2]= σ^2 (S^2 is not define for n=1.). All parts of the theorem are now proved. You are expected to be familiar with the details of this proof.

The theorem shows that $\sigma/\sqrt{n}$ is the standard deviation of the sampling distribution of $\overline{X}$ as an estimator of μ. That is, measures the variability of possible estimates *2* about the 'true' population mean p. A sample estimate of this variability is $s/\sqrt{n}$

called the *(estimated)* standard error *of* the (sample) mean. MINITAB refers to this statistic as SEMEAN, given by the DESCRIBE command.
As the sample size increases, but with the sample still random, the variability or uncertainty in our estimate of μ decreases monotonically with a limit of zero, i.e. knowledge without uncertainty as n$\rightarrow \infty$.

We can show using some rather difficult results in probability theory that At $\bar{X} \rightarrow \infty \;\; as \;\; n \rightarrow \infty$ with probability 1, with the intuitive interpretation that we are certain to arrive at the 'true' value as the sample size increases indefinitely. This is the Strong Law of Large Numbers and is not discussed further here. Theorem 1 allows us to prove (proof is not required in this course) only that

$$P[|\bar{X} - \mu| < \in] \rightarrow 1 \; as \; n \rightarrow \infty$$

of real numbers giving the probability that $\bar{X}$ is within $\in$ of μ has a limit of unity. This is called the Weak Law of Large Numbers.

Another important result for statistical inference for large sample sizes is the **Central Limit Theorem**, which says that as $n \rightarrow \infty$ the sampling distribution of $\bar{X}$ tends to a Normal distribution with the same mean and variance. We saw this demonstrated *empirically* using MINITAB. The importance of this result is that we do not need to know the form or type of the original population distribution if our sample size is sufficiently large. We can use instead the Normal distribution for statistical inference with the knowledge that the probabilities we calculate will be good approximations to the true (but generally unknown) probabilities. Recall in chapter the Normal approximations to the Binomial and Poisson distributions. See the following chapter for details of the tests of hypotheses and methods of estimation such as confidence intervals. These are the techniques statistical inference applies to real data. Using tic symbol "' to mean 'is distributed as' , we write the Central Limit Theorem

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{approximately for large n.}$$

Then using the properties of the Normal distribution (swe can say that for large n,

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > u\right)$$

can be found approximately (using say, Table 1) for any specified value u *without* knowing the original form of the population.

If, however, we do know the form of the population *and* it follows a Normal distribution, then for any sample size n > 1 it can be shown that

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Thus $$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0;1)$$

has a sampling distribution which is Standard Normal for any n (Table 3).
As the population standard deviation σ is often unknown, replacing it by the corresponding sample quantity *s* changes the sampling distribution. However, provided the underlying population is Normal, it can be shown that

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

has a 'Student's t-distribution' with v 'degrees of freedom', where $v = n - 1$ (named after W. S. Gossett, who took the pseudonym 'Student'). The percentiles of this distribution are given in Table 2; thus $t_v(0.25)$ is the 75th percentile or upper quartile whose value for different ii is given by the second column of figures in the main body of Table 4. These percentiles will be used extensively in Chapter 3 for statistical inference on Normal populations.

Another distribution which arises from random samples of Normal populations is the 'chi-square' distribution, whose percentage points are given in Table 5. It can be shown that

$$V = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

the chi-square distribution with $n - 1$ degrees of freedom, whatever the value of $\overline{X}$. Using some rather tricky distribution theory we can derive the *t*- distribution mentioned in (2.4) above. Yet another distribution is the (Fisher) F-distribution with percentage points in Table 6. The F and χ^2 distributions are used for statistical inference on the variances of Normal populations as well as for wider application in Goodness-of-Fit tests (Chapters 6 and 7).

Note:
Using MINITAB it is possible to check these distributional results empirically by generating a sufficient .number of random samples from a Normal population.

5.1 Exercise

Use appropriate MINITAB commands to generate 1000 random samples of size 16 from an exponential population with mean and variance 1.

Give histograms of the approximate sampling distributions of the sample mean for samples of size 1, 2, 4, 8 and 16. Comment on the approach to Normality shown by these data.
Demonstrate empirically the three results of Theorem 1 by using the commands MEAN and STDEV.