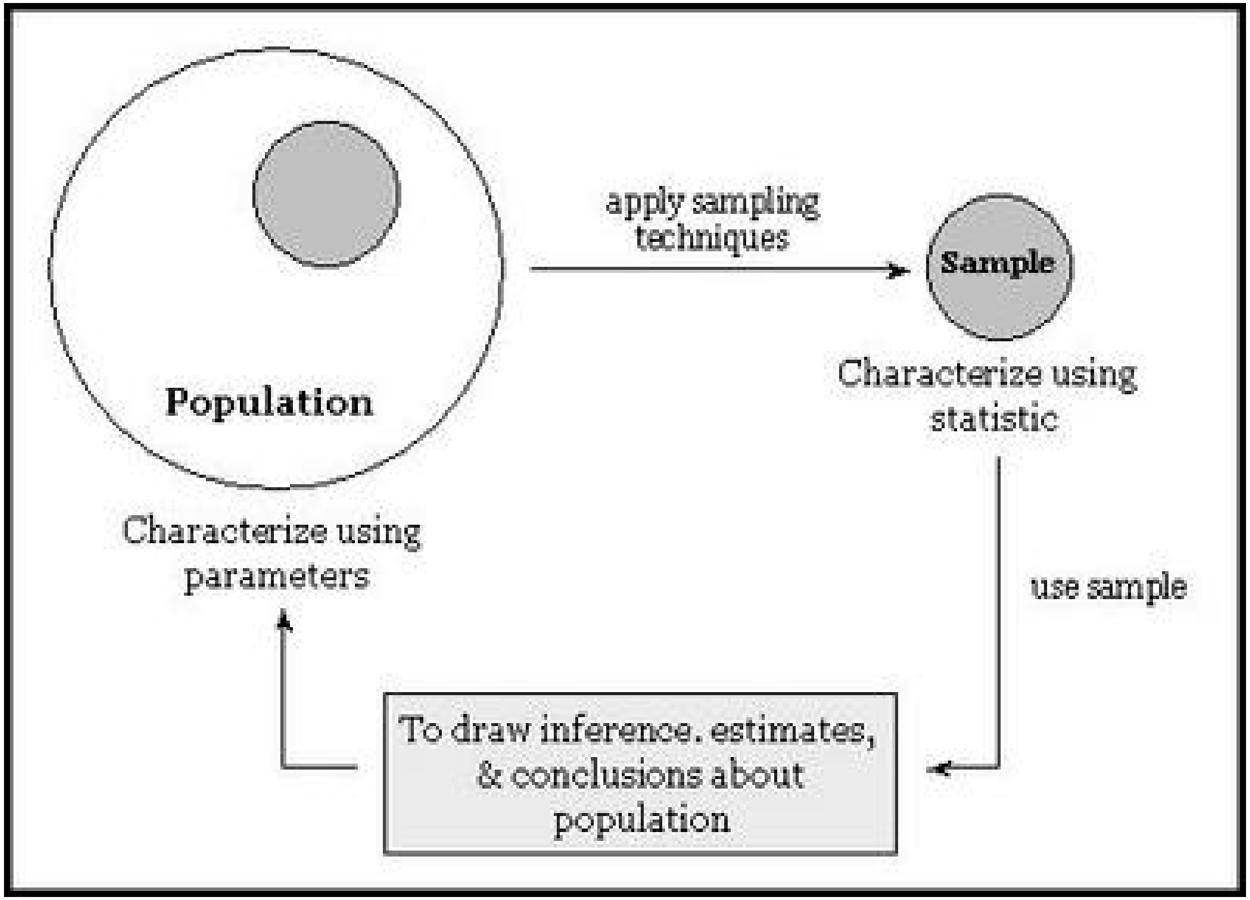
Statistical Estimation

Inferential Statistics



Source: http://www.discover6sigma.org/post/2005/12/statistics-simplified/

Main Objectives

- Estimation: To estimate the parameters on the basis of sample observations through a statistic.
- Hypothesis Testing: To compare these parameters among themselves on the basis of observations and their estimates.

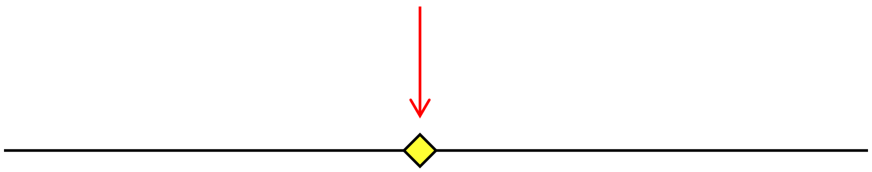
Introduction

- The objective of estimation is to determine the ***approximate value*** of a population parameter on the basis of a sample statistic.

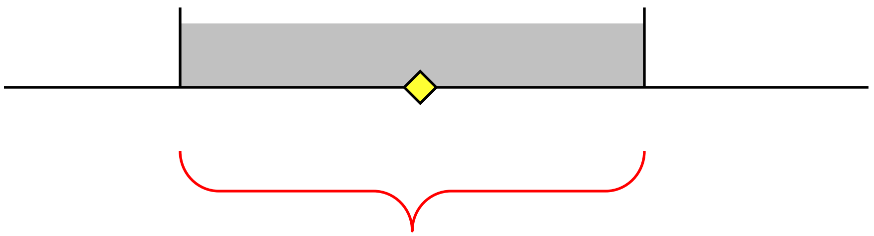
- E.g., the sample mean ($\bar{X}$) is employed to ***estimate*** the population mean (μ).

Types of Estimators

- Point Estimation:
  - A single number



- Interval Estimation
  - provides additional information about variability



Criteria of a Good Estimator

- Unbiasedness: Suppose $\hat{\theta}$ is an estimator of parameter θ.

$$E[\hat{\theta}] - \theta = 0$$

- Efficiency: smaller standard error.
- Consistency: As *n* increases $\hat{\theta} \to \theta$
- Sufficiency: extracts information from the sample.

Point Estimation

- A ***point estimator*** draws inferences about a population by estimating the value of an unknown parameter using a single value or point.
- We can estimate the population mean, μ using sample mean

$$\bar{x} = \frac{\Sigma x}{n}$$

Point Estimation

- We saw earlier that point probabilities in continuous distributions were virtually zero.

- How much uncertainty is associated with a point estimate of a population parameter?

Interval Estimation

- An ***interval estimator*** draws inferences about a population by estimating the value of an unknown parameter using an interval.

- That is we say (with some percent certainty or confidence) that the population parameter of interest is between some lower and upper bounds.

Example

- For example, suppose we want to estimate the mean summer income of a class of business students. For n=25 students, $\bar{x}$ is calculated to be 400 $/week.

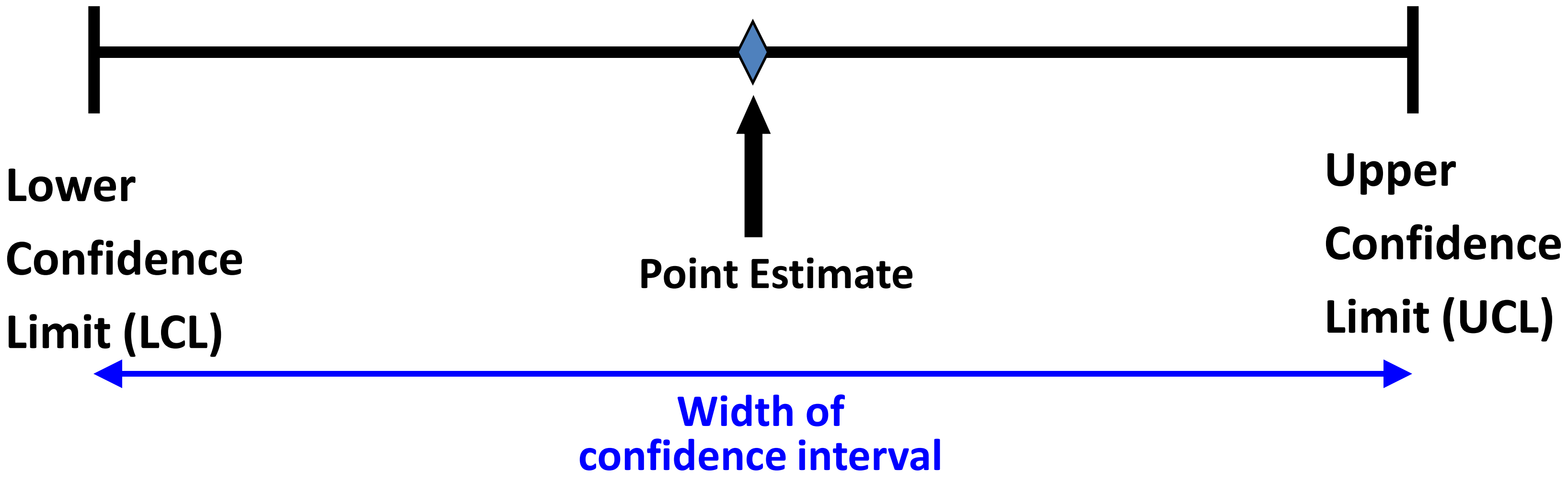
point estimate

interval estimate

- An alternative statement is:

The mean income is ***between*** 380 and 420 $/week with 95% confidence.

Confidence Interval

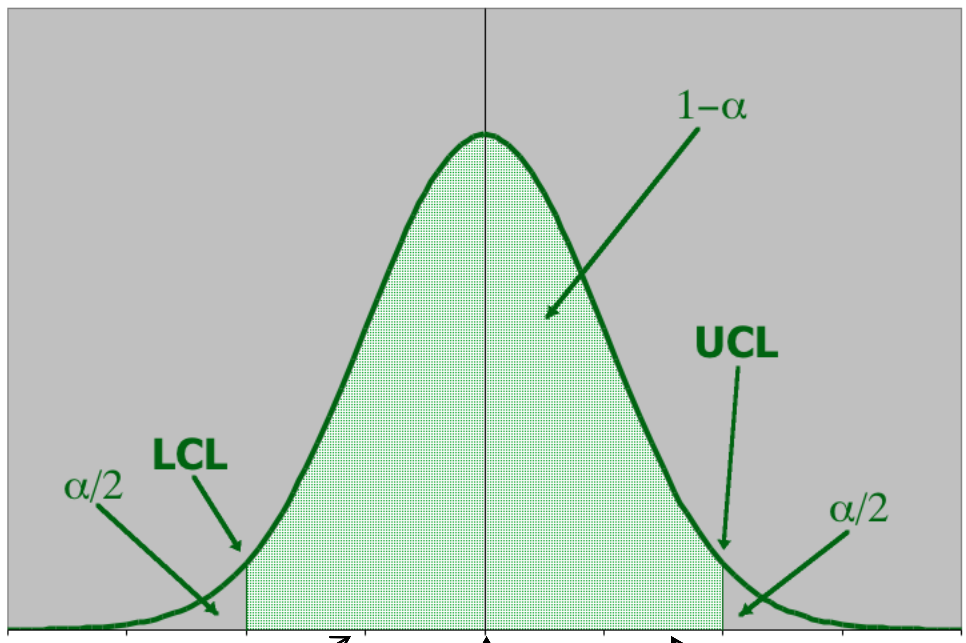


Level of Confidence

- Probability that the unknown population parameter falls within the interval
- Denoted $(1 - \alpha)\ \% =$ level of confidence

  α is probability that the parameter is **Not** within the Interval.
- If we say that the population mean, μ falls within the interval a and b with 95% confidence (i.e., $\alpha = 0.5$), then mathematically, $P(a < \mu < b) = 0.95$

Graphical Representation



… μ may be here…

…or here

…or possibly even here…

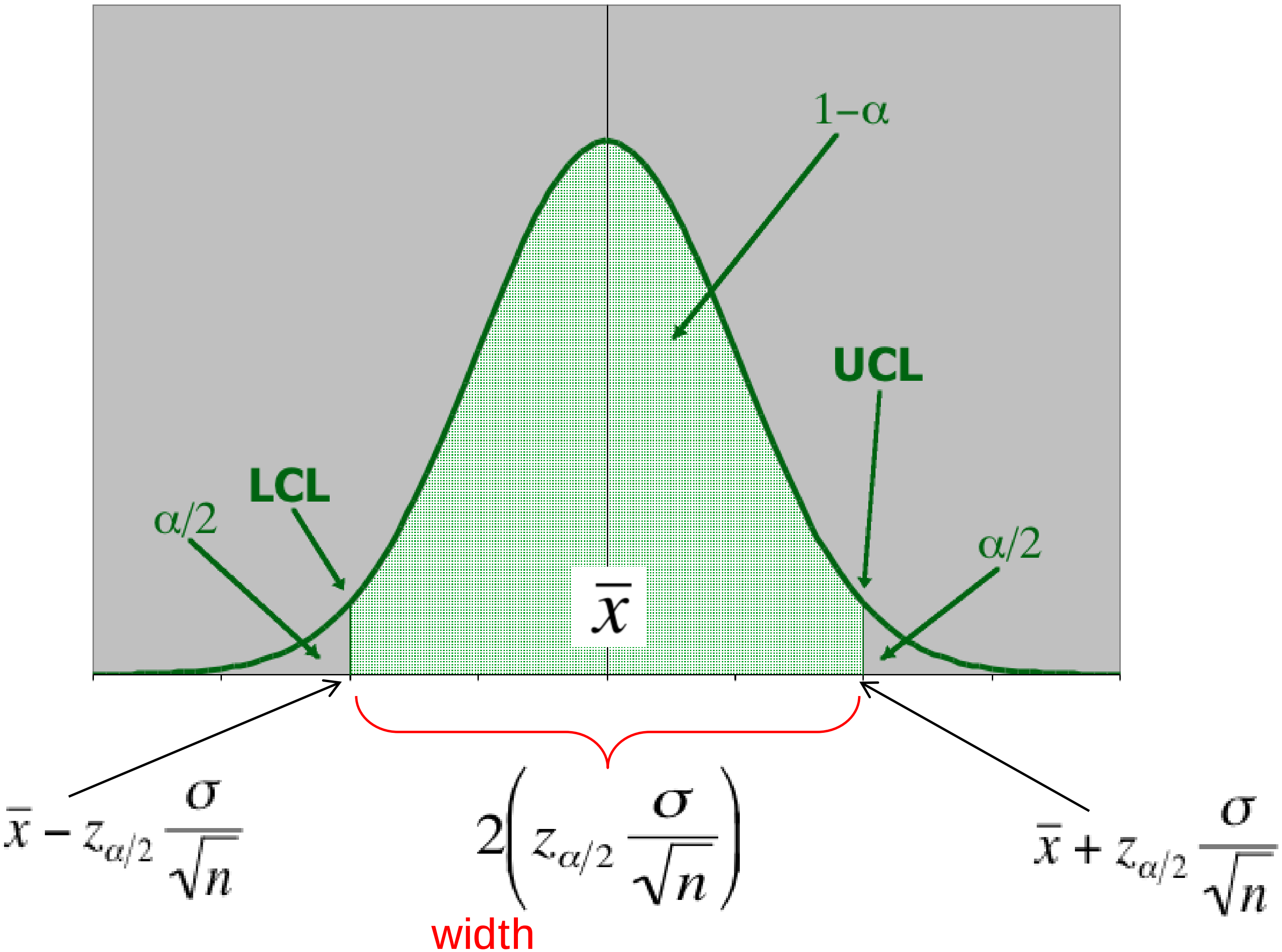
Estimating μ from Large Samples When σ is Known

- the *probability* that the interval:

$$\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \left\{\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

contains the population mean μ is 1 – α. This is the *interval estimator for* μ with (1 – α)% confidence.

Graphical Representation



Commonly Used Confidence Levels

$1-\alpha$	α	$\alpha/2$	$z_{\alpha/2}$
.90	.10	.05	$z_{.05} = 1.645$
.95	.05	.025	$z_{.025} = 1.96$
.98	.02	.01	$z_{.01} = 2.33$
.99	.01	.005	$z_{.005} = 2.575$

Example

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms. Determine a 95% confidence interval for the true mean resistance of the population.
- Given: $X \sim N(\mu, 0.35^2)$; $n = 11$, $\overline{X} = 2.20$ ohm

  $1 - \alpha = 0.95 \Rightarrow \alpha/2 = 0.025$

Solution

- Solution: $\overline{X} \pm Z \frac{\sigma}{\sqrt{n}}$

$$= 2.20 \pm 1.96\,(0.35/\sqrt{11})$$

$$= 2.20 \pm 0.2068$$

$$1.9932 \le \mu \le 2.4068$$

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean

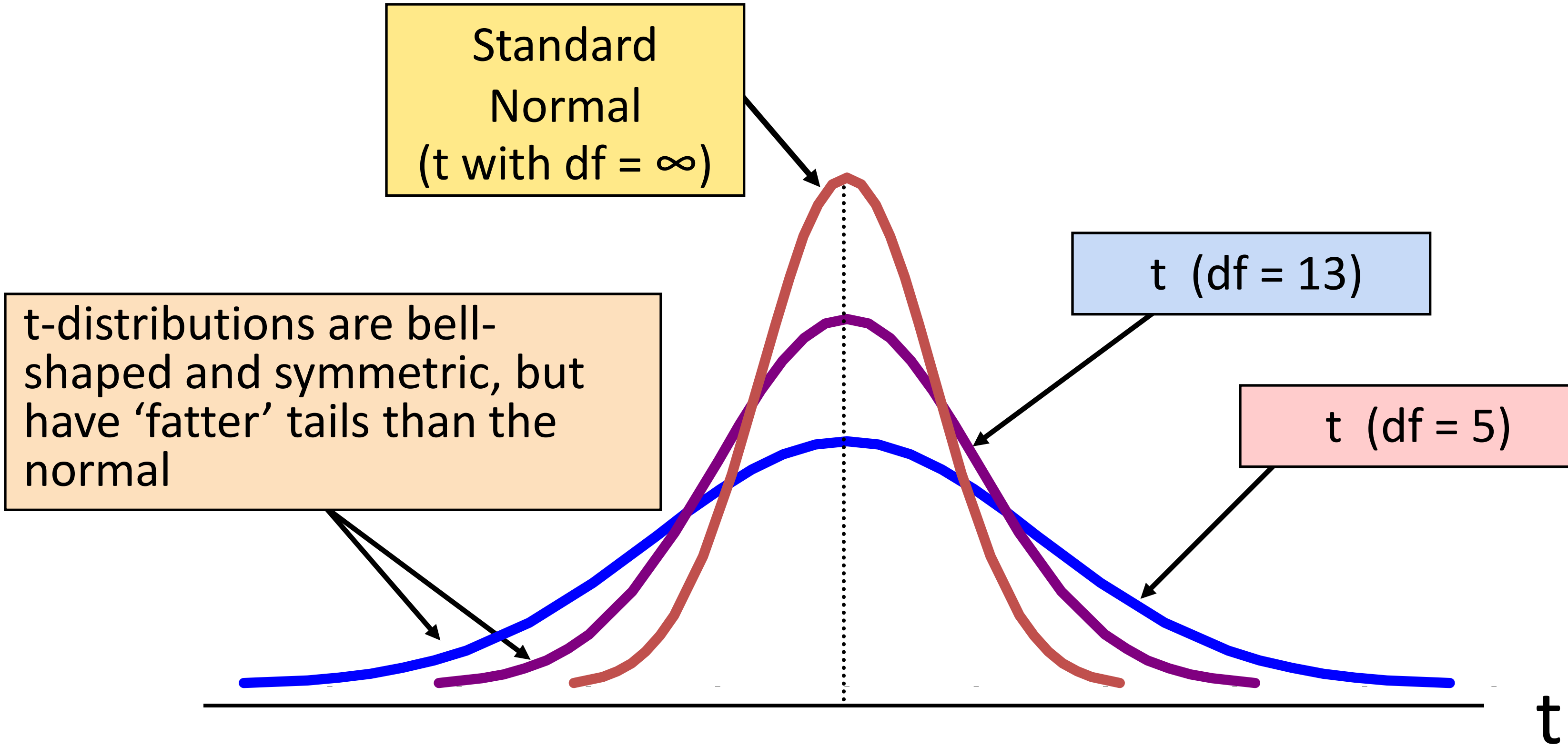
Estimating μ from Small Samples

- When sample size is 30 or less normal distribution is not the appropriate sampling distribution.
- When sample size is 30 or less and population standard deviation is unknown, t-distribution (or Student's t-distribution) is more appropriate.
- Interval Estimate:

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} = \left\{ \bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \right\}$$

Student's t-distribution

Note: $t \longrightarrow Z$ as n increases



Degrees of Freedom (*df*)

- Number of observations that are free to vary after sample statistic has been calculated

Example:

Sum of 3 Numbers Is 6

X_1 = 1 (or Any Number)

X_2 = 2 (or Any Number)

X_3 $\underline{= \mathbf{3}}$ (Cannot Vary)

Sum = 6

degrees of freedom = n -1

= 3 -1

= 2

Example

- A random sample of n = 25 taken from a normal population has $\bar{X}$ = 50 and S = 8. Form a 95% confidence interval for μ.

1 - α = 0.95 => α = .05

df = n – 1 = 24, so $t_{\alpha/2,\,n-1} = t_{0.025,24} = 2.064$

[In the Students' t table, 2.064 is the t value corresponding to α = .05 and df = 24.]

The confidence interval is

$$\bar{X} \pm t_{\alpha/2,\,n-1}\frac{S}{\sqrt{n}} = 50 \pm (2.0639)\frac{8}{\sqrt{25}}$$

= [46.698 , 53.302]

Confidence Intervals for the Population Proportion, π

- An interval estimate for the population proportion (π) can be calculated by adding and subtracting an allowance for uncertainty to the sample proportion (p)

Confidence Intervals for the Population Proportion, π

- the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- We will estimate this with sample data:

$$\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals for the Population Proportion, π

- Upper and lower confidence limits for the population proportion are calculated with the formula

$$p \pm Z\sqrt{\frac{p(1-p)}{n}}$$

- where
  - Z is the standard normal value for the level of confidence desired
  - p is the sample proportion
  - n is the sample size

Example

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

$$p \pm Z\sqrt{p(1-p)/n}$$

$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$

$$= 0.25 \pm 1.96\,(0.0433)$$

$$0.1651 \leq \pi \leq 0.3349$$

Determination of Sample Size

- The required sample size needed to estimate a population parameter to within a selected margin of error (e) using a specified level of confidence (1 - α) can be computed

- The margin of error is also called sampling error
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

Determining Sample Size for the Mean

Sampling error
(margin of error)

$$\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} \rightarrow e = Z\frac{\sigma}{\sqrt{n}}$$

Now solve for n
to get

$$\boxed{n = \frac{Z^2 \sigma^2}{e^2}}$$

Determining Sample Size for the Mean

- To determine the required sample size for the mean, you must know:

  – The desired level of confidence $(1 - \alpha)$, which determines the critical Z value

  – The acceptable sampling error, e

  – The standard deviation, σ

Example

- If σ = 45, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{Z^2\ \sigma^2}{e^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220 (Always round-up)**