LECTURE 05

ESTIMATION - BIAS - MEAN SQUARE ERROR (MSE)

Outline of today's lecture:

I. Introduction

- Purpose of statistics is to permit the user to make an inference about a population based on information contained in a sample.

- An estimate may be given in two distinct forms:
  1. Point Estimate
  2. Interval Estimate

- **Definition 8.1** An *estimator* is a rule that tells how to calculate the value of an estimate based on the measurements contained in a sample.

II. Bias and Mean Square Error of Point Estimators

Definition 8.2 Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an *unbiased* estimator if $E(\hat{\theta}) = \theta$. Otherwise $\hat{\theta}$ is said to be *biased*.

- *Note that* $E(\hat{\theta})$ *equals to a constant, it is not a random variable.*

- We want a point estimator to be unbiased. In other words, we would like the mean or expected value of the distribution of estimates to equal the parameter estimated.

Definition 8.3 The *bias* of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$

- *Note that,* $B(\hat{\theta})$ *is not a random variable since* $E(\hat{\theta})$ *is not a random variable*

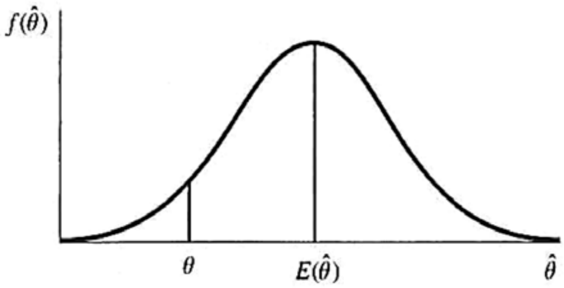


Figure 1 Sampling Distribution for a *positively biased* estimator

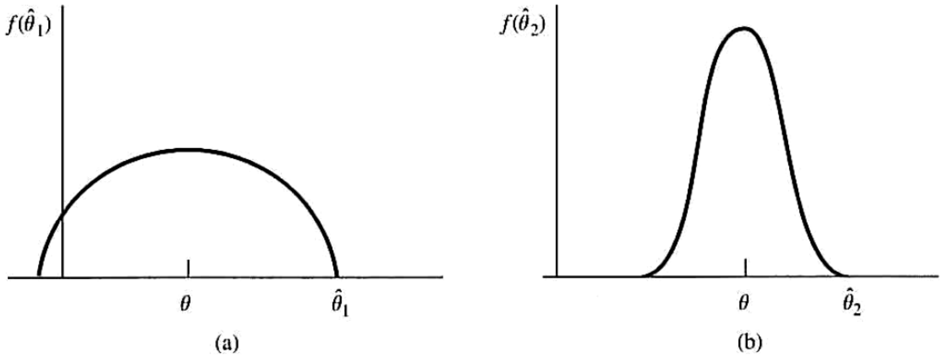


Figure 1 Sampling Distributions for two *unbiased* estimators: (a) estimator with large variation; (b) estimator with small variation

- We would prefer that our estimator have the type of distribution indicated in Figure 8.3(b) because the smaller variance guarantees that, in repeated sampling, a higher fraction of values of $\hat{\theta}_2$ will be "close" to θ.

- Thus, in addition to preferring *unbiasedness*, we want the variance of the distribution of the estimator $V(\hat{\theta})$ to be as small as possible.
    - *Given two unbiased* estimators of parameter θ, and all other things being equal, we *would select* the estimator with the *smaller variance*.

- To characterize the goodness of a point estimator, we might employ the expected value of $(\hat{\theta} - \theta)^2$, the average of the square of the distance between the estimator and its target parameter.

Definition 8.4 The *mean square error* of a point estimator $\hat{\theta}$ is the expected value of $(\hat{\theta}-\theta)^2$:

$$MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right]$$

- Hence, the mean square error of an estimator $\hat{\theta}$, $MSE(\hat{\theta})$ is a function of both its *variance* and its *bias*.

The MSE is given by: $MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right]$

We know that $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ which implies $\theta = E(\hat{\theta}) - B(\hat{\theta})$. Thus:

$$MSE(\hat{\theta}) = E\left\{\left[\hat{\theta} - E(\hat{\theta}) + B(\hat{\theta})\right]^2\right\}$$

$$MSE(\hat{\theta}) = E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]^2 + \left[B(\hat{\theta})\right]^2 + 2B(\hat{\theta})\left[\hat{\theta} - E(\hat{\theta})\right]\right\}$$

$$MSE(\hat{\theta}) = E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]^2\right\} + \left[B(\hat{\theta})\right]^2 + 2B(\hat{\theta}).E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]\right\}$$

$$MSE(\hat{\theta}) = E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]^2\right\} + \left[B(\hat{\theta})\right]^2 + 2B(\hat{\theta}).E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]\right\}$$

$$MSE(\hat{\theta}) = \underbrace{E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]^2\right\}}_{V(\hat{\theta})} + \left[B(\hat{\theta})\right]^2 + 2.B(\hat{\theta}).\cancel{\left\{E(\hat{\theta}) - E(\hat{\theta})\right\}}$$

which yields:

$$MSE(\hat{\theta}) = V(\hat{\theta}) + \left[B(\hat{\theta}) \right]^2$$

- In particular, we often seek unbiased estimators with relatively small variances.
- Below, we will consider some common and useful unbiased point estimators.

- Sample Mean, $\bar{Y}$

$$E\left(\bar{Y}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$$

$$E\left(\bar{Y}\right) = \frac{1}{n}E\left(Y_1 + Y_2 + \ldots + Y_n\right)$$

$$E\left(\bar{Y}\right) = \frac{1}{n}\left[\underbrace{E\left(Y_1\right)}_{\mu} + \underbrace{E\left(Y_2\right)}_{\mu} + \ldots + \underbrace{E\left(Y_n\right)}_{\mu}\right]$$

$$E\left(\bar{Y}\right) = \frac{1}{n}n\mu$$

$$E\left(\bar{Y}\right) = \mu$$

Hence $\bar{Y}$ is an unbiased estimator of μ.

- Sample Variance, S^2

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}$$

$$E(S^2) = \frac{E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]}{E(n-1)} = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]$$

Hence we need to find $E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]$.

For this purpose let us rearrange $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ as follows:

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y}\right]^2 = \sum_{i=1}^{n}\left[Y_i - \mu - \overline{Y} + \mu\right]^2$$

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y}\right]^2 = \sum_{i=1}^{n}\left[\left(Y_i - \mu\right) - (\overline{Y} - \mu)\right]^2$$

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y}\right]^2 = \sum_{i=1}^{n}\left[\left(Y_i - \mu\right)^2 + (\overline{Y} - \mu)^2 - 2\left(Y_i - \mu\right)(\overline{Y} - \mu)\right]$$

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y}\right]^2 = \sum_{i=1}^{n}\left(Y_i - \mu\right)^2 + \sum_{i=1}^{n}(\overline{Y} - \mu)^2 - 2(\overline{Y} - \mu)\sum_{i=1}^{n}\left(Y_i - \mu\right)$$

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y}\right]^2 = \sum_{i=1}^{n}\left(Y_i - \mu\right)^2 + n(\overline{Y} - \mu)^2 - 2(\overline{Y} - \mu)\underbrace{\sum_{i=1}^{n}\left(Y_i - \mu\right)}_{\substack{\sum Y_i - \sum \mu \\ = n.\overline{Y} - n.\mu \\ = n(\overline{Y} - \mu)}}$$

$$\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2=\sum_{i=1}^{n}\left(Y_i-\mu\right)^2+n(\bar{Y}-\mu)^2-2n(\bar{Y}-\mu)^2$$

$$\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2=\sum_{i=1}^{n}\left(Y_i-\mu\right)^2-n(\bar{Y}-\mu)^2$$

Taking the expectation of both sides:

$$E\left(\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2\right)=E\left(\sum_{i=1}^{n}\left(Y_i-\mu\right)^2\right)-n\underbrace{E\left((\bar{Y}-\mu)^2\right)}_{Var(\bar{Y})}$$

$$E\left(\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2\right)=E\left(\sum_{i=1}^{n}\left(Y_i-\mu\right)^2\right)-n\frac{\sigma^2}{n}$$

$$E\left(\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2\right)=\left[\underbrace{E\left(Y_1-\mu\right)^2}_{\sigma^2}+\underbrace{E\left(Y_2-\mu\right)^2}_{\sigma^2}+\ldots+\underbrace{E\left(Y_n-\mu\right)^2}_{\sigma^2}\right]-n\frac{\sigma^2}{n}$$

$$E\left(\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2\right)=n\sigma^2-\sigma^2$$

$$E\left(\sum_{i=1}^{n}\left[Y_i-\bar{Y}\right]^2\right)=(n-1)\sigma^2$$

Thus:

$$E(S^2)=\frac{E\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]}{E(n-1)}=\frac{1}{n-1}E\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]$$

$$E(S^2)=\frac{E\left[\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]}{E(n-1)}=\frac{1}{n-1}(n-1)\sigma^2$$

$$E(S^2)=\sigma^2$$

which implies that $S^2=\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}$ is an unbiased estimator of σ^2.

Therefore, the estimator $S^{*2}=\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n}$ is not an unbiased estimator of σ^2. That is why, we use $S^2=\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}$.

- $\overline{Y}_1-\overline{Y}_2$

If the random samples 1 and 2 are independent:

$$E\left(\overline{Y}_1-\overline{Y}_2\right)=E(\overline{Y}_1)-E(\overline{Y}_2)$$

$$E\left(\overline{Y}_1-\overline{Y}_2\right)=\mu_1-\mu_2$$

Hence $\left(\overline{Y}_1-\overline{Y}_2\right)$ is an unbiased estimator of $\mu_1-\mu_2$.

- *What about the variance?*

$$Var\left(\overline{Y}_1-\overline{Y}_2\right)=Var(\overline{Y}_1)+Var(\overline{Y}_2)-2Cov(\overline{Y}_1,\overline{Y}_2)$$

If random samples 1 and 2 are independent, $Cov(\overline{Y}_1, \overline{Y}_2) = 0$

$$Var\left(\overline{Y}_1 - \overline{Y}_2\right) = Var(\overline{Y}_1) + Var(\overline{Y}_2)$$

$$Var\left(\overline{Y}_1 - \overline{Y}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

III. Evaluating the Goodness of a Point Estimator

Tchebysheff's Theorem If μ and σ are the mean and the standard deviation of a random variable Y, and $\sigma \neq 0$, then for any positive constant *k*:

$$P\left(\left|Y - \mu\right| < k\sigma\right) \geq 1 - \frac{1}{k^2}$$

In other words, the probability that *Y* will take on a value within *k* standard deviation of the mean is at least $1 - \frac{1}{k^2}$

- Tchebysheff 's theorem applies to any data set, *not only normally distributed data sets*.
  - What is important is the fact that this prediction of Tchebysheff's Theorem is the worst case scenario and often there is a greater probability.

Example 1 Assume *k*=2, then $P\left(|Y-\mu|<2\sigma\right)\geq 1-\frac{1}{2^2}=0.75$. At least 75% of the values of Y will fall within 2 standard deviations of the mean.

Example 2 For *k*=3, we see that $1-\frac{1}{k^2}=1-\frac{1}{9}=\frac{8}{9}=0.89$, which implies that 89% of the values of Y will fall within *k*=3 standard deviations of the mean.

- One way to measure the *goodness of any point estimation* procedure is in terms of the *distances* between the *estimates it generates* and the *target parameter*.
  - This quantity, which varies randomly in repeated sampling, is called the *error of estimation*.
  - Naturally we would like the error of estimation to be as small as possible.

Definition 8.5 The error of estimation ε is the distance between an estimator and its target parameter. That is, $\varepsilon=\left|\hat{\theta}-\theta\right|$.

- Suppose that we want to find the value of *b* such that $P(\varepsilon<b)=0.90$.

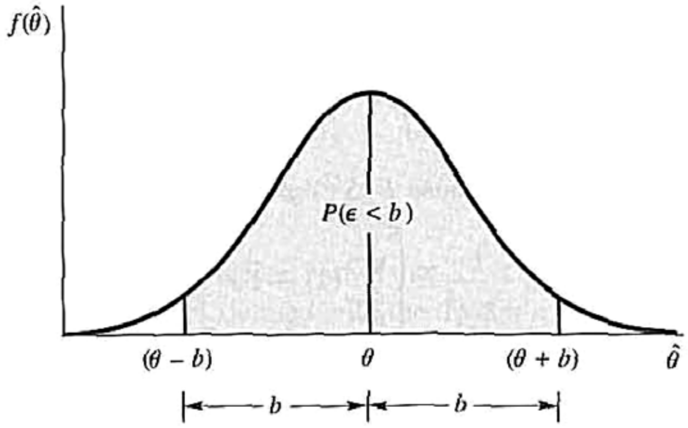


Figure 2 Sampling Distribution of a Point Estimator $\hat{\theta}$

- If we know the probability distribution of $\hat{\theta}$, we can easily seek a value *b* such that $\int_{\theta-b}^{\theta+b} f(\hat{\theta})d\hat{\theta} = 0.90$

- Whether we know the probability distribution of $\hat{\theta}$ or not, *<u>if $\hat{\theta}$ is unbiased</u>* we can find an approximate bound on ε by Tchebysheff's Theorem.

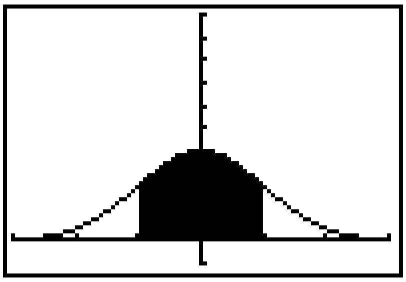
    - For example, for *k*=2, $\varepsilon = |\hat{\theta} - \theta|$ will be less than $2\sigma_{\hat{\theta}}$ with probability at least 0.75

- When the data values seem to have a normal distribution, or approximately so, we can use a much easier theorem or *rule* than Tchebysheff's →

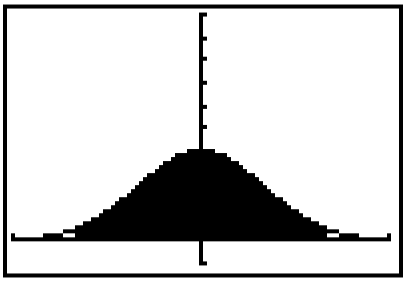
Emprical Rule The "empirical rule" states that in cases where the distribution is normal, the following statements are true:

- Approximately 68% of the data values will fall within 1 standard deviation of the mean.
- Approximately 95% of the data values will fall within 2 standard deviations of the mean.
- Approximately 99.7% of the data values will fall within 3 standard deviations of the mean.

Graphically, this corresponds to the area under the curve as shown below for 1 and 2 standard deviations. *The empirical rule is often stated simply as 68-95-99.7.* Note how this ties in with the range rule of thumb, by stating that 95% of the data usually falls within two standard deviations of the mean.



Data within 1σ of their mean (68%)



Data within 2σ of their mean (95%)

Example A comparison of the durability of two types of automobile tires was obtained by road-testing samples of $n_1 = n_2 = 100$ tires of each type. The number of miles until wear-out was recorded, where wear-out was defined as the number of miles until the amount of remaining tread reached a pre-specified small value. The measurements for the two types of tires were obtained independently and the following means and variances were computed:

$$\bar{y}_1 = 26,400 \qquad \bar{y}_2 = 25,100$$
$$s_1^2 = 1,440,000 \qquad s_2^2 = 1,960,000$$

Estimate the difference in mean miles to wear-out, and place a two-standard-error bound on the error of estimation.

Solution

The point estimate of $(\mu_1 - \mu_2)$ is

$\bar{y}_1 - \bar{y}_2 = 26,400 - 25,100 = 1,300$ miles. The standard error of the estimator is $\sigma_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

We must know σ_1^2 and σ_2^2, or we must have good approximate values for them, in order to calculate $\sigma_{(\bar{y}_1 - \bar{y}_2)}$. We can use the information coming from the current sample data by using the unbiased estimators:

$$\hat{\sigma}_i^2 = S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}i\right)^2, \qquad i=1,2.$$

These estimates will be adequate if the sample sizes are reasonably large, say, $n_i \geq 30$, for i=1,2. The calculated values of S_i^2, based on the two wear tests, are $s_1^2 = 1,440,000$ and $s_2^2 = 1,960,000$. Substituting these values for σ_1^2 and σ_2^2 in the formula for $\sigma_{(\bar{y}_1 - \bar{y}_2)}$, we have:

$$\sigma_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{1,440,000}{100}+\frac{1,960,000}{100}}$$

$$\sigma_{(\bar{y}_1-\bar{y}_2)} = \sqrt{34,000} = 184.4 \text{ miles.}$$

Consequently, we estimate the difference in mean wear to be 1300 miles, and we expect the error of estimation to be less than $2\sigma_{(\bar{y}_1-\bar{y}_2)}$ or 368.8 miles, with a probability of approximately 0.95.