1/9/20

(General) Preliminary Stuff:

We use capital letters to denote random variables
We use lower case letters to denote values

A parameter is a value derived from a random variable

A statistic is a value derived from a sample of observations from a random variable

We use greek letters to denote parameters
We use english letters to denote statistics

$X$ - random variable

$x_1, x_2, \ldots, x_n$ : denotes a sample of $n$ observations from $X$ (fixed, but unknown, values)

$X_1, X_2, \ldots, X_n$: denotes a sample of $n$ random observations from $X$ ($n$ random variables)

The random sample of size $n$ from $X$ is $\{X_1, X_2, \ldots, X_n\}$ where these random variables are independent of one another and each one is identically distributed to the common distribution (r.v.) $X$

We say $\{X_i\}_{i=1}^{n}$ is iid to $X$

$E[X] = \mu$ = mean of the r.v. $X$ (parameter)

$\bar{x} = \frac{\sum x_i}{n}$ = mean of sample values (statistic)

$\bar{x}$ is commonly used to approximate $\mu$, but we could use other statistics like the "median" or the "mode".

$Var(X) = \sigma^2$ = variance of the r.v. $X$ (parameter)

~~is commonly approximated~~

$Var(X) = \sigma^2 = E[(X-\mu)^2]$ is commonly approximated by

1) $$"S_n^2" = \frac{\sum (x_i - \bar{x})^2}{n} \quad \text{(statistic)}$$

2) $$"S_{n-1}^2" = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{(statistic)}$$

**Remark**: $Var(X) = E[(X-\mu)^2]$

$\mu = E[X]$ → constant

$$= E[X^2 - 2\mu \cdot X + \mu^2]$$

$$= E[X^2] - E[2\mu \cdot X] + E[\mu^2]$$

$$= E[X^2] - 2\mu \cdot \underbrace{E[X]}_{\mu} + \mu^2$$

$$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2$$

$$\therefore Var(X) = E[X^2] - (E[X])^2$$

values, $\bar{x}$ is an estimate of $\mu$

$s_n^2$ & $s_{n-1}^2$ are estimates of $\sigma^2$

These have corresponding estimators (r.v.'s) that are defined by the random processes used to get the estimates

E.g. $\bar{x} = \dfrac{\sum x_i}{n}$ (value) estimate

$\bar{X} = \dfrac{\sum X_i}{n}$ (r.v.) estimator

Then $E[\bar{x}] = \bar{x}$ ($\bar{x}$ is just a fixed number)

$$E[\bar{X}] = E\left[\frac{\sum X_i}{n}\right] = \frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\cancel{n} \cdot E[X]}{\cancel{n}} = \mu$$

We say $\bar{X}$ is an unbiased estimator of $\mu$ since

$$E[\bar{X}] = \mu$$

$$Var(\bar{x}) = 0$$

$$Var(\bar{X}) = Var\left(\frac{\sum X_i}{n}\right) = \frac{Var(X_1 + X_2 + \cdots + X_n)}{n^2} = \frac{n \cdot Var(X)}{n^2}$$

$$\therefore Var(\bar{X}) = \frac{Var(X)}{n}$$

$$s_n^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$ is an estimate for $\sigma^2$

$$\underline{\underline{S}}_n^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$ is the corresponding estimator

↳ double underline means capital letter (r.v.)

Likewise $$s_{n-1}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$ is an estimate for $\sigma^2$

$$\underline{\underline{S}}_{n-1}^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$ is the corresponding estimator

Remarks:

value (estimate)

$$E[s_n^2] = s_n^2$$

estimate (value)

$$Var(s_{n-1}^2) = 0$$

Fact: $E[\underline{\underline{s}}_n^2] \neq \sigma^2$ ($\underline{\underline{s}}_n^2$ is a biased estimator of $\sigma^2$)

Actually $$E[\underline{\underline{s}}_n^2] = \frac{n-1}{n} \cdot \sigma^2$$

Since $\lim_{n \to \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$, then

$\underline{\underline{s}}_n^2$ is called asymptotically unbiased