Sampling Distribution of Sample Mean

Session 3

06/03/2021

Sampling distribution of a Statistic(contd…)

- The set of values of the statistic so obtained , one for each sample, constitutes what is called the sampling distribution of the statistic.
- For example, the values $t_1, t_2, \ldots t_k$ determine the sampling distribution of the statistic t.
- In other words, the statistic t may be regarded as a random variable which can take values $t_1, t_2, \ldots t_k$ and we can compute the various statistical measures like mean, variance, skewness, kurtosis etc for its distribution.

Mean and variance of sampling distribution of t

- Mean and variance of sampling distribution of t is given by
- $\bar{t} = \frac{1}{k}(t_1 + t_2 + \cdots + t_k) = \frac{1}{k}\sum_{i=1}^{k} t_i$

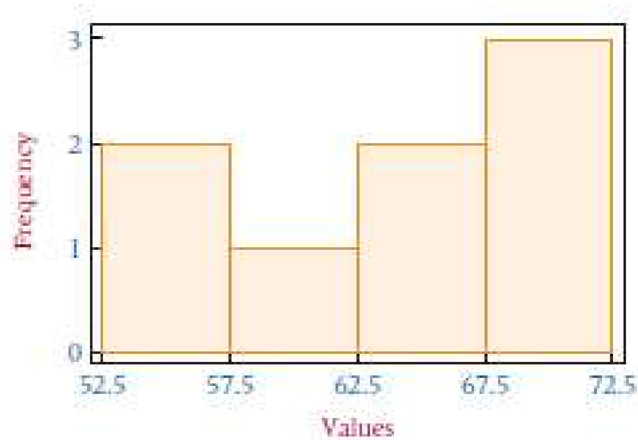
and

- Var (t) =$\frac{1}{k}\left((t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \cdots + (t_k - \bar{t})^2\right)$

$$= \frac{1}{k}\sum_{i=1}^{k}(t_i - \bar{t})^2$$

Illustration

- Suppose a small finite population consists of only *N* = 8 numbers:
- 54 ,55, 59, 63, 64, 68, 69, 70
- Using an Excel-produced histogram, we can see the shape of the distribution of this population.



- Suppose we take all possible samples of size n = 2 from this population with replacement.

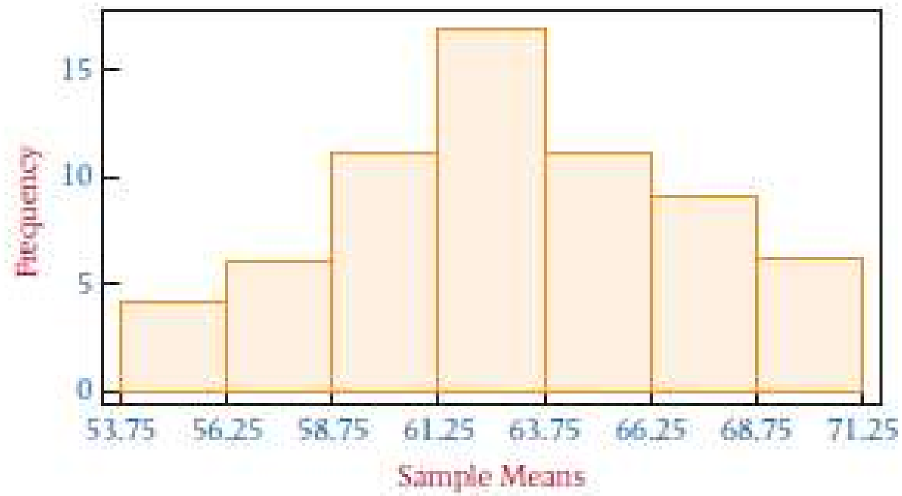
The result is the following pairs of data.

(54,54)	(55,54)	(59,54)	(63,54)
(54,55)	(55,55)	(59,55)	(63,55)
(54,59)	(55,59)	(59,59)	(63,59)
(54,63)	(55,63)	(59,63)	(63,63)
(54,64)	(55,64)	(59,64)	(63,64)
(54,68)	(55,68)	(59,68)	(63,68)
(54,69)	(55,69)	(59,69)	(63,69)
(54,70)	(55,70)	(59,70)	(63,70)
(64,54)	(68,54)	(69,54)	(70,54)
(64,55)	(68,55)	(69,55)	(70,55)
(64,59)	(68,59)	(69,59)	(70,59)
(64,63)	(68,63)	(69,63)	(70,63)
(64,64)	(68,64)	(69,64)	(70,64)
(64,68)	(68,68)	(69,68)	(70,68)
(64,69)	(68,69)	(69,69)	(70,69)
(64,70)	(68,70)	(69,70)	(70,70)

The means of each of these samples follow.

54	54.5	56.5	58.5	59	61	61.5	62
54.5	55	57	59	59.5	61.5	62	62.5
56.5	57	59	61	61.5	63.5	64	64.5
58.5	59	61	63	63.5	65.5	66	66.5
59	59.5	61.5	63.5	64	66	66.5	67
60	61.5	63.5	65.5	66	68	68.5	69
61.5	62	64	66	66.5	68.5	69	69.5
62	62.5	64.5	66.5	67	69	69.5	70

Again using an Excel-produced histogram, we can see the shape of the distribution of these sample means.



- Comment?

Unbiased estimate for population Mean(μ) and Variance(σ^2)

... values ~~of the given variable from the distribution.~~

14·8·1. Unbiased Estimate for Population Mean (μ) and Variance (σ^2). Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a large population $X_1, X_2, \ldots, X_N$ (of size N) with mean μ and variance σ^2. Then the sample mean ($\bar{x}$) and variance (s^2) are given by :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \text{and} \quad s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Now
$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$

Since x_i is a sample observation from the population X_i, ($i = 1, 2, \ldots, N$) it can take any one of the values $X_1, X_2, \ldots, X_N$ each with equal probability $1/N$.

$$\therefore \quad E(x_i) = \frac{1}{N}X_1 + \frac{1}{N}X_2 + \ldots + \frac{1}{N}X_N = \frac{1}{N}(X_1 + X_2 + \ldots + X_N) = \mu \qquad \ldots(1)$$

Hence
$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}(\mu) = \frac{1}{n}n\mu \quad \Rightarrow \quad E(\bar{x}) = \mu \qquad \ldots(14\cdot6)$$

Thus, *the sample mean ($\bar{x}$) is an unbiased estimate of the population mean (μ).*

Unbiased estimate for population Mean(μ) and Variance(σ^2)

Now $\quad E(s^2) = E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\} = E\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2\right) = \frac{1}{n}\sum_{i=1}^{n}E(x_i^2) - E(\bar{x}^2) \quad \ldots(2)$

We have $V(x_i) = E[x_i - E(x_i)]^2 = E(x_i - \mu)^2$, $\quad$ [From (1)]

$$= \frac{1}{N}\left[(X_1-\mu)^2 + (X_2-\mu)^2 + \ldots + (X_N-\mu)^2\right] = \sigma^2 \quad \ldots(3)$$

Also $\quad V(x) = E(x^2) - [E(x)]^2 \quad \Rightarrow \quad E(x^2) = V(x) + [E(x)]^2 \quad \ldots(4)$

In particular $\quad E(x_i^2) = V(x_i) + [E(x_i)]^2 = \sigma^2 + \mu^2 \quad \ldots(5)$

Also from (4), we obtain $\quad E(\bar{x}^2) = V(\bar{x}) + [E(\bar{x})]^2$

But $\quad V(\bar{x}) = \frac{\sigma^2}{n}$, where σ^2 is the population variance. $\quad$ [*cf.* § 14·8·2]

$\therefore \quad E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2 \quad$ [Using (14·6)] $\quad \ldots(5a)$

Unbiased estimate for population Mean(μ) and Variance(σ^2)

Substituting from (5) and (5a) in (2) we get

$$E(s^2) = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right)$$

$$= \frac{1}{n}n(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right) = \left(1-\frac{1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2 \qquad \ldots(14\cdot7)$$

Since $E(s^2) \neq \sigma^2$, *sample variance is not an unbiased estimate of population variance.*

From (14·7), we get $\quad \frac{n}{n-1}E(s^2) = \sigma^2 \quad \Rightarrow \quad E\left(\frac{ns^2}{n-1}\right) = \sigma^2$

$$\Rightarrow \quad E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} = \sigma^2, \quad i.e., \quad E(S^2) = \sigma^2 \qquad \ldots(14\cdot8)$$

where $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad \ldots(14\cdot8a)$

$\therefore$ *S^2 is an unbiased estimate of the population variance σ^2.*

Key Learnings

- Sample mean $(\bar{x})$ is an unbiased estimate of Population Mean (μ)
- Sample variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is NOT an unbiased estimate of population variance σ^2.
- Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of population variance σ^2.