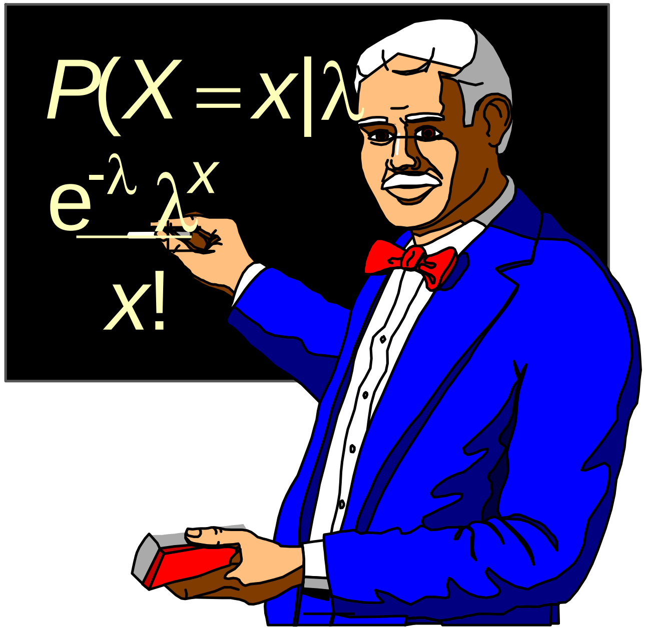
Phys 443
Computational Physics

Estimate



Types of Inference

- Whether you are estimating parameters or testing hypotheses, statistical methods are important because they provide:
  - **Methods for making the inference**
  - **A numerical measure of the goodness or reliability of the inference**

Point estimation

Sample data is used to estimate parameters of a population

Statistics are calculated using sample data.

Parameters are the characteristics of population data

sample mean

$\bar{x}$

Sample SD

s

estimates

Population mean

μ

Population SD

σ

Notation hint: population notation is mostly greek; sample latin.

Point estimation

An **estimator** is a rule, usually a formula, that tells you how to calculate the estimate based on the sample.

– **Point estimation:** A single number is calculated to estimate the parameter.

Point Estimator of Population Mean

An point estimate of population mean, μ , is the sample mean

$$\bar{x} = \frac{\sum x_i}{n}$$

A sample of heigths of 34 male freshman students was obtained.

185	161	174	175	165	178	182	149	177
170	151	176	197	154	183	184	189	168
168	170	178	180	167	177	166	166	176
184	179	155	148	180	194	176		

If one wanted to estimate the true mean of all male freshman students, you might use the sample mean as a point estimate for the true mean.

Sample mean = $\bar{x}$ = 173.00

Point Estimation of Population Proportion

An point estimate of population mean, p, is the sample proportion, where x is the number of successes in the sample. $\hat{p} = x / n$

A sample of 200 students at a large university is selected to estimate the proportion of students that wear contact lens. In this sample 47 wear contact lens.

$$\hat{p} = 47 / 200 = .235$$

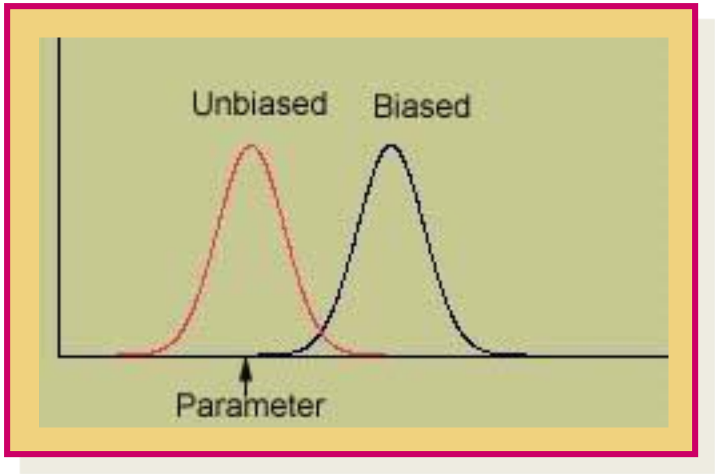
Properties of Point Estimators

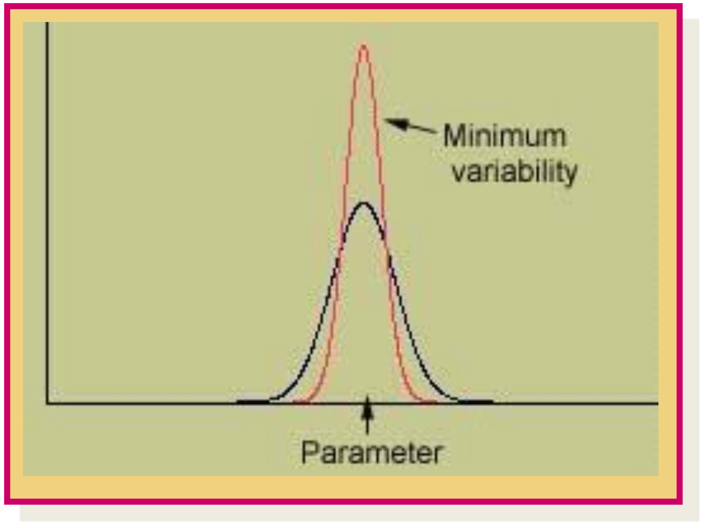
- Since an estimator is calculated from sample values, it varies from sample to sample according to its **sampling distribution.**
- An **estimator** is **unbiased** if the mean of its sampling distribution equals the parameter of interest. <u>It does not systematically overestimate or underestimate the target parameter.</u>
- Both sample mean and sample proportion are unbiased estimators of population mean and proportion. The following sample variance is an unbiase estimator of population variance.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Properties of Point Estimators

Of all the **unbiased** estimators, we prefer the estimator whose sampling distribution has the **smallest spread** or **variability**.





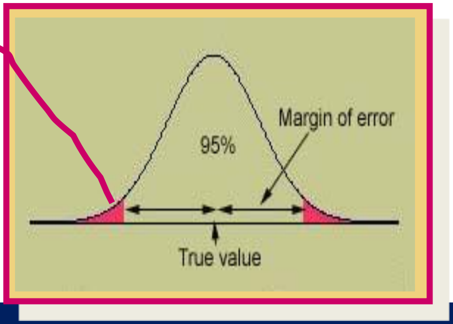
Interval Estimators/Confidence Intervals

- Confidence intervals depend on sampling distributions
- The shape of sampling distributions depend on sample sizes.
- For large sample sizes, central limit theorem applies which allow us to use **normal distributions**
- For small sample sizes, we need to learn a new distribution

The margin of error

- We assume that the sample sizes are large
- From the Central Limit Theorem, the sampling distributions of $\overline{X}$ and $\hat{p}$ will be **approximately Gaussian** under certain assumptions.
- For ***unbiased*** estimators with normal sampling distributions, 95% of all point estimates will lie within 1.96 standard deviations of the parameter of interest.
- **Margin of error:** provides a upper bound to the difference between a particular estimate and the parameter that it estimates. It is calculated as

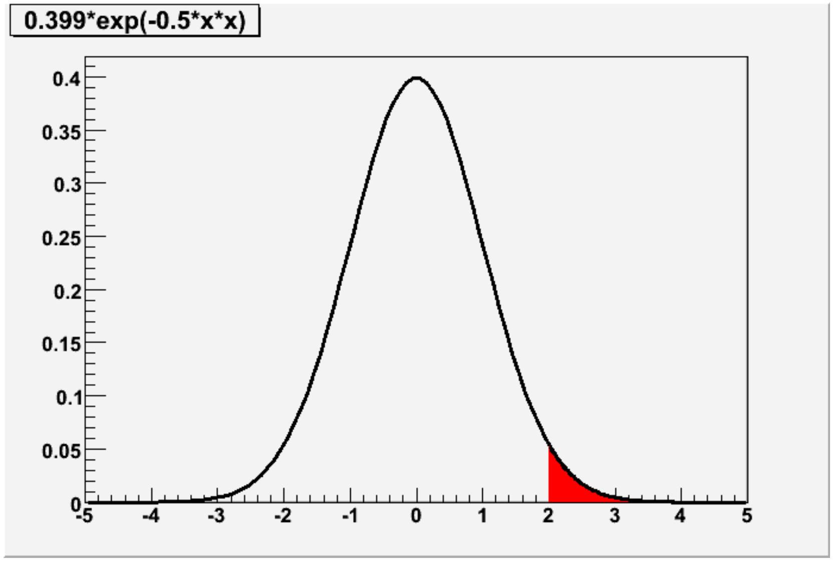
$1.96 \times$ std error of the estimator



The Gaussian Distribution(Remainder)

By far the most useful distribution is the Gaussian (normal) distribution:

$$P(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



68.27% of area within $\pm 1\sigma$
95.45% of area within $\pm 2\sigma$
99.73% of area within $\pm 3\sigma$

Mean = μ, Variance=σ^2
Note that width scales with σ.

Area out on tails is important---use lookup tables or cumulative distribution function.

In plot to left, red area ($>2\sigma$) is 2.3%.
90% of area within $\pm 1.645\sigma$
95% of area within $\pm 1.960\sigma$
99% of area within $\pm 2.576\sigma$

Estimating Means and Proportions

- For a quantitative population,

$$\text{Point estimator of population mean } \mu : \bar{x}$$

$$\text{Margin of error } (n \geq 30) : \pm 1.96 \frac{s}{\sqrt{n}}$$

- For a binomial population,

$$\text{Point estimator of population proportion p} : \hat{\text{p}} = \text{x/n}$$

$$\text{Margin of error} : \pm 1.96 \sqrt{\frac{\hat{\text{p}}\hat{\text{q}}}{\text{n}}}$$

$$\textit{Assumption}: \text{ np} > 5 \text{ and nq} > 5; \quad \text{or } 0 < \text{ p} \pm 2\sqrt{\frac{\text{pq}}{\text{n}}} < 1$$

Example: Gaussian

- A homeowner randomly samples 64 homes similar to her own and finds that the average selling price is 250,000 TL with a standard deviation of 15,000 TL. Estimate the average selling price for all similar homes in the city.

$$\text{Point estimator of}: \bar{x} = 250,000$$

$$\text{Margin of error}: \pm 1.96 \frac{s}{\sqrt{n}} = \pm 1.96 \frac{15,000}{\sqrt{64}} = \pm 3675$$

Example: binomial

- A quality control technician wants to estimate the proportion of soda cans that are underfilled. He randomly samples 200 cans of soda and finds 10 underfilled cans.

$$n = 200 \qquad p = \text{proportion of underfilled cans}$$

$$\text{Point estimator of } p: \hat{p} = x/n = 10/200 = .05$$

$$\text{Margin of error}: \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = \pm 1.96\sqrt{\frac{(.05)(.95)}{200}} = \pm .03$$

Confidence Interval

- Create an interval (a, b) so that you are fairly sure that the parameter lies between these two values.
- "Fairly sure" means "with high probability", measured using the confidence coefficient, $1-\alpha$.

Usually, $1-\alpha = .90, .95, .99$

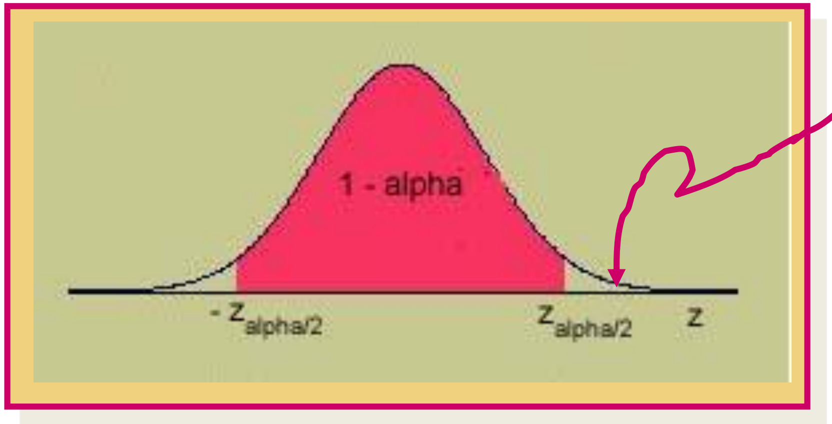
- For large-Sample size,

100($1-\alpha$)% Confidence Interval:

Point Estimator $\pm\, z_{a/2}$SE

To Change the Confidence Level

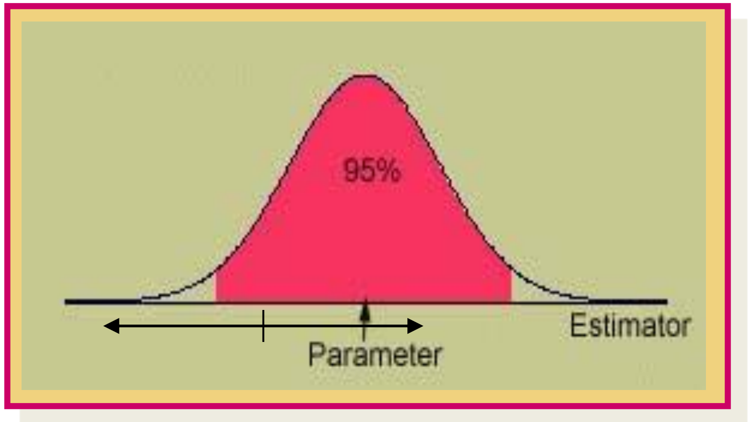
- To change to a general confidence level, $1-\alpha$, pick a value of z that puts area $1-\alpha$ in the center of the z distribution.



Tail area $\alpha/2$	$z_{\alpha/2}$
.05	1.645
.025	1.96
.005	2.58

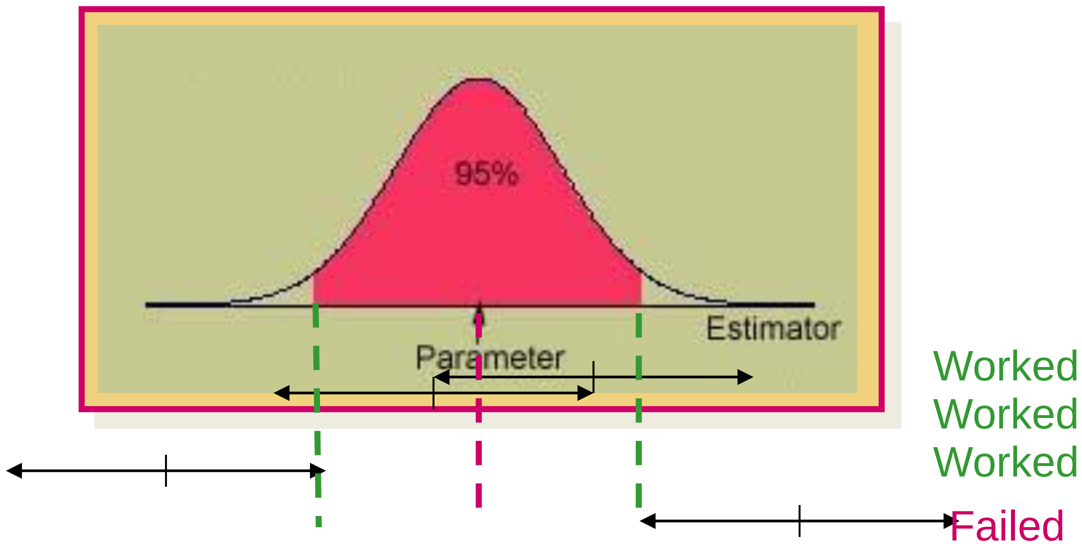
- Suppose $1-\alpha = .95$,

95% of the intervals constructed in this manner will enclose the population mean



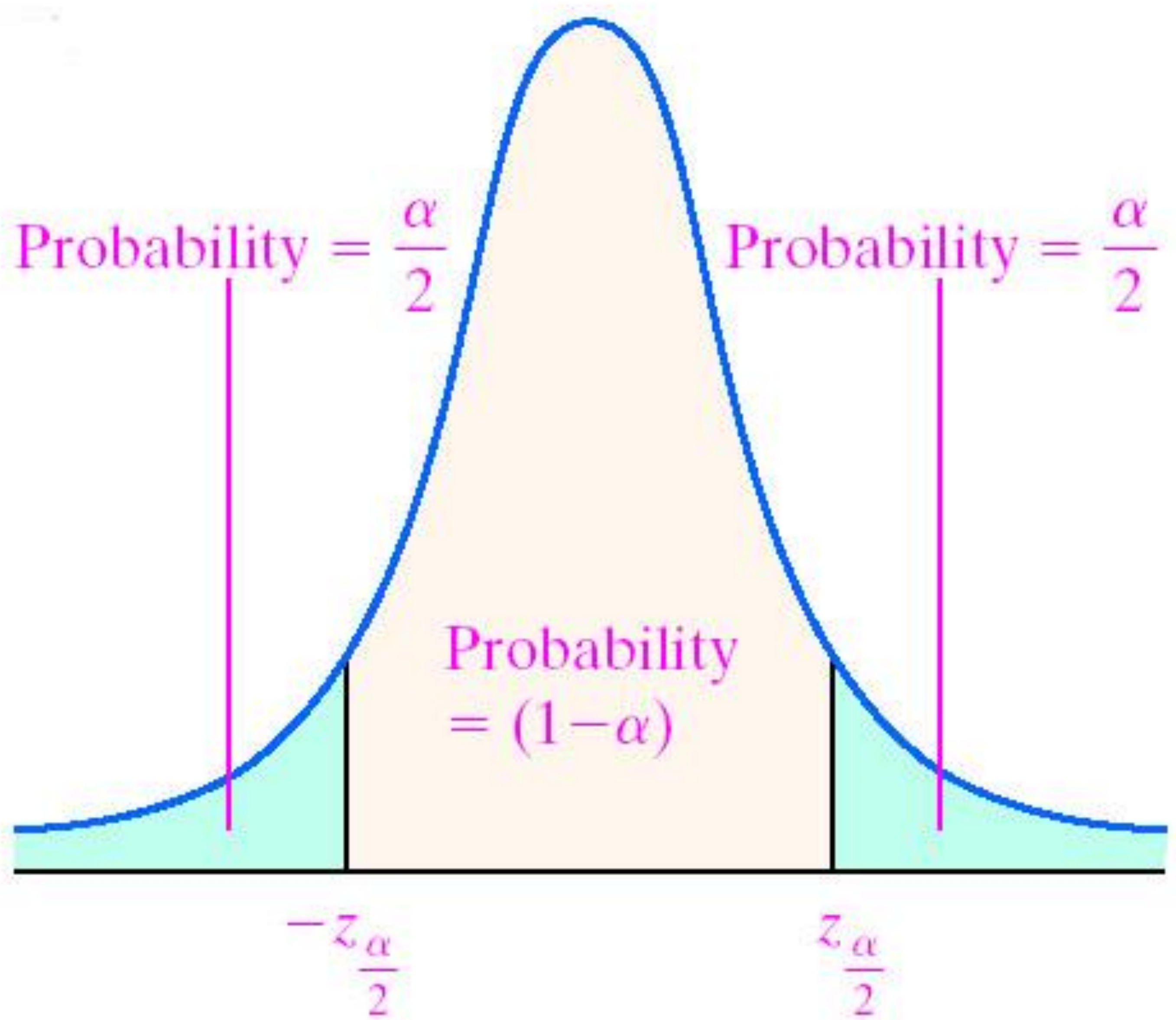
To Change the Confidence Level

- Since we don't know the value of the parameter, consider Point Estimator ± 1.96SE which has a variable center.



- Only if the estimator falls in the tail areas will the interval fail to enclose the parameter. This happens only 5% of the time.

Confidence Level



Interpretation of A Confidence Interval

- A confidence interval is calculated from one given sample. It either covers or misses the true parameter. **Since the true parameter is unknown, you'll never know which one is true.**
- **If independent samples are taken repeatedly from the same population**, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter.
- The confidence level associated with a confidence interval is the success rate of the confidence interval.

Confidence Intervals for Means and Proportions

- For a quantitative population,

Confidence interval for a population mean μ :

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

- For a binomial population,

Confidence interval for a population proportion p :

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example

A random sample of n = 50 males showed a mean average daily intake of dairy products equal to 756 grams with a standard deviation of 35 grams. Find a 95% confidence interval for the population average μ.

$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} \Rightarrow 756 \pm 1.96 \frac{35}{\sqrt{50}} \Rightarrow 756 \pm 9.70$$

$$\text{or } 746.30 < \mu < 765.70 \text{ grams.}$$

Example

Find a 99% confidence interval for μ, the population average daily intake of dairy products for men.

$$\bar{x} \pm 2.58 \frac{s}{\sqrt{n}} \Rightarrow 756 \pm 2.58 \frac{35}{\sqrt{50}} \Rightarrow 756 \pm 12.77$$

The interval must be wider to provide for the increased confidence that is does indeed enclose the true value of μ.

Example: binomial

Of a random sample of n = 150 college students, 104 of the students said that they had played on a soccer team. Estimate the proportion of college students who played soccer in their youth with a 90% confidence interval.

$$\hat{p} \pm 1.645\sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow \frac{104}{150} \pm 1.645\sqrt{\frac{.69(.31)}{150}}$$

$$\Rightarrow \quad .69 \pm .06 \quad \text{or} \quad .63 < p < .75.$$

Estimating the Difference between Two Means

- Sometimes we are interested in comparing the means of two populations.
  - The average scores for students taught with two different teaching methods.
  - Observable differences between the brains of *male* and *female*

To make this comparison,

A random sample of size n_1 drawn from population 1 with mean μ_1 and variance σ_1^2.

A random sample of size n_2 drawn from population 2 with mean μ_2 and variance σ_2^2.

Notations - Comparing Two Means

	Mean	Variance	Standard Deviation
Population 1	μ_1	σ_1^2	σ_1
Population 2	μ_2	σ_2^2	σ_2

	Sample size	Mean	Variance	Standard Deviation
Sample from Population 1	n_1	$\overline{x}_1$	s_1^2	s_1
Sample from Population 2	n_2	$\overline{x}_2$	s_2^2	s_2

Estimating the Difference between Two Means

- We compare the two averages by making inferences about μ_1-μ_2, the difference in the two population averages.
    - If the two population averages are the same, then μ_1-μ_2 = 0.
    - The best estimate of μ_1-μ_2 is the difference in the two sample means,

$$\boxed{\bar{x}_1 - \bar{x}_2}$$

The Sampling Distribution $\bar{x}_1 - \bar{x}_2$

1. The mean of $\bar{x}_1 - \bar{x}_2$ is $\mu_1 - \mu_2$, the difference in the population means.

2. The standard deviation of $\bar{x}_1 - \bar{x}_2$ is $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

3. If the sample sizes (both n_1 and n_2) are large, the sampling distributi on of $\bar{x}_1 - \bar{x}_2$ is approximat ely normal, and standard deviation can be estimated as $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Estimating μ_1-μ_2

For large samples, point estimates and their margin of error as well as confidence intervals are based on the standard normal (z) distribution.

Point estimate for $\mu_1 - \mu_2 : \bar{x}_1 - \bar{x}_2$

Margin of Error : $\pm 1.96\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Assumption :

Both $n_1 \geq 30$ and $n_2 \geq 30$

Example

Avg Daily Intakes	Men	Women
Sample size	50	50
Sample mean	756	762
Sample Std Dev	35	30

- Compare the average daily intake of dairy products of men and women using a 95% confidence interval.

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\Rightarrow (756 - 762) \pm 1.96\sqrt{\frac{35}{50} + \frac{30}{50}} \quad \Rightarrow \quad -6 \pm 12.78$$

$$\text{or } -18.78 < \mu_1 - \mu_2 < 6.78.$$

Example

$$-18.78 < \mu_1 - \mu_2 < 6.78$$

- Could you conclude, based on this confidence interval, that there is a difference in the average daily intake of dairy products for men and women?
- The confidence interval contains the value $\mu_1-\mu_2= 0$. Therefore, it is possible that $\mu_1 = \mu_2$. You would not want to conclude that there is a difference in average daily intake of dairy products for men and women.

Estimating the Difference between Two Proportions

- Sometimes we are interested in comparing the proportion of "successes" in two binomial populations.
  - The germination rates of untreated seeds and seeds treated with a fungicide.
  - The proportion of male and female voters who favor a particular candidate for governor.
- To make this comparison,

A random sample of size n_1 drawn from binomial population 1 with parameter p_1.

A random sample of size n_2 drawn from binomial population 2 with parameter p_2.

Notations - Comparing Two Proportions

	Sample size	Sample Proportion	Sample Variance	Standard Deviation
Sample from Population 1	n_1	$\hat{p}_1 = \frac{x_1}{n_1}$	$\frac{\hat{p}_1 \hat{q}_1}{n}$	$\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n}}$
Sample from Population 2	n_2	$\hat{p}_2 = \frac{x_2}{n_2}$	$\frac{\hat{p}_2 \hat{q}_2}{n}$	$\sqrt{\frac{\hat{p}_2 \hat{q}_2}{n}}$

Estimating the Difference between Two Means

- We compare the two proportions by making inferences about p_1-p_2, the difference in the two population proportions.

  - If the two population proportions are the same, then p_1-$p_2 = 0$.
  - The best estimate of p_1-p_2 is the difference in the two sample proportions,

$$\boxed{\hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}}$$

The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

1. The mean of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$, the difference in the population proportion s.

2. The standard deviation of $\hat{p}_1 - \hat{p}_2$ is $\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$.

3. If the sample sizes (both n_1 and n_2) are large, the sampling distributi on of $\hat{p}_1 - \hat{p}_2$ is approximat ely normal, and stanard deviation can be estimated as

$$SE = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}.$$

Estimating $\hat{p}_1 - \hat{p}_2$

For large samples, point estimates and their margin of error as well as confidence intervals are based on the standard normal (z) distribution.

Point estimate for p_1-p_2 : $\hat{p}_1 - \hat{p}_2$

Margin of Error : $\pm 1.96\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$

Confidence interval for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Assumption : both n_1 and n_2 are sufficient ly large so that

$$-1 \le \hat{p}_1 - \hat{p}_2 \pm 2SE \le 1$$

Example

Youth Soccer	Male	Female
Sample size	80	70
Played soccer	65	39

- Compare the proportion of male and female college students who said that they had played on a soccer team using a 99% confidence interval.

$$(\hat{p}_1 - \hat{p}_2) \pm 2.58\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$\Rightarrow \quad (\frac{65}{80} - \frac{39}{70}) \pm 2.58\sqrt{\frac{.81(.19)}{80} + \frac{.56(.44)}{70}} \quad \Rightarrow \quad .25 \pm .19$$

$$\text{or } .06 < p_1 - p_2 < .44.$$

Example

$$.06 < p_1 - p_2 < .44$$

- Could you conclude, based on this confidence interval, that there is a difference in the proportion of male and female college students who said that they had played on a soccer team?
- The confidence interval does not contains the value $p_1 - p_2 = 0$. Therefore, it is not likely that $p_1 = p_2$. You would conclude that there is a difference in the proportions for males and females.

A higher proportion of males than females played soccer in their youth.

If the sample size is small?

- Point estimators remain the same
- There are small sample interval estimators / confidence intervals for
  - μ, the mean of a normal population
  - $\mu_1 - \mu_2$, the difference between two normal population means

The Sampling Distribution of the Sample Mean

- When we take a sample from a normal population, the sample mean $\bar{x}$ has a normal distribution for **any sample size *n***, and

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

has a standard normal distribution.

- But if σ is unknown, and we must use *s* to estimate it, the resulting statistic **is not normal**.

$$\frac{\bar{x} - \mu}{s / \sqrt{n}} \text{ is not normal!}$$

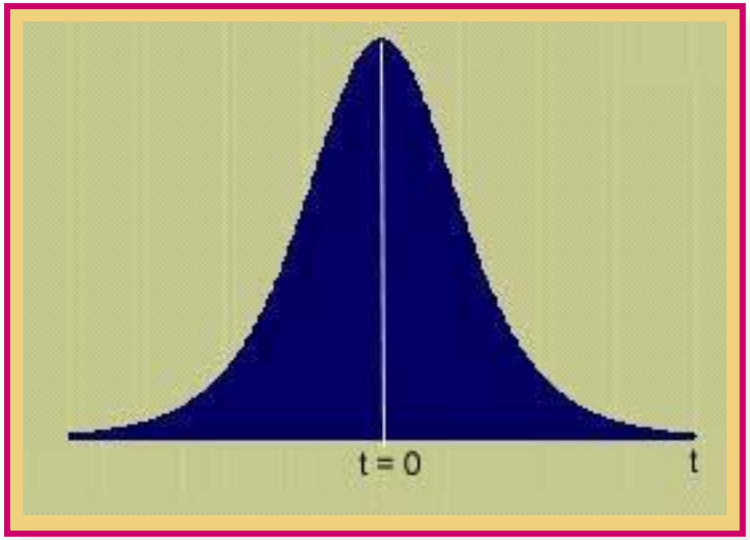
Student's t Distribution

- Fortunately, this statistic does have a sampling distribution that is well known to statisticians, called the **Student's t distribution,** with ***n*-1 degrees of freedom**.

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

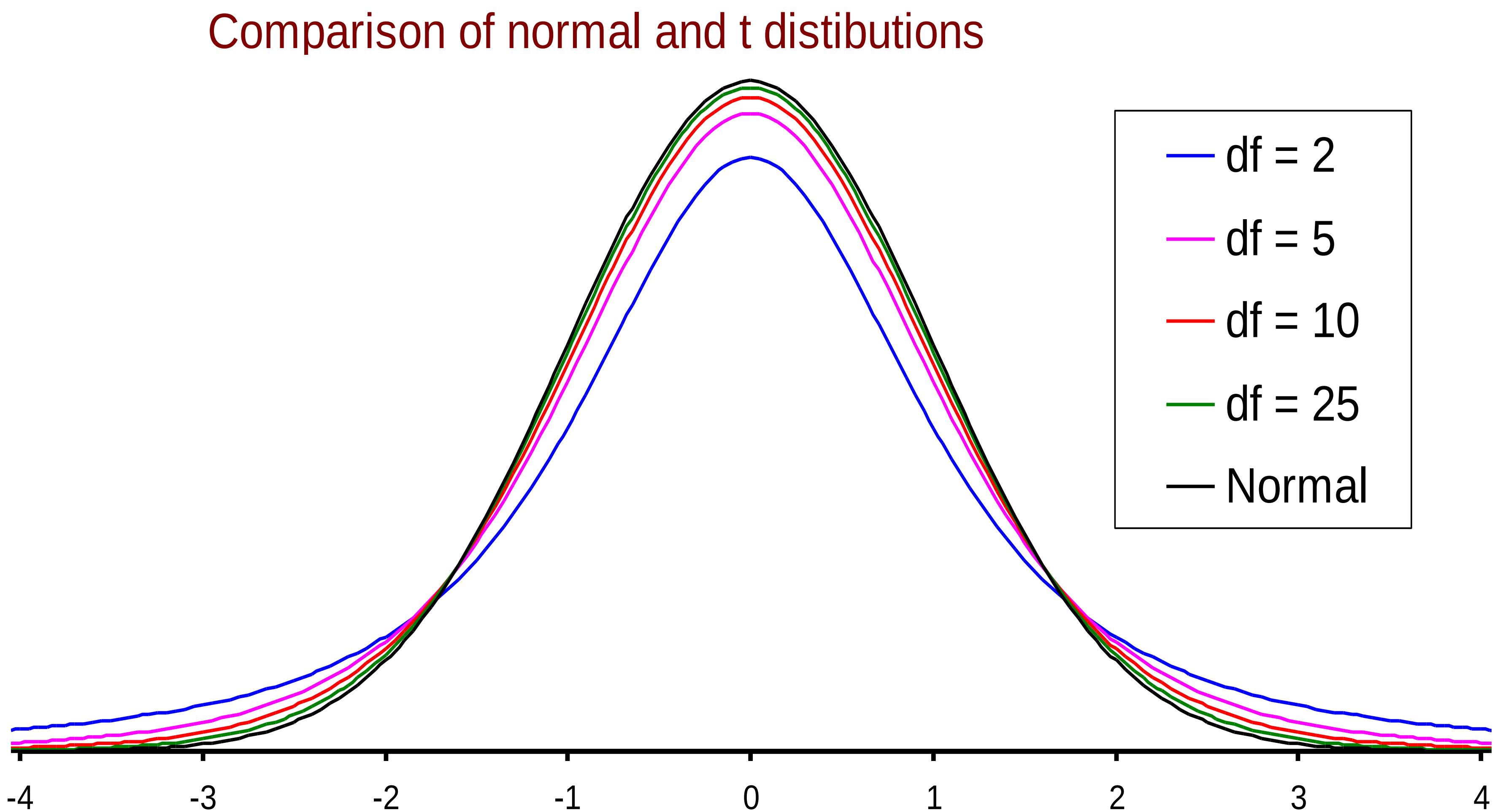
- We can use this distribution to create estimation procedures for the population mean μ.

Properties of Student's *t*



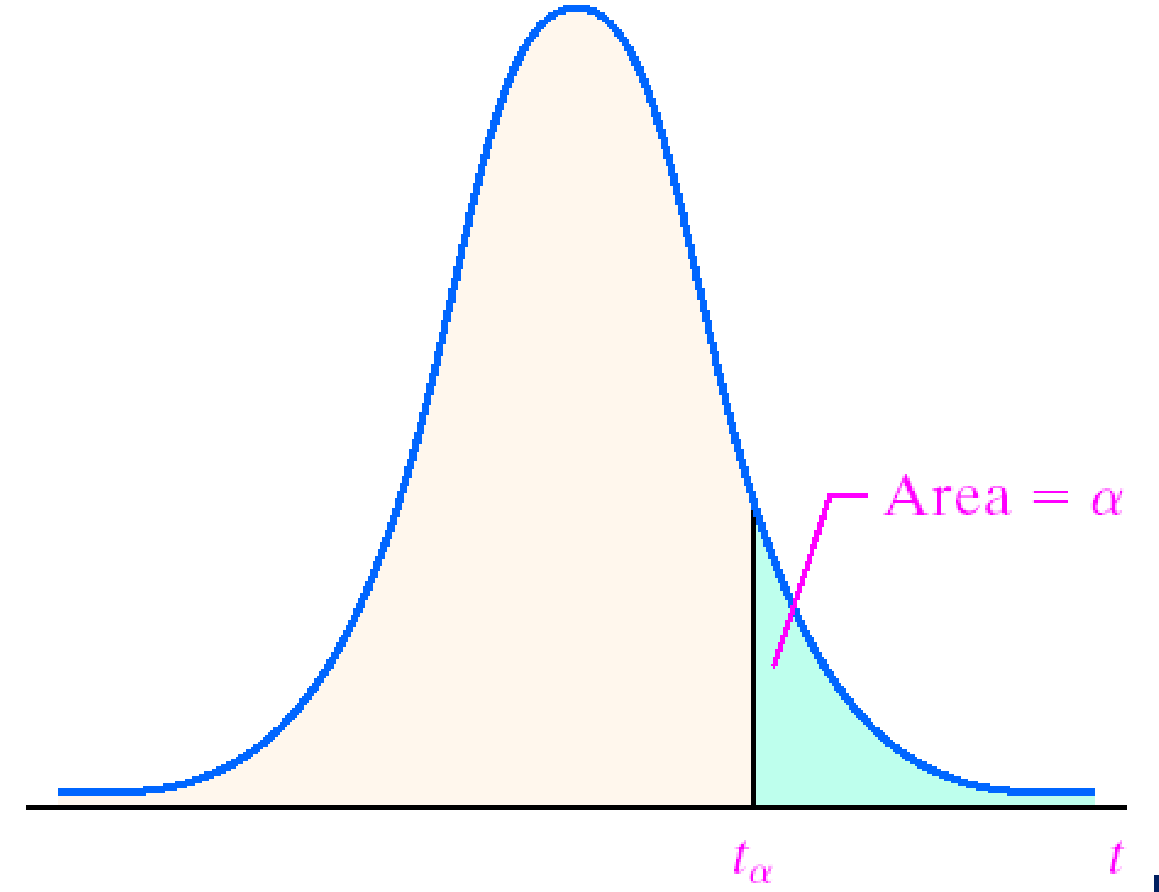
- Mound-shaped and symmetric about 0.
- More variable than *z*, with "heavier tails"
- Shape depends on the sample size ***n*** or the **degrees of freedom, *n*-1.**
- **As *n* increases the shapes of the *t* and *z* distributions become almost identical.**

Properties of Student's *t*



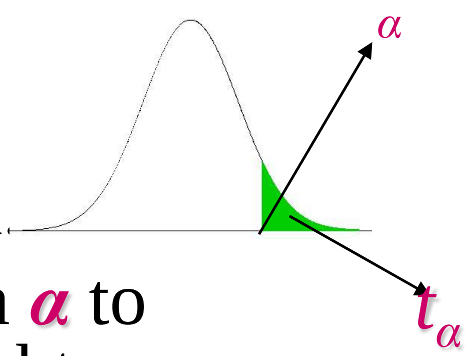
Using t-distribution

The notation t_α is the t-value such that the area under the standard normal curve to the right is α. The figure illustrates the notation.



Using t-Table

- Table 4 gives the values of t that cut off certain critical values in the right tail of the t distribution



- Use index ***df*** and the appropriate tail area α to find t_α, the value of t with area α to its right.

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602

For a random sample of size $n = 10$, find a value of t that cuts off .025 in the right tail.

Row = $df = n - 1 = 9$

Column subscript = α = .025

$t_{.025} = 2.262$

Table 4

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Small Sample Confidence Interval for Population Mean μ

Small - Sample $(1-\alpha)100\%$ confidence interval of the population mean μ is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the value of t that cuts off area $\alpha/2$ in the right tail of a t - distribution with $df = n-1$.

Assumption: population must be normal

Example

Randomly sampling 16 students for their GPA, you get a sample mean of 3.0 and sample std. deviation (s) of .4

Identify an interval which will contain the true population mean 95% of the time.

Calculate standard dev. of mean: $= \frac{s}{\sqrt{n}} = \frac{.4}{\sqrt{16}} = .10$

$$\boxed{t = \left(\bar{X} - \mu\right) / \quad \frac{s}{\sqrt{n}} \quad rewriting \quad \mu = \bar{X} \pm t \frac{s}{\sqrt{n}}}$$

the interval 3 ±(2.145*.1)=3±.21 This is a confidence interval from 2.79 to 3.21. 95% of the time this interval will contain the mean.

Example

Ten randomly selected students were each asked to list how many hours of television they watched per month. The results are

82	66	90	84	75
88	80	94	110	91

Find a 90% confidence interval for the true mean number of hours of television watched per month by students.

Example

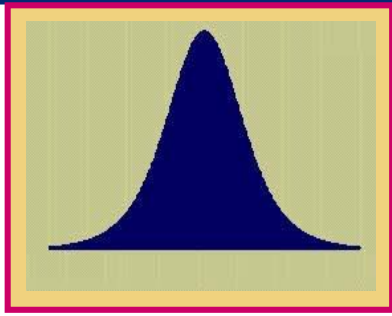
Calculating the sample mean and standard deviation we have $n = 10, \bar{x} = 86,$ and $s = 11.766$

We find the critical t value of 1.833 by looking on the t table in the row corresponding to df = 9, in the column with label $t_{.050}$. The 90% confidence interval for μ is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 86 \pm (1.833) \frac{11.842}{\sqrt{10}} = 86 \pm 6.86$$

$$(79.14,\ 92.86)$$

Estimating the Difference between Two Means



- You can also create a 100(1-α)% confidence interval for $\mu_1 - \mu_2$.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\text{with } s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

and $t_{\alpha/2}$ is the critical value of t with degrees of freedom $n_1 + n_2 - 2$

Remember the three assumptions:

1. Original populations normal
2. Samples random and independent
3. Equal population variances

$$\sigma_1^2 = \sigma_2^2$$

In Class Exercise

House Number	Selling Price (in kTurkishLira)	*X*: House Size (in m^2)
1	895	200
2	799	148
3	831	205
4	569	125
5	666	180
6	825	143
7	1263	275
8	793	165
9	1199	243
10	876	202
11	1126	220
12	1208	219
13	785	123
14	743	140
15	748	167

- Writing a python code:
- Estimate the average selling price for all similar homes in the city.
  - Using selling prices
  - Using prices/House size
- Find a 95% confidence interval for the true mean