STATISTICAL METHODS IN ECONOMICS-II

LESSON: POINT ESTIMATES FOR POPULATION MEAN, VARIANCE AND PROPORTIONS: SINGLE SAMPLE AND TWO SAMPLES

LESSON DEVELOPER: KAMLESH AGGARWAL AND NIDHI AGGARWAL

COLLEGE/DEPARTMENT: DEPARTMENT OF ECONOMICS, SPM COLLEGE ABD NATA SUNDARI COLLEGE UNIVERSITY OF DELHI

TABLE OF CONTENTS

Reference: *Jay L .Devore : Probability and Statistics for Engineering and the Sciences, 8th Edition.*

POINT ESTIMATES FOR POPULATION MEAN, VARIANCE AND PROPORTIONS: SINGLE SAMPLE AND TWO SAMPLES

Learning Objectives

After completing study of this chapter we will be able to make a reasonably precise inference about the population parameters like mean, variance and proportion on the basis of sample data. We will also be able to make an inference about the difference between the means, variances and proportions of two different population distributions on the basis of samples collected from each of these populations. We will also be able to have an idea about the accuracy of the above estimates.

1.Basic Concepts

Point estimate is a single number determined from a sample and is used to estimate the population value. By implication, the term estimate refers to the actual sample result which is used to represent the parameter being estimated. If the average age based on a random sample of size n = 36 is 65 years, the sample mean $\bar{x}$= 65 years is an estimate of the parameter μ and the statistic $\bar{X}$ its estimator.

Clearly, a point estimate is normally different from the actual value of the parameter for the simple reason that a point estimate is derived from a random sample and the value of the point estimate varies from sample to sample. So while reporting the value of a point estimate, we should also give some indication of its precision or error. The best indicator is standard error of the estimator used. The standard error of an estimator $\hat{\theta}$ is its standard deviation which can be denoted by $\sigma_{\hat{\theta}}$. It is the size of an average deviation between $\hat{\theta}$ and θ. If we use estimated values of some unknown parameters, then we call it estimated standard error and denote it by $\hat{\sigma}_{\hat{\theta}}$ or by $S_{\hat{\theta}}$.

Now we will show the computation of point estimates and their standard error for population mean, variance and proportion for a single sample. We will also extend these computation methods to situations involving the means, proportions and variances of two different population distributions.

2. Point Estimates for Population Mean

We often need some idea about the average value of the relevant population. For example, we might be interested in knowing the average daily sales of soft drinks in Delhi. Similarly, sometimes we want to make an inference about the difference in average values of two different populations. Not only we want to estimate these parameter values but we also like to have an idea about the precision of our estimates. Now we will discuss methods for computing these estimates and their precision.

2.1 Methodology

Sample arithmetic mean, $\bar{X}$; sample median, $\tilde{X}$; sample k% trimmed mean, $\bar{X}_{tr(k)}$ and average of the two extreme observations in the sample, $\bar{X}_e$, can all be used as estimators of population mean μ. However, when there is more than one estimator, the best estimator is the one which gives an estimate closer to the actual value of μ which will depend on the sampling distribution of the estimator. However, the sampling distribution of the estimator itself depends on the distribution of the population from which the sample is drawn. In particular,

1) If we draw a random sample from a normal population, then $\bar{X}$ is the best among the four estimators ($\bar{X}, \tilde{X}, \bar{X}_e$ and $\bar{X}_{tr(k)}$), since its variance is least among all unbiased estimators. An estimator $\hat{\theta}$ is called an unbiased estimator of population parameter θ if $E(\hat{\theta}) = \theta$.

2) If we draw a random sample from a Cauchy distribution,

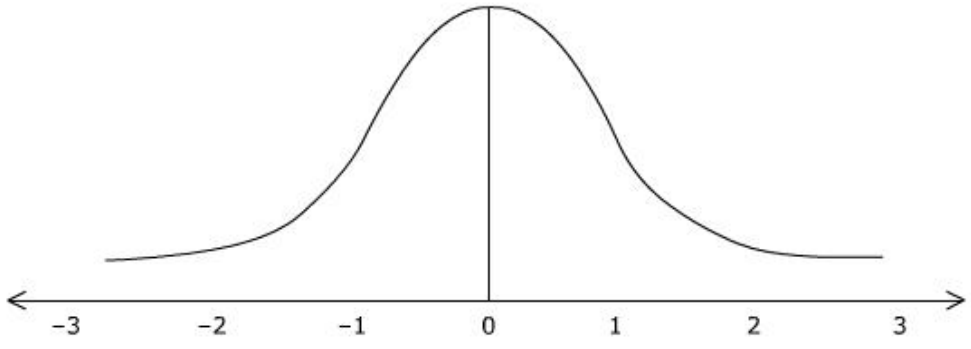


Figure 1 : Cauchy Distribution

then $\bar{X}$ and $\bar{X}_e$ are bad estimators for μ, while $\tilde{X}$ is reasonably good. $\bar{X}$ is bad as it is very sensitive to extreme observations, and due to heavy tails of the Cauchy distribution it is very likely that a few such observations appear in any sample.

3) If we draw a random sample from a uniform distribution, then $\bar{X}_e$ is the best estimator. $\bar{X}_e$ is very sensitive to extreme observations but such observations are unlikely to appear in any sample as uniform distribution does not have any tails.

4) The trimmed mean is not best in any of these three situations. However it is quite good in all three. Hence, $\bar{X}_{tr(k)}$ with small trimming percentage is called a **robust estimator** i.e. one that performs reasonably well for a wide variety of population distributions.

So both i.e. distribution of population and sampling distribution of estimator are important to decide which estimator is best for a given situation. Now we will show some important results assuming that the population is normal.

Let X_1, X_2,-----X_n be a random sample from a normal population with mean μ and variance σ^2, then $\hat{\mu} = \bar{X}$ is the best estimator of μ. It can be shown that the expected value of $\bar{X}$ is μ, so $\bar{X}$ is an unbiased estimator of μ.

Proof: Since the sample mean is defined as

$$\bar{X} = \frac{X_1 + X_2 + \text{------} + X_n}{n}$$

Hence

$$E(\bar{X}) = \frac{1}{n}[E(X_1) + E(X_2) + \cdots + E(X_n)] = \frac{1}{n}(n\mu) \qquad [\text{since } E(X_i) = \mu \text{ for } i = 1,2 \ldots\ldots\ldots n]$$

$$= \mu \text{ as desired.}$$

Further it can be shown that if the value of σ is known, the standard error of $\bar{X}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Proof:

$We\ have \bar{X} = \frac{X_1}{n} + \frac{X_2}{n} + \cdots + \frac{X_n}{n}\ .Then\ since X_1, X_2, - - - - - X_n\ are\ independent \{hence\ cov(X_i, X_j = 0)\}\ and\ have\ variance\ \sigma^2, we\ have$

$$Var(\bar{X}) = \frac{1}{n^2}Var(X_1) + \cdots + \frac{1}{n^2}Var(X_n)$$

$$= \frac{1}{n^2}\sigma^2 + \ldots\ldots\ldots\ldots\ldots + \frac{1}{n^2}\sigma^2 \qquad [\text{since } V(X_i) = \sigma^2 \text{ for } i = 1,2 \ldots\ldots\ldots n]$$

$$= n\left(\frac{1}{n^2}\sigma^2\right)$$

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

Hence $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

If we do not know the value of , then we substitute the estimate $\hat{\sigma}$ = s into $\sigma_{\bar{x}}$ and denote the estimated standard error by $\hat{\sigma}_{\bar{x}} = s_{\bar{x}} = \frac{s}{\sqrt{n}}$.

Now we extend the above methods to problems which deal with the means of two different population distributions. For instance, if μ_1 denotes true average Rockwell hardness for heat -treated steel specimens and μ_2 denotes true average hardness for cold-rolled specimens, then an investigator might wish to use samples of hardness observations from each type of steel as a basis for calculating an estimate of $\mu_1 - \mu_2$, the difference between the two true average hardnesses. Assuming that

(1) $X_1, X_2, \ldots\ldots X_m$ is a random sample from a distribution with mean μ_1 and variance $\sigma_1{}^2$, and

(2) $Y_1, Y_2, \ldots\ldots Y_n$ is a random sample from a distribution with mean μ_2 and variance $\sigma_2{}^2$, and

(3) The X and Y samples are independent of one another.

It can be shown that $\bar{X} - \bar{Y}$, the difference between the two sample means can be used as natural estimator of $\mu_1 - \mu_2$, the difference between the corresponding means of two different population distributions. The expected value of $\bar{X} - \bar{Y}$ is equal to $\mu_1 - \mu_2$, so **$\bar{X} - \bar{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$.**

Proof: $E(\bar{X} - \bar{Y}) = E\left[\left(\frac{X_1 + X_2 + \ldots\ldots + X_m}{m}\right) - \left(\frac{Y_1 + Y_2 + \ldots\ldots + Y_n}{n}\right)\right]$

$$= \left[\frac{1}{m}\{E(X_1) + E(X_2) + \ldots\ldots + E(X_m)\} - \frac{1}{n}\{E(Y_1) + E(Y_2) + \ldots\ldots + E(Y_n)\}\right]$$

$= \frac{1}{m}(m\mu_1) - \frac{1}{n}(n\mu_2)$ [since $E(X_i) = \mu_1$ for i=1,2…. m and $E(Y_i) = \mu_2$ for i=1,2…. n]

$= \mu_1 - \mu_2$ as desired.

Further it can be shown that the **standard deviation of $\bar{X} - \bar{Y}$ is $\sigma_{(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1{}^2}{m} + \frac{\sigma_2{}^2}{n}}$**

Proof: Since X and Y samples are independent, so $\bar{X}$ and $\bar{Y}$ will be independent quantities implying that $Cov(\bar{X}\ \bar{Y}) = 0$. Hence the variance of the difference between the two sample means is the sum of $V(\bar{X})$ and $V(\bar{Y})$:

$$V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}$$

The standard deviation of $\bar{X} - \bar{Y}$ is the square root of this expression. Hence

$$\boldsymbol{\sigma_{(\bar{X} - \bar{Y})} = \sqrt{\frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}}}$$

The sample variances must be used when ${\sigma_1}^2$ and ${\sigma_2}^2$ are unknown.

2.2 Solved Examples

Example 2.1:

We examine each one of the 150 newly typed pages and record the number of mistakes per page (the pages are supposed to be free of mistakes). We observe the following data:

Number of mistakes per page	0	1	2	3	4	5	6	7
Observed frequency	18	37	42	30	13	7	2	1

Let X=the number of mistakes on a randomly chosen page. Also assume that X follows a Poisson distribution with parameter μ.

a) Find an unbiased estimator of μ and compute the estimate for the data.

b) What is the standard error of your estimator? Compute the estimated standard error.

Solution:

a. An unbiased estimator of μ is given by sample mean, $\bar{X}$, since

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \ldots\ldots + X_n}{n}\right) = \frac{1}{n}[E(X_1) + E(X_2) + \ldots\ldots\ldots + E(X_n)]$$

$$= \frac{1}{n}(n\mu) \qquad [\text{since } E(X_i) = \mu \text{ for } i = 1,2\ldots\ldots\ldots n]$$

$$= \mu \text{ as desired.}$$

$$\text{Estimate} = \bar{x} = \frac{0(18)+1(37)+2(42)+3(30)+4(13)+5(7)+6(2)+7(1)}{150} = 2.11$$

b. Let the standard deviation of our estimator, $\bar{x}$, be denoted by $\sigma_{\bar{x}}$

Now $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\mu}{n}}$ (since $\sigma^2 = \mu$ for X Poisson)

Substituting the estimated value of μ i.e. $\bar{x}$ to compute the estimated standard error, we get

$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{2.11}{150}}$ =0.1186.

Example 2.2:

If X_1,X_2.....X_n constitute a random sample from a population with the mean μ, what condition must be imposed on the constants a_1,a_2.....a_n, so that $a_1X_1 + a_2X_2$.....+a_nX_n
is an unbiased estimator of μ?

Solution:

$a_1X_1 + a_2X_2$.....+a_nX_n is an unbiased estimator of μ if E($a_1X_1 + a_2X_2$.....+a_nX_n)= μ

Now E [$a_1X_1 + a_2X_2$.....+a_nX_n]

=a_1E(X_1)+a_2E(X_2)+.....+a_nE(X_n)

= $a_1\mu + a_2\mu$ +.....+$a_n\mu$ [since E(X_i)= μ for i=1,2…….. n]

= (a_1 + a_2+.....+a_n)μ

= μ only if (a_1 + a_2+.....+a_n)=1

So a_1 + a_2 +.....+ a_n should be equal to one for $a_1X_1 + a_2X_2$ .....+ a_nX_n to be an unbiased estimator of μ.

Example 2.3:

Independent random samples of size n_1 and n_2 are taken from a normal population with the mean μ and the variance σ^2. If n_1=25, n_2=50, $\overline{x_1}$ = 27.6 and $\overline{x_2}$ =38.1, find an unbiased estimator of μ.

Solution:

An unbiased estimator of μ is given by $\hat{\mu} = \frac{n_1\overline{x_1}+n_2\overline{x_2}}{n_1+n_2} = \frac{(25)(27.6)+(50)(38.1)}{25+50}$ = 34.6

It is unbiased since E($\hat{\mu}$)= μ as shown below:

E($\hat{\mu}$)= $\frac{n_1}{n_1+n_2}$ E($\overline{X_1}$) + $\frac{n_2}{n_1+n_2}$ E($\overline{X_2}$)

=$\frac{n_1}{n_1+n_2}$ (μ) + $\frac{n_2}{n_1+n_2}$ (μ) (since E($\overline{X_i}$)= μ for i = 1,2)

= μ as desired.

Example 2.4:

A sample of 20 measurements each on flexural strength (MPa) for concrete beams of a certain type and cylinders respectively gave the following results.

Beams:	5.9	7.2	7.3	6.3	8.1	6.8	7.0	7.6	6.8	6.5
	7.9	9.0	8.2	8.7	7.8	9.7	7.4	7.7	11.6	11.3
Cylinders:	6.1	5.8	7.8	7.1	7.2	9.2	6.6	8.3	7.0	8.3
	7.8	8.1	7.4	8.5	8.9	9.8	9.7	14.1	12.6	11.2

Before obtaining data we denote the beam strengths by $X_1, X_2 \ldots\ldots\ldots X_n$ and the cylinder strengths by $Y_1, Y_2 \ldots\ldots\ldots Y_n$. Suppose that the X_i's are drawn from a population with mean μ_1 and standard deviation σ_1. Similarly Y_i's are drawn from another population with mean μ_2 and standard deviation σ_2. Also assume that Y_i's are independent of the X_i's.

a) Prove that an unbiased estimator of $\mu_1 - \mu_2$ is given by $\bar{X} - \bar{Y}$ .Compute the estimate for the above data.

b) What is the variance and standard error of your estimator in part (a)? Compute the estimated standard error.

Solution:

a. $\bar{X} - \bar{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$ if $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$.

Now $E(\bar{X} - \bar{Y}) = E\left[\left(\frac{X_1 + X_2 + \ldots\ldots + X_n}{n}\right) - \left(\frac{Y_1 + Y_2 + \ldots\ldots + Y_n}{n}\right)\right]$

$= \left[\frac{1}{n}\{E(X_1) + E(X_2) + \ldots\ldots + E(X_n)\} - \frac{1}{n}\{E(Y_1) + E(Y_2) + \ldots\ldots + E(Y_n)\}\right]$

$= \frac{1}{n}(n\mu_1) - \frac{1}{n}(n\mu_2)$ [since $E(X_i) = \mu_1$ and $E(Y_i) = \mu_2$ for $i = 1,2 \ldots\ldots\ldots n$]

$= \mu_1 - \mu_2$ as desired.

To find an estimate for the given data, we first compute $\bar{X}$ and $\bar{Y}$.

Table: Calculations for mean, variance

X	Y	X^2	Y^2
5.9	6.1	34.81	37.21
7.2	5.8	51.84	33.64
7.3	7.8	53.29	60.84
6.3	7.1	39.69	50.41
8.1	7.2	65.61	51.84
6.8	9.2	46.24	84.64
7.0	6.6	49.00	43.56
7.6	8.3	57.76	68.89
6.8	7.0	46.24	49.00
6.5	8.3	42.25	68.89
7.9	7.8	62.41	60.84
9.0	8.1	81.00	65.61

8.2	7.4	67.24	54.76
8.7	8.5	75.69	72.25
7.8	8.9	60.84	79.21
9.7	9.8	94.09	96.04
7.4	9.7	54.76	94.09
7.7	14.1	59.29	198.81
11.6	12.6	134.56	158.76
11.3	11.2	127.69	125.44
$\sum X$ =**158.8**	$\sum Y$**=171.5**	$\sum X^2$**=1304.3**	$\sum Y^2$**=1554.73**

$\bar{X}= \frac{1}{20}\sum_{i=1}^{20} X_i=\frac{158.8}{20}$ = 7.94

$\bar{Y}= \frac{1}{20}\sum_{i=1}^{20} Y_i=\frac{171.5}{20}$ = 8.575

$So\ \hat{\mu}_1\text{-}\hat{\mu}_2 = \bar{X} - \bar{Y}$= 7.94 - 8.575 = -0.635

b. Var($\bar{X} - \bar{Y}$)= Var ($\bar{X}$)+ Var ($\bar{Y}$) $[Cov(\bar{X}\,\bar{Y}) = 0\ since\ \bar{X}\ and\ \bar{Y}\ are\ independent\ random\ variables]$

Now Var ($\bar{X}$)=$\frac{{\sigma_1}^2}{n}$ and Var ($\bar{Y}$)=$\frac{{\sigma_2}^2}{n}$

So ${\sigma^2}_{(\bar{X}-\bar{Y})}$= $\frac{{\sigma_1}^2}{n}$+$\frac{{\sigma_2}^2}{n}$ and hence $\sigma_{(\bar{X}-\bar{Y})}$=$\sqrt{\frac{{\sigma_1}^2}{n}+\frac{{\sigma_2}^2}{n}}$

To compute the estimated standard error we will first have to compute standard deviation, S, for both X and Y variables. Now $S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$ = $\sqrt{\left(\frac{\sum X_i^2}{n}-\bar{X}^2\right)\left(\frac{n}{n-1}\right)}$

Substituting the values, we get

S_x = $\sqrt{\left\{\frac{1304.3}{20}-(7.94)^2\right\}\left(\frac{20}{19}\right)}$ =1.512

Similarly $S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$ = $\sqrt{\left(\frac{\sum Y_i^2}{n}-\bar{Y}^2\right)\left(\frac{n}{n-1}\right)}$

= $\sqrt{\left\{\frac{1554.73}{20}-(8.575)^2\right\}\left(\frac{20}{19}\right)}$ = 2.104.

Hence $\hat{\sigma}_{(\bar{X}-\bar{Y})}$ = $\sqrt{\frac{2.286}{20}+\frac{4.427}{20}}$ = 0.579.

3. Point Estimates for Population Variance

Inferences regarding a population variance σ^2 or standard deviation σ are mostly needed to find an estimate of the precision of various point estimates. Similarly sometimes we face

problems where comparison of two population variances (or standard deviations) is required. Now we will discuss methods for computing these estimates.

3.1 Methodology

If the population is normal then we can use the following result concerning the sample variance S^2 to draw inferences about a population variance.

Assuming that the population is normally distributed $\hat{\sigma}^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$ is an unbiased estimator of σ^2.

Proof: $E(\hat{\sigma}^2)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$

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

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

Then, since $E[(X_i-\mu)^2]=\sigma^2$(given) and $E[(\bar{X}-\mu)^2]=\frac{\sigma^2}{n}$ as shown below

$$we\ have \bar{X}=\frac{X_1}{n}+\frac{X_2}{n}+\cdots+\frac{X_n}{n}\ .Then\ since X_1, X_2, - - - - - X_n\ are\ independent\ \{hence\ cov(X_i, X_j=0)\}$$

$$and\ have\ variance\ \sigma^2, we\ have\ Var(\bar{X})=\frac{1}{n^2}Var(X_1)+\cdots+\frac{1}{n^2}Var(X_n)=n\left(\frac{1}{n^2}\sigma^2\right)=\frac{\sigma^2}{n}$$

it follows that $E(\hat{\sigma}^2)=\frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2-n\frac{\sigma^2}{n}\right]=\frac{(n-1)}{n-1}\sigma^2=\sigma^2$ as desired.

Now suppose we want to compare the variances of two different populations. Assuming that the populations under investigation are normal, $\frac{{S_1}^2}{{S_2}^2}$ can be used as a point estimator of $\frac{{\sigma_1}^2}{{\sigma_2}^2}$.

3.2 Solved Examples

Example 3.1:

Given a random sample of size n from a population which has the known mean μ and the finite variance σ^2, show that $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is an unbiased estimator of σ^2.

Solution:

We know that $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$is an unbiased estimator of σ^2 if $E\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right\}=\sigma^2$

Now $E\ \{\frac{1}{n}\sum_{i=1}^{n}E(X_i-\mu)^2\}=\ \frac{1}{n}\sum_{i=1}^{n}\sigma^2$ {since we are given that $E(X_i-\mu)^2\ =\sigma^2$}

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

$$=\sigma^2 \text{ as desired.}$$

Example 3.2:

Consider a hypothetical normal population comprising only three values 2,5 and 8. Draw all possible samples of size 2 and calculate the mean $\bar{x}$ and variance

$s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2]$ for each sample. Examine whether the statistics are unbiased for the corresponding parameters. Show that $\hat{s}^2=\frac{1}{n-1}[(x_1-\bar{x})^2+(x_2-\bar{x})^2]$ is an unbiased estimator of population variance. Also calculate the variance of sampling distribution of mean $\bar{X}$ and verify that variance of $\bar{X}$ is equal to $\frac{\sigma^2}{n}$.

Solution :

Table : Calculations for mean and variance of samples

Sr.No	Sample values	Sample totals	Sample mean $(\bar{x})$	Sample variance (s^2)	Sample Variance $(\hat{s}^2)$	$(\bar{x}-\mu)^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	2,2	4	2	0	0	9
2	2,5	7	3.5	2.25	4.5	2.25
3	2,8	10	5	9	18	0
4	5,2	7	3.5	2.25	4.5	2.25
5	5,5	10	5	0	0	0
6	5,8	13	6.5	2.25	4.5	2.25
7	8,2	10	5	9	18	0
8	8,5	13	6.5	2.25	4.5	2.25
9	8,8	16	8	0	0	9
			$\sum\bar{x}$=45	$\sum s^2$=27	$\sum\hat{s}^2$=54	$\sum(\bar{x}-\mu)^2$=27

So there are 9 possible samples of size 2 as shown in column (2). The mean $\bar{x}=\frac{x_1+x_2}{2}$ and variance $s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2]$ for each sample are shown in columns (4) and (5) respectively.

To examine whether the statistics are unbiased for the corresponding parameters, we will first have to calculate the mean and variance for the population.

Population mean $=\mu=\frac{2+5+8}{3}=5$

Population variance $=\sigma^2=\frac{(2-5)^2+(5-5)^2+(8-5)^2}{3}=6$

Now the statistics are unbiased if

$E(\bar{x})=\mu$ and $E(s^2)=\sigma^2$

Substituting the values $E(\bar{x})=\frac{\sum\bar{x}}{n}=\frac{45}{9}=5$

Since $E(\bar{x})=\mu=5$ so $\bar{X}$ is an unbiased estimator of population mean.

Now E(s^2)= $\frac{\sum s^2}{n}$ = $\frac{27}{9}$ =3

Since E(s^2)= 3 ≠ σ^2= 6, so s^2 is not an unbiased estimator of population variance. However it can be shown that $\hat{s}^2 = \frac{1}{n-1}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2]$ is an unbiased estimator of population variance σ^2. We find that E($\hat{s}^2$)= $\frac{\sum \hat{s}^2}{n}$ = $\frac{54}{9}$ =6 = σ^2.

Variance of sampling distribution of mean $\bar{X}$ is denoted by

$\sigma^2_{\bar{x}}$= $\frac{\sum s^2}{n}$ = $\frac{27}{9}$ =3=$\frac{\sigma^2}{n}$ = $\frac{6}{2}$ =3, hence verified that $\sigma^2_{\bar{x}}$ =$\frac{\sigma^2}{n}$

Question 3.3:

Suppose each side of a square plot has length μ . So area of the plot will be μ^2. Since value of μ is unknown so we take n independent measurements $X_1, X_2 ..... X_n$ of the length. Assume that each X_i has mean μ and variance σ^2.

a. Show that $\bar{X}^2$ is a biased estimator for μ^2.

b. What value of k will make the estimator $\bar{X}^2$- kS^2 unbiased for μ^2, where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$?

Solution:

a. Since $\bar{X}$ is a random variable, so E($\bar{X}^2$)=V($\bar{X}$)+$[E(\bar{X})]^2$=$\frac{\sigma^2}{n}$ + $\mu^2 \neq \mu^2$

So $\bar{X}^2$ is not an unbiased estimator for μ^2.

b. For $\bar{X}^2$- kS^2 to be an unbiased estimator for μ^2, E($\bar{X}^2$- kS^2) should be equal to μ^2

Now E($\bar{X}^2$- kS^2)=E($\bar{X}^2$)- kE(S^2)

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

= $\frac{\sigma^2}{n}$+μ^2 - $k(\frac{n-1}{n})\sigma^2$ $\left[\text{because } \mathrm{E}(S^2) = (\frac{n-1}{n})\sigma^2\right]$

= μ^2 only if k=$\frac{1}{n-1}$.

Example 3.4:

A sample of 10 television tubes produced by a company showed that the mean lifetime is 1200 hours and the standard deviation is 100 hours.

a. Calculate the mean of the population of all television tubes produced by this company.

b. Compute the standard deviation of the population of all television tubes produced by this company.

c. If the same results are obtained for 30, 50 and 100 television tubes, estimate the mean and the standard deviation of the population.

d. What can you conclude about the relation between sample standard deviation and estimates of population standard deviation for different sample sizes?

Solution:

a. We can use sample mean $\bar{X}$ as an estimator of population mean μ. So $\hat{\mu}$=1200 hours.

b. We can use sample standard deviation defined as $\hat{S} = S\sqrt{\frac{n}{n-1}}$ as an estimator of population standard deviation σ. So $\hat{\sigma} = 100\sqrt{\frac{10}{9}}$ =105.4 hours.

c. The estimate for the population mean will remain same i.e. 1200 hours in all cases. However, the estimate for population standard deviation will differ for different sample sizes. If sample size is 30 then $\hat{\sigma} = 100\sqrt{\frac{30}{29}}$=101.7 hours. If sample size is 50 then $\hat{\sigma} = 100\sqrt{\frac{50}{49}}$=101 hours. If sample size is 100 then $\hat{\sigma} = 100\sqrt{\frac{100}{99}}$=100.5 hours.

d. As sample size increases, estimates of population standard deviation come closer and closer to sample standard deviation.

Example 3.5:

Suppose use of a certain type of pesticide increases average yield per acre by μ_1 with variance σ^2, whereas the use of second type of pesticide increases average yield per acre by μ_2 with the same variance σ^2.Let S_1^2 and S_2^2 denote the unbiased estimators of population variances of yields based on sample sizes n_1 and n_2 respectively, of the two pesticides. Show that the pooled estimator $\hat{\sigma}^2 = \left(\frac{n_1-1}{n_1+n_2-2}\right)(S_1^2)+\left(\frac{n_2-1}{n_1+n_2-2}\right)(S_2^2)$ is an unbiased estimator of σ^2.

Solution:

The pooled estimator $\hat{\sigma}^2$ is an unbiased estimator of σ^2 if $E(\hat{\sigma}^2)=\sigma^2$

Now $E(\hat{\sigma}^2) = \left(\frac{n_1-1}{n_1+n_2-2}\right)E(S_1^2)+\left(\frac{n_2-1}{n_1+n_2-2}\right)E(S_2^2)$

$= \left(\frac{n_1-1}{n_1+n_2-2}\right)\sigma^2+\left(\frac{n_2-1}{n_1+n_2-2}\right)\sigma^2$ [since $E(S_i^2)=\sigma^2$ for $i=1,2$]

$= \frac{(n_1+n_2-2)}{n_1+n_2-2}\sigma^2$

$= \sigma^2$ as desired.

Example 3.6:

Using data and calculations of example 2.4 compute a point estimate of the ratio $\frac{\sigma_1}{\sigma_2}$ of the two standard deviations.

Solution:

A point estimate of the ratio of the two standard deviations, $\frac{\sigma_1}{\sigma_2}$, is given by

$$\frac{\hat{\sigma}_1}{\hat{\sigma}_2} = \frac{S_x}{S_y} = \frac{1.512}{2.104} = 0.719.$$

4. Point Estimates for Population Proportions

Inferences concerning population proportion for specified characteristics are often required by the policymakers. Similarly sometimes estimates regarding differences in proportions of two different populations are needed for policy decisions. Now we will discuss methods for estimating these population parameters and also give an expression for estimating the reliability of the estimates.

4.1 Methodology

Suppose a random sample of size n is taken from a population and it is found that the number of "successes" is X. Now we can use $\hat{p} = \frac{X}{n}$, the sample fraction of "successes" as an estimator of p. $E(\hat{p}) = p$ (unbiasedness) and $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$

Proof: $E(\hat{p}) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$

$$\sigma_{\hat{p}} = \sqrt{V(\frac{X}{n})} = \sqrt{V(X)/n^2} = \sqrt{npq/n^2} = \sqrt{pq/n}$$ as desired.

Now we extend these methods to situations involving the proportions of two different population distributions. Let p_1 denote the true proportion of nickel-cadmium cells produced under current operating conditions that are defective because of internal shorts, and let p_2 represent the true proportion of cells with internal shorts produced under modified operating conditions. If the rationale for the modified conditions is to reduce the proportion of defective cells, a quality engineer would want to use sample information as a basis for calculating an estimate of $p_1 - p_2$.

Suppose that a sample of size m is selected from the first population and independently a sample of size $\boldsymbol{n}$ is selected from the second one. Let X denote the number of successes in the first sample and Y be the number of successes in the second. Independence of the two samples implies that X and Y are independent. Provided that the two sample sizes are much smaller than the corresponding population sizes, X and Y can be regarded as having binomial distributions. The natural estimator for $p_1 - p_2$, the difference in population proportions, is the corresponding difference in sample proportions $X/m - Y/n$.

$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$

so $\hat{p}_1 - \hat{p}_2$ is an unbiased estimator of $p_1 - p_2$, and

$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}$ (where $q_i = 1 - p_i$)

Proof: Since $E(X) = mp_1$ and $E(Y) = np_2$,

So $E(\hat{p}_1 - \hat{p}_2) = E\left(\frac{X}{m} - \frac{Y}{n}\right) = \frac{1}{m} E(X) - \frac{1}{n} E(Y) = \frac{1}{m} mp_1 - \frac{1}{n} np_2 = p_1 - p_2$ as desired.

Similarly, since $V(X) = mp_1 q_1$ and $V(Y) = np_2 q_2$, and X and Y are independent,

So $V(\hat{p}_1 - \hat{p}_2) = V\left(\frac{X}{m} - \frac{Y}{n}\right) = V\left(\frac{X}{m}\right) + V\left(\frac{Y}{n}\right) = \frac{1}{m^2} V(X) + \frac{1}{n^2} V(Y) = \frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}$ as desired.

4.2 Solved Examples

Example 4.1:

A sample of 20 students of XYZ College gave the following information on the brand of calculator used (F = Fiamo, O = Orpat, C = Citizen, S= Sharp):

F	F	O	F	C	F	F	S	C	O
S	S	F	O	C	F	F	F	O	F

a. Estimate the true proportion of all such students who used a Fiamo calculator.

b. Of the 10 students who used a Fiamo calculator, 4 had graphing calculators. Estimate the proportion of students who do not use a Fiamo graphing calculator.

Solution:

a. An estimate of the true proportion of all such students who used a Fiamo calculator is given by $\hat{p} = \frac{x}{n} = \frac{10}{20} = 0.5$ where x is the number of favourable cases i.e. number of students who used Fiamo calculator and n is the total number of cases i.e. total number of students.

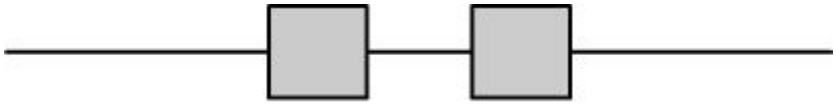
b. Using the same method as above, an estimate of the proportion of students who do not use a Fiamo graphing calculator is given by $\hat{p} = \frac{x}{n} = \frac{16}{20} = 0.8$ where x is the number of students who do not use a Fiamo graphing calculator.

Example 4.2:

A sample of 80 components is taken from a large factory and it is found that 68 components are not defective.

a. Estimate the proportion of all such components that are not defective.

b. Suppose now we randomly select two of these components and connect them in series, as shown here to construct a system.



The system will function if both components are not defective. Give a point estimate of the proportion of properly working systems?

Solution:

Let p denote the probability that a component works properly and P denote the probability that the system works properly. Then

a. Estimate of the proportion of all such components that are not defective is given by

$$\hat{p} = \frac{x}{n} = \frac{68}{80} = 0.85.$$

where x is the number of favourable cases i.e. number of components that are not defective and n is the total number of cases i.e. total number of components sampled.

b. A point estimate of the proportion of systems that work properly is given by

$$(\hat{p}) = \frac{{}^{68}C_2}{{}^{80}C_2} = \frac{(68)(67)}{(80)(79)} = 0.721.$$

Example 4.3:

A sample of 10 measurements of the weights of female students at xyz university gave the following results.

Student	1	2	3	4	5	6	7	8	9	10
Weight(kg)	40	45	47.6	48.2	52.8	57	52.5	52	59	49

Assume that the population weight is normally distributed. Find

a. Two unbiased estimators of population mean and make efficiency comparisons.

b. An unbiased estimator of population variance.

c. A point estimate of the proportion of all such female students whose weight exceeds 53kg.

Solution:

a. Since population is normally distributed so both sample mean and median are unbiased estimators of population mean

$\bar{X}= \frac{\sum X_i}{n} = \frac{503.1}{10} =50.31.$

To calculate median our first step is to arrange weights of students either in increasing or decreasing order as follows.

40 45 47.6 48.2 49 52 52.5 52.8 57 59

Since the total number of students is 10, so median weight would be the weight of $\left(\frac{10+1}{2}\right)$th student i.e. the average of the weights of 5th and 6th student which is $\frac{49+52}{2}= 50.5.$

To make efficiency comparisons we should compare MSEs of mean and median. Since both are unbiased so MSE of $\bar{X}$ and $\tilde{X}$ is equal to V($\bar{X}$) and V($\tilde{X}$) respectively. We know that for a normal distribution V($\bar{X}$)= $\frac{\sigma^2}{n}$ and V($\tilde{X}$)= $1.57\frac{\sigma^2}{n}$.

Since V($\bar{X}$)< V($\tilde{X}$),so mean is more efficient estimator of population mean.

b. An unbiased estimator of population variance is given by

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

$= \frac{(40-50.31)^2+(45-50.31)^2+(47.6-50.31)^2+(48.2-50.31)^2+(52.8-50.31)^2+(57-50.31)^2+(52.5-50.31)^2+(52-50.31)^2+(59-50.31)^2+(49-50.31)^2}{9}$

$= \frac{282.128}{9}= 31.35$

c. Since the number of students in the sample whose weight exceeds 53kg is two, so a point estimate for the population proportion of all such students whose weight exceeds 53kg is given by $\hat{p}=\frac{x}{n} = \frac{2}{10}$ where x is the number of favourable cases i.e. number of students whose weight exceeds 53kg and n is the total number of cases i.e. total number of students sampled.

Example 4.4:

Consider a random sample of 16 observations on plywood thickness. The observations are following:

.88 .88 .83 1.09 1.04 1.12 1.29 1.31

1.49 1.48 1.59 1.65 1.62 1.76 1.71 1.83

Assume that plywood thickness follows a normal distribution.

a. Calculate an estimate of the average value of plywood thickness. Which estimator did you use?

b. Calculate a point estimate of the median of the plywood thickness distribution, and state which estimator you used.

c. Calculate a point estimate of the value that separates the largest 10% of all values in the thickness distribution from the remaining 90%, and state which estimator you used.

d. Estimate P(X<1.5) i.e. the proportion of all thickness values less than 1.5.

e. What is the estimated standard error of the estimator that you used in part (b)?

Solution:

a. A point estimate of the mean value of plywood thickness is

$$\bar{X} = \frac{\sum x_i}{n} = \frac{.83+.88+.88+1.04+1.09+1.12+1.29+1.31+1.48+1.49+1.59+1.62+1.65+1.71+1.76+1.83}{16} = 1.348$$

We used sample mean as an estimator because for a normal distribution $\bar{X}$ is MVUE of population mean

b. A point estimate of the median of the plywood thickness distribution is $\bar{X}$=1.348 because for a normal distribution mean, median and mode are all equal. The reason for using sample mean, $\bar{X}$, rather than sample median, $\tilde{X}$, is that Var($\bar{X}$) is less than Var($\tilde{X}$). So $\bar{X}$ as an estimator of population median is more reliable.

c. To calculate a point estimate of the value that separates the largest 10% of all values in the thickness distribution from the remaining 90%, we can make use of the fact that for a normal distribution $\bar{X}$ + 1.28σ is the value that separates the largest 10%of all values from the remaining 90%. So we can use **$\bar{X}$ + 1.28$\hat{\sigma}$ as an estimator** where $\bar{X}$ is sample mean and $\hat{\sigma}$ is sample standard deviation defined as $\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$. Now our next step is to compute

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} = \left[\frac{\sum X^2}{n} - (\bar{X})^2\right]\left(\frac{n}{n-1}\right) = \left[\frac{30.7981}{16} - (1.348)^2\right]\left(\frac{16}{15}\right) = 0.114965.$$

so $\hat{\sigma} = \sqrt{0.114965} = 0.339$ and $\bar{X} + 1.28\hat{\sigma} = 1.348+1.28(0.339)=1.7819$.

So the X value 1.7819 separates the largest 10% from the remaining 90%.

d. To estimate P(X<1.5), we can first standardize X variable and find Z value defined as $\left(\frac{X-\bar{X}}{\sigma}\right)$. Then use this Z value to find the area (probability) under the standard normal curve. Now $Z=\frac{X-\bar{X}}{\sigma} = \frac{1.5-1.348}{0.339} = 0.448$. .

The area under the standard normal curve corresponding to Z=0.448 is 0.6736. So P(X<1.5)= 0.6736.

e. The estimated standard error of the estimator, $\bar{X}$, that we used in part (b) is given by $\frac{\hat{\sigma}}{\sqrt{n}}$ = $\frac{0.339}{\sqrt{16}}$=0.08475.

Example 4.5:

A random sample of male students of size n_1 is taken from XYZ University and it is found that X_1 scored more than 70% in their final exams. Another sample of female students of size n_2 from the same University showed that X_2 scored more than 70%. Let p_1 denotes the probability that a male student scores more than 70% and p_2 denotes the probability that a female student scores more than 70% in their final exams.

a. Show that $\left(\frac{X_1}{n_1}\right)$ - $\left(\frac{X_2}{n_2}\right)$ is an unbiased estimator for$p_1 - p_2$.

b. Find an expression for the standard error of your estimator in part (a).

c. What is the use of $x_1$and x_2 in estimating the standard error of your estimator?

d. If $n_1 = n_2 = 200$,x_1=127,and x_2=176, compute a point estimate for $p_1 - p_2$ and also give an estimate of its standard error.

Solution:

a. To show that $\left(\frac{X_1}{n_1}\right)$ - $\left(\frac{X_2}{n_2}\right)$ is an unbiased estimator for $p_1 - p_2$, we will have to prove that

$\mathrm{E}\left[\left(\frac{X_1}{n_1}\right) - \left(\frac{X_2}{n_2}\right)\right] = p_1 - p_2$

Now $\mathrm{E}\left[\left(\frac{X_1}{n_1}\right) - \left(\frac{X_2}{n_2}\right)\right] = \frac{1}{n_1}\mathrm{E}(X_1) - \frac{1}{n_2}\mathrm{E}(X_2)$

$= \frac{1}{n_1}(n_1 p_1) - \frac{1}{n_2}(n_2 p_2)$ [since $\mathrm{E}(X_i) = n_i p_i$ for i $= 1, 2$]

$= p_1 - p_2$ as desired.

b. The standard error of the estimator in part (a) is given by

$S.E_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{Var\left[\left(\frac{X_1}{n_1}\right) - \left(\frac{X_2}{n_2}\right)\right]}$

$= \sqrt{\frac{1}{n_1^2}Var(X_1) + \frac{1}{n_2^2}Var(X_2)}$ (Covariance is zero as $X_1 and X_2$ are independent random variables)

$= \sqrt{\frac{1}{n_1^2}n_1 p_1 q_1 + \frac{1}{n_2^2}n_2 p_2 q_2}$ (Since Var (X_i)=$n_i p_i q_i$ for i=1,2)

$= \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

c. We will use the observed values x_1 andx_2 to estimate the standard error of our estimator by using $\frac{X_1}{n_1}$ for $\hat{p}_1$ and $\frac{X_2}{n_2}$ for $\hat{p}_2$.

d. Substituting the values for X_1, X_2,n_1 andn_2, we get an estimate of $\hat{p}_1$- $\hat{p}_2$ as

$\hat{p}_1$- $\hat{p}_2$ = $\frac{127}{200}$ - $\frac{176}{200}$ = - $\frac{49}{200}$ = -0.245

Using the relevant values, we get an estimate of the standard error of the estimator as follows $\hat{\sigma}_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$

$= \sqrt{\frac{(.635)(.365)}{200} + \frac{(.88)(.12)}{200}}$

$= \sqrt{.001159 + .000528}$

= .041.

PRACTICE QUESTIONS

Q.1 Consider a random sample X_1..... X_n from the pdf

$f(x;\theta)$ = .5(1+θx) -1 ≤ x ≤ 1

where -1 ≤ θ ≤ 1. Show that $\hat{\theta}$ = 3$\bar{X}$ is an unbiased estimator of θ.

Q.2 If X_1, X_2..... X_n constitute a random sample from a normal populations with μ = 0, show that $\sum_{i=1}^{n} \frac{X_i^2}{n}$ is an unbiased estimator of σ^2.

Q.3 A random sample of size 65 was taken to estimate the mean annual income of 1000 families and the mean and S. D. were found to be Rs. 6300 and Rs. 9.5 respectively. Find an estimate for the population mean. Also calculate its standard error.

Q.4 A sample of 150 bulbs of brand A showed an average life of 1800 hrs with a standard deviation of 15 hrs. Another sample of 100 bulbs of brand B showed an average life of 1500 hrs with a standard deviation of 11 hrs. Find an estimate for the difference of the mean life of the population of A and B brand bulbs. Also calculate the standard error of the estimate.

Q.5 According to the mendelian law of segregation in genetics, when certain type of peas are crossed, the probability that the plant yields a yellow pea is 3/4 and that it yields a green pea is 1/4. For a plant yielding 400 peas, find the standard error of the proportion of yellow peas.

Q.6 An insecticide of brand A was sprayed to kill mosquitos of a container having 150 mosquitoes. It was found that 100 of the mosquitoes were killed. When another container having 170 mosquitoes of the same type was sprayed with brand B, 130 mosquitoes were killed. Find an estimate for the difference in the effectiveness of the two brands of insecticides. Also calculate the standard error of the estimate.

Q.7 The marketing manager of a large company conducted a sample survey in two states, Bihar and Orissa, taking 400 sample salesman in each case. The main findings of research are given in the following table;

State	Average Sales Per Day	Standard Deviation
Bihar	Rs. 2500	Rs. 400
Orissa	Rs. 2200	Rs. 500

Find an estimate for the difference in average per day sales of the salesmen in two states. Also calculate the standard error of the estimate.

Q.8 The following results were obtained from two samples each drawn from two different populations A and B;

Population	A	B
Sample	I	II
Sample size	n_1 = 16	n_2 = 9
Sample S. D.	s_1= 3	s_2 = 2

Find an estimate for the ratio of the population variances for brand A and B i.e. $\frac{\sigma_1^2}{\sigma_2^2}$.

Q.9 A population consists of numbers 4, 5, 8, 10, 13. Enumerate all possible samples of size 3 which can be drawn from the population without replacement and show that the mean of the sampling distribution of the sample means is equal to the population mean.

Calculate the variance of the sampling distribution of the sample mean and show that it is less than the population variance.

Q.10 A builder is considering two different areas of a large western state as sites for primary school. Of 50 households surveyed in one area, the proportion of households having primary school going children was 0.52. Similarly, of 45 households surveyed in another area, the proportion of households having primary school going children was 0.48. Find an estimate for the difference in the proportions of primary school going children in the two areas of the state? Also calculate the standard error of the estimate?