لملمة أفكار lec 15 :

confidence Interval (CI)

: The range of values used to estimate the population paramete

$$CI = \text{Point estimate} \pm (\text{critical value})(\text{standard error})$$

Point estimate ⇓ the sample statistic estimating the population parameter of interest.

critical value ⇓ table value: t table ↙ ↘ z table

standard error ⇓ standard deviation of the point estimate.

مثلا لو بدنا estimating μ (mean)

population parameter → sample statistic

$\mu \longrightarrow \bar{X}$

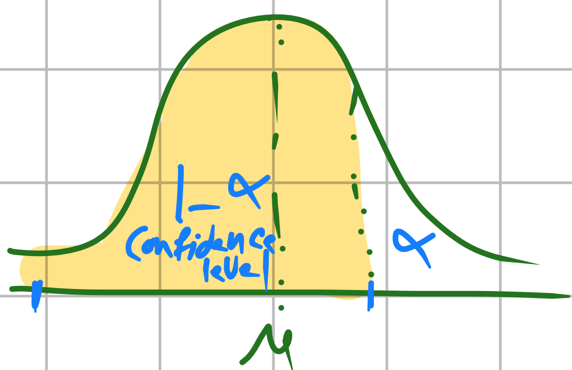
confidence level $(1-\alpha)$

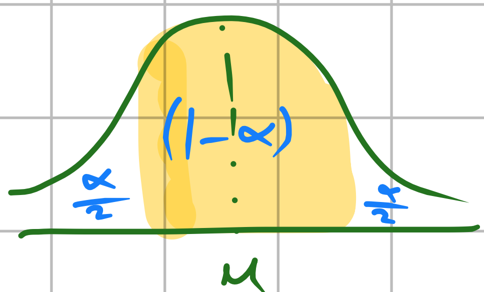
: The Probability that the (CI) actually contains the population parameter.

$\alpha = 0.1 \leftarrow \alpha - 1 = 90\%$

$\alpha = 0.05 \leftarrow \alpha - 1 = 95\%$

$\alpha = 0.01 \leftarrow \alpha - 1 = 99\%$





CI

Part ①

CI for a population mean (μ) when:

1) σ unknown
2) $n < 30$
3) normal distribution

Ⓩ is not appropriate for these conditions.

(CI will not capture the population mean)

⇓

t is appropriate

- قيم t أكبر.
- CI أعرض.

⇓

Can capture the population mean.

لو كان حجم العينة كبير نعتبر أنّ $s \approx \sigma$, ولكن صغر حجم العينة يفسر ← ولو اعتبرناهم نفس الاشي ؟ يصير عنا CI is not quite wide enough ⊕ extra error

Part Ⓐ

CI for a population mean (μ): when:

1) σ is known
2) normally distributed
   - ما حكاها بيرجع العباره ↓ $n > 30$ يعني.
   - حكاها بالسؤال بيرجع العبارة

(sample mean) $\bar{X}$: best point estimate for the population mean

$\bar{x}$ to estimate μ

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.
- 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.

CI captures the population mean ↓ sample mean is a good estimator to the population mean

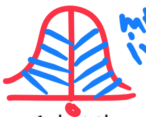
CI doesn't capture the population mean. ↓ sample mean is poorly estimate the population mean.

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

point estimate — critical value — standard error

The t Distribution

- Properties of the t distribution:
  - It has a mean of 0 $\mu = 0$
  - 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. $s^2 > 1$
  - The variable t ranges from $-\infty$ 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). degree of freedom
  - 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.



↑n يزداد اقتراب variance الى 1.

↑ degree of freedom (n-1) يزداد التقرب من ال normal distribution

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, (n-1)

  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)

↑df يزداد قرب شكله من شكل ال normal.

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.

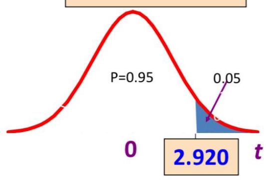
$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{(n-1)}}$$

Student's t Table

Let: n = 3
df = n - 1 = 2
confidence level: 90%
α = 0.10
α/2 = 0.05
P=1- α/2

df	P 0.9	0.95	0.975
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182

The body of the table contains t values, not probabilities



The t Distribution

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

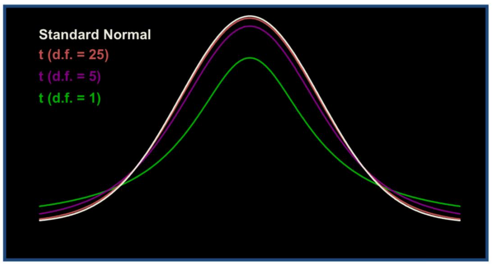
$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

- 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 8.0

Let $X_1 = 7$
Let $X_2 = 8$
What is 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.**

Comparison of Selected *t* Distributions to the Standard Normal



↑df

t distribution shape زاد قرب من normal distribution shape.

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).

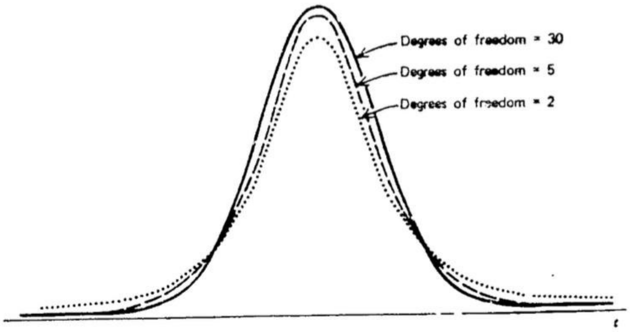


Figure 6.3.1 The t distribution for different degrees-of-freedom values.

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

Examples → Part (A)

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)?

$n = 25$

$\bar{x} = \mu = 2.86$

$\sigma = 0.15$

(normal distribution)

طالب بـ : CI

90% CI

$$\bar{x} \pm Z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}$$

Z ليش استخدمنا

① σ is known ✓

② normal ✓

$\alpha = 0.1$

$1 - \alpha = 0.9$

$Z_{(1-\frac{\alpha}{2})}$

$Z_{(1-\frac{0.1}{2})}$

$Z_{(0.95)}$ جدول Z

$= 1.645$

$$2.86 \pm (1.645) \frac{0.15}{\sqrt{25}}$$

upper limit $= 2.909$

lower limit $= 2.811$

$CI = (2.811, 2.909)$

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:

Step 2:

$CI = \bar{X} \pm Z \ast \frac{s}{\ldots}$

s

$n = 49$

$\mu = \bar{x} = 90$

$s = 20$

σ مش معلومه ولكن لأن حجم العينة كبير

بنعتبر $s = \sigma$

σ is unknown

نرجع لل sample size

$n < 30$ → t جدول

$n \geq 30$ → Z جدول

$49 > 30$

اذا نعتبر normal

ونرجع لل Z جدول

$$90 \pm (1.96)\frac{20}{(7)}$$

$1 - \alpha = 0.95$

$\alpha = 0.05$

$Z(1 - \frac{\alpha}{2})$

$Z(1 - \frac{0.05}{2})$

$Z(0.975)$

$= 1.96$

من جدول Z

Lower = 84.4

upper = 95.6

CI (84.4, 95.6)

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.

$n=15$

$s^2=144$

$s=12$

$n=15$

$s=12$

$\bar{x}=84.3$

$84.3 \pm (2.578)\frac{12}{\sqrt{15}}$

$CI\ (76.3\ ,\ 92.3)$

$Z\left(1-\frac{0.01}{2}\right)$

$= Z(0.995)$

$= 2.578$

Examples → Part (B)

$n = 20$

- 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

$\text{mean} = \bar{X} = \frac{\text{مجموع القيم}}{\text{عددهم}} = 319.75$

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

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

1. σ is unknow
2. sample size is small
   $20 < 30$
3. normal distribution

$\underbrace{\qquad\qquad}_{t \text{ dis.}}$

$t(1 - \frac{\alpha}{2})$ $(1-\alpha) = 0.95$ $\alpha = 0.05$

$t(0.975)$

df is 19 $t(0.975) = 2.093$

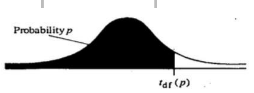
$n-1$
$20-1 = 19$

$$\bar{X} \pm t_{(1-\frac{\alpha}{2})} \frac{S}{\sqrt{n}}$$

$$319.75 \pm (2.093) \frac{8.2}{\sqrt{20}}$$

$L = 315.91$

$U = 323.59$



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	[illegible]	15.89	31.82	63.67	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	[illegible]	4.849	6.965	9.925	14.09	22.32	31.60
3	0.765	0.978	1.250	1.638	2.353	[illegible]	3.482	4.541	5.841	7.453	1.215	12.92
4	0.741	0.941	1.190	1.533	2.132	[illegible]	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	[illegible]	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	[illegible]	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	[illegible]	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	[illegible]	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	[illegible]	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	[illegible]	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	[illegible]	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	[illegible]	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	[illegible]	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	[illegible]	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	[illegible]	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	[illegible]	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	[illegible]	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	[illegible]	2.214	2.552	2.878	3.197	3.610	3.922
19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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	[illegible]	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

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)?

① normal

② $s = 21$

③ $n < 30$

t dist.

$n = 20$

mean $= \bar{x} = 95$

$t_{(1-\frac{\alpha}{2})}$

$t_{(1-\frac{0.01}{2})}$

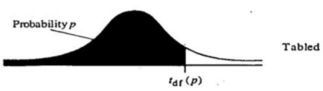
$= t_{(0.995)}$

df is 19

$(0.995) = 2.861$

$$\bar{x} \pm t_{(1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}$$

$$95 \pm (2.861) \frac{21}{\sqrt{20}}$$



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

d.f.	.75	.80	.85	.90	.95	.975	.98	.99	.995	.9975	.999	.9995
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	[illegible]	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	[illegible]	14.09	22.32	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	[illegible]	7.453	1.215	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	[illegible]	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	[illegible]	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	[illegible]	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	[illegible]	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	[illegible]	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	[illegible]	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	[illegible]	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	[illegible]	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	[illegible]	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	[illegible]	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	[illegible]	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	[illegible]	[illegible]	3.610	3.922
19	[illegible]	[illegible]	[illegible]	1.328	[illegible]	[illegible]	[illegible]	[illegible]	2.861	[illegible]	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	[illegible]	[illegible]	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

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

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

or

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

$$df = n - 1$$

هاي لو σ غير موجودة

⊕ $n < 30$

+ normal وكان بالبداية

Probability Interpretation of the Level of Confidence

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

هاي لما σ موجودة

⊕

normal distribution → سواء حكى بنص العبارة أو من عندي $n \ge 30$

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

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

$S \simeq \sigma$

or

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

هاي لو σ غير موجودة

⊕

$n \ge 30$

هون بنقدر نعتبر $(S \simeq \sigma)$