Date:
/ /

لجان الدُّفعات

Soul Academy

تفريغ إحصاء صيدلاني

موضوع المحاضرة: lecture 15

رقم المحاضرة: فاينال

إعداد الصيدلانيه: نور مصطفى

Estimating a Single Population Mean: Point Estimation and Interval Estimation (Confidence Interval)

Dr Muna Oqal

lecture 15

Estimating a Single Population Mean: Point Estimation and Interval Estimation (Confidence Interval)

- If we wish to estimate the mean of some normally distributed population (μ), we would draw a random sample of size n from the population and compute the sample mean ($\overline{\mathbf{x}}$) which can be used as a point estimate of the population mean (μ). الـ mean الي حسبناه وعملناله point estimation
- Although the sample mean is a good estimator of the population mean, we know that random sampling inherently involves chance and the sample mean cannot be expected to be equal to the population mean.
- It is more meaningful to estimate μ by an interval that communicates information regarding the probable magnitude of μ, this interval is called the confidence interval.

اذا كان عنا population وهي normally distributed واحنا بدنا ناخذ random sample من هاي population
ونحسب الـmean، هسا الـmean الي حسبناه وبدنا نعمل عليه estimation بنسميه point estimation

Definition

Point Estimate

A single value (or point) used to approximate a population parameter

	Population Parameter	تقريباً	Best Point Estimate
Proportion	p	$\approx$	$\hat{p}$
Mean	μ	$\approx$	$\bar{x}$
Std. Dev.	σ	$\approx$	s

Definition

Confidence Interval:

mean $\pm$ CI ← مثلاً

- It is abbreviated as **CI**

95% CI بيقولوا

- The **range** (or **interval**) of values used to estimate the a population parameter. The general formula for constructing a confidence interval is given as follows:

CI = Point Estimate ± (Critical Value)(Standard Error)

Where:

- Point estimate is the sample statistic estimating the population parameter of interest.
- Critical value is a table value based on the sampling distribution of the point estimate and the desired confidence level.
- Standard error is the standard deviation of the point estimate.

Definition

Confidence Level : 1 – α /1

The **probability** that the confidence interval actually contains the population parameter.

The most common confidence levels used are **90%**, **95%**, **99%** المستخدمة بالعادة

one sd 2sd 3sd

90% : **α**=0.1 95% : **α**=0.05 99% : **α**=0.01.

$\alpha = 1-90\%$ $\alpha = 1-95\%$ $\alpha = 1-99\%$

$CI = 1-\alpha \longrightarrow \alpha = 1 - CI$

Objectives

- **In this section, we will learn how:**

To construct and interpret confidence interval estimates for the population mean.

Two cases for the confidence intervals of the population mean (μ)

- When population standard deviation σ is known
- When population standard deviation σ is unknown.

سيغما

Estimating a Population mean μ (σ known)

Objective

CI

Find the **confidence interval** for a population mean **μ** when **σ** is known

Requirements

(1) The population standard deviation **σ is known**

(2) One or both of the following:

1 The population is **normally distributed**

or

2 **n** > 30

* الطريقة مش مهم بالاكتشافات إلي قبل، انه كيف اعرف اذا normal او لا.

Best Point Estimation

The **best point estimate** for a **population mean** μ (**σ** known) is the **sample mean** $\bar{x}$

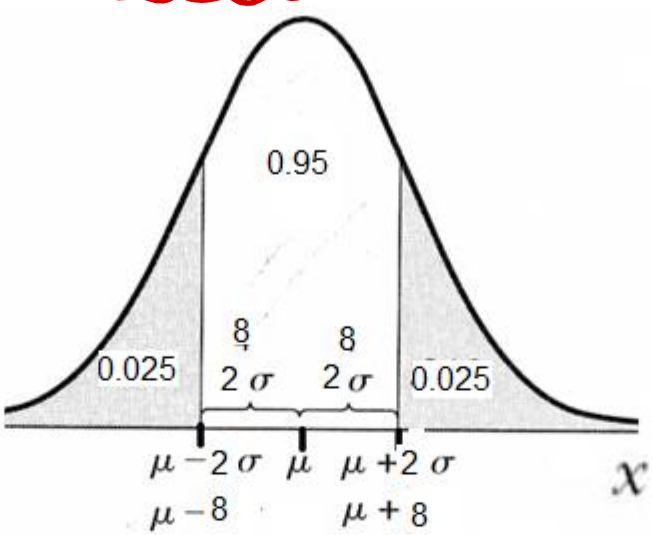
Best point estimate : $\bar{x}$

Confidence Interval for Population Mean

- Suppose that the mean (μ) a population of normal distribution is **UNKNOWN** . But there is **good estimate** of the **population variance** (KNOWN). If we wish to estimate the mean of such population, we would draw a random sample of size n from the population and compute its mean , which can be used as a point estimate of μ.
- E.g. Let us imagine all possible samples (n=4) were taken from a population normally distributed with unknown μ and known variance of 16 to give sampling distribution of the same mean of population (μ) and $\sigma_{\bar{x}} = 2$

$\sigma^2 \longrightarrow \sigma = 4$

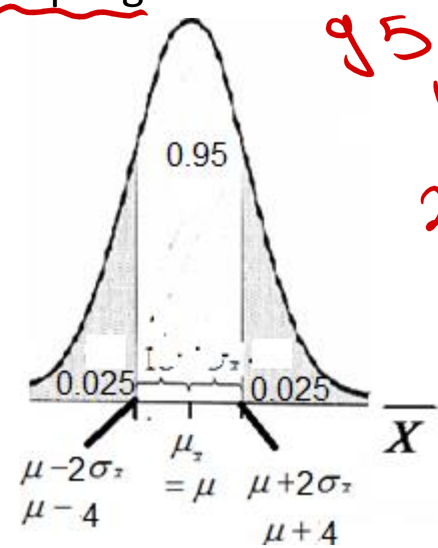
Population distribution



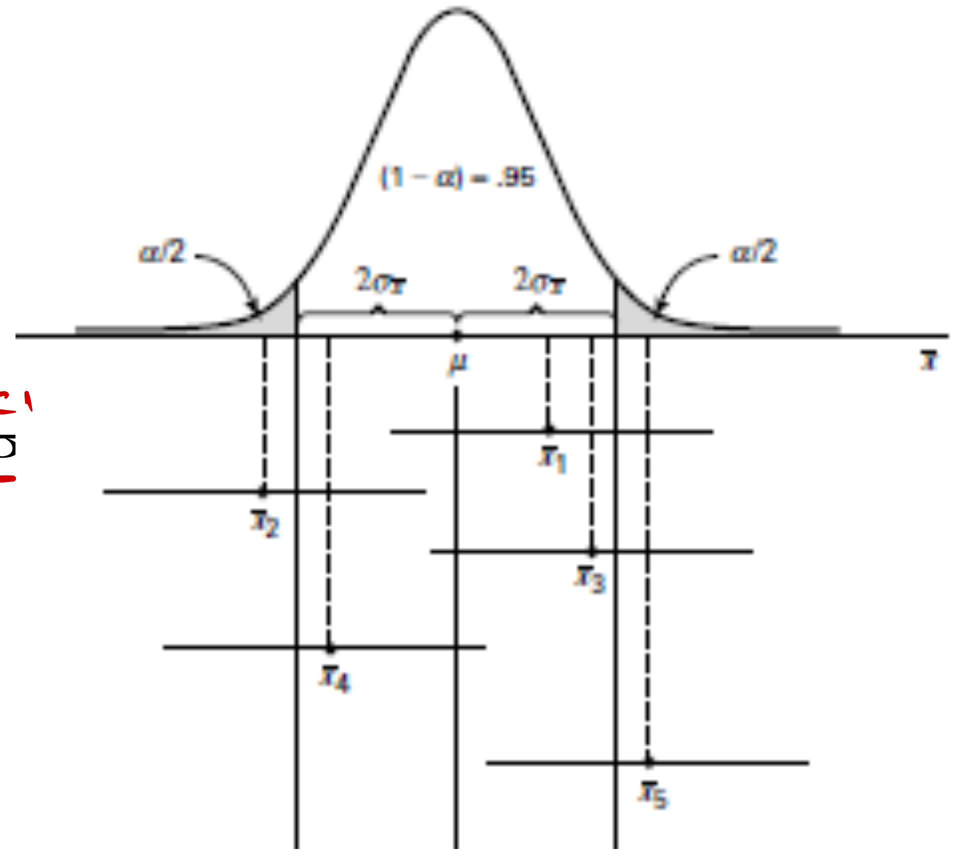
Sampling distribution

95% = CI عالفرض *

2σ



- The **probability** that a sample would have a **mean within 2σ** of the sampling distribution is **0.95**. This the same as what is the percent of samples of all possible samples that give means within 2σ of the sampling distribution.
- For such samples if we construct CI intervals for each mean as $\bar{x} \pm 2\sigma$ **95% of these intervals would capture or overlap with the population mean (μ).**
- For any random sample we are **95% confident that** $\bar{x} \pm 2\sigma_{\bar{x}}$ **captures the population mean (μ).** Thus the confidence level is 0.95 and this constructed interval is called <u>**confidence interval.**</u>

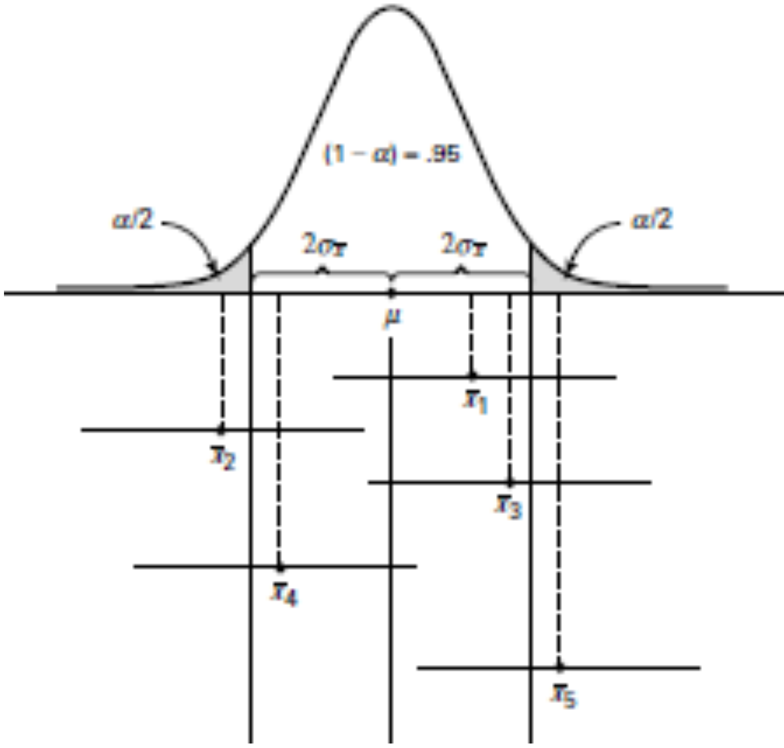


The probability that a random sample has a mean outside $\mu \pm 2\sigma_{\bar{x}}$ of the sampling distribution is 0.05(α), because area above =area below = (α/2)=0.025.

This the same as what is the percent of all possible samples which give means higher or lower than $\mu \pm 2\sigma_{\bar{x}}$

For such samples if you constructed interval for each sample mean as $\bar{x} \pm 2\sigma_{\bar{x}}$ the confidence intervals will not capture or overlap with the population mean (μ).

Accordingly, there is 5% (α) chance that a random sample will not contain the population mean within it confidence interval at confidence level of 0.95. Such samples are unusual samples and can poorly estimate the population mean.



α Is called Type 1 error. α =1-confidence level

Or confidence level =1-α

.

Summary: when a sample is taken from a population and its mean is to be used to estimate the population mean. There is two chances or probabilities as follows

- The sample mean is within certain standard deviation around the population mean at certain level of confidence. E.g. within 2σ if confidence level is 0.95, at this level the probability the sample mean to be within 2σ around the population= mean is 0.95.If this was the real case, then fair estimation of the population mean is obtained.

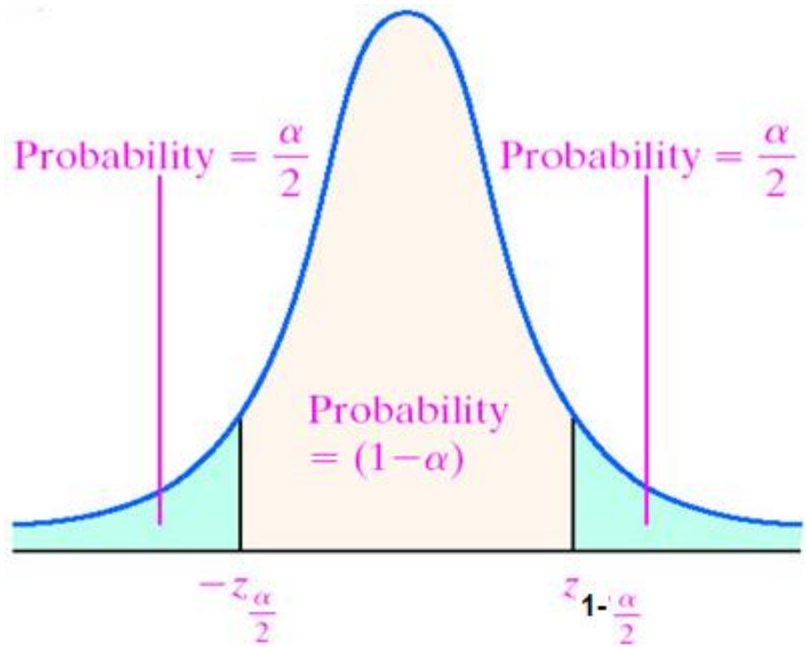
  within interval

- The sample happened to be one of those rare samples, of which its confidence interval does not capture the population mean, and thus it would poorly estimate the mean of the population.

  # تكون sample في هذه الحالة كتير ضعيفة
  و مابتمثل population

Confidence Interval for μ (σ Known)

$Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail



Where $Z_{(1-\alpha/2)}$ is the value of z to the left of which lies 1- α/2 and to the right of which lies 1- α/2 of the area under its curve.

- **Assumptions**
  - **Population standard deviation σ is known**
  - **Population is normally distributed**
  - **If population is not normal, use large sample (n > 30)**

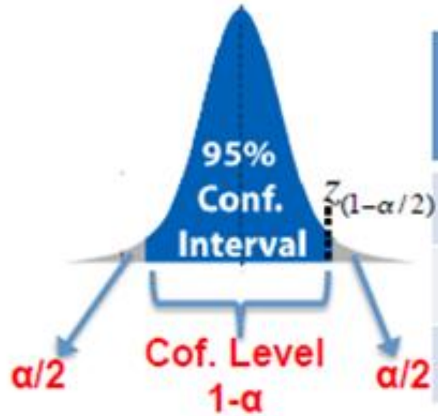
$$\bar{X} \pm z_{(1-\alpha/2)} \sigma_{\bar{x}} =$$

$$\bar{X} \pm Z_{(1-\alpha/2)} * \frac{\sigma}{\sqrt{n}} =$$

$$\bar{X} \pm Z_{(1-\alpha/2)} * SE$$

estimator ∓ (reliability coefficient) x (standard error)

Common Levels of Confidence!!



α	Conf. Level 1-α	$z_{(1-\alpha/2)}$
10%(0.1)	90% (0.9)	$z_{0.95} = 1.645$
5%(0.05)	95% (0.95)	$z_{0.975} = 1.96$
1%(0.01)	99% (0.99)	$z_{0.995} = 2.578$

$$90\%\, CI\, of\ \mu = \overline{X} \pm 1.645\left(\sigma_x/\sqrt{n}\right)$$

$$95\%\, CI\, of\ \mu = \overline{X} \pm 1.96\left(\sigma_x/\sqrt{n}\right)$$

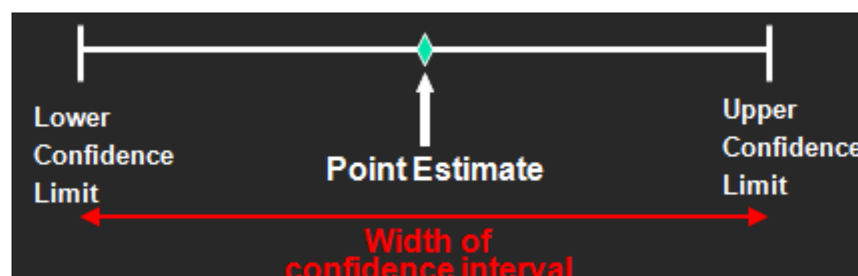
Any z can be used (from table)

$$99\%\, CI\, of\ \mu = \overline{X} \pm 2.578\left(\sigma_x/\sqrt{n}\right)$$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.97128	.97193	.97257	.97320	.97381	.97441	$\boxed{.97500}$	.97558	.97615	.97670

Width of CI is $2 \times Z_{1-\alpha/2} * \dfrac{\sigma}{\sqrt{n}}$

* السؤال ڤي سلايد ١١

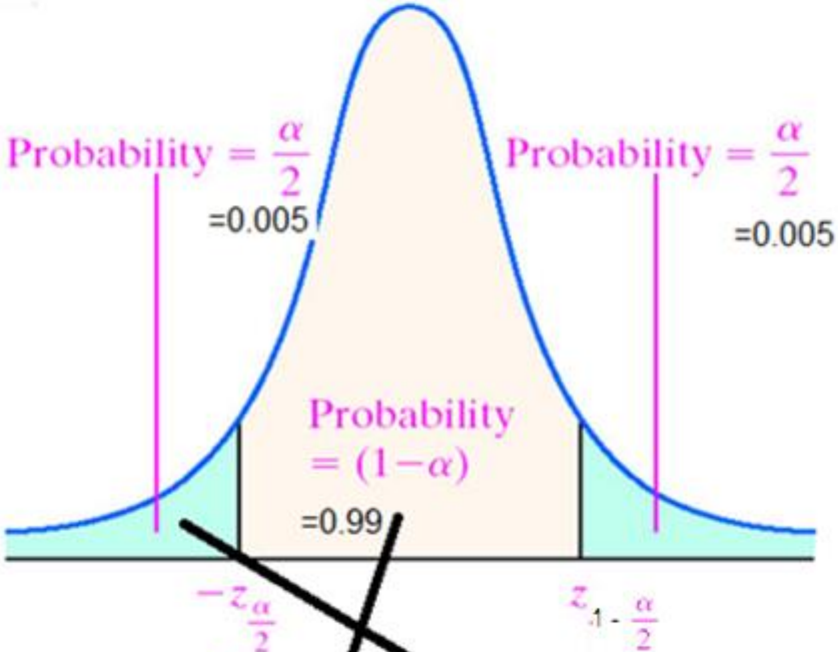


$Z_{1-\alpha/2}$ of 0.95 confidence is $z_{0.95}$ with areas above of 0.05 and below of 0.95, which is 1.645 as shown in the table below

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$$\boxed{z_{0.95} = (1.64+1.65)/2 = 1.645}$$

α =1-confidence level (0.99) = 0.01



Area below $z_{1-\frac{\alpha}{2}}$ =0.99 +0.005=0.995 OR

Area below $z_{1-\frac{\alpha}{2}} = 1 - \frac{\alpha}{2}$ = 1-0.005 = 0.995

$Z_{1-\alpha/2}$ of 0.99 confidence is $z_{0.995}$ with areas above of 0.005 and below of 0.995, which is 2.578 as shown in the table below

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60

Example

A random sample of **25** pharmacy students selected from HU had a grade point average with a **mean of 2.86**. Past studies have shown that the **standard deviation is 0.15** and the population is **normally distributed**. Construct a **90% confidence** interval fro the population mean grade point average (μ)?

Solution: we have normal distribution and a known σ. The (1-α)100% = 90% confidence interval for the population mean (μ) can be constructed as follows:

Step 1:

$(1 - \alpha)100\% = 90\%$

$1 - \alpha = 0.9$, so $\alpha = 0.1$, and $\alpha/2 = 0.1/2 = 0.05$

$1 - \alpha/2 = 1 - 0.05 = 0.95$ so $Z_{0.95} = 1.645$

Continued

Step 2:

$$CI = \overline{X} \pm Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

0.03

Lower limit = 2.86 – (1.645)(0.15/$\sqrt{25}$)

= 2.86 – 0.04935

= 2.811

Upper limit = 2.86 + 0.04935

= 2.909

Then CI = (L, U) = (2.811, 2.909)

Conclusion

With 90% confidence we can say that the mean grade point average for all students in the population (μ) is between 2.811 and 2.909.

Example

We wish to estimate the average number of heartbeats per minute for a certain population (μ). The **average** number of heartbeats per minute for a random sample of **size 49** subjects was found to be **90** with a **standard deviation of 20.** Find the 95% confidence interval for the population mean(μ)?

Step 1:

we have normal distribution (Population), unknown σ (S=10), and sample size 49 which is > 30.

The (1-α)100% = 95% confidence interval for the population mean (μ) can be constructed as follows:

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

1- α = 0.95, so α = 0.05, and α /2 = 0.05/2 = 0.025

$1 - \alpha/2 = 1 - 0.025 = 0.975$ so $Z_{0.975}$ = 1.96

Step 2:

$$CI = \bar{X} \pm Z_{1-\alpha/2} * \frac{S}{\sqrt{n}}$$

$$\text{Lower limit} = 90 - (1.9)(10/\sqrt{49})$$
$$= 90 - 2.8$$
$$= \mathbf{87.2}$$
$$\text{Upper limit} = 90 + 2.8$$
$$= \mathbf{92.8}$$

Then CI = (L, U) = (87.2, 92.8)

Conclusion

With 95% confidence we can say that the mean number of heartbeats per minute for all subjects in the population (μ) is between 87.2 and 92.8

Confidence Interval for Population Mean

- A physical therapist wished to estimate, with 99% confidence, the mean maximal strength of a particular muscle in a certain group of individuals. Assuming that strength scores are approximately normally distributed with a variance of 144. a sample of 15 subjects who participated in the study yielded a mean of 84.3. μ

$\sigma = 12$

- The z value corresponding to a confidence level of 0.99 is found to be 2.578. → جاهزة
- The standard error is → $\sigma_{\bar{x}}$ $\qquad s / \sqrt{N}$ $\qquad 12/\sqrt{15} = 3.0984$
- The 99% CI for μ is $\qquad 84.3 \pm 2.578(3.0984)$

Go to previous slide on finding $Z_{1-\alpha/2}$

Point estimate → 84.3 ± 8 ← width of confidence interval = 8×2=16

Lower confidence limit → $(76.3 - 92.3)$ ← Upper confidence limit

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

- Problem: Standard error is unknown (σ is also a parameter). It is estimated by replacing σ with its point estimate from the sample data:

→ of σ

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

95% confidence interval for μ :

$$\bar{X} \pm 1.96(\sigma_{\bar{x}} \quad \text{which is} \quad \bar{X} \pm 1.96\frac{s}{\sqrt{n}}$$

This works ok for "large *n*," because *s* then a good estimate of σ (and CLT applies). **But for small *n*, replacing σ by its estimate *s* introduces extra error, and CI is not quite wide enough unless we replace *z*-score by a slightly larger "*t*-score."**

The t Distribution

- The t-distribution is used instead of the standard normal distribution (z) to estimate the population mean (μ) if:

1. The population is normal
2. The normal population standard deviation σ is unknown.
3. The sample size (n) is small ($n < 30$).

The t Distribution

- Properties of the t distribution:
  - It has a mean of 0
  - It is symmetrical about the mean
  - In general it has a variance greater than 1, but the variance approaches 1 as the sample size becomes larger.
  - The variable t ranges from $-\infty \text{ to } +\infty$
  - The t distribution is a family of distributions, since there is a different distribution for each sample value of n-1 (the divisor used in computing s^2).
  - Compared to the normal distribution, the t distribution is less peaked in the center and has higher tails.
  - The t distribution approaches the normal distribution as n-1 approaches infinity.

Comparison of Selected *t* Distributions to the Standard Normal

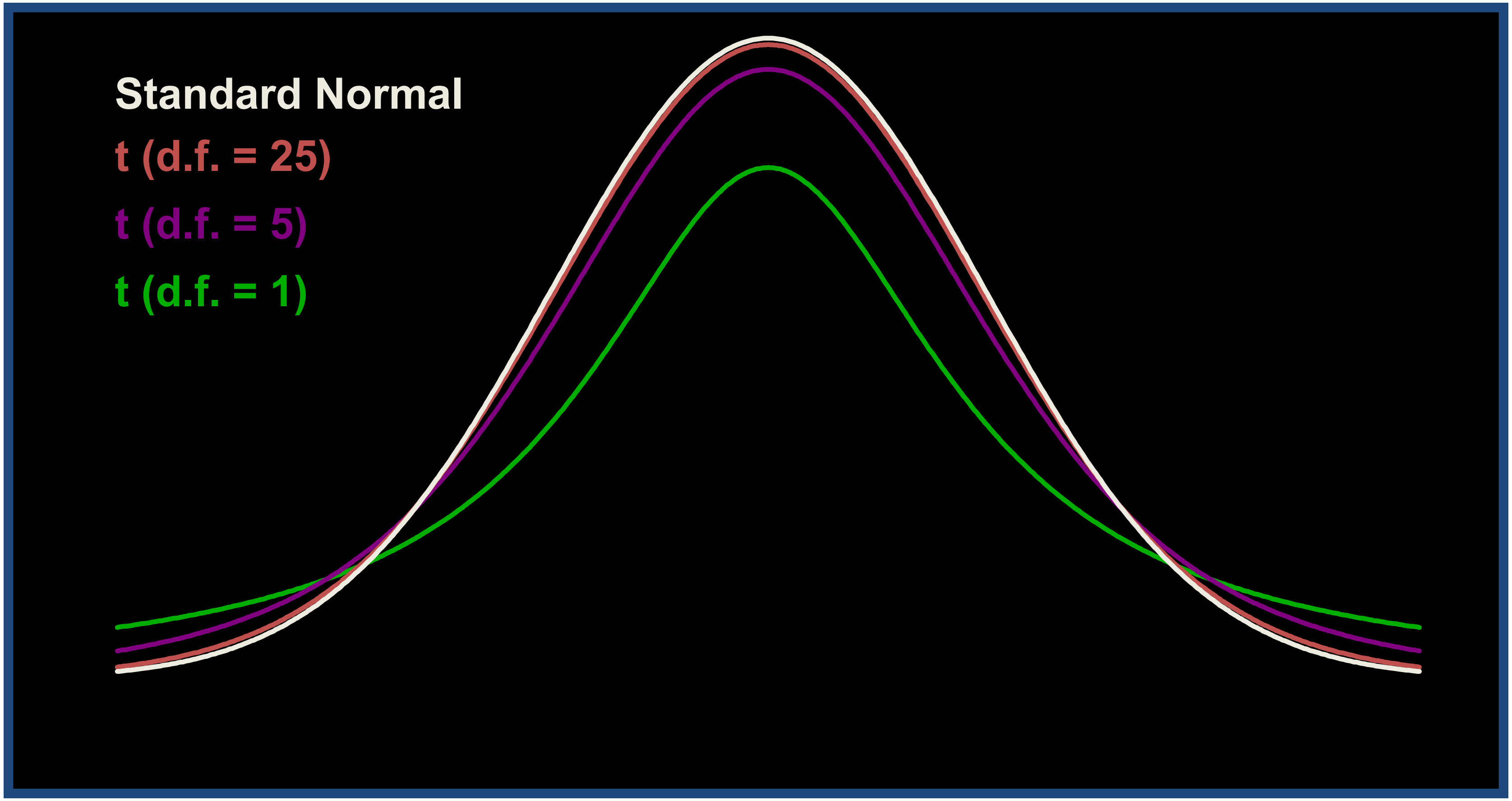
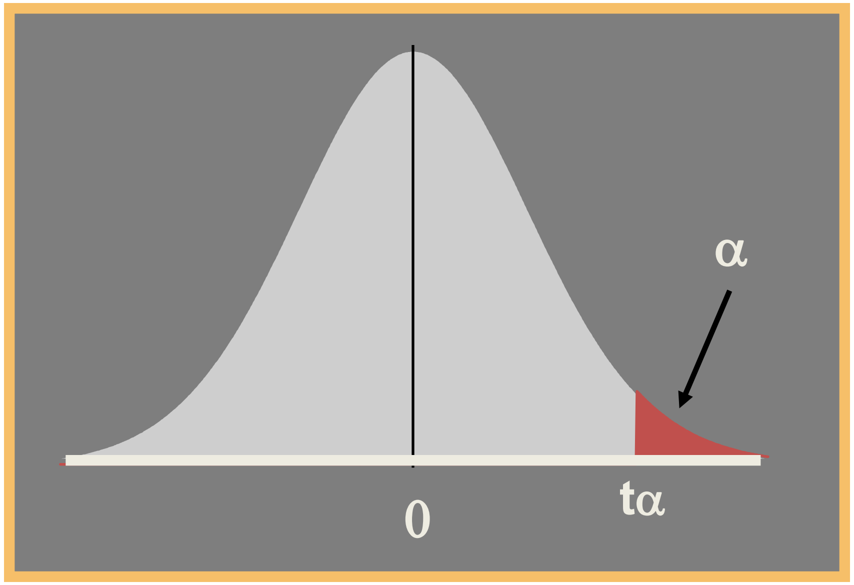


Table of Critical Values of *t*

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	[illegible]	**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



Confidence Intervals for population of **unknown** mean and **unknown** variance

- It is the usual case that the population variance as well as the population mean are unknown.
- As a result, although the z statistic is normally distributed, we can not use this fact because σ is unknown and thus standard deviation of the sampling distribution ($\sigma_{\bar{x}}$)cannot be calculated.
- So we may use the sample standard deviation to replace σ.

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

The t Distribution

- When s is used to replace σ, instead of z-scores t-values are calculated as follows

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

- T-scores has normal distribution called Student's t distribution. For t-distribution **degrees or freedom** is calculated as **n-1**, which is the denominator of sample standard deviation

- df: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is $\boxed{8.0}$

Let $X_1 = 7$
Let $X_2 = 8$
What is $\mathbf{X_3}$? **X_3 must be 9. n-1 numbers (d,f.) can be assumed to be any (2 in the example), but one number cannot and depends on the mean.**

Student's t Table

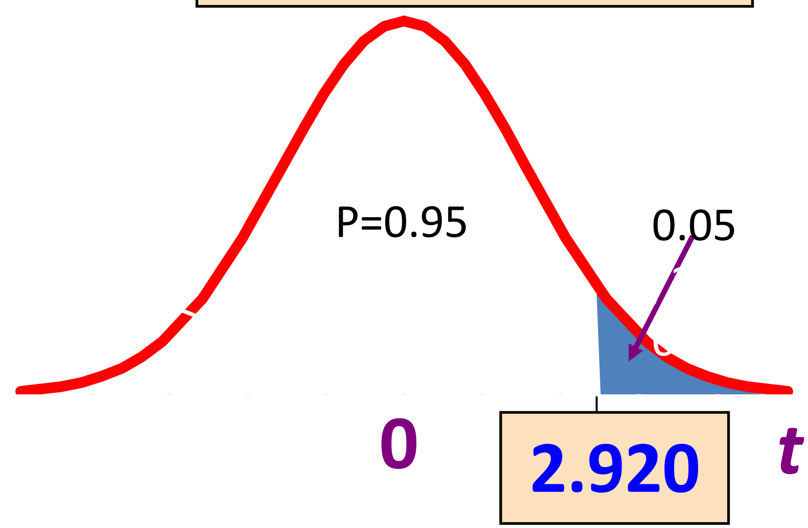
	P		
df	0.9	**0.95**	0.975
1	3.078	6.314	12.706
2	1.886	$\boxed{\textbf{2.920}}$	4.303
3	1.638	2.353	3.182

The body of the table contains t values, not probabilities

Let: n = 3
df = *n* - 1 = 2
confidence level: 90%

$\alpha = 0.10$
$\alpha/2 = 0.05$

$P=1-\alpha/2 = 0.95$



Confidence Interval for μ (σ Unknown)

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)
- Use Student's t Distribution
- Confidence Interval Estimate:

بدل z حطينا t

$$\overline{X} \pm t_{1-\alpha/2} \frac{S}{\sqrt{n}}$$

(where $t_{\alpha/2}$ is the critical value of the t distribution with n -1 degrees of freedom and an area of $\alpha/2$ in each tail)

estimator ∓ (reliability coefficient) x (standard error)

The *t* distribution (*Student's t*)

- Bell-shaped, symmetric about 0
- Standard deviation a bit larger than 1 (slightly thicker tails than standard normal distribution, which has mean = 0, standard deviation = 1)
- Precise shape depends on **degrees of freedom** (*df*). For inference about mean,

  $$df = n - 1$$

- Gets narrower and more closely resembles standard normal distribution as *df* increases

  (nearly identical when $df > 30$)

- CI for mean has margin of error $t(se)$,

  (instead of z(se) as in CI for proportion)

Part of a *t* table

	Confidence Level			
	90%	95%	98%	99%
df	t.050	t.025	t.010	t.005
1	6.314	12.706	31.821	63.657
10	1.812	2.228	2.764	3.169
16	1.746	2.120	2.583	2.921
30	1.697	2.042	2.457	2.750
100	1.660	1.984	2.364	2.626
infinity	1.645	1.960	2.326	2.576

في infinity تتساوى مع Z & t

df = $\propto$ corresponds to standard normal distribution

Confidence Interval for μ (σ Unknown)

- 20 tablets were chosen randomly from a batch. Their weights in mg were

300 321 306 321 310 322 315 325 316 323 316 325 317 325 319 327 320 331 320 336. Provide a 95 % CI for the estimated mean

-Estimate the mean =319.75

$$s = \sqrt{\frac{\sum_{i-1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}} = 8.2$$

-df is 19 so $t_{0.975}$=2.093 (t-table)

$$95\% \text{CI of } \mu = \overline{X} \pm \overset{t}{2.093}\left(s/\sqrt{n}\right)$$

$$95\% \text{CI of } \mu = 319.75 \pm \overset{t}{2.093}\left(8.2/\sqrt{20}\right)$$

315.91 to 323.59

Table A.2 Selected Percentiles of t-Distributions

Tabled values are $t_{df}(p)$

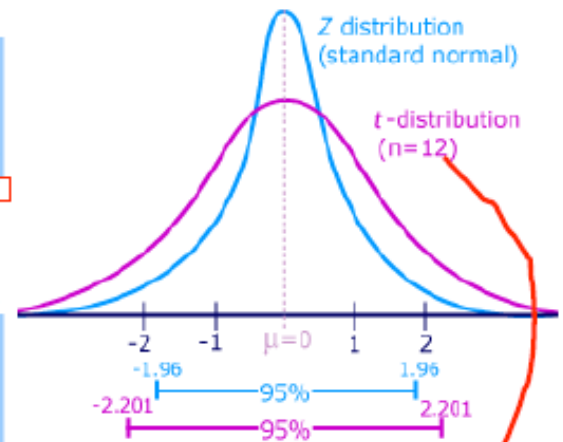
d.f.	.75	.80	.85	.90	.95	.975	.98	.99	.995	.9975	.999	.9995
						Probability p						
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.67	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.32	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	1.215	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
70	0.678	0.847	1.044	1.294	1.667	1.994	2.093	2.381	2.648	2.899	3.211	3.435
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
90	0.677	0.846	1.042	1.291	1.662	1.987	2.084	2.368	2.632	2.878	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
500	0.675	0.842	1.038	1.283	1.648	1.965	2.059	2.334	2.586	2.820	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291

هاد الجدول حيعطى في الامتحان

t Table

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

عند infinity



D.F=n-1=11

Note: differentiate between cum. Probability and confidence level

Degrees of freedom (ν)	Amount of area in one tail (α)							
	0.0005	**0.001**	**0.005**	**0.010**	**0.025**	**0.050**	**0.100**	**0.200**
1	636.6192	318.3088	63.65674	31.82052	12.70620	6.313752	3.077684	1.376382
2	31.59905	22.32712	9.924843	6.964557	4.302653	2.919986	1.885618	1.060660
3	12.92398	10.21453	5.840909	4.540703	3.182446	2.353363	1.637744	0.978472
4	8.610302	7.173182	4.604095	3.746947	2.776445	2.131847	1.533206	0.940965
5	6.868827	5.893430	4.032143	3.364930	2.570582	2.015048	1.475884	0.919544
6	5.958816	5.207626	3.707428	3.142668	2.446912	1.943180	1.439756	0.905703
7	5.407883	4.785290	3.499483	2.997952	2.364624	1.894579	1.414924	0.896030
8	5.041305	4.500791	3.355387	2.896459	2.306004	1.859548	1.396815	0.888890
9	4.780913	4.296806	3.249836	2.821438	2.262157	1.833113	1.383029	0.883404
10	4.586894	4.143700	3.169273	2.763769	2.228139	1.812461	1.372184	0.879058
11	4.436979	4.024701	3.105807	2.718079	2.200985	1.795885	1.363430	0.875530
12	4.317791	3.929633	3.054540	2.680998	2.178813	1.782288	1.356217	0.872609
13	4.220832	3.851982	3.012276	2.650309	2.160369	1.770933	1.350171	0.870152
14	4.140454	3.787390	2.976843	2.624494	2.144787	1.761310	1.345030	0.868055
15	4.072765	3.732834	2.946713	2.602480	2.131450	1.753050	1.340606	0.866245
16	4.014996	3.686155	2.920782	2.583487	2.119905	1.745884	1.336757	0.864667
17	3.965126	3.645767	2.898231	2.566934	2.109816	1.739607	1.333379	0.863279

Selected t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z or t (∞ d.f.)
0.80	1.372	1.325	1.310	1.28
0.90	1.812	1.725	1.697	1.645
0.95	2.228	2.086	2.042	1.96
0.99	3.169	2.845	2.750	2.58

Note: t approaches Z as n increases

* بمعنى انه t distribution بتعطيني Sd اكبر وفترة اعرض .

The t Distribution

- Calculating the confidence intervals using t-distribution is similar to calculating them for Z distribution, with one key difference

- Unlike the Z confidence intervals, which always use 1.96 and 2.576 for the 95% and 99% confidence intervals, respectively, the confidence interval (CI) uses a t value that will vary depending on the degrees of freedom

Example

df = 19

In a random sample of **20** customers at a given pharmacy in Jordan, the **mean** waiting time to get service is **95 seconds**, and the standard deviation is **21 seconds**. Assume the wait times are normally distributed, then construct a **99% confidence** interval for the mean wait time of all customers (μ)?

Solution: we have

1. Normal distribution (population)
2. The standard deviation (σ)is unknown (S=21)
3. The sample size (n) is small (n=20 <30)

The (1-α)100% = 99% confidence interval for the population mean (μ) can be constructed as follows:

Step 1:

two tail

(1 – α)100% = 99%

1- α = 0.99, so α = 0.01, and α /2 = 0.01/2 = 0.005, df= n-1= 20-1=19 ,

$\mathbf{t_{(0.005,\ 19)} = 2.861}$

or 1 – α/2 = 1 – 0.005 = 0.995 so df = 10 and $t_{0.995}$ will give same answer.

one tail

Z-distribution versus t-distribution

- Because for a population of **known** σ, since constant σ is used to calculate z-scores, one z-distribution is obtained with a standard deviation of 1.

- t-scores differ according to sample size of df and s paticulrly for small sample sizes , and thus different t-distributions are obtained of different n or different degrees of freedom (df). **So,** the t distribution is a family of distributions, since there is a different distribution for each sample value of n-1 (the divisor used in computing s^2).

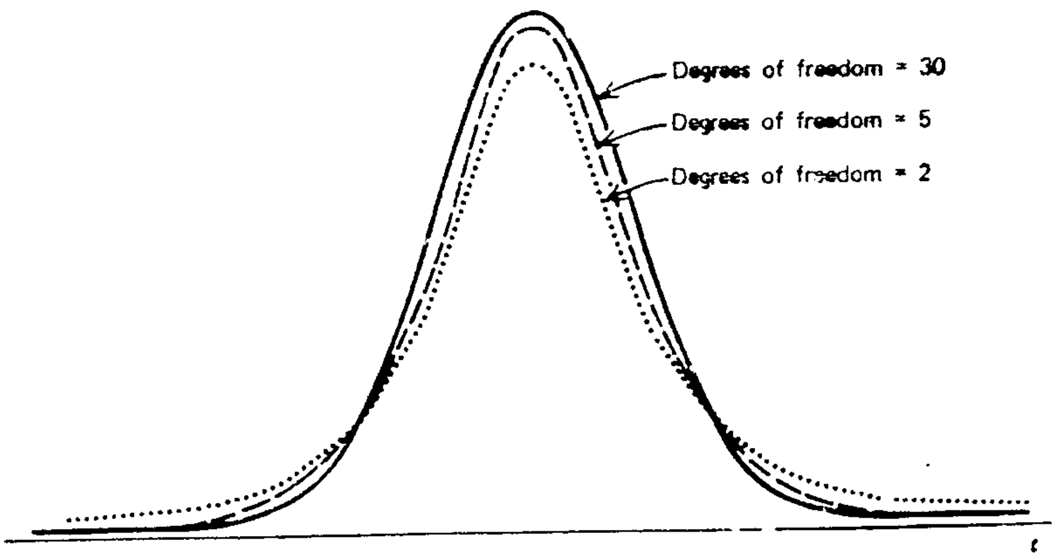


Figure 6.3.1 The t distribution for different degrees-of-freedom values.

As df is smaller the t distribution is less peaked in the center and has fatter tails. Thus standard deviation for the distribution is higher than 1, but it gets closer to 1 as df Or n is increased.

Z- versus t-table

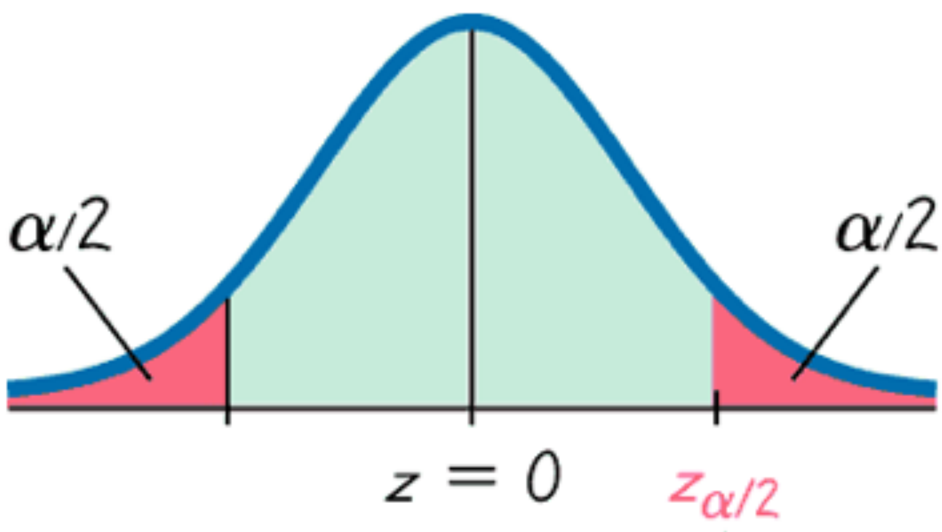
Z-table	T-table
Only one parameter: z-score	Two parameters: cumm.prop and D.F
We calculate z-score and then we find probability (AUC)	We decide on the cumm.prop (from Conf. Level) and then we find the t-score (reliability coefficient)
Always entries less than the corresponding entries for t-tables at same probability	When n approach ∞, t-table entries match the z-table entries
For the z-curve: $\mu=0$; $\sigma=1$	For the t-curve: $\mu=0$; $\sigma>1$ and approach 1 when n is large

Summary and Highlights

Normal Dist. Critical Values

For a population mean μ (σ known), the critical values are found using z-scores on a standard normal distribution

The standard normal distribution is divided into three regions: middle part has area $1 - \alpha$ and two tails (left and right) have area $\alpha/2$ each:

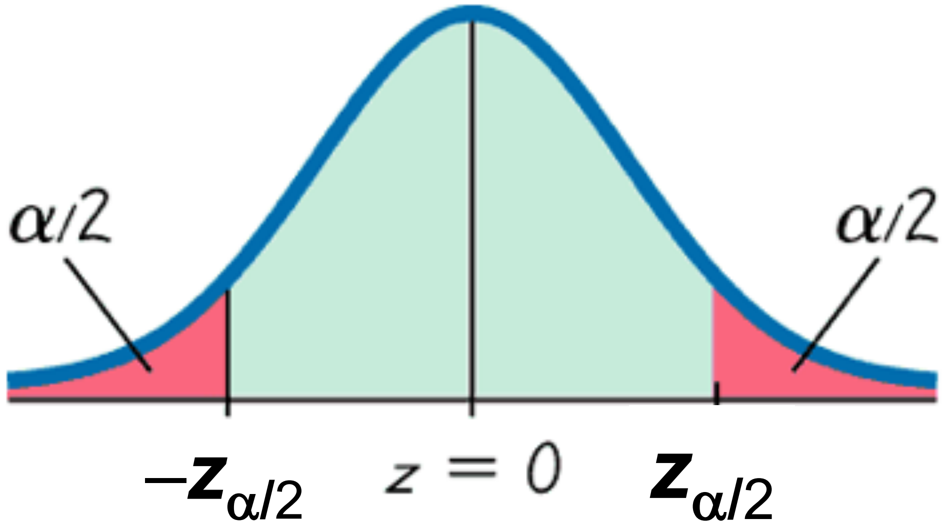


Normal Dist. Critical Values

The z-scores $z_{\alpha/2}$ and $-z_{\alpha/2}$ separate the values:

Likely values (middle interval)

Unlikely values (tails)

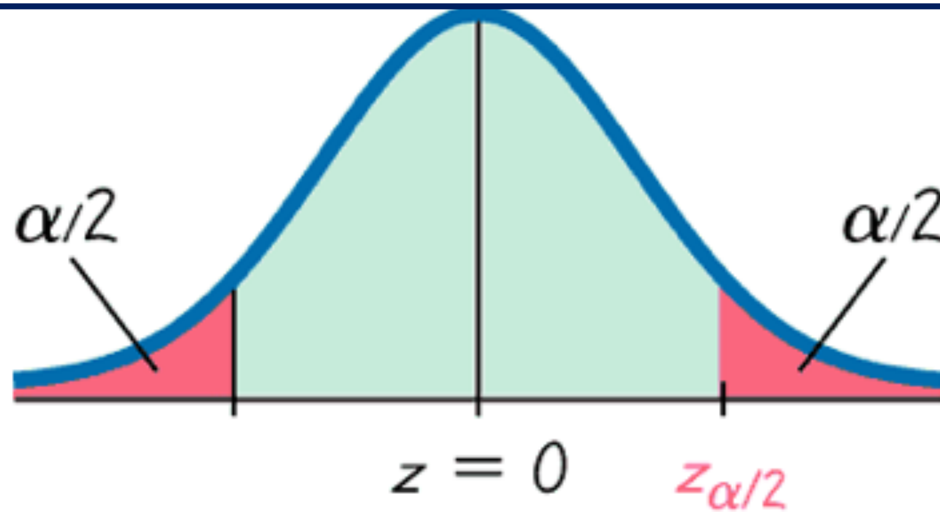


Normal Dist. Critical Values

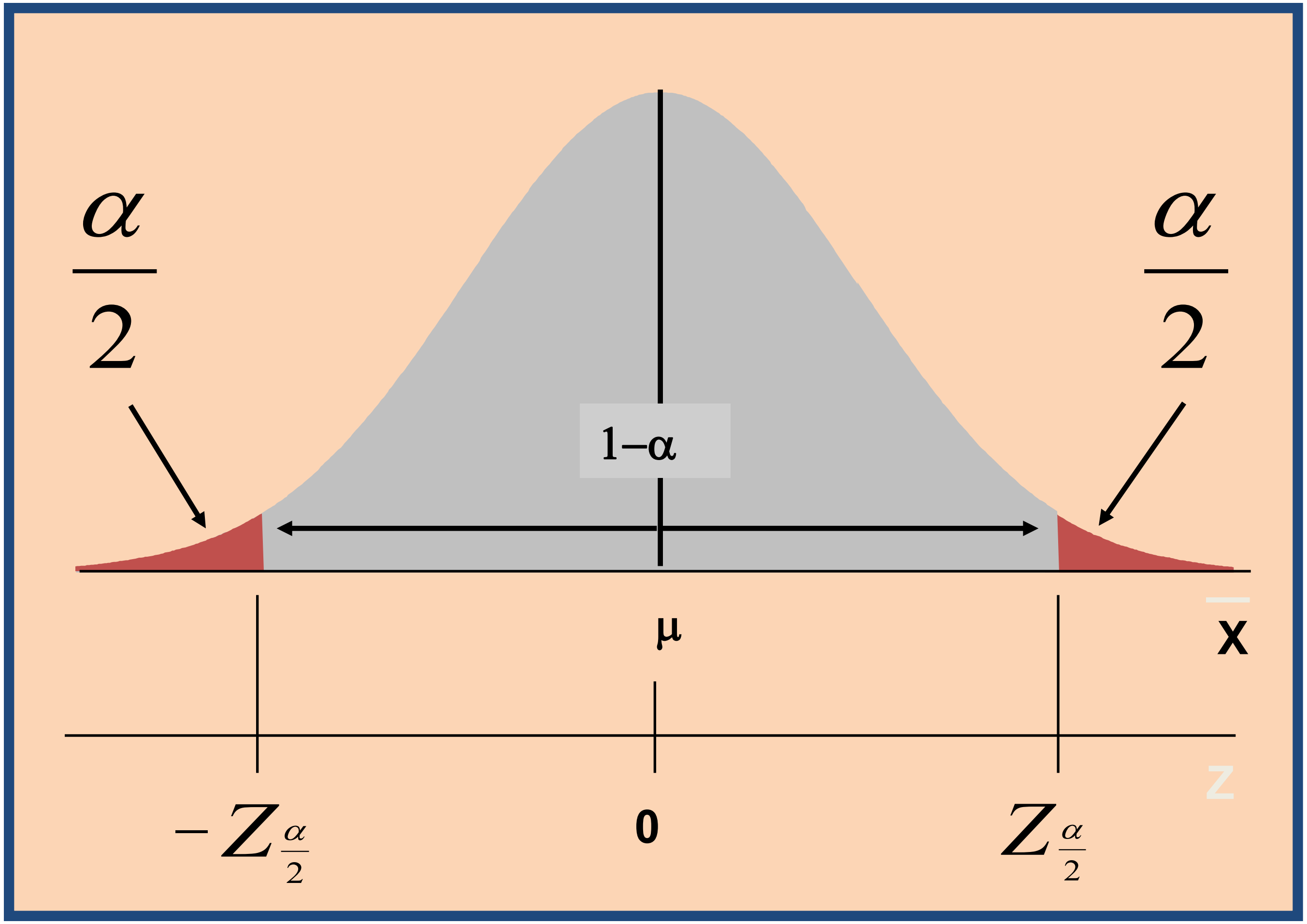
The value $\mathbf{z}_{\alpha/2}$ separates an area of $\alpha/2$ in the **right tail** of the z-dist.

The value $-\mathbf{z}_{\alpha/2}$ separates an area of $\alpha/2$ in the **left tail** of the z-dist.

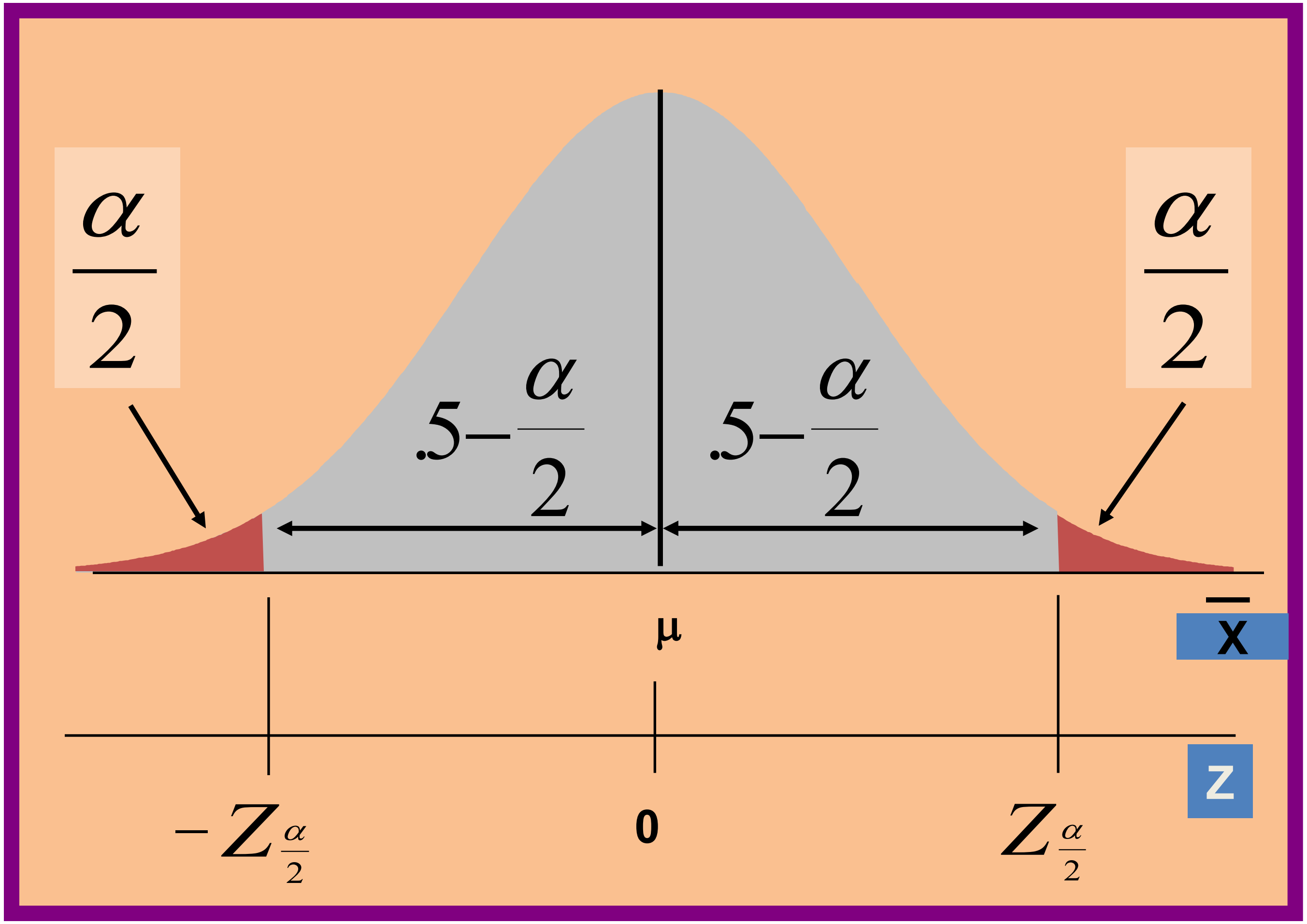
The subscript $\alpha/2$ is simply a reminder that the *z*-score separates an area of $\alpha/2$ in the tail.



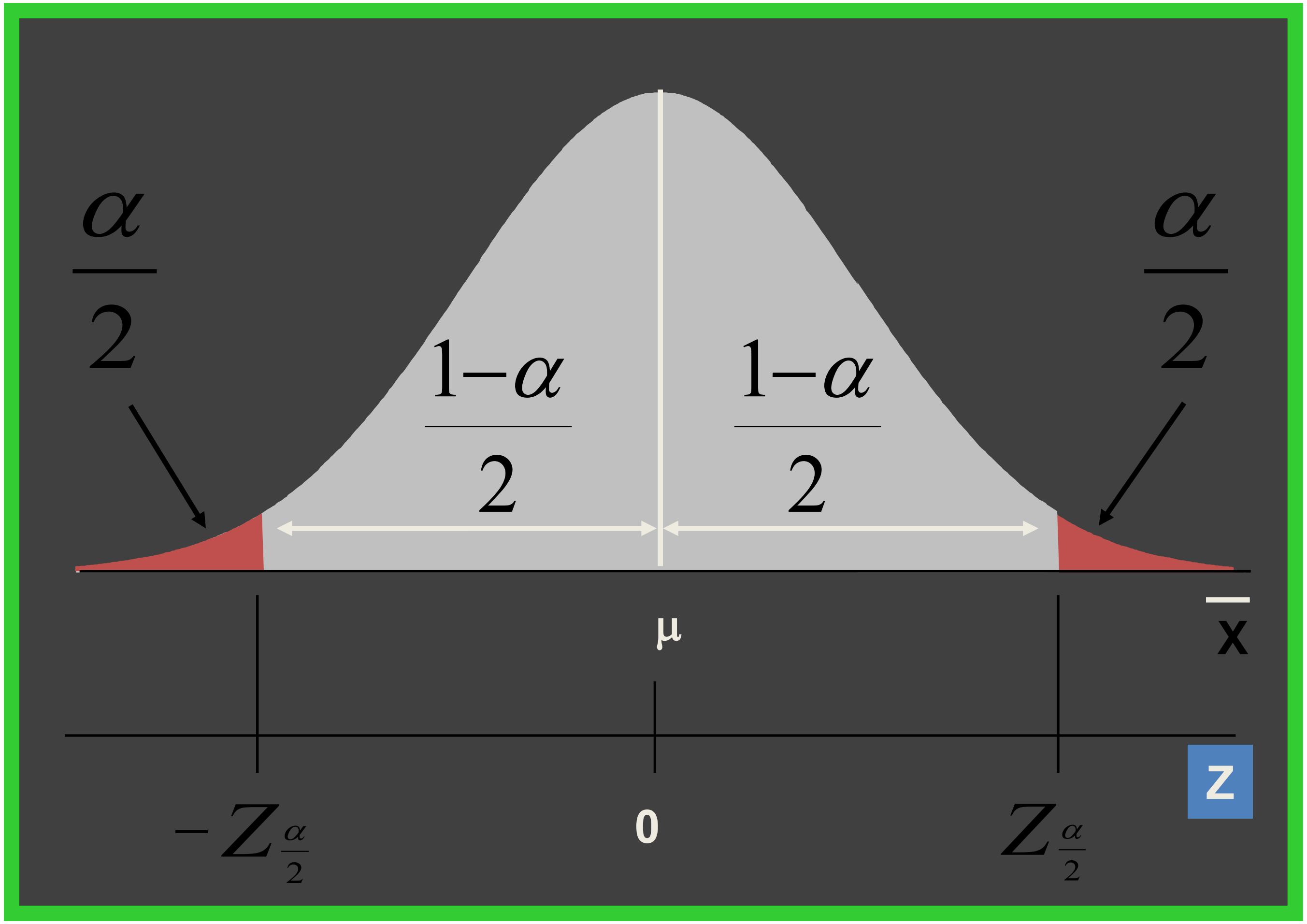
Distribution of Sample Means for (1-α)% Confidence



Z Scores for Confidence Intervals in Relation to α



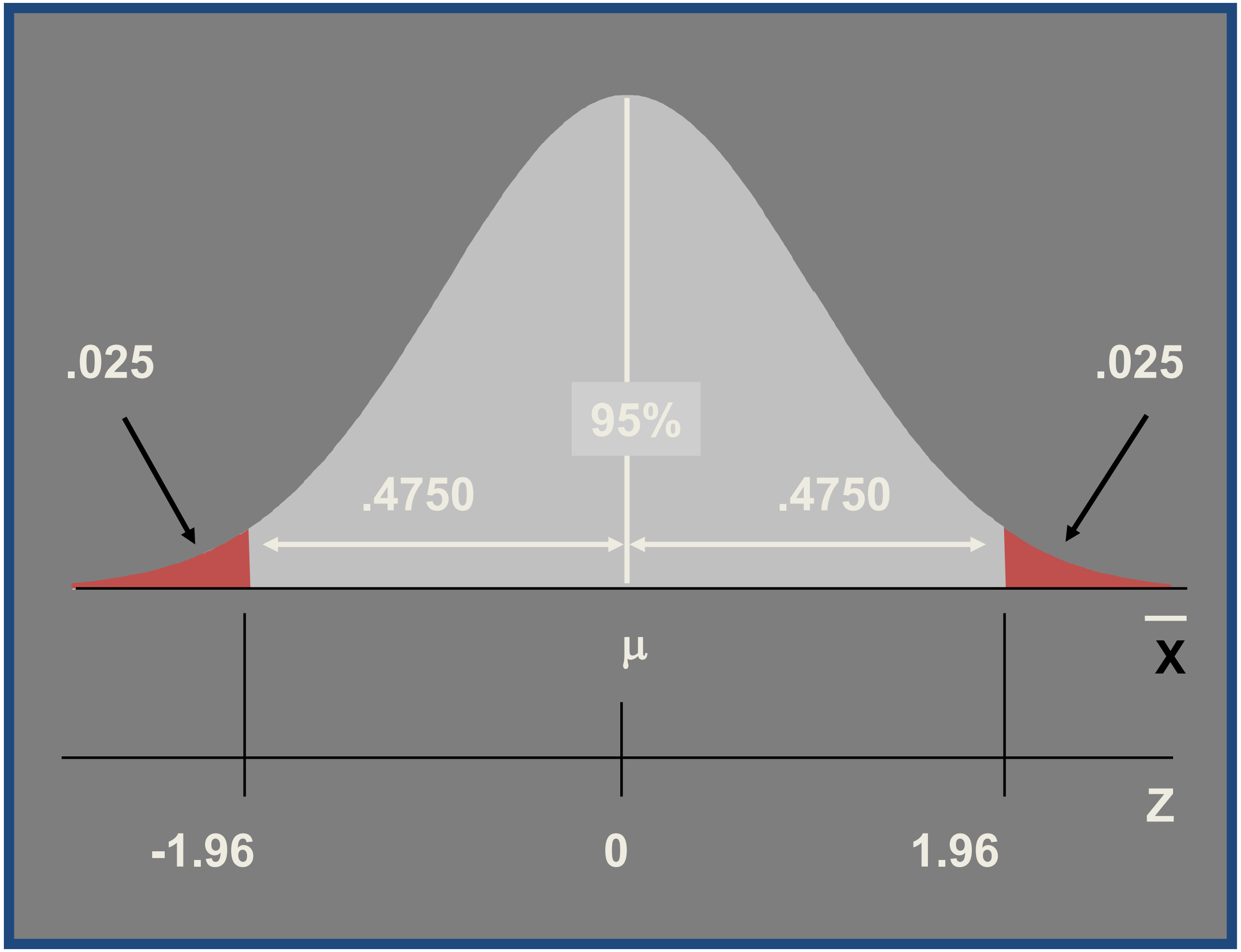
Distribution of Sample Means for (1-α)% Confidence



Probability Interpretation of the Level of Confidence

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

Distribution of Sample Means for 95% Confidence



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

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

or

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

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

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

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

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

$$or$$

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

$$df = n - 1$$

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