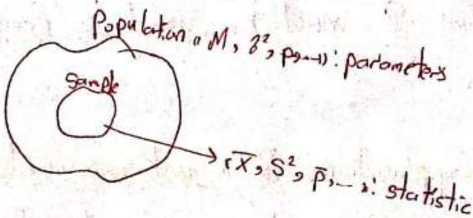
*Ch8: Interval Estimation:-

- Inferential statistics: using samples to predict a characteristics of the population.



- Sample statistic: numerical measures based on samples.
- Population parameter: numerical measures based on population.

- A point estimator is a sample statistic used to estimate a population parameter.

- $\bar{X}$ is a point estimator of the population mean μ.
- S is a point estimator of σ.
- The sample proportion $\bar{p}$ is a point estimator of the population proportion p.

→ Because a point estimator can't be expected to provide the exact value of the population parameter, an interval estimate is computed by adding and subtracting a value, called the margin of error, to the point estimate.

*Sec 8.1: Population mean; σ Known.

To develope an interval estimate of the population mean, either the population standard deviation σ or the sample standard deviation S must be used to compute the margin of error.

→ In most applications σ is unknown and S is used.

→ In some applications, large amounts of relevant historical data are available, and can be used to estimate σ.

- If σ is Known, then the margin of error «E»

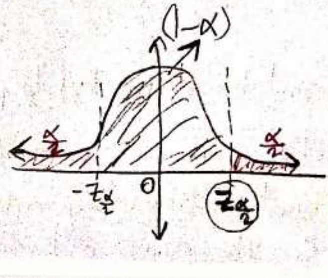
$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$;

σ: population standard deviation.

n: sample size

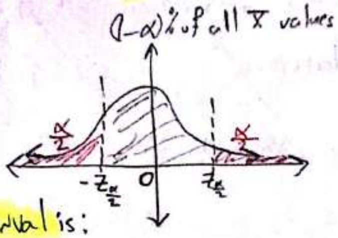
α: the significance level.

$Z_{\frac{\alpha}{2}}$: the z-value providing an area of $\frac{\alpha}{2}$ in the upper tail of the standard normal distribution



- The general form of an interval estimate of the population mean μ is:

$$\bar{x} \mp E = (\bar{x} - E \, , \bar{x} + E) \quad ;$$



$\bar{x}$: the sample mean.

E: the margin of error.

A $(1-\alpha)\%$, Confidence interval is:

$$(\bar{x} - E \, , \bar{x} + E) \quad ; \quad E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} .$$

where $(1-\alpha)$ is the confidence level.

- Ex: If $\alpha = 10\%$ «significance level»

$\longrightarrow$ the confidence level «$1-\alpha$» $= 90\%$

* A $(1-\alpha)\%$ C. I :

we are $(1-\alpha)\%$ confident that the true mean is between $\bar{x} - E$ and $\bar{x} + E$, and there is $\alpha\%$ chance that we are wrong.

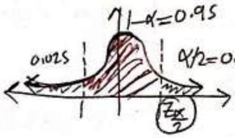
-Ex: To estimate the weekly time spent on Facebook by students, we took a sample of 65 students. Assume $\bar{X} = 34$ and $\sigma = 8.5$. Find a 95% confidence interval for the true mean.

$\bar{X} = 34$, $\sigma = 8.5$ «Known», $n = 65$,

C.L: $95\% = 0.95 = 1 - \alpha$

$\therefore$ Significance level $\alpha = 0.05$

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$



$1 - \frac{\alpha}{2} = 0.975$

	0.06
1.9	0.975

$\therefore Z_{\frac{\alpha}{2}} = 1.96$

$$\rightarrow E = 1.96 \left(\frac{8.5}{\sqrt{65}}\right) = 2.066 \cong 2.07$$

A 95% C.I is :-

$$(34 - 2.07, 34 + 2.07) = (31.93, 36.07)$$

We are 95% confident that the true mean is between 31.93 and 36.07

- Sec 8.2: Population mean, σ Unknown.

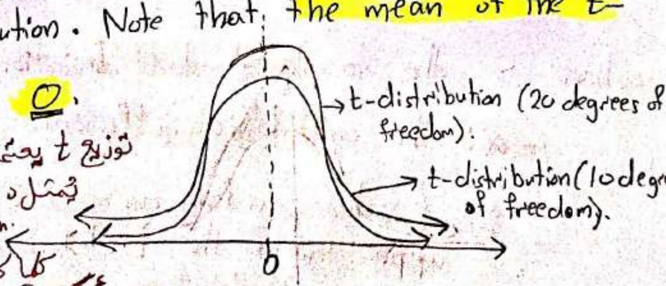
• When developing an interval estimate of a population mean we usually don't have a good estimate of the population standard deviation. In these case, we must use the sample to estimate μ and σ.

عندما تكون σ غير معروفة نستخدم s.

• When s is used to estimate σ, the margin of error and the interval estimate for μ are based on a probability distribution known as t-distribution.

إذا استخدمنا s فإن margin of error يعتمد على توزيع t.

• The t-distribution is a family of similar probability distributions, with a specific t distribution depending on a parameter known as the degrees of freedom. A t-distribution with more degrees of freedom exhibits less variability and more closely resembles the standard normal distribution. Note that the mean of the t-distribution is 0.



توزيع t يعتمد على parameter يمثل درجات الحرية degree of freedom

كلما كانت درجات الحرية أكبر يكون اتساع أقل.

- As the degrees of freedom increase, the t-distribution approaches the standard normal distribution. كلما زادت درجات الحرية يقترب شكل الـ t-dist من التوزيع الطبيعي "جدول Z".

If the degrees of freedom exceed 100, the infinite degrees of freedom row can be used to approximate the t-value, and the standard normal Z-value provides a good approximation to the t-value. That is,

$$t_{\frac{\alpha}{2}} = Z_{\frac{\alpha}{2}} \quad \text{if the degree of freedom} = \infty.$$

أي أن قيمة $t_{\frac{\alpha}{2}} = Z_{\frac{\alpha}{2}}$ إذا كانت درجات الحرية $= \infty$.

- Margin of Error and the interval Estimate "σ unknown"

$$E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

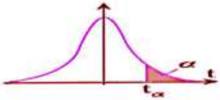
- Interval estimate of a population mean; σ unknown:

$$\bar{X} \mp t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} = \left(\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} , \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right).$$

where S: the sample standard deviation.

$1-\alpha$: the confidence coefficient.

$t_{\frac{\alpha}{2}}$: t value providing an area of $\frac{\alpha}{2}$ in the upper tail of the t distribution with $n-1$ degrees of freedom.



T-Distribution Table

df	α = 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.660
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.96	2.326	2.576	3.09	3.291

- Ex: The following sample data are from a normal population: 10, 8, 12, 15, 13, 11, 6, 5.

a) What is the point estimate of the population mean

$\rightarrow \bar{X} = 10$

b) What is the point estimate of the population standard deviation?

$\rightarrow S = 3.46$

c) What is the 95% confidence interval for the population mean?

σ is unknow.

$\alpha = 5\% = 0.05$, $\bar{X} = 10$, $S = 3.46$, $n = 8$

$\frac{\alpha}{2} = 0.025$ $\qquad$ $df = n-1 = 7$

$\rightarrow E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

$= 2.365\left(\frac{3.46}{\sqrt{8}}\right)$

$= 2.89$

df	$\alpha/2$ 0.025
7	$\boxed{2.365}$

A 95% C.I is: $(\bar{X} - E, \bar{X} + E) = (10 - 2.89, 10 + 2.89)$
$= (7.11, 12.89)$

We are 95% confident that the true mean (μ) is between 7.11 and 12.89

-Ex: A simple random sample with $n=64$, provided a sample mean of 22.5 and a sample standard deviation of 4.4.

$\bar{X}=22.5$, $S=4.4$ « σ is unknown » , $n=64$

a) Develop a 90% confidence interval for the population mean?

$\alpha=10\%=0.1$, $df=63$, $\frac{\alpha}{2}=0.05$

$$E=t_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}=1.669\left(\frac{4.4}{\sqrt{64}}\right)$$

	0.05
63	$\boxed{1.669}$

$\therefore \boxed{E=0.913}$

A 90% CI is: $(22.5-0.913, 22.5+0.913)$

$=(21.587, 23.413)$.

We are 90% confident that the true mean (μ) is between 21.587, 23.413.

b) Develop a 95% confidence interval for μ.

$\alpha=5\%=0.05$, $\frac{\alpha}{2}=0.025$, $df=63$

$$E=t_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}=1.998\left(\frac{4.4}{\sqrt{64}}\right)=1.1$$

	0.025
63	$\boxed{1.998}$

A 95% C.I is: $(22.5-1.1, 22.5+1.1)$

$=(21.4, 23.6)$.

- Note that:

① when confidence level increases, the margin of error increases and the confidence interval becomes longer.

$$C.L \uparrow \to E \uparrow \to C.I \underset{\text{longer}}{\uparrow}$$

② when the sample size increases, the margin of error decreases and the confidence interval becomes shorter.

$$n \uparrow \to E \downarrow \to C.I \underset{\text{shorter}}{\downarrow}$$

③ when the significance level increases, the margin of error decreases and the confidence interval becomes shorter.

$$\alpha \uparrow \to E \downarrow \to C.I \underset{\text{shorter}}{\downarrow}$$

- Summary: ① If $n \geq 30 \uparrow$ and σ is unknown, and the pop. is normal, then we can use Z or t.

لأن كلما زادت درجات الحرية يصبح جدول Z قريباً على t.

② If $n < 30$ and σ is unknown, then $E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$.

أي نستخدم فقط t.

③ If σ is Known, then we use $E = Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ «بغض النظر عن قيمة n»

لماذا الحالت σ معروفة
فإننا نستخدم دائماً جدول Z
ولكن للتسهيل علينا في إيجاد
قيمة Z فإنه بإمكاننا نعوض

$$Z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2}} \text{ at } \infty$$

حجم المجتمع كامل "لكن ما بلزم"

-Ex: 200 students are enrolled in statistic class.

حجم العينة

After the first examination, a random sample of 6 papers was selected. The grades were 61, 75, 94, 76, 70, 80. Provide a 99% confidence interval for the mean grade of all students in the class.

σ unknown "الانحراف المعياري للمجتمع كامل غير معطى"

$\overline{X} = 76$, $S = 10.97$, $\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$

$df = 5$.

$\longrightarrow t_{0.005} = 4.032$

$E = 4.032 \left(\frac{10.97}{\sqrt{6}} \right) = 18.06$

$\therefore$ A 99% C.I is $(76 - 18.06, 7 + 18.06)$

$= (57.94, 94.06)$.

-Ex: to estimate the weekly times spent on facebook by students, we took a sample of 65 students. Assume, $\overline{x} = 34$, $\sigma = 8.5$. Find a 95% confidence interval for the true mean. $\overline{X} = 34$, $\sigma = 8.5$ "Known"

$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$\alpha = 0.05 \longrightarrow \frac{\alpha}{2} = 0.025$ لإيجاد قيمة $Z_{\frac{\alpha}{2}}$ بإمكاننا
استخدام قيمة $t_{\frac{\alpha}{2}}$ عند ∞.

$$Z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2},\infty} = 1.96 \rightarrow E = 1.96\left(\frac{8.5}{\sqrt{65}}\right) = 2.07$$

$\therefore$ A 95% C.I is :-

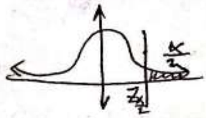
$$(34 - 2.07 \,,\, 34 + 2.07) = (31.93 \,,\, 36.07).$$

- Sec 8.3: Determining the sample size:

In this section, we describe how to choose a sample size large enough to provide a desired margin of error.

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \longrightarrow n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$$

$$; \quad Z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2},\infty}$$



E = margin of error

$\sigma \longrightarrow$ 1) the estimate of the population standard deviation. (from historical data).

2) the sample standard deviation from the preliminary sample.

3) Use Judgment or a «best guess»:

$$\frac{\text{largest value} - \text{smallest value}}{4} = \frac{\text{Range}}{4}.$$

- Ex. Find the minimum sample size needed to acheive 95% confidence interval with margin of error of 10. Assume $\sigma = 40$.

$\sigma = 40, \ \alpha = 5\%, \ E = 10$

$$\rightarrow n = \left(\frac{Z_{\frac{\alpha}{2}}\,\sigma}{E}\right)^2 \quad ; \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = t_{0.025,\infty} = 1.96$$

$\therefore \quad n = \left(\frac{1.96(40)}{10}\right)^2 = 61.47 \longrightarrow \therefore n = 62$ or more

roundup.

Ex: The range for a set of data is estimated to be 36. At 95% confidence, how large a sample would provide a margin of error of 3?

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\, \sigma}{E}\right)^2 \quad ; \quad Z_{\frac{\alpha}{2}} = t_{0.025,\infty} = 1.96$$

$$\sigma \cong Range/4 = 36/4 = 9$$

$$E = 3$$

$\rightarrow n = \left(\frac{1.96(9)}{3}\right)^2 = 34.57 \longrightarrow \therefore n = 35$ or more

*Ch9: Hypothesis Tests: اختبار الفرضيات

In this chapter, we continue the discussion of statistical inference by showing how hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.

← فحص قراءة خاصة بالمجتمع «فرضية» هل هي مقبولة أم لا
هذا من ضمن الإحصاء الاستدلالي.

→ In hypothesis testing, We begin by making an assumption about a population parameter, it is called the null hypothesis, and is denoted by H_0.

→ We then define another hypothesis, called the alternative hypothesis, which is the opposite of the null hypothesis, and is denoted by H_a

→ The hypothesis testing uses data from a sample to test the two statements indicated by H_0 and H_a.

- Sec 9.1: Developing Null and Alternative Hypotheses:

كتابة الفرضيات: ممكن أ- يكون الفحص لل μ او م
لكن سنأخذ فقط ال μ.

We have 3 possible forms of hypotheses H_0 and H_a

① $H_0: \mu \geq \mu_0$
$H_a: \mu < \mu_0$ — one tailed test « lower tailed test »

② $H_0: \mu \leq \mu_0$
$H_a: \mu > \mu_0$ — one tailed test « upper tailed test »

③ $H_0: \mu = \mu_0$
$H_a: \mu \neq \mu_0$ — two tailed test.

- Note that the equality always appears in the null hypothesis H_0.

→ Asking whether the user is looking for evidence to support $\mu < \mu_0$, $\mu > \mu_0$ or $\mu \neq \mu_0$ will help determine H_a.

- Ex: Currently the average weekly crimes in a neighborhood of Chicago city is 5 per week. A community activist from the neighborhood believes that increasing police patrol in that neighborhood will reduce the average weekly crimes. Develop the appropriate hypotheses to test the activist's belief.

$\rightarrow$ $H_0: M \geq 5$ « lower tailed test »
$H_a: M < 5$

- Ex: The average starting salary for MPA graduates is 1000 JOD. Recently, the demand on MPA graduates increased. Set up the hypotheses to test whether the demand increase on MPA graduates have increased the average starting salary for MPA graduates.

$H_0: M \leq 1000$ vs $H_a: M > 1000$ «upper tailed test»

- Ex: The average daily expenditure of BZU's students is 45 NIS. Due to recent inflations, it is believed that this average has changed. Set up the appropriate hypothese.

$H_0: M = 45$ vs $H_a: M \neq 45$

* The hypothesis testing uses data from a sample to test H_0 and H_a.

$\rightarrow$ Based on sample, we either reject H_0 or accept H_0.

we will follow the practice of concluding don't reject H_0, This conclusion is preffered over accept H_0.

- Sec 9.2: Type I and Type II Errors:

The hypothesis testing procedure lead to the acceptance of H_0 when H_0 is true and the rejection of H_0 when H_a is true, the correct conclusions are not always possible. Because hypothesis tests are based on sample information, we must allow for the possibility of errors.

		H_0 True	H_a True
Conclusion	Accept H_0	Correct conclusion	Type II Error
	Reject H_0	Type I Error	Correct conclusion

1) Type I Error: we reject H_0, when it is true.

2) Type II Error: we accept H_0, when it is false.

3) Level of significance: the probability of making a type I error when the null hypothes is true, « α »

- Sec 9.3 : Population Mean, σ Known.

In this section, we show how to conduct a hypothesis test about a population mean for the σ Known case. p.

The methods presented in this section are exact if the sample is selected from a population that is normally distributed.

* A procedure based on sample information to determine whether the hypothesis is reasonable statment; «σ Known»

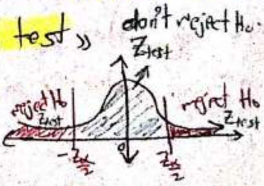
1. State the null hypothesis H_0 and the alternative hypothesis H_1
2. Select a level of significance. $\alpha = P(\text{type I error})$
3. Identify a statistic test. «if σ is Known»

$$Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad ; \quad \begin{aligned} &\bar{X}: \text{sample mean.} \\ &\mu_0: \text{the hypothesized value.} \\ &\sigma: \text{population standard deviation.} \\ &n: \text{sample size.} \end{aligned}$$

4. Formulate a rejection rule , «critical value approach»

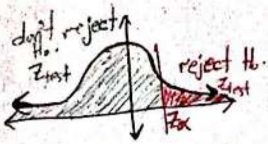
→ $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$ « 2 tailed test »

reject H_0 if $Z_{test} \geq Z_{\frac{\alpha}{2}}$ or $Z_{test} \leq -Z_{\frac{\alpha}{2}}$



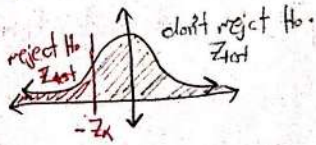
→ $H_0: \mu \leq \mu_0$ vs $H_a: \mu > \mu_0$ «Upper tailed test»

reject H_0 if $Z_{test} \geq Z_\alpha$.



→ $H_0: \mu \geq \mu_0$ vs $H_a: \mu < \mu_0$ «lower tailed test».

reject H_0 if $Z_{test} \leq -Z_\alpha$



OR ⑤ Identify a rejection rule using p-value approach.

* p-value: is a probability that provides a measure of the evidence against the null hypothesis provided by the sample. Smaller p-values indicate more evidence against H_0.

→ p-value $= P(Z > Z_{test})$ «upper tailed test»

→ p-value $= P(Z < Z_{test})$ «lower tailed test»

→ p-value $= 2P(Z > Z_{test})$ «2 tailed test and Z_{test} +ve»

p-value $= 2P(Z < Z_{test})$ «2 tailed test, and Z_{test} −ve»

If p-value $\leq \alpha$, we reject H_0.

If p-value $> \alpha$, we don't reject H_0.

- Ex: Suppose a high school principle claims that the mean SAT score in math at his school is better than 550. A random sample of 72 students finds a mean score of 574. Assume the population standard deviation is 100. Is the principal's claim valid? using $\alpha = 0.05$.

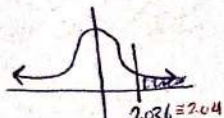
$$n = 72, \quad \bar{X} = 574, \quad \sigma = 100.$$

$\rightarrow H_0: \mu \leq 550 \quad$ vs $\quad H_a: \mu > 550$

$$Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{574 - 550}{100/\sqrt{72}} = 2.036.$$

1) If we want to test the claim using p-value:-

$$\begin{aligned} \text{p-value} &= P(Z > Z_{test}) \\ &= P(Z > 2.036) \\ &= 1 - 0.9793 \\ &= 0.0207 \end{aligned}$$



P from the table:-

$\rightarrow P(Z < \overset{2.04}{2.036}) = 0.9793$

Now:

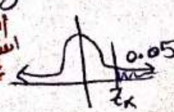
$$\text{p-value} = 0.207 < 0.05 = \alpha$$

$\therefore$ reject H_0.

2) If we want to test the claim using critical value

$$\begin{aligned} \text{critical value} &= Z_\alpha \\ &= t_\alpha \;; df = \infty \\ &= 1.645 \end{aligned}$$

أسهل من استخدام Z



$Z_{test} \overset{??}{\geq} Z_{\alpha}$

$2.036 \geq 1.645$ $\therefore$ reject H_0.

The evidence supports the claim that the mean SAT math score is better than 550.

- Sec 9.4 : Population Mean, σ Unknown.

In this section, we describe how to conduct hypothesis tests about a population mean for the σ unknown case. Because the σ unknown case corresponds to situations in which an estimate of the population standard deviation can't be developed prior to sampling, the sample must be used to develop an estimate of both μ and σ.

- A procedure based on sample information to determine whether the hypothesis is reasonable statment :- (σ unknown)

1. State H_0 and H_a.
2. Select a significance level α. $\alpha = P(\text{type 1 error})$.
3. Identify a statistic test : « σ unknown »

$$t_{test} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

; $\bar{X}$: sample mean.

μ_0 : the hypothesized value.

s : sample standard deviation.

n : sample size.

4 Formulate a rejection rule using critical values:-

→ $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$ «2-tailed test»

critical values: $t_{\frac{\alpha}{2}}$, $-t_{\frac{\alpha}{2}}$ with $df = n-1$.

reject H_0 if $t_{test} \geq t_{\frac{\alpha}{2}}$ or $t_{test} \leq -t_{\frac{\alpha}{2}}$.

→ $H_0: \mu \geq \mu_0$ vs $H_a: \mu < \mu_0$ «lower tailed test»

critical values: $-t_{\alpha}$; $df = n-1$

reject H_0 if $t_{test} \leq -t_{\alpha}$.

→ $H_0: \mu \leq \mu_0$ vs $H_a: \mu > \mu_0$ «Upper tailed test»

critical value: t_{α} ; $df = n-1$

reject H_0 if $t_{test} \geq t_{\alpha}$.

يمكن أيضاً تحديد منطقة الرفض من خلال p-value ولكن لا نريد التطرق لها في هذا section.

* Summary of Hypothesis Tests About a population mean:

1. Develope H_0 and H_a.
2. Specify the level of significance.
3. Collect the sample data ($\bar{x}$, n, s)

Lower tailed test:

$H_0: \mu \geq \mu_0$ vs $H_a: \mu < \mu_0$

→ if σ Known:-

test statistic ($Z_{test} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$)

Rejection region using:-

① critical value(s): $-Z_\alpha = -t_{\alpha,\infty}$

reject H_0 if $Z_{test} \leq -Z_\alpha$

or

② p-value: $p(Z < Z_{test})$

reject H_0 if p-value $\leq \alpha$

2- tailed test

$H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$

→ if σ Known:-

$Z_{test} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Rejection region:-

① critical values: $Z_{\frac{\alpha}{2}}, -Z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2},\infty}$

reject H_0 if $|Z_{test}| \geq Z_{\frac{\alpha}{2}}$

or

② p-value:

if $Z_{test} > 0: 2P(Z > Z_{test})$

if $Z_{test} < 0: 2P(Z < Z_{test})$

reject H_0 if p-value $\leq \alpha$

Upper tailed test

$H_0: \mu \leq \mu_0$ vs $H_a: \mu > \mu_0$

→ if σ Known:-

$Z_{test} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Rejection region:-

① critical value:- Z_α

reject H_0 if $Z_{test} \geq Z_\alpha$, $Z_\alpha = t_{\alpha,\infty}$

② p-value:- $P(Z > Z_{test})$

reject H_0 if p-value $\leq \alpha$

«lower»

→ if σ is unkown:-

test statistic:

$$t_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Rejection region using critical value :

$-t_\alpha$; $df = n-1$

reject H_0 if $t_{test} \leq -t_\alpha$

«2-tailed»

→ if σ is unknown

$$t_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Rejection region using critical value(s):-

$t_{\frac{\alpha}{2}}, -t_{\frac{\alpha}{2}}$; $df = n-1$

reject H_0 if $t_{test} \geq t_{\frac{\alpha}{2}}$ or $t_{test} \leq -t_{\frac{\alpha}{2}}$

«Upper»

→ if σ is unknown

$$t_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Rejection region using critical value(s):-

t_α ; $df = n-1$

reject H_0 if $t_{test} \geq t_\alpha$

— Examples:

1. In the past, the average age of employees of a large corporation has been 40 years. Recently, the company has been hiring older individuals. In order to determine whether there has been an increase in the average age of all employees, a sample of 100 employees was selected. The average age in the sample was 42 years. Assume that the population standard deviation is 12 years. Let $\alpha = 0.05$.

a) State the null and alternative hypotheses.

$$H_0: \mu \leq 40 \quad \text{vs} \quad H_a: \mu > 40$$

Upper tailed test

b) Compute the test statistic.

$$\mu_0 = 40, \quad n = 100, \quad \bar{x} = 42, \quad \sigma = 12 \text{ «Known»}$$

$$\rightarrow Z_{test} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{42 - 40}{12/\sqrt{100}} = 1.67$$

1.6666

c) Find the critical value(s). $\alpha = 0.05$

$\rightarrow$ critical value $= Z_\alpha$

$$Z_{0.05} = t_{0.05} \quad df = \infty$$
$$= 1.645$$

$$Z_{test} \overset{??}{\geq} Z_\alpha$$

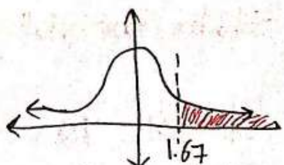
$\rightarrow 1.67 \geq 1.645$ ✓ $\therefore$ reject H_0.

$\therefore$ The mean age of all employees is more than 40 years.

d) Find the p-value. «upper tailed»

$\rightarrow$ p-value $= P(Z \geq Z_{test})$

$$= P(Z > 1.67)$$
$$= 1 - 0.9525$$
$$= 0.0475$$



from the table:- $P(Z < 1.67) = 0.9525$

$$\text{p-value} \overset{??}{\leq} \alpha$$
$$0.0475 \leq 0.05$$

$\rightarrow$ reject H_0.

② The following sample provides the waiting times (in minutes) of 10 random customers at a certain bank: 35, 40, 44, 60, 32, 17, 50, 40, 35, 36.

Assume the times come from a normal distribution. Let μ denote the population mean of the waiting time. The bank claims that a customer would wait at most 30 minutes. We would like to test this claim. using 5% level of significance.

$H_0: \mu \leq 30$ vs $H_a: \mu > 30$

upper tailed test « δ unknown »

① $\bar{X} = 38.9$ « we find it using SD mode:-

$\boxed{\text{mode}}$ $\boxed{2}$ 35 $\boxed{M+}$ 40 $\boxed{M+}$

---- 36 $\boxed{M+}$

→ $\boxed{\text{shift}}$ $\boxed{2}$ $\boxed{1}$ $\boxed{=}$

② $S = 11.37$ → $\boxed{\text{shift}}$ $\boxed{2}$ $\boxed{3}$ $\boxed{=}$

$n = 10$, $\alpha = 0.05$

$$t_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{38.9 - 30}{11.37/\sqrt{10}} = 2.475$$

critical value: t_α ; $df = n-1 = 9$

$\therefore t_\alpha = t_{0.05} = 1.833$

$t_{test} \overset{?}{\geqslant} t_\alpha$

$2.475 \geqslant 1.833$ ✓

$\therefore$ reject H_0 using $\alpha = 0.05$

↳ reject the bank claims.

3 The high school athletic director is asked if football players are doing as well academically as the other student athletes. We know from a a previous study that the average GPA for the student athletes is 3.10. After an initiative to help improve the GPA of student athletes, the athletic director randomly samples 20 football players and finds that the average GPA of the sample is 3.18 with a sample standard deviation of 0.54. Is there a significance evidence that the average GPA for the student change from 3.10? (Use .5% significance level).

$n = 20$, $\bar{X} = 3.18$, $S = 0.54$, σ unknown.

$H_0: \mu = 3.10$ vs $\mu \neq 3.10$ «2-tailed test»

$$\rightarrow t_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{3.18 - 3.10}{0.54/\sqrt{20}} = 0.66$$

critical values: $t_{\frac{\alpha}{2}}$, $-t_{\frac{\alpha}{2}}$ $df = n-1 = 19$

	$\alpha/2 = 0.025$
19	$\boxed{2.093}$

$\therefore \quad t_{\frac{\alpha}{2}, df} = t_{0.025, 19} = 2.093$

$-t_{\frac{\alpha}{2}, df} = -t_{0.025, 19} = -2.093$

$\longrightarrow \quad t_{test} \overset{??}{\geq} t_{\frac{\alpha}{2}}$

$0.66 \overset{??}{\geq} 2.093 \quad \times$

$\therefore$ don't reject H_0