APPLIED STATISTICS

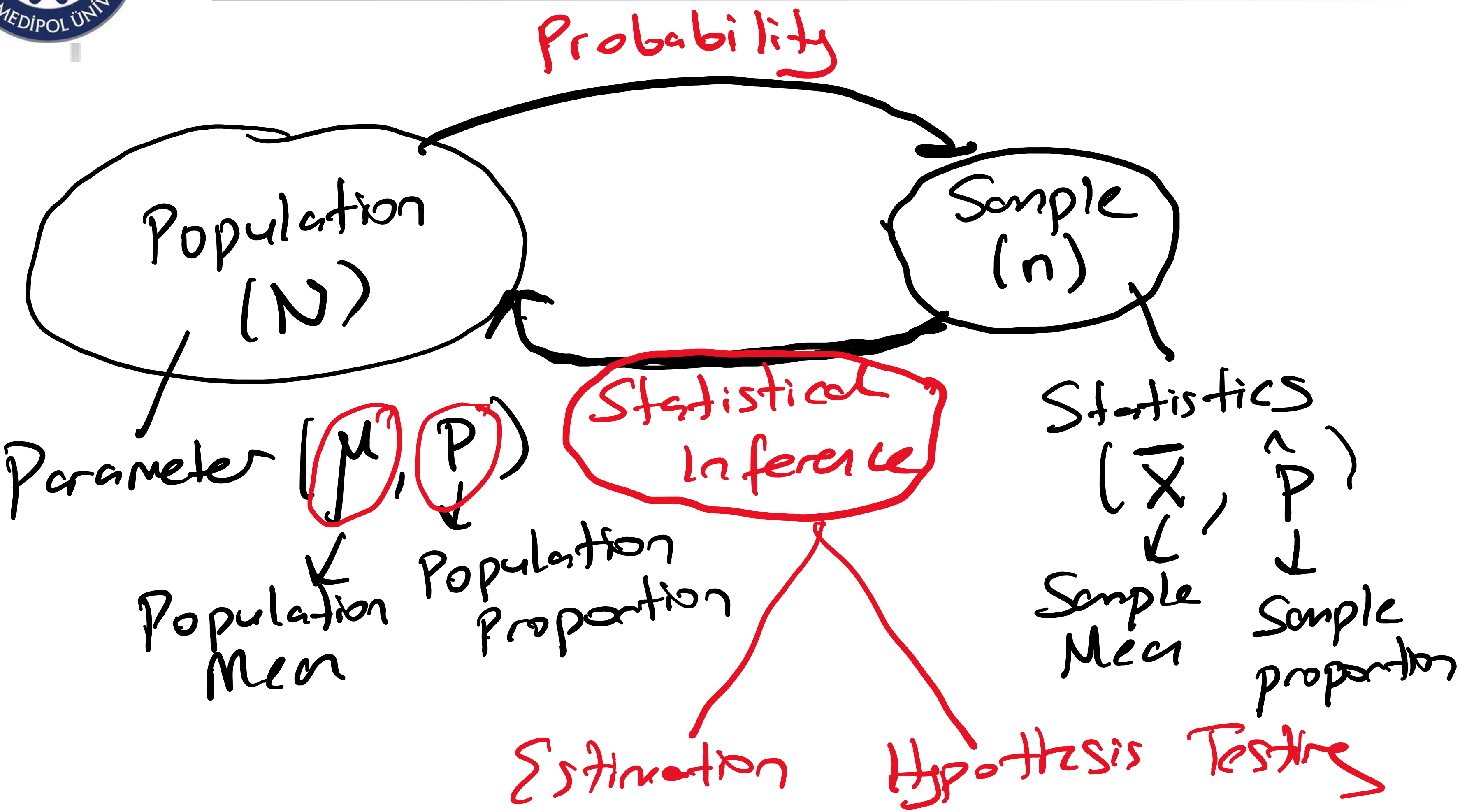
Confidence Interval Estimation
on a Single Sample

WEEK 3

OUTLINE

- Confidence Intervals for the Population Mean, μ
  - when Population Variance **σ^2 is Known** → Case I
  - when Population Variance σ^2 **is Unknown** → Case II
- Confidence Intervals for the Population Proportion, $\hat{p}$ (large samples)

Fundamental relationship between probability and statistical inference



Statistical Inference

Estimation

* Estimation of Population Parameters (μ, P)
* Accuracy of estimation

Tests of Hypothesis

* Accuracy of our statistical decision

H_0: Drug has no effect

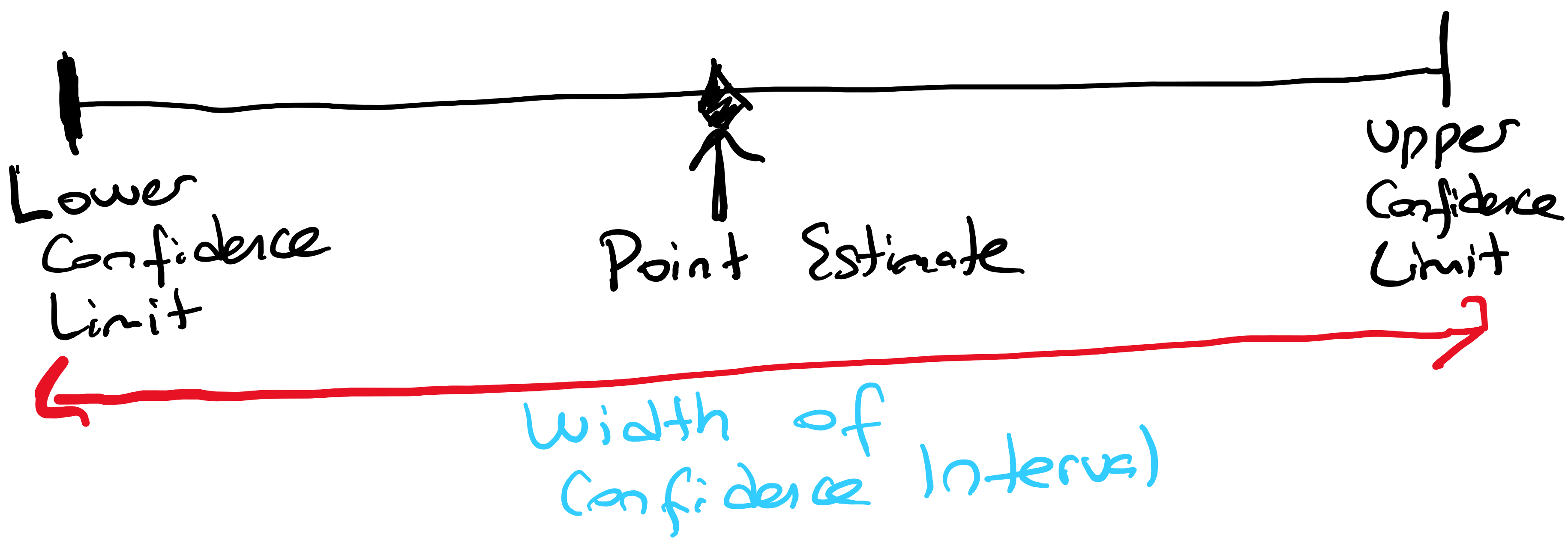
H_1 or H_a: Drug has effect

Definitions

* An estimator of a population parameter is
  - a random variable that depends on sample information
  - whose value provides an approximation to this unknown parameter.
* A specific value of that random variable is called an estimate.

Point and Interval Estimates

- A point estimate is a single number.
- A confidence interval provides addition information about variability



Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{x}$
Proportion	P	$\hat{p}$

Confidence Intervals

* An interval estimate provides more information about population characteristic than does a point estimate.
* Such interval estimates are called confidence intervals.

Confidence Interval Estimate

* An interval gives a range of values:

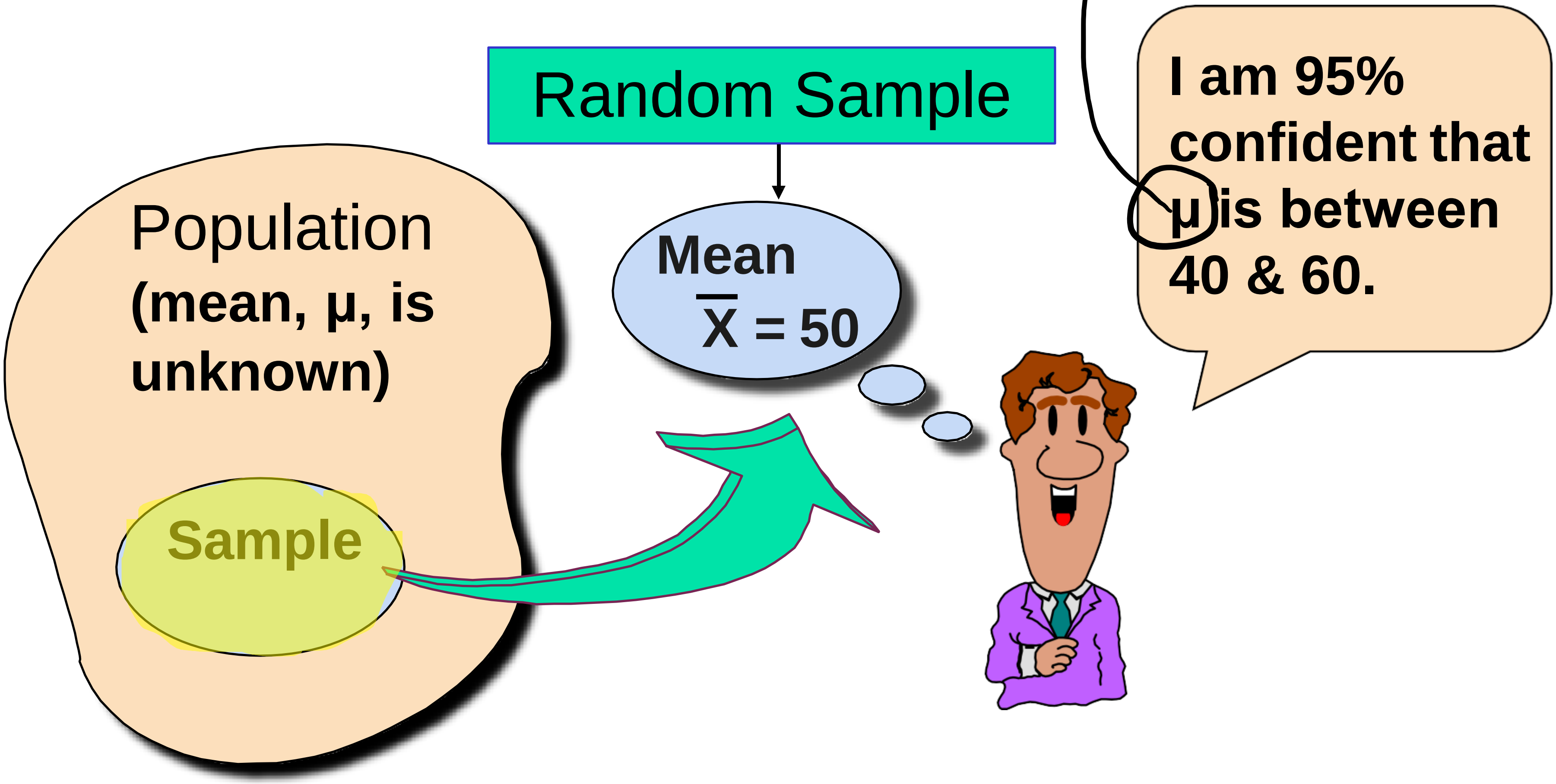
→ takes into consideration variation in sample statistics from sample to sample.

→ gives information about closeness to unknown population parameters

→ stated in terms of level of confidence

Estimation Process

Actual population value Mean



Confidence Level, $(1-\alpha)$

* Suppose confidence level = $\boxed{95\%}$
* Also written as $(1-\alpha) = 0.95$
* From repeated samples, 95% of all the confidence intervals that can be constructed will contain the unknown true (actual) parameter.

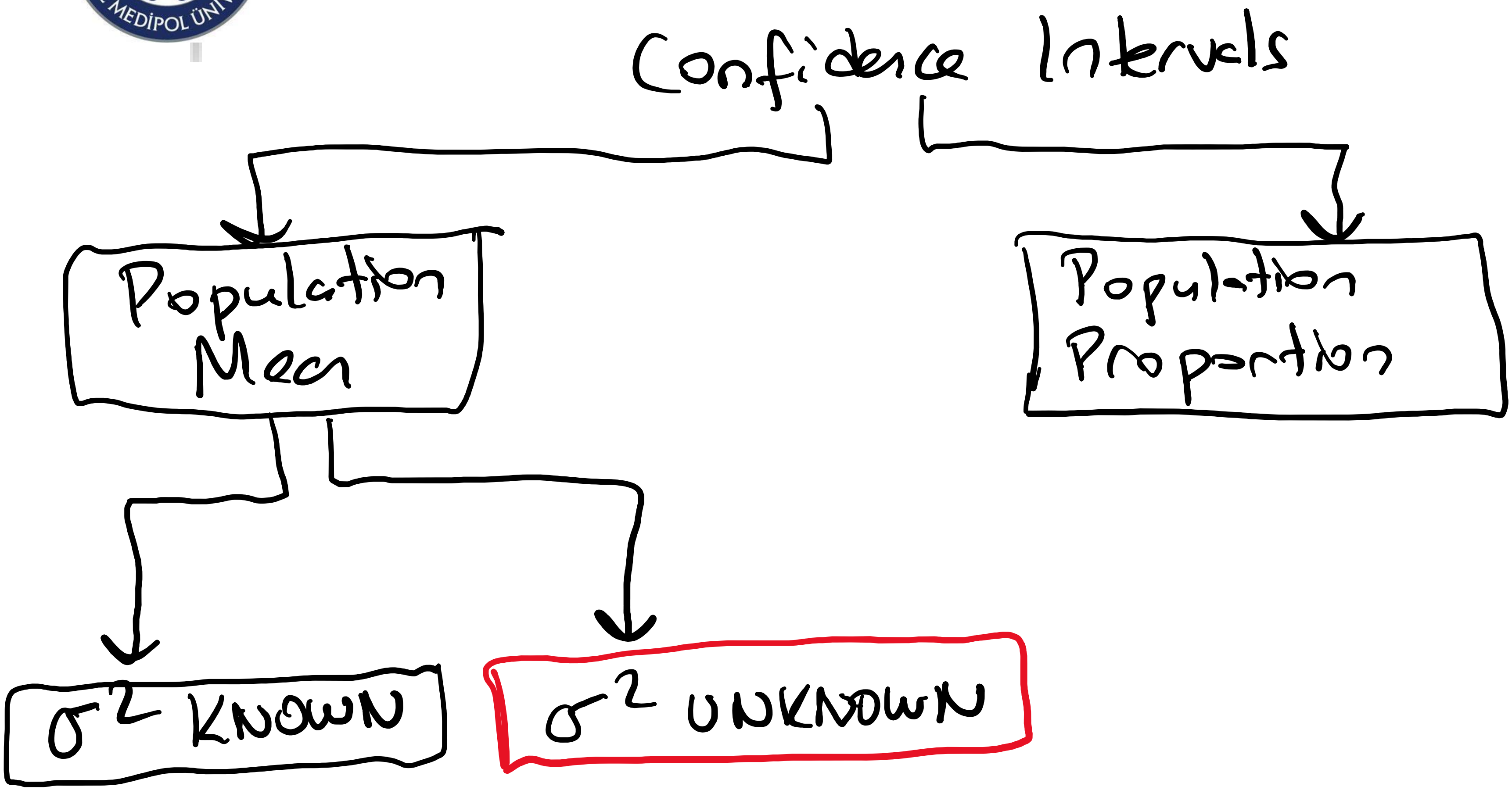
General Formula

Point Estimate ± (Reliability Factor) (Standard Error)

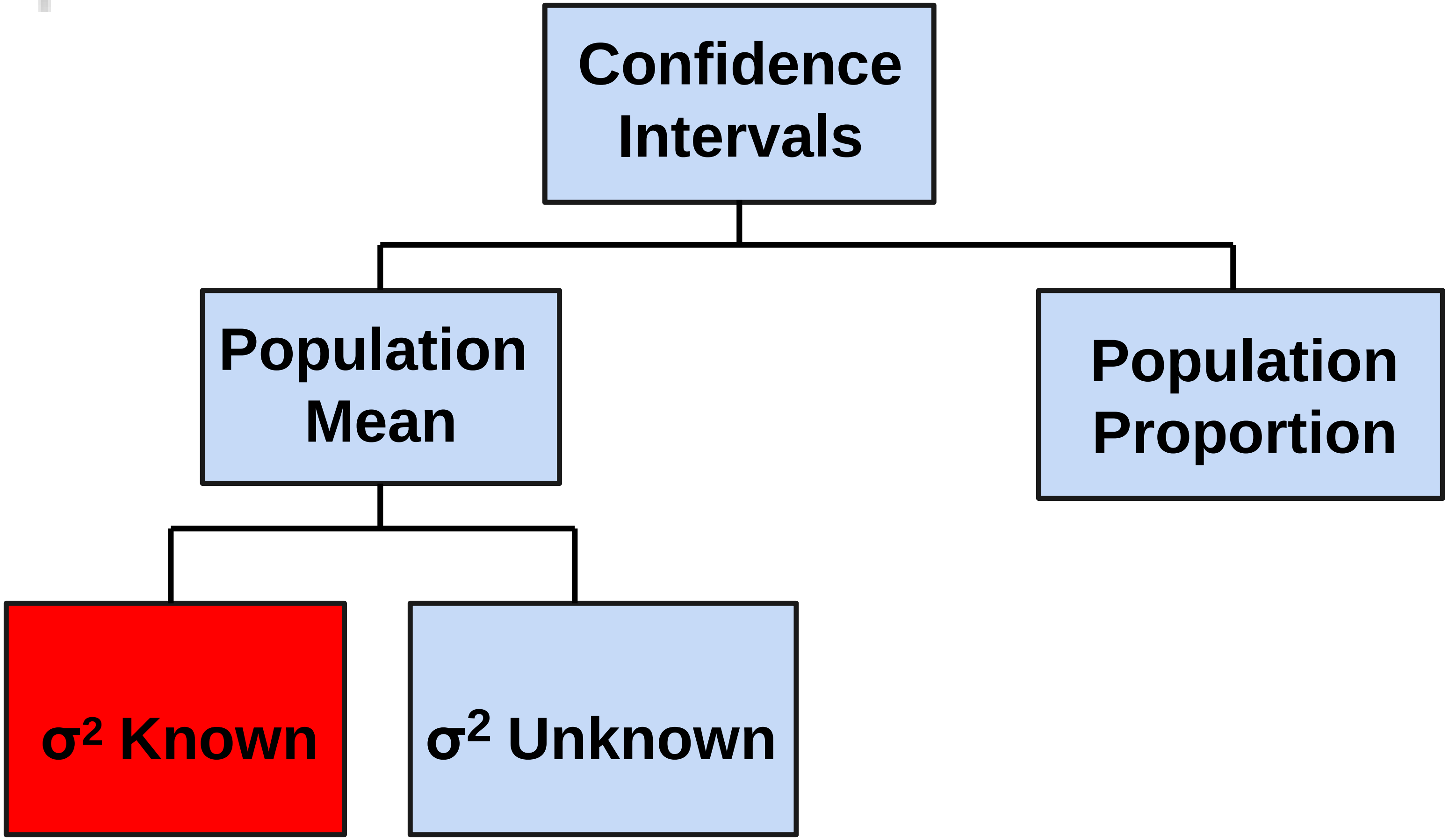
Note that:

* The value of the reliability factor depends on the desired level of confidence

Confidence Intervals



Confidence Intervals



Confidence Interval for μ (σ^2 Known) → Case I

Assumptions:

- Population variance σ^2 is known.
- Population is normally distributed.
- If population is not normal, use

large sample.

→ Confidence Interval Estimate:

$$\boxed{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

where $Z_{\alpha/2}$ is the normal dist. value for a probability of $\alpha/2$ in each tail

Margin of Error (ME)

The confidence interval

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Can also be written as

$$\bar{x} \pm \text{Margin of Error (ME)}$$

$$\boxed{ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

- The interval width, W, is equal to twice the margin of error.

$$\boxed{W = 2 \times ME}$$

Reducing the Margin of Error

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The margin of error can be reduced (↓) if

→ the population standard deviation can be reduced (σ ↓)

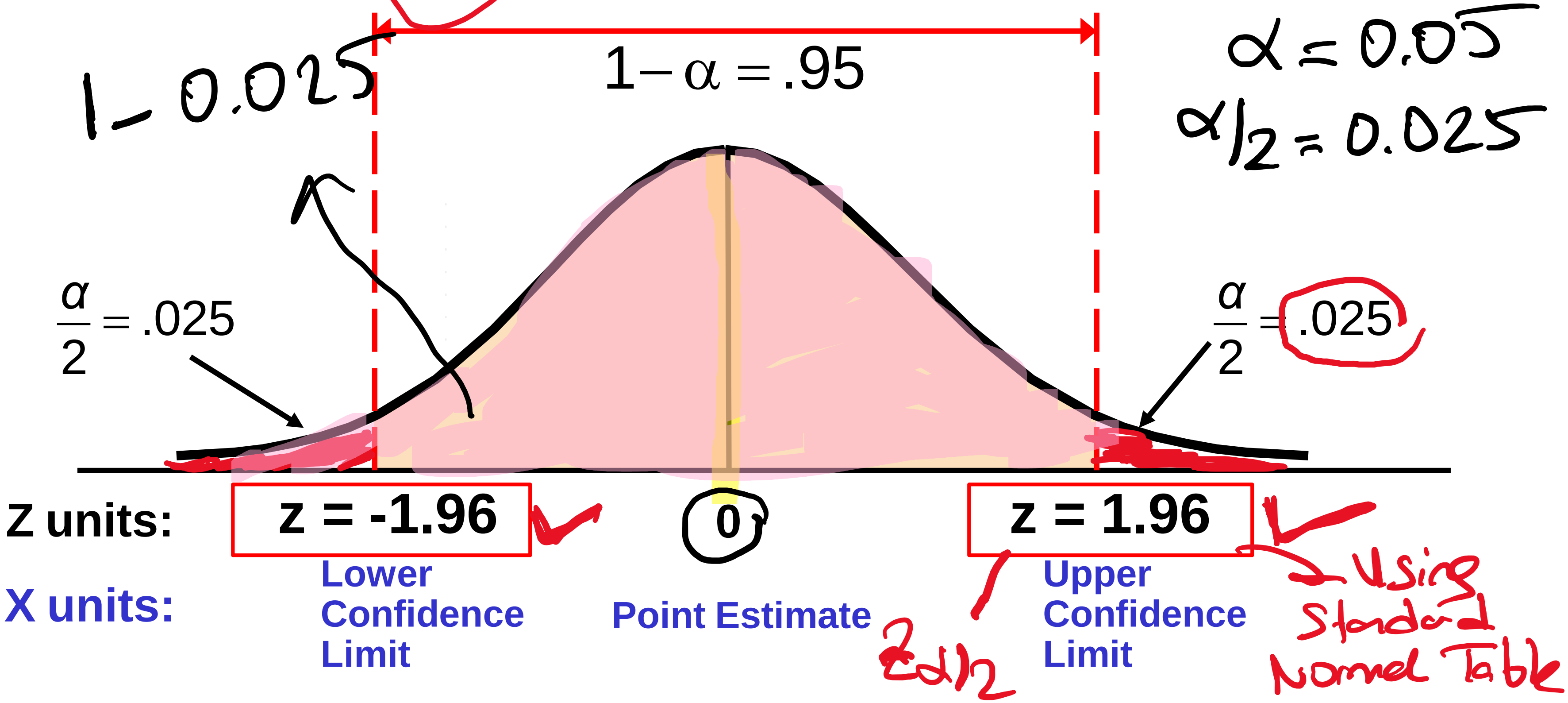
→ the sample size is increased (n ↑)

→ the confidence level is decreased, $(1-\alpha)$ ↓

Finding the Reliability Factor, $z_{\alpha/2}$

- Consider a 95% confidence interval:

$1 - \alpha = 0.95$



- Find $z_{.025} = \pm 1.96$ from the standard normal distribution table

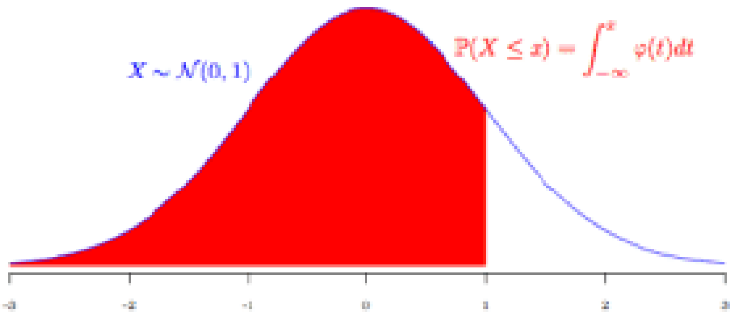
Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

90%
95%
99%

Confidence Level	Confidence Coefficient, $1-\alpha$	$Z_{\alpha/2}$ value
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.58
99.8%	.998	3.08
99.9%	.999	3.27

Using Standard Normal Table



$z_{\alpha/2} = ?$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

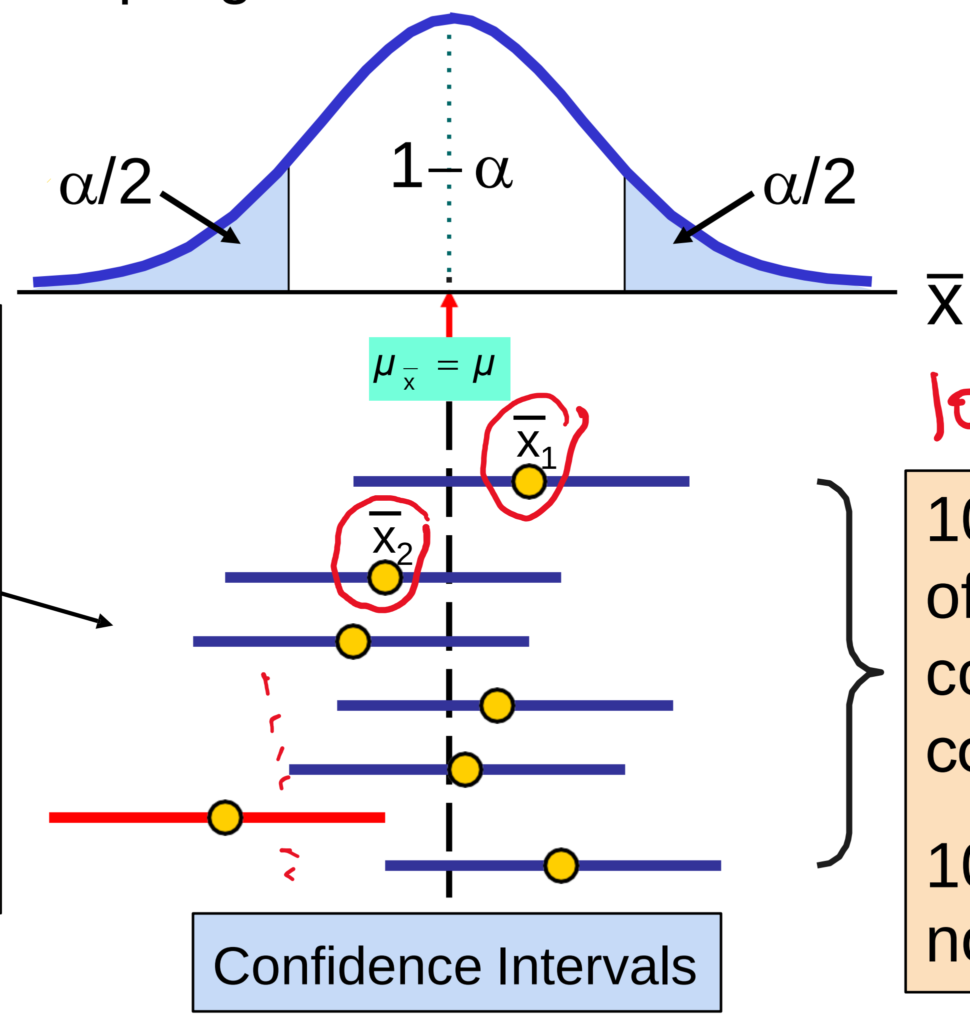
$1 - 0.025$

$= 0.975$

$$\begin{array}{r} 1.9 \\ +0.06 \\ \hline 1.96 \end{array}$$

Intervals and Level of Confidence

Sampling Distribution of the Mean



$100(1-\alpha)\%$

Intervals extend from

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

to

$$\bar{x} + z\frac{\sigma}{\sqrt{n}}$$

$100(1-\alpha)\%$ of intervals constructed contain μ;

$100(\alpha)\%$ do not.

Question 1:

$n = 11$

- 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.

Solution:

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow 2.20 - 1.96 \cdot \frac{0.35}{\sqrt{11}} < \mu < 2.20 + 1.96 \cdot \frac{0.35}{\sqrt{11}}$$

$$1 - \alpha = 0.95 \rightarrow z_{\alpha/2} = 1.96$$

$$\boxed{1.9932 < \mu < 2.4068}$$

$1-\alpha = 0.95$

Solution

(continued)

$$\boxed{1.9932 < \mu < 2.4068}$$

Interpretation:

* We are 95% confident that the true mean resistance is between 1.9932 and 2.4068.

* Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean.

Confidence Intervals

- Confidence Intervals
  - Population Mean ✓
    - σ^2 Known
    - σ^2 Unknown → Case II
  - Population Proportion

Student's t-Distribution

* Consider a random sample of n observations

- with mean $\bar{X}$ and standard dev. S
- from a normally distributed population

with mean μ.

$$\boxed{t = \frac{\bar{X} - \mu}{S / \sqrt{n}}}$$

follows the Students t-distribution with $(n-1)$ degrees of freedom

Confidence Interval for μ (σ^2 Unknown)

- If the population standard deviation σ is unknown, we can substitute the sample standard deviation, S.
- So we use t-distribution instead of normal distribution

Confidence Interval for μ (σ Unknown)

(continued)

Assumptions:

- Population standard deviation is unknown.
- Population is normally distributed
- If population is not normal, use large sample,

Use Student's t_ distribution

$$\boxed{\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}}$$

where $t_{n-1,\alpha/2}$ is the critical value of the t-dist. with $n-1$ d.f. and an area of $\alpha/2$ in each tail.

Student's t Distribution

- The t- value depends on degrees of freedom (d.f.)

→ Number of observations that are free to vary after sample mean has been calculated

$$\boxed{d.f = n - 1}$$

d.f = ν = degrees of freedom

df = n−1

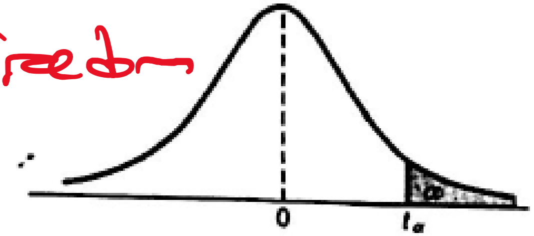


Table A.4 Critical Values of the t-Distribution

v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
				α			
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145

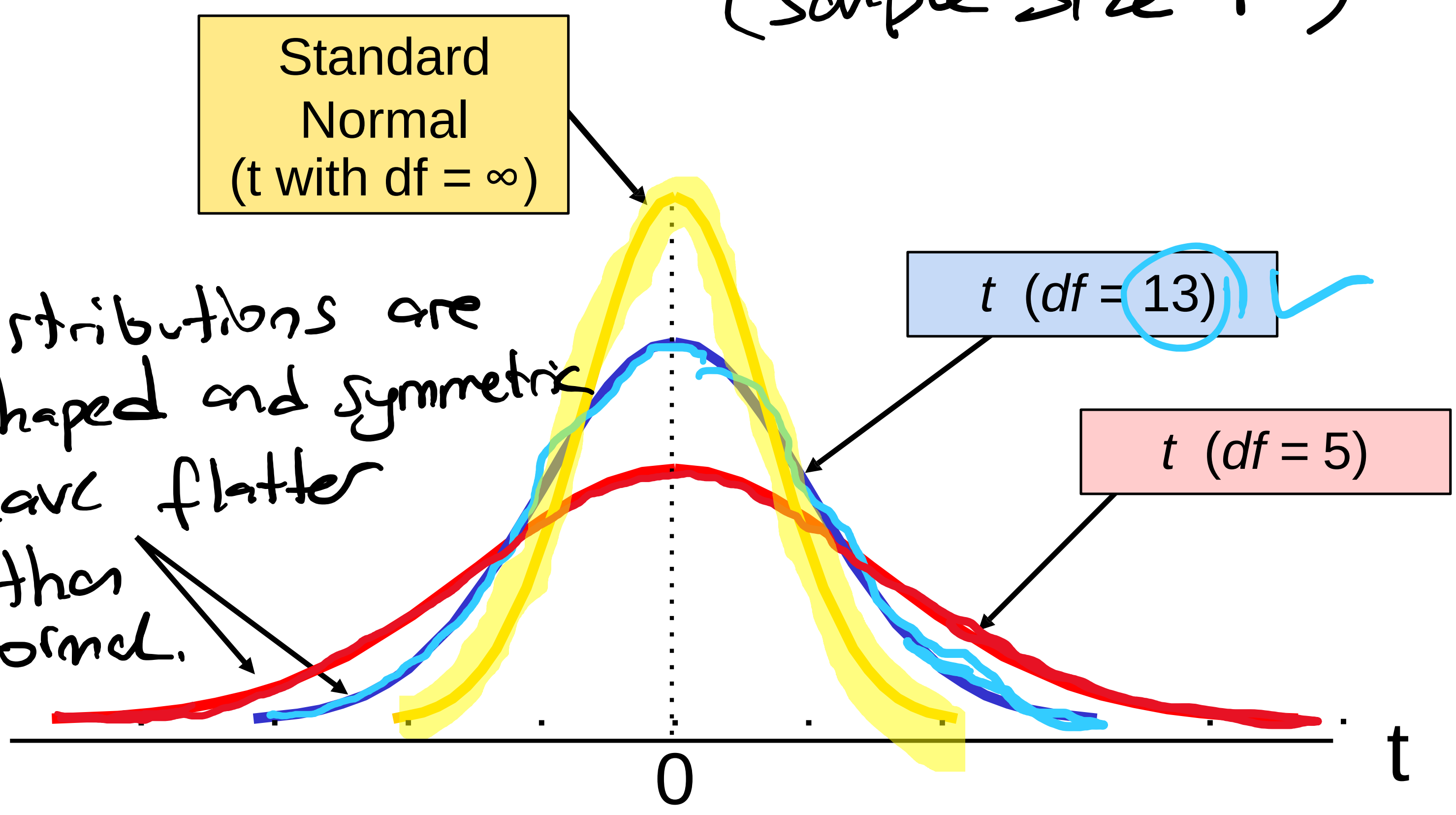
$$t_{n-1,\alpha/2} \Rightarrow t_{2,\,0.05} = \boxed{2.920}$$

Student's t Distribution

Note: $t \longrightarrow Z$ as n increases

(sample size ↑)

t-distributions are bell-shaped and symmetric but have flatter tails than the normal.



Student's t Table

ν

	Upper Tail Area α		
df	.10	.05	.025
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182

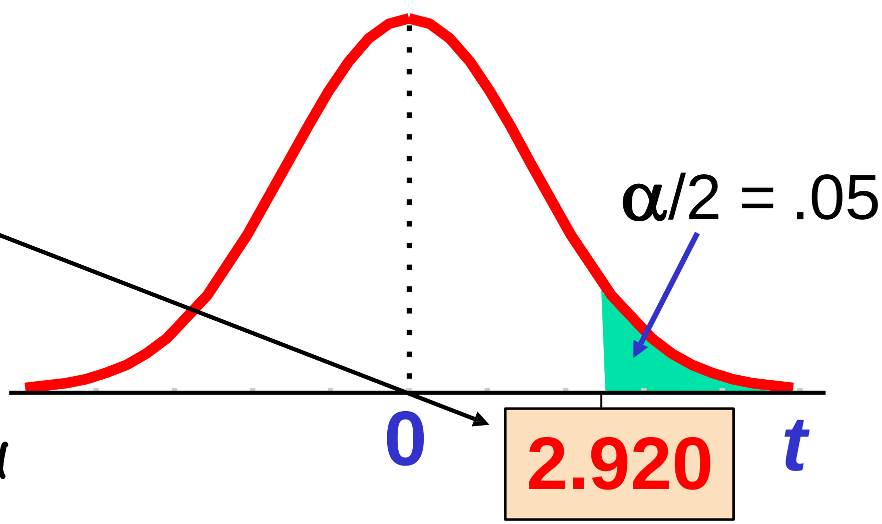
Example:

Let: n = 3

df = n - 1 = 2

$\boxed{\alpha = .10}$

$\alpha/2 = .05$



The body of the table contains t-values, not probabilities

t-distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z
.80	1.372	1.325	1.310	1.282
.90	1.812	1.725	1.697	1.645
.95	2.228	2.086	2.042	1.960
.99	3.169	2.845	2.750	2.576

Note: $t \longrightarrow Z$ as n increases

Question 2:

A random sample of n = 25 has x̄ = 50 and s = 8. Form a 95% confidence interval for μ

Sample standard dev. is given !

$$\bar{x} - \boxed{t_{n-1,\alpha/2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + \boxed{t_{n-1,\alpha/2}} \frac{s}{\sqrt{n}}$$

$d.f = n-1 = 25-1 = 24$

$1-\alpha = 0.95 \rightarrow \alpha = 0.05, \ \alpha/2 = 0.025$

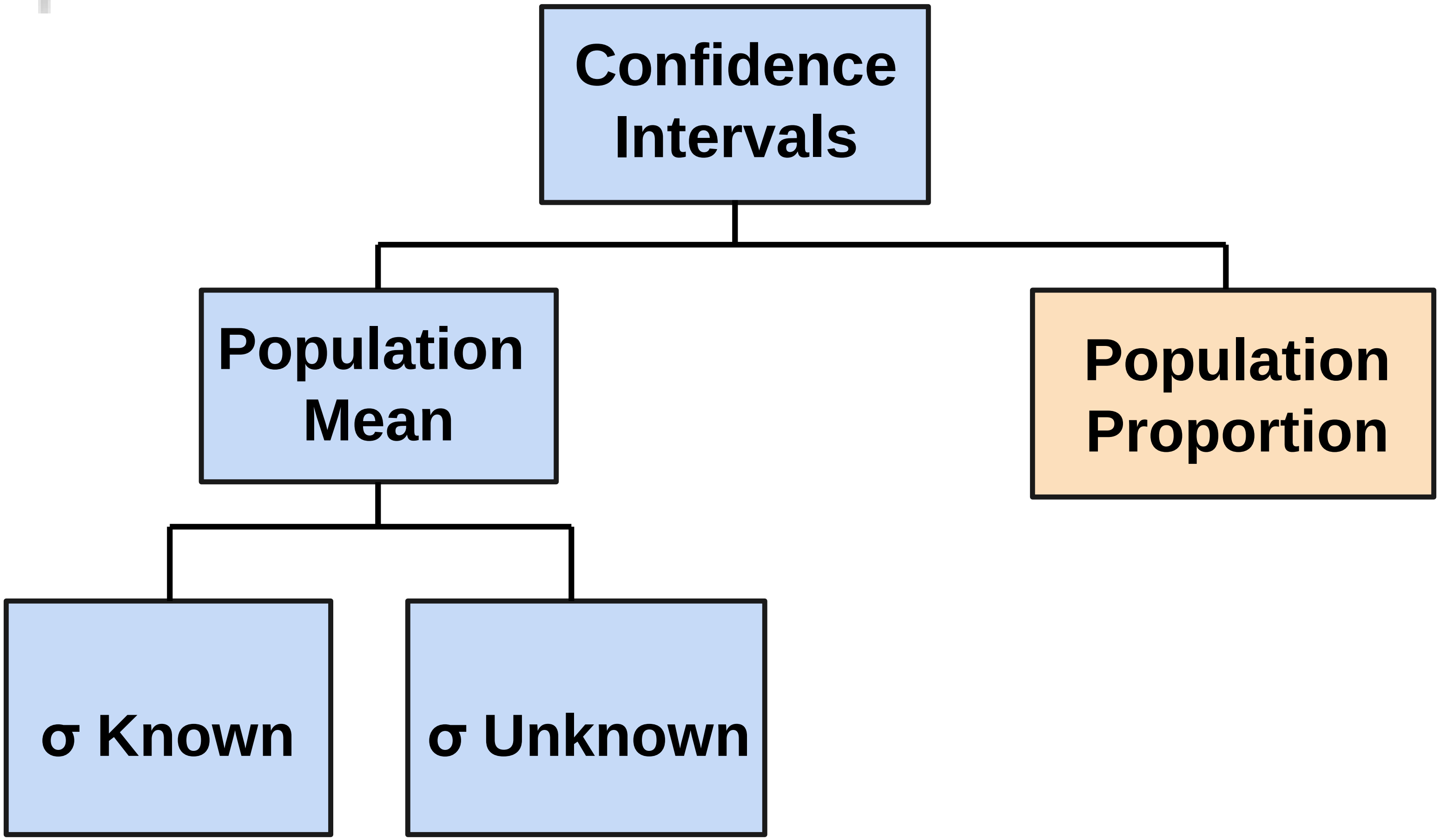
$t_{24,0.025} = 2.064$

Solution

$$50 - 2.064\,\frac{8}{\sqrt{25}} < \mu < 50 + 2.064 \cdot \frac{8}{\sqrt{25}}$$

$$\boxed{46.698 < \mu < 53.302}$$

Confidence Intervals



Confidence Intervals for the Population Proportion, p

- An interval estimate for the population proportion (P) can be calculated by adding an allowance for uncertainty to the sample proportion ($\hat{p}$)

$X \sim \text{Binomial}(n, p) \} N($

Normal Approximation to Binomial

If n is large enough, an approximation to $B(n,p)$ is given by the normal distribution

$$B(n,p) \xrightarrow[n \text{ large}]{} N(np, np(1-p))$$

The normal approximation is often used in statistical inference

Check
$$\boxed{\begin{array}{l} n.p > 5 \\ n.q > 5 \end{array}}$$

$(1-p) = q$

Confidence Intervals for the Population Proportion, p

(continued)

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

$$\sigma_P = \sqrt{\frac{P(1-P)}{n}}$$

- We will estimate this with sample data:

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

Note that: (Proof !)

$$\mu_{\hat{p}} = E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{\cancel{n} \cdot p}{\cancel{n}} = p$$

$$\sigma^2_{\hat{p}} = Var\left(\frac{X}{n}\right) = \frac{n \cdot p \cdot q}{n^2} = \boxed{\frac{p \cdot q}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} \Rightarrow * \; \sigma_{\hat{p}} = \sqrt{\frac{p \cdot (1-p)}{n}} \; *$$

$$\hat{p} = \frac{X}{n}$$

$\hat{p}$: point estimator of proportion p

Confidence Interval Endpoints

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

$$\boxed{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

where

- $z_{\alpha/2}$: Standard normal value for the level of confidence desired.
- $\hat{p}$: sample proportion
- n : sample size

Question 3:

- 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 = prob. of success?

$P = \frac{25}{100}$

$n.p > 5$? $\rightarrow 100.\left(\frac{25}{100}\right) = 25 > 5$

$n.q > 5$? $\rightarrow 100\left(\frac{75}{100}\right) = 75 > 5$

Solution

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

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

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \frac{25}{100} - 1.96\sqrt{\frac{0.25(0.75)}{100}} < P < \frac{25}{100} + 1.96\sqrt{\frac{0.25(0.75)}{100}}$$

$$\boxed{0.1651 < P < 0.3349}$$

Interpretation

* We are 95 % confident that the true percentage of left-handers in the population is between $\boxed{16.51\ \% \text{ and } 33.49\ \%}$

* Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95 % of intervals formed from samples of size 100 in this manner will contain the true proportion