MATH E-156 Mathematical Statistics

Harvard Extension School

Dmitry Kurochkin

Fall 2020
Lecture 9

Contents

Contents

Definition of Sample Mean and Sample Variance

Def.

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (iid) random variables. Then we define:

- *sample mean* as

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

- *sample variance* as

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \bar{X}\right)^2$$

Contents

Distribution of Sample Mean

Claim

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

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

Contents

Independence of Sample Mean and Sample Variance

Thm.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\bar{X} \;\text{ and }\; (X_1 - \bar{X},\, X_2 - \bar{X}, \ldots,\, X_n - \bar{X}) \;\text{ are independent.}$$

Independence of Sample Mean and Sample Variance

Thm.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\bar{X} \text{ and } (X_1 - \bar{X},\, X_2 - \bar{X}, \ldots,\, X_n - \bar{X}) \text{ are independent.}$$

Corollary

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\bar{X} \text{ and } S^2 \text{ are independent.}$$

Contents

Distribution of Sample Variance

Thm.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Distribution of Sample Variance

Thm.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Corollary

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Contents

Population Parameters: Mean and Variance

Def.

Let population of size N consist of $x_1, x_2, \ldots, x_N$ values. Then we define:

- *population mean* as

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- *population variance* as

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

Population Parameters: Mean and Variance

Claim

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu$$

Contents

Definition of SRS, $X_1, X_2, \ldots, X_n$

Def.

Sampling from a population of size N in way that each of the $\binom{N}{n}$ possible samples of size n taken without replacement has the same probability of occurrence, is called *simple random sampling*.

We denote observed sampling values by

$$X_1, X_2, \ldots, X_n.$$

Remark

1. Note that for each $i \in \{1, 2, \ldots, n\}$,
   - X_i is not same as x_i, the i-th population value
   - X_i is a random variable (r.v.)
   - X_i is not independent of X_j, $j \in \{1, 2, \ldots\}$

Probability Distributions of Sampling Values, $P(X_i = x)$

Lemma

Denote the distinct values assumed by the population members by $\xi_1, \xi_2, \ldots, \xi_m$, and denote the number of population members that have the value ξ_k by n_k, where $k \in \{1, 2, \ldots, m\}$.

Then, given $i \in \{1, 2, \ldots n\}$, X_i is a discrete random variable with the following probability mass function:

$$P(X_i = \xi_k) = \frac{n_k}{N} \quad \text{for} \quad k \in \{1, 2, \ldots, m\}.$$

Also, for each $i \in \{1, 2, \ldots, n\}$,

$$\mathrm{E}[X_i] = \mu$$
$$\mathsf{Var}(X_i) = \sigma^2$$

Contents

Definition of Sample Mean of SRS, $\bar{X}$

Def.

Let $X_1, X_2, \ldots, X_n$ denote the sample values. We define the *sample mean* as

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

Remark

1. Note that for each $i \in \{1, 2, \ldots, n\}$,
   - X_i is not same as x_i, the i-th population value
   - X_i is a random variable (r.v.)
   - X_i is not independent of X_j, $j \in \{1, 2, \ldots\}$
2. Note that
   - $\bar{X}$ is a random variable (r.v.); distribution of $\bar{X}$ will be referred to as *sampling distribution* of $\bar{X}$
   - the sampling distribution of $\bar{X}$ determines how accurately $\bar{X}$ estimates the population mean, μ

Expected Value of Sample Mean, $\mathrm{E}[\bar{X}]$

Thm.

Assuming simple random sampling from a population with the population mean μ,

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

Covariance between Two Sampling Values, Cov(X_i, X_j)

Lemma

For simple random sampling (i.e. without replacement),

$$P(X_i = \xi_l | X_j = \xi_k) = \begin{cases} \frac{n_l}{N-1}, & \text{if } k \neq l, \\ \frac{n_l - 1}{N-1}, & \text{if } k = l, \end{cases}$$

and

$$\text{Cov}(X_i, X_j) = -\frac{\sigma^2}{N-1} \quad \text{if} \quad i \neq j.$$

Variance of Sample Mean, Var($\bar{X}$)

Thm.

Assuming simple random sampling from a population with the population variance σ^2,

$$\begin{aligned}\mathrm{Var}(\bar{X}) &= \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\\ &= \frac{\sigma^2}{n}\underbrace{\left(1-\frac{n-1}{N-1}\right)}_{\substack{\text{finite population}\\ \text{correction}}}\end{aligned}$$

Contents

Expectation of $\hat{\sigma}^2$

Thm.

Assuming simple random sampling from a population with the population variance σ^2,

$$\mathrm{E}[\hat{\sigma}^2] = \sigma^2 \left(\frac{n-1}{n}\right) \frac{N}{N-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

Remark

If

$$\mathrm{E}[\hat{\sigma}^2] \neq \sigma^2,$$

we will say that $\hat{\sigma}^2$ is *biased* estimator of σ^2.

Expectation of $s^2_{\bar{X}}$

Corollary

Assuming simple random sampling from a population with the population variance σ^2,

$$s^2_{\bar{X}} = \frac{\hat{\sigma}^2}{n}\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

is an unbiased estimate of Var$(\bar{X})$, that is,

$$\mathrm{E}[s^2_{\bar{X}}] = \mathsf{Var}(\bar{X}).$$

Remark

Note that

$$s^2_{\bar{X}} = \frac{S^2}{n}\Big(1 - \underbrace{\frac{n}{N}}_{\text{sampling fraction}}\Big), \quad \text{where} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$