Confidence Levels, Intervals and T-test

Learning Objectives

- Know the difference between point and interval estimation.
- Estimate a population mean from a sample mean for large sample sizes.
- Estimate a population mean from a sample mean for small sample sizes.
- Estimate a population proportion from a sample proportion.
- Estimate the minimum sample size necessary to achieve given statistical goals.

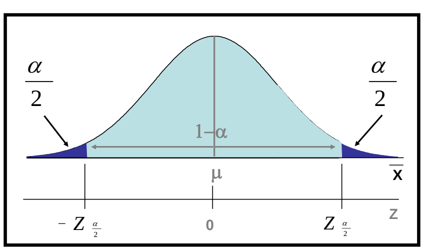
Statistical Estimation

- **Point estimate** -- the single value of a statistic calculated from a sample
- **Interval Estimate** -- a range of values calculated from a sample statistic(s) and standardized statistics, such as the *Z*.
  - Selection of the standardized statistic is determined by the sampling distribution.
  - Selection of critical values of the standardized statistic is determined by the desired level of confidence.

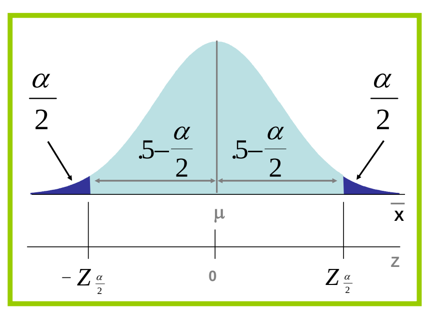
Confidence Interval to Estimate μ when n is Large

- Point estimate $\bar{X} = \frac{\sum X}{n}$
- Interval Estimate $\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$ *or* $\bar{X} - Z \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z \frac{\sigma}{\sqrt{n}}$

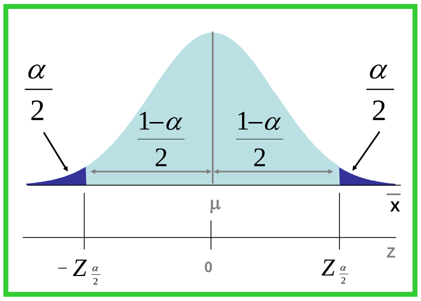
Distribution of Sample Means for (1-α)% Confidence level α - Significance Level



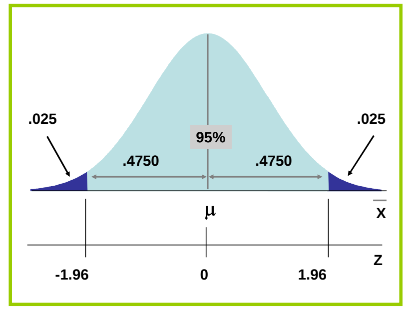
Z Scores for Confidence Intervals in Relation to α



Distribution of Sample Means for (1-α)% Confidence



Distribution of Sample Means for 95% Confidence



Z Values for Some of the More Common Levels of Confidence

Confidence Level	Z Value
90%	1.645
95%	1.96
98%	2.33
99%	2.575

Probability Interpretation of the Level of Confidence

$$\Pr ob[\overline{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}] = 1 - \alpha$$

Exercise: 95% Confidence Interval for μ

$$\overline{X} = 4.26, \ \sigma = 1.1, \ and \ \ n = 60.$$

$$\overline{X} - Z\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z\frac{\sigma}{\sqrt{n}}$$

$$4.26 - 1.96\frac{1.1}{\sqrt{60}} \le \mu \le 4.26 + 1.96\frac{1.1}{\sqrt{60}}$$

$$4.26 - 0.28 \le \mu \le 4.26 + 0.28$$

$$3.98 \le \mu \le 4.54$$

Confidence Interval to Estimate μ when n is Large and σ is Unknown

$$\overline{X} \pm Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

or

$$\overline{X} - Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

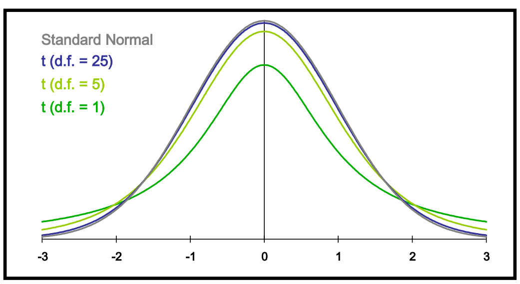
Estimating the Mean of a Normal Population: Small *n* and Unknown σ

- The population has a normal distribution.
- The value of the population standard deviation is unknown.
- The sample size is small, $n < 30$.
- *Z* distribution is not appropriate for these conditions
- *t* distribution is appropriate

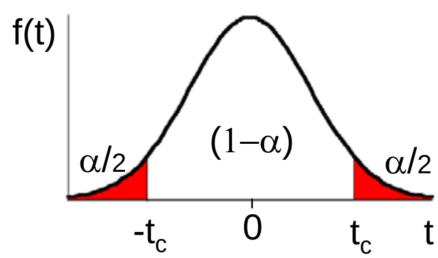
The *t* Distribution

- A family of distributions -- a unique distribution for each value of its parameter, degrees of freedom (d.f.)
- Symmetric, Unimodal, Mean = 0, Flatter than a *Z*
- *t* formula $$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Comparison of Selected *t* Distributions to the Standard Normal

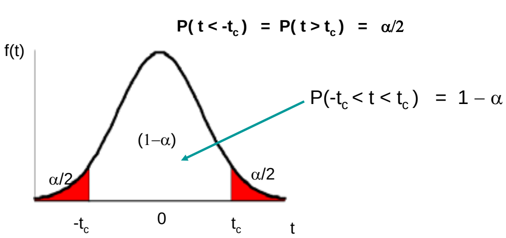


The t distribution



red area = rejection region for 2-sided test

Probability Statements



Excel: P =TDIST(t_c, n-1, #_of_tails)
Inverse: tc =TINV(a, n-1)

Inference about a population MEAN when the population STDEV Is Unknown

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- When the sampled population is normally distributed, the t statistic is Student t distributed with n-1 degrees of freedom.
- Confidence Interval: $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ where $t_{\alpha/2,n-1}$ is the $\alpha/2$ quantile of the Student t-distribution with n-1 degrees of freedom.

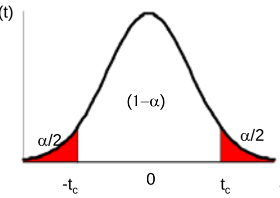
Checking the required conditions

- In deriving the test and confidence interval, we have made two assumptions:
  - (i) the sample is a random sample from the population;
  - (ii) the distribution of the population is normal.
- The t test is robust – the results are still approximately valid as long as
  - (i)the population is not extremely non-normal.
  - (ii) or, if the sample size is large.

Table of Critical Values of t

α/2

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.327	2.576



Excel: t_c =TINV(α, n-1)

Confidence Intervals for μ of a Normal Population: Small n and Unknown σ

$$\bar{X} \pm t \frac{S}{\sqrt{n}}$$

or

$$\bar{X} - t \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t \frac{S}{\sqrt{n}}$$

$$df = n - 1$$

Ex: Find 99% Confidence Interval

$$\bar{X} = 2.14, \; S = 1.29, \; n = 14, \; df = n - 1 = 13$$

$$\frac{\alpha}{2} = \frac{1 - .99}{2} = 0.005$$

$$t_{.005,13} = 3.012$$

$$\bar{X} - t \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t \frac{S}{\sqrt{n}}$$

$$2.14 - 3.012 \frac{1.29}{\sqrt{14}} \le \mu \le 2.14 + 3.012 \frac{1.29}{\sqrt{14}}$$

$$2.14 - 1.04 \le \mu \le 2.14 + 1.04$$

$$1.10 \le \mu \le 3.18$$

Determining Sample Size when Estimating μ

- Z formula $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- Error of Estimation (tolerable error) $E = \bar{X} - \mu$
- Estimated Sample Size $n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2}{E^2} = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{E} \right)^2$

Estimated σ $\sigma \approx \frac{1}{4} range$

Ex: Sample Size when Estimating μ

$$E = 1, \quad \sigma = 4$$

$$90\% \text{ confidence} \Rightarrow t = 1.645$$

$$n = \frac{t_{\frac{\alpha}{2}}^2 \sigma^2}{E^2} = \frac{(1.645)^2 (4)^2}{1^2} = 43.30 \text{ or } 44$$

Example

- Companies that sell groceries over the Internet are called e-grocers. Customers enter their orders, pay by credit card, and receive delivery by truck. A potential e-grocer analyzed the market and determined that to be profitable the average order would have to exceed $85. To determine whether an e-grocer would be profitable in one large city, she offered the service and recorded the size of the order for a random sample of customers. Can we infer from the data that the e-grocery will be profitable in this city at significance level 0.05?