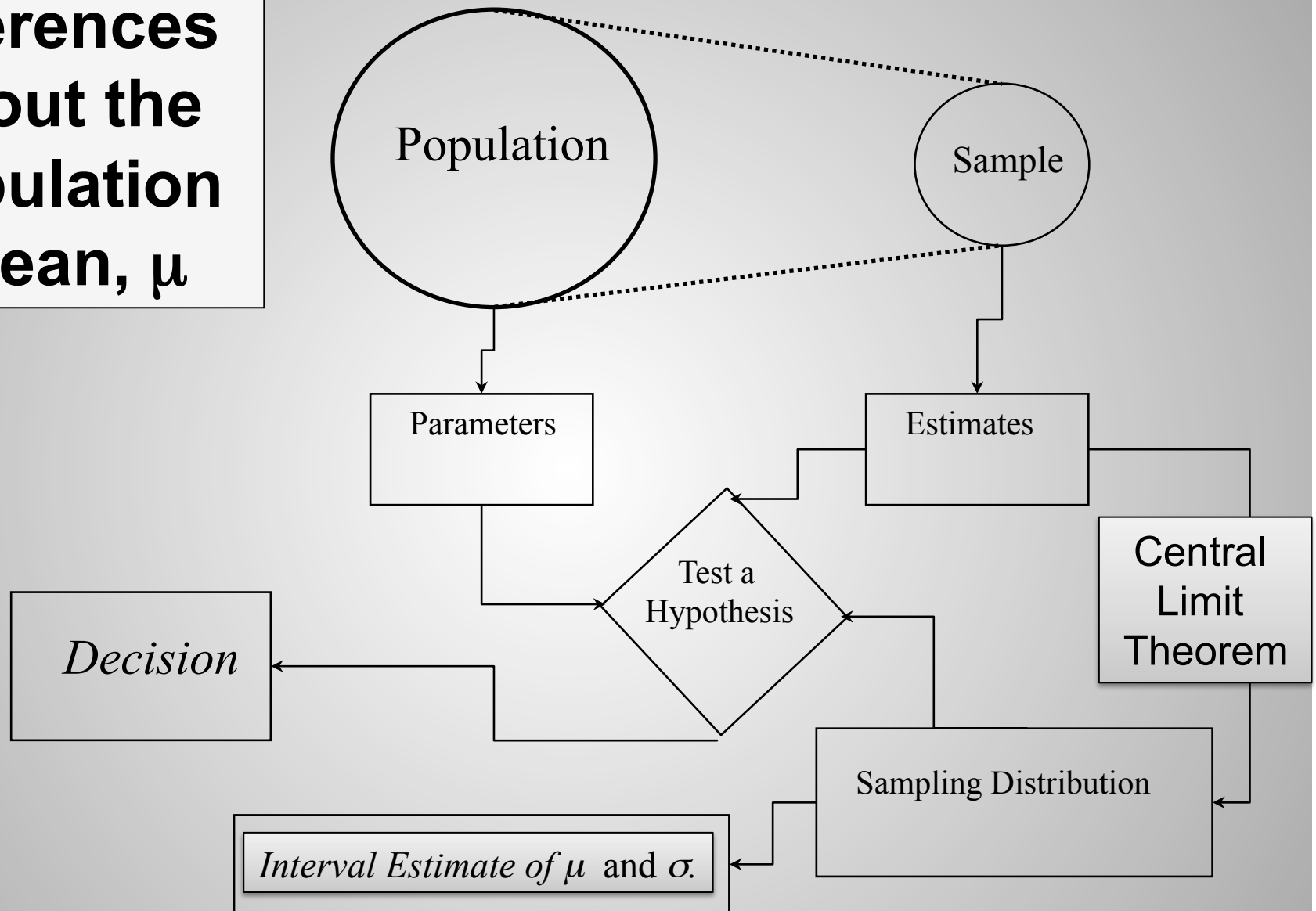


One Sample Mean Inference (Chapter 5)

Points to discuss:

- How to estimate the mean from a normal distribution and compute a 95% confidence interval for the true unknown population mean.
- How to statistically test whether the population sampled could have a specified population mean.
- The five parts of a statistical test.
- Type I and Type II errors associated with testing.
- In general how sample size is determined.
- Inferences about the median.

Inferences about the population mean, μ



Estimation of the Population Mean, μ .

Point Estimate of μ is the sample mean:

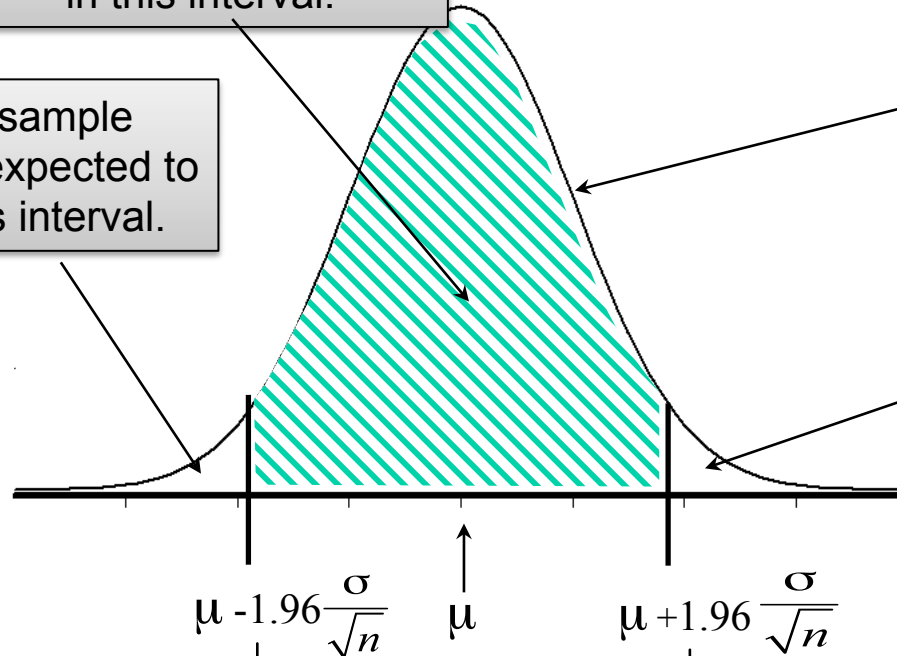
The Central Limit Theorems tell us the shape of the sampling distribution

95% of sample means expected to lie in this interval.

2.5% of sample means expected to lie in this interval.

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

2.5% of sample means expected to lie in this interval.



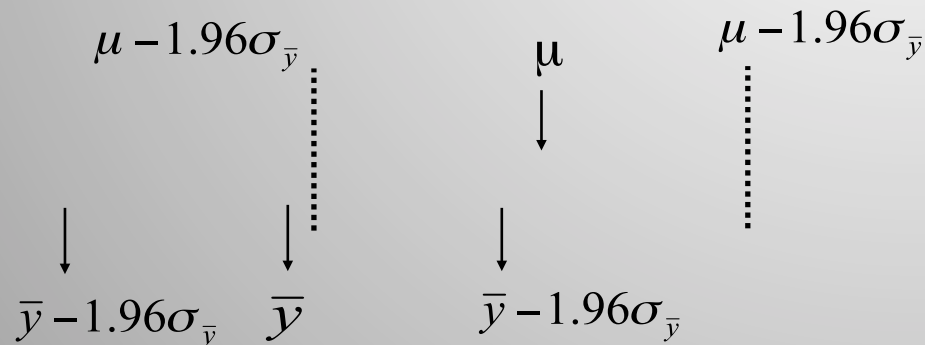
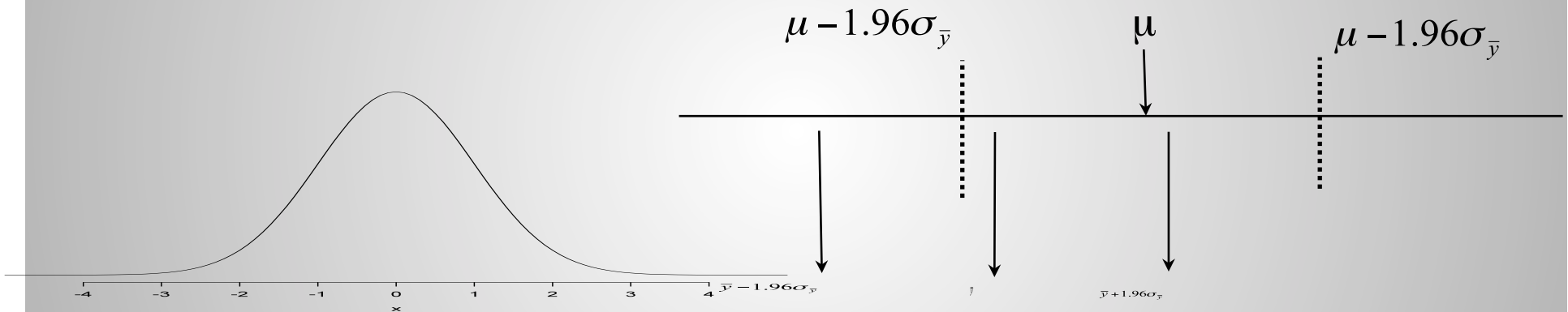
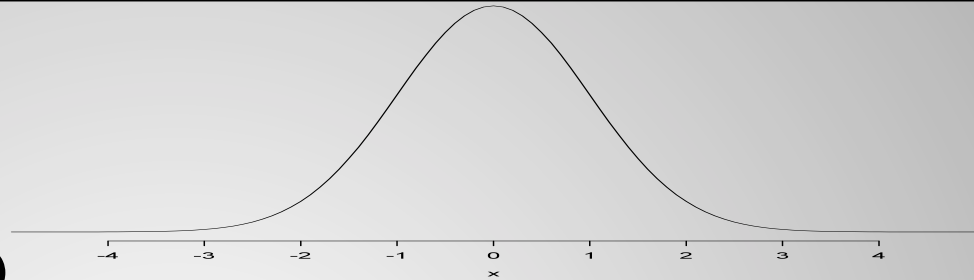
$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

Interval Estimate of μ is the confidence interval:

Interval Estimate of the Population Mean

The event

\bar{y} in $(\mu - 1.96\sigma_{\bar{y}}, \mu + 1.96\sigma_{\bar{y}})$

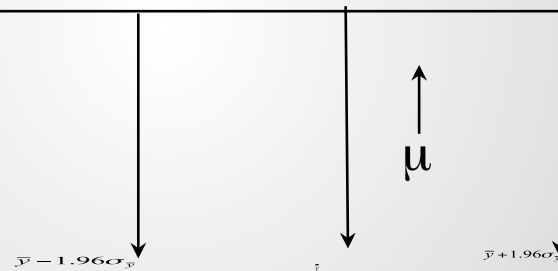
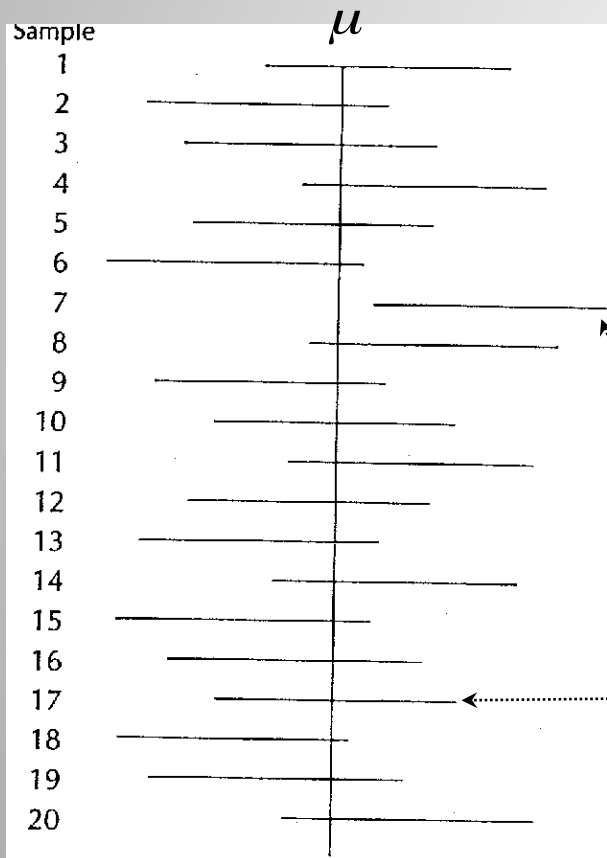
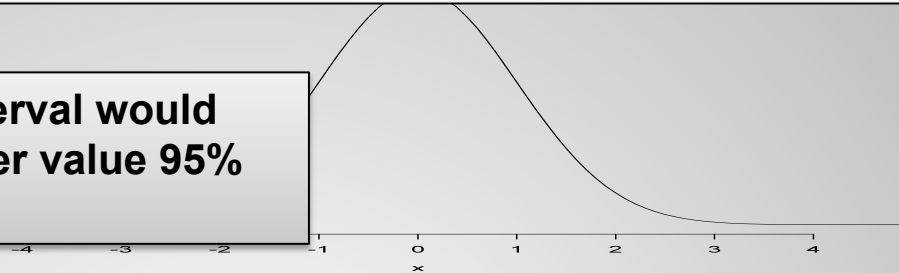


is equivalent to the event

μ in $(\bar{y} - 1.96\sigma_{\bar{y}}, \bar{y} + 1.96\sigma_{\bar{y}})$

Meaning of a 95% Confidence Interval

In repeated sampling, the interval would encompass the true parameter value 95% of the time.



5% of the intervals don't include the true mean.
95% of the intervals include the true mean.

Critical Value for Confidence Interval

95% Confidence Interval

$$\bar{y} \pm 1.96\sigma_{\bar{y}}$$

99% Confidence Interval

$$\bar{y} \pm 2.58\sigma_{\bar{y}}$$

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

(1- α)100% Confidence Interval

$$\bar{y} \pm z_{\alpha/2}\sigma_{\bar{y}}$$

- The $\alpha/2$ critical value of the standard normal distribution.
- The value of the standard normal distribution for which $P(z > z_{\alpha/2}) = \alpha/2$.

Read from Table 1 in
Ott & Longnecker

Common Critical Values

| Confidence Coefficient ($1-\alpha$) | Area to left of z-value $1-\alpha/2$ | Area to right of z-value $\alpha/2$ | Corresponding z-value $z_{\alpha/2}$ |
|--|--|---|--|
| .90 | .95 | .05 | 1.645 |
| .95 | .975 | .025 | 1.96 |
| .98 | .99 | .01 | 2.33 |
| .99 | .995 | .005 | 2.58 |

Using an Estimate of σ

What if we don't know the value of σ ?

We simply replace σ with its estimate s in the confidence interval equation.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

However, the resulting sampling distribution is no longer normal, but follows a **t distribution** instead. This involves replacing the normal critical value, $z_{\alpha/2}$, with a critical value from a t distribution, $t_{\alpha/2, n-1}$ (more on this later).

If n is large (30 or more), it turns out that because of the Central Limit Theorem the normal is a good enough approximation to the t, and we can therefore use $z_{\alpha/2}$, after all.

Example

$$n=50 \quad \bar{y}=105 \quad s=11$$

95% CI for μ

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$z_{\alpha/2} = 1.96$$

$$\bar{y} \pm z_{\alpha/2} \sigma / \sqrt{n} \quad \text{or} \quad \bar{y} \pm z_{\alpha/2} s / \sqrt{n}$$
$$105 \pm (1.96) \left(\frac{11}{\sqrt{50}} \right)$$
$$(101.95, 108.05)$$

In a large amount of repeated sampling, 95 out of 100 intervals constructed this way would include the true mean.

99% CI for μ

$$\alpha = 0.01$$

$$\alpha/2 = 0.005$$

$$z_{\alpha/2} = 2.58$$

$$\bar{y} \pm z_{\alpha/2} \sigma / \sqrt{n} \quad \text{or} \quad \bar{y} \pm z_{\alpha/2} s / \sqrt{n}$$
$$105 \pm (2.58) \left(\frac{11}{\sqrt{50}} \right)$$
$$(100.98, 109.01)$$

In a large amount of repeated sampling, 99 out of 100 intervals constructed this way would include the true mean.

Sample Size Determination

The CI equation can be turned around to allow computation of the **sample size** needed to estimate the true mean with a specified level of precision and confidence.

Need three things:

- 1) Target level of confidence: $(1-\alpha)100\%$
- 2) Tolerable margin of error (half of CI width): E
- 3) Estimate of population standard deviation: σ

Typical Confidence Levels

90%

95%

99%

Target Precision

- i) Prior Knowledge
- ii) Pilot Sample
- iii) Expert Guess (range/4)

100(1- α)% CI for μ

$$\bar{y} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Equate this to E.

Equate the CI precision to the target precision and solve for n.

$$Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = E \quad \Rightarrow \quad Z_{\alpha/2} \frac{\sigma}{E} = \sqrt{n} \quad \Rightarrow \quad \left(Z_{\alpha/2} \frac{\sigma}{E} \right)^2 = n$$

Precision can be specified relative to σ .

Sample size!

Sample Size Example

$$\sigma = 2 \text{ oz}$$

$$E = \pm 0.25$$

$$(1 - \alpha)100\% = 95\%$$

$$\frac{\alpha}{2} = .025 \quad Z_{\alpha/2} = 1.96$$

$$\left(\frac{1.96 \cdot 2}{.25} \right)^2 = (4 \cdot 2 \cdot 1.96)^2 = 246 \quad (\text{round to next highest integer})$$

$$\left(Z_{\alpha/2} \frac{\sigma}{E} \right)^2 = n$$

Take home message: If you have some idea of the underlying variability in the population and you know roughly how precise you wish the mean estimate to be you can compute the required sample size.