STAT 319 – Probability and Statistics For Engineers

Lecture 6

Fundamentals of Sampling Distributions and Point Estimations

Engineering College, Hail University, Saudi Arabia

6.1 Definitions

Definitions

A population: Consist of the totality of observations with which we are concerned

A sample is a subset of a population

Let X_1 , X_2 , … , X_n be n independent random variables, each having the same probability distribution f(x). We then define X_1 , X_2 , … , X_n to be random sample of size n from the population f(x).

Random Samples

The rv's $X_1,\ldots,X_n$ are said to form a (simple *random sample* of size n if

1. The X_i's are independent rv's.
2. Every X_i has the same probability distribution.

Statistic

A *statistic* is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result.

or

Any function of the random variables constituting of random sample called a statistic.

If X_1 , X_2 , … , X_n represent a random sample of size n, then:

1) the sample mean is defined by the statistic

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

2) the sample variance is defined by the statistic

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

6.2 Sampling Distributions and the Central Limit Theorem

Sample Mean

Let $\boldsymbol{X_1, \ldots, X_n}$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

$$1.\ E\left(\overline{X}\right) = \mu_{\overline{X}} = \mu$$

$$2.\ V\left(\overline{X}\right) = \sigma^2_{\overline{X}} = \sigma^2 / n$$

In addition, with $T_o = X_1 + \ldots + X_n$,
$E\left(T_o\right) = n\mu,\ V\left(T_o\right) = n\sigma^2$, and $\sigma_{T_o} = \sqrt{n}\,\sigma.$

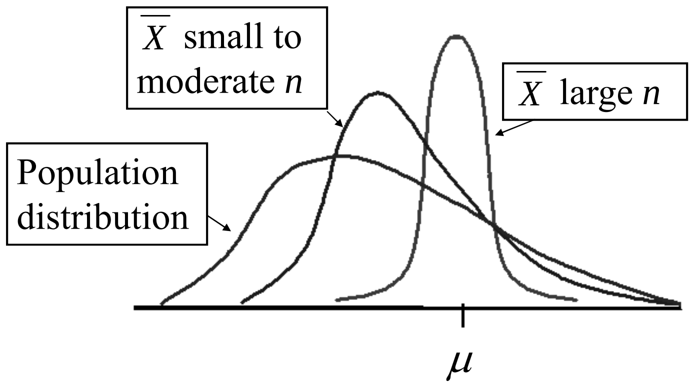
Normal Population Distribution

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean value μ and standard deviation σ. Then for any n, $\overline{X}$ is normally distributed with mean μ and standard deviation $\sigma_{\overline{X}} = \sigma / \sqrt{n}$.

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean value μ and variance σ^2. Then if n sufficiently large, $\overline{X}$ has approximately a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2 / n$, and T_o also has approximately a normal distribution with $\mu_{T_o} = n\mu$, $\sigma_{T_o} = n\sigma^2$. The larger the value of n, the better the approximation.

The Central Limit Theorem



If $n > 30$, the Central Limit Theorem can be used.

Developing the Distribution Of the Sample Mean

Developing a Sampling Distribution

- **Assume there is a population …**
- Population size N=4
- Random variable, x, is age of individuals
- Values of x: 18, 20, 22, 24 (years)

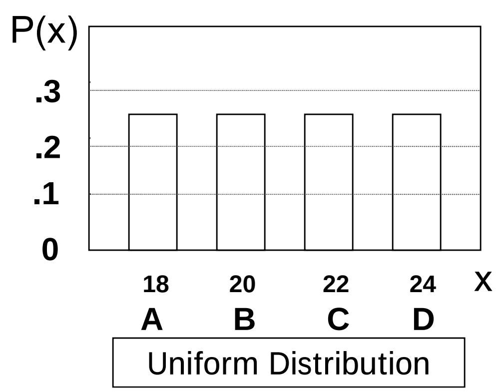
Developing a Sampling Distribution

(continued)

Summary Measures for the Population Distribution:

$$\mu = \frac{\sum x_i}{N} = \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = 2.236$$



Developing a Sampling Distribution

(continued)

Now consider all possible samples of size n=2

1st Obs	2nd Observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Developing a Sampling Distribution

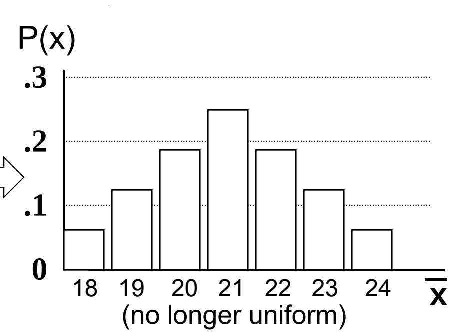
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



Developing a Sampling Distribution

(continued)

Summary Measures of this Sampling Distribution:

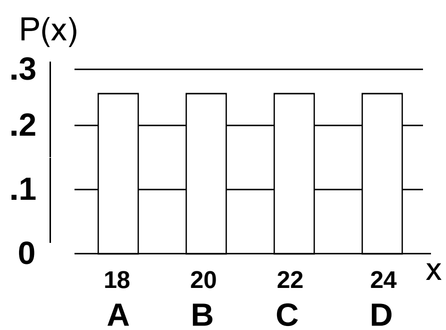
$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21$$

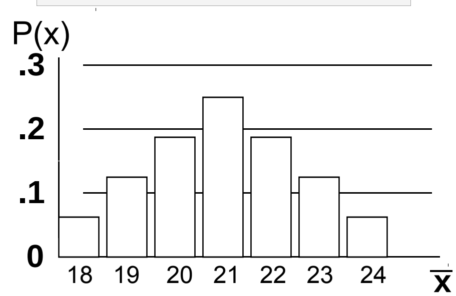
$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu_{\bar{x}})^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Comparing the Population with its Sampling Distribution

Population N = 4	Sample Means Distribution n = 2
$\mu = 21 \quad \sigma = 2.236$	$\mu_{\bar{x}} = 21 \quad \sigma_{\bar{x}} = 1.58$





Central Limit Theorem

a Normal Population

If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\overline{X}$ is also normally distributed with:

$$\mu_{\bar{x}} = \mu$$

and

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

Central Limit Theorem

a Normal Population

- Z value for the sampling distribution of $\bar{x}$

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

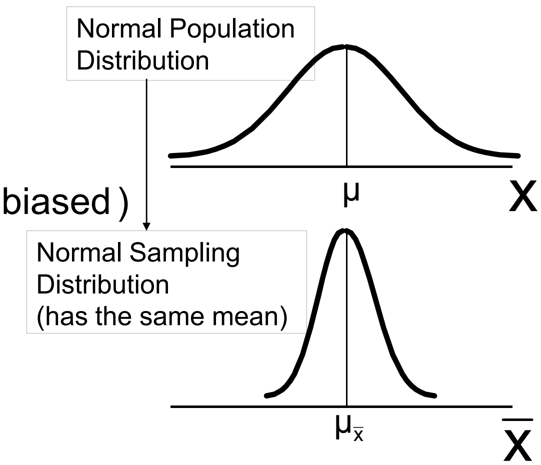
Where

$\bar{x}$	the sample mean
μ	the population mean
σ	population standard deviation
n	sample size

Sampling Distribution Properties

$$\mu_{\bar{x}} = \mu$$

(i.e. $\bar{x}$ is unbiased)

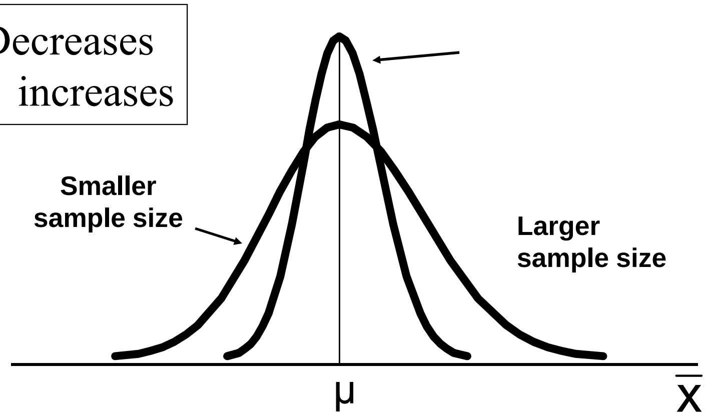


Sampling Distribution Properties

(continued)

- For sampling **with replacement**

$\sigma_{\bar{x}}$ Decreases As n increases



If the Population is **not** Normal

We can apply the Central Limit Theorem:

Even if the population is not normal,

…sample means from the population will be approximately normal as long as the sample size is large enough

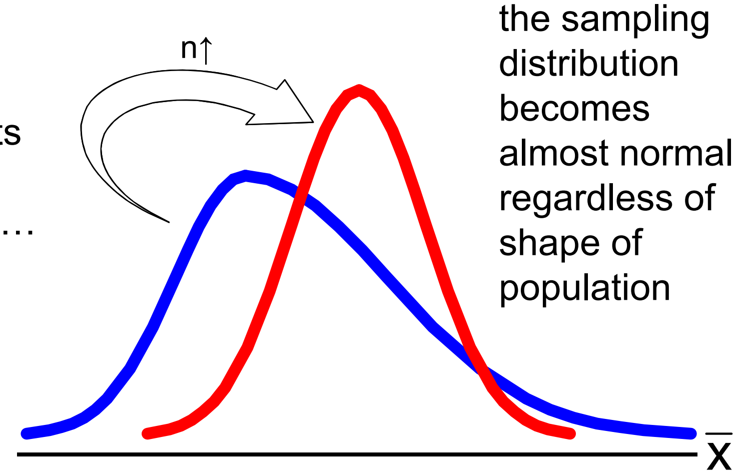
…and the sampling distribution will have

$$\mu_{\bar{x}} = \mu$$

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

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

If the Population is **not** Normal

(continued)

Sampling distribution properties:

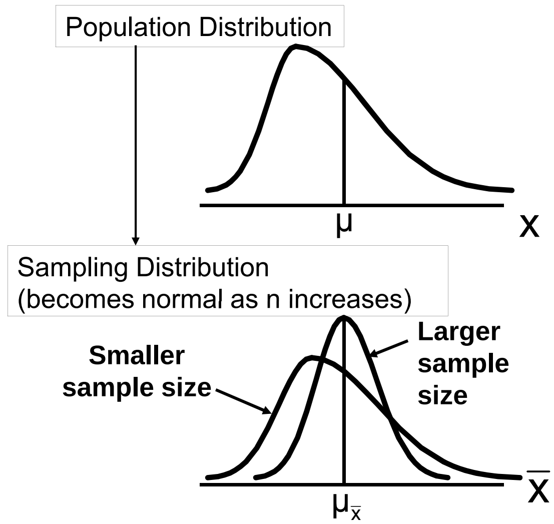
Central Tendency

$$\mu_{\bar{x}} = \mu$$

Variation

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

(Sampling with replacement)



Central Limit Theorem

The *Central Limit Theorem* states that for sufficiently large sample sizes ($n \geq 30$), regardless of the shape of the population distribution, if samples of size n are randomly drawn from a population that has a mean μ and a standard deviation σ, the samples' means $\overline{X}$ are approximately normally distributed. If the populations are normally distributed, the samples' means are normally distributed regardless of the sample sizes. The implication of this theorem is that for sufficiently large populations, the normal distribution can be used to analyze samples drawn from populations that are not normally distributed, or whose distribution characteristics are unknown.

Example 6.1

In an Electronics company that manufactures circuit boards, the average imperfection (defects) on a board is ***μ = 5*** *with a standard deviation of* ***σ = 2.34*** *when the production process is under statistical control.*

A random sample of ***n = 36*** *circuit boards has been taken for inspection and a mean of x = 6 defects per board was found.*

What is the probability of getting a value of ***x ≤ 6*** *if the process is under control?*

Solution 6.1

Because the sample size is greater than 30, the Central Limit Theorem can be used in this case even though the number of defects per board follows a Poisson distribution.

Therefore, the distribution of the sample mean *x is approximately normal with the standard deviation :*

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.34}{\sqrt{36}} 0.39$$

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{6-5}{0.39} = \frac{1}{0.39} = 2.56$$

The result *Z = 2.56 corresponds to 0.4948 on the table of normal curve areas:*

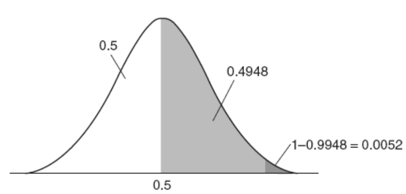


Table I Cumulative Standard Normal Distribution (continued)

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du$$

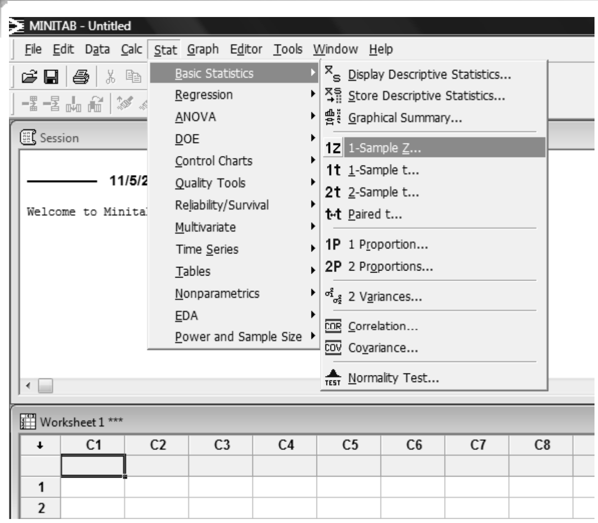
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856	0.0
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345	0.1
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092	0.2
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732	0.3
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933	0.4
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405	0.5
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903	0.6
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236	0.7
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267	0.8
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913	0.9
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143	1.0
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977	1.1
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475	1.2
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736	1.3
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888	1.4
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083	1.5
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486	1.6
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273	1.7
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621	1.8
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705	1.9
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691	2.0
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738	2.1
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989	2.2
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576	2.3
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613	2.4
⇨ 2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201	2.5
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427	2.6
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365	2.7
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074	2.8
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605	2.9
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999	3.0
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289	3.1
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499	3.2
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650	3.3
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758	3.4
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835	3.5
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888	3.6
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925	3.7
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950	3.8
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967	3.9

$Z=2.56$

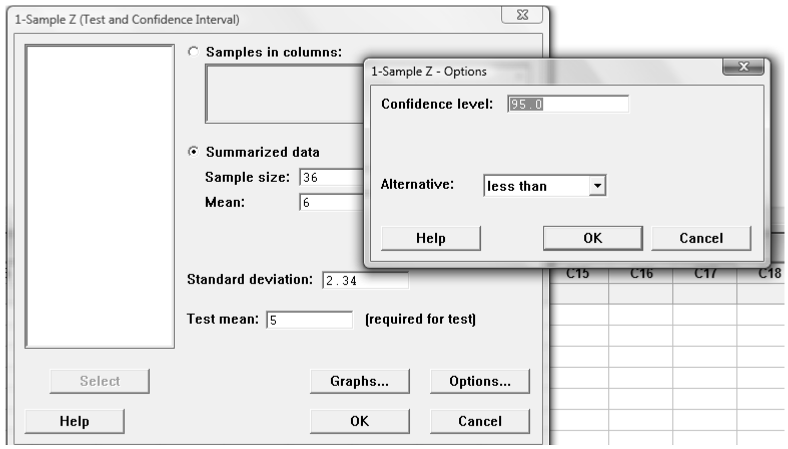
$P(z \le 2.56) = 0.9948$

$$P(\bar{x} \le 6) = 0.9948$$

Minitab Solution Example 6.1



Minitab Solution Example 6.1



Minitab *Continued*

One-Sample Z

```
Test of mu = 5 vs < 5
The assumed standard deviation = 2.34


                                95%
                              Upper
 N     Mean   SE Mean         Bound     Z      P
36  6.00000   0.39000       6.64149  2.56  0.995
```

Example 6.2

An Electronics company that manufactures resistors, the average resistance is ***μ = 100 Ohms*** *with a standard deviation of* ***σ = 10****.*

Find the probability that a random sample of ***n = 25*** *resistors will have an average resistance less than 95 Ohms*

Example 6.3

The average number of parts that reach the end of a production line defect-free at any given hour of the first shift is 372 parts with a standard deviation of 7.

What is the probability that a random sample of 34 different productions' first-shift hours would yield a sample mean between 369 and 371 parts that reach the end of the line defect-free?

Example 6.4

Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.

What is the probability that the sample mean is between 7.8 and 8.2?

6.4 Point Estimation

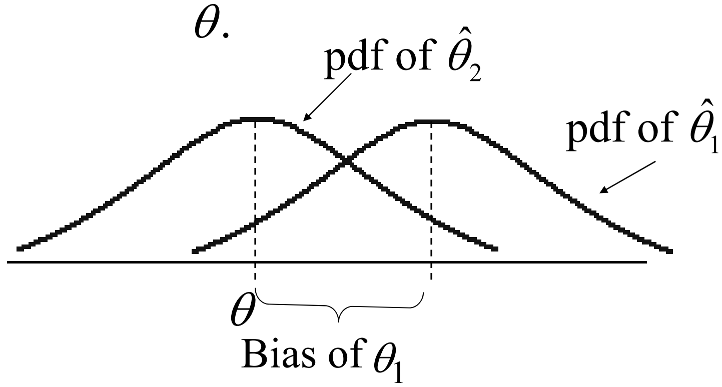
Point Estimator

A ***point estimator*** of a parameter θ is a single number that can be regarded as a sensible value for θ. A point estimator can be obtained by selecting a suitable statistic and computing its value from the given sample data.

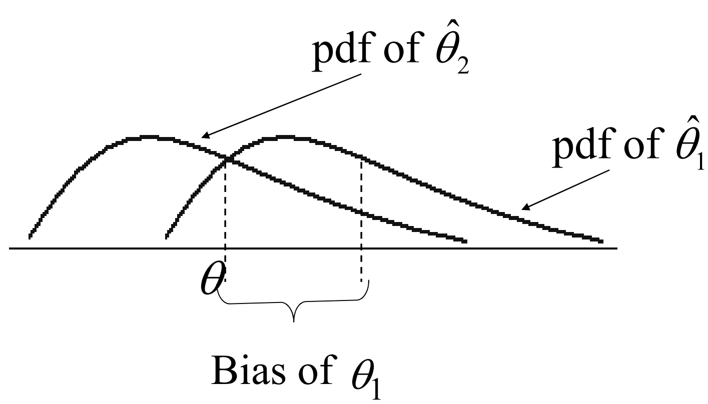
Unbiased Estimator

A ***point estimator*** $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is not biased, the difference $E(\hat{\theta}) - \theta$ is called the ***bias*** of $\hat{\theta}$.

The pdf's of a biased estimator $\hat{\theta}_1$ and an unbiased estimator $\hat{\theta}_2$ for a parameter θ.



The pdf's of a biased estimator $\hat{\theta}_1$ and an unbiased estimator $\hat{\theta}_2$ for a parameter θ.



Some Unbiased Estimators

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ, then $\overline{X}$ is an unbiased estimator of μ.

When X is a binomial rv with parameters n and p, the sample proportion $\hat{p} = X / n$ is an unbiased estimator of p.

Some Unbiased Estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the estimator

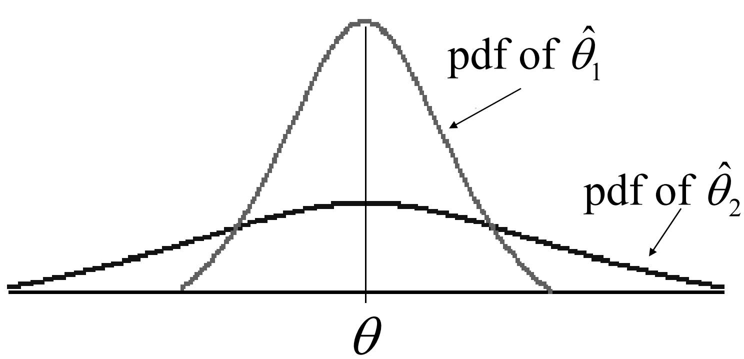
$$\hat{\sigma}^2 = S^2 = \frac{\sum \left(X_i - \overline{X} \right)^2}{n-1}$$

is an unbiased estimator.

Principle of Minimum Variance Unbiased Estimation (MVUE)

Among all estimators of θ that are unbiased, choose the one that has the minimum variance. The resulting $\hat{\theta}$ is called the ***minimum variance unbiased estimator (MVUE)*** of θ.

Graphs of the pdf's of two different unbiased estimators



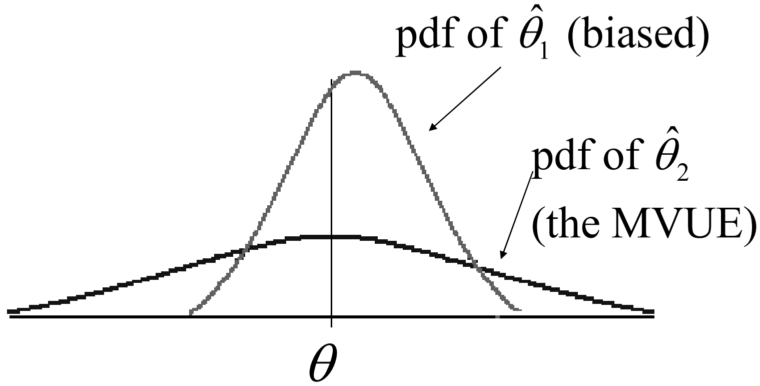
MVUE for a Normal Distribution

Let $X_1, X_2,\ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ

Then the estimator $\hat{\mu} = \overline{X}$ is the MVUE for μ

.

A biased estimator that is preferable to the MVUE



Standard Error

The ***standard error*** of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. If the standard error itself involves unknown parameters whose values can be estimated, substitution into $\sigma_{\hat{\theta}}$ yields the ***estimated standard error*** of the estimator, denoted

$$\hat{\sigma}_{\hat{\theta}} \text{ or } s_{\hat{\theta}}.$$

Confidence Intervals

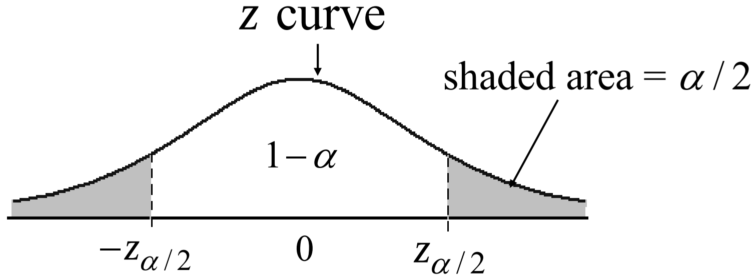
An alternative to reporting a single value for the parameter being estimated is to calculate and report an entire interval of plausible values – a ***confidence interval*** **(CI).** A ***confidence level*** is a measure of the degree of reliability of the interval.

95% Confidence Interval

If after observing $X_1 = x_1, \ldots, X_n = x_n$, we compute the observed sample mean $\overline{x}$, then a ***95% confidence interval*** for μ the mean of normal population can be expressed if σ known as:

$$\left(\overline{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Other Levels of Confidence



$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha$$

Other Levels of Confidence

A $100(1-\alpha)\%$ confidence interval for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}},\, \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$$

Sample Size

The general formula for the sample size *n* necessary to ensure an interval width *w* is

$$n = \left(z_{\alpha/2}\cdot\frac{\sigma}{w}\right)^2$$