Confidence Interval Estimate Mean (σ Unknown and n < 30)

Confidence Interval Mean (σ Unknown)

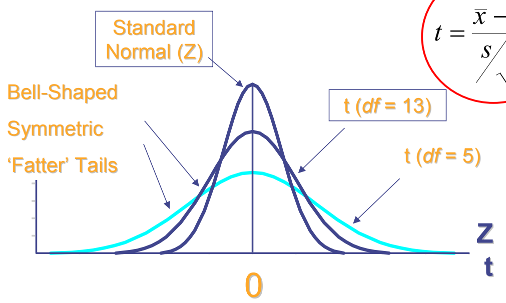
1. Assumptions
   - Population Standard Deviation Is Unknown
   - Population Must Be **Normally Distributed if sample size is relatively small.**
2. Use Student's t Distribution
3. Confidence Interval Estimate

$$\left(\bar{X} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}},\ \bar{X} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \right)$$

$$\boxed{\bar{X} \pm t_{\alpha/2,\,n-1} \cdot \frac{s}{\sqrt{n}}}$$

Student's t Distribution

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



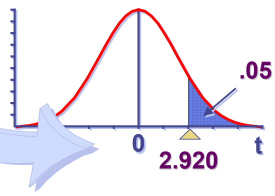
Student's t Table

Area in upper tail

df	0.10	0.05	0.025
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182

t values

For a 90% C.I. and n = 3:
$df = n - 1 = 2$
α = .10
$\alpha/2$ =.05



Jennifer visited Las Vegas, and was still thinking about statistics. She took a random sample of 5 visitors who played game during the day she took the sample and asked them how much money they won that day. The data is the following, in dollars. Negative number represents loss.

62.00, -116.00, -90.00, -220.00, -165.50

Assume that the amount of money won by visitor at Las Vegas from gambling is **normally distributed**. Find a **90% confidence interval** to estimate the average amount of money visitors win per day for visitors at Las Vegas.

$$\boxed{\bar{X} \pm t_{\alpha/2,\,n-1} \cdot \frac{s}{\sqrt{n}}}$$

Sample mean = –105.90

Sample standard deviation = 106.20

$t_{.05}$ = 2.132 (or Confidence interval 90%)

The 90% confidence interval is

$$-105.90 \pm 2.132 \times \frac{106.2}{\sqrt{5}}$$

$-105.90 \pm 101.26 \quad \Leftrightarrow \quad (-207.16, -4.64)$

SPSS

One-sample t-Test

(*with Unknown Variance* σ^2)

In practice, population variance is unknown most of the time. The sample standard deviation **s^2 is used instead for σ^2**. If the random sample of size n is from a normal distributed population and if the null hypothesis is true, the test statistic (standardized sample mean) will have a t-distribution with degrees of freedom $n-1$.

$$\textbf{Test Statistic:}\quad t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$$

7

I. State Hypothesis

One-side test example:

If one wish to test whether the body temperature is less than 98.6 or not.

H_0: $\mu = 98.6$ v.s. H_a: $\mu < 98.6$

(Left-sided Test)

8

II. Test Statistic

If we have a random sample of size 16 from a normal population that has a mean of 98.32°F, and a sample standard deviation 0.4. The test statistic will be a *t*-test statistic and the value will be: (standardized score of sample mean)

$$\textbf{Test Statistic:}\quad t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{98.32-98.6}{\frac{0.4}{\sqrt{16}}} = \frac{-0.28}{0.1} = -2.8$$

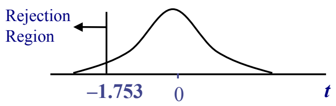
Under null hypothesis, this *t*-statistic has a *t*-distribution with degrees of freedom n – 1, that is, 15 = 16 – 1.

9

III. Decision Rule

Critical Value Approach:

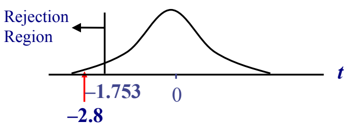
To test the hypothesis at α level 0.05, the **critical value** is $-t_\alpha = -t_{0.05} = -1.753$.



Descion Rule: Reject null hypothesis if $t < -1.753$

10

IV. Conclusion



Decision Rule:

If $t < -1.753$, we reject the null hypothesis.

Conclusion: Since $t = -2.8 < -1.753$, we reject the null hypothesis. There is sufficient evidence to support the research hypothesis that the average body temperature is less than 98.6°F.

11

P-value Approach

III. Decision Rule

Decision Rule: Reject null hypothesis if *p*-value < α.

12

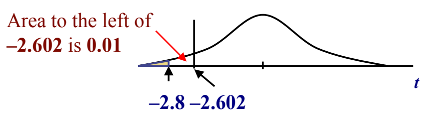
p-value Calculation

***p*-value corresponding the test statistic:**

For t test, unless computer program is used, *p*-value can only be approximated with a range because of the limitation of *t*-table.

***p*-value** = $P(T<-2.8)$ **<** $P(T<-2.602)$ = **0.01**

Since the area to the left of –2.602 is .01, the area to the left of –2.8 is definitely less than 0.01.



13

IV. Conclusion

Decision Rule:

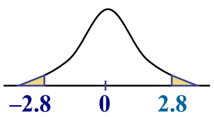
If *p*-value < 0.05, we reject the null hypothesis.

Conclusion: Since *p*-value < 0.01 < 0.05, we reject the null hypothesis. There is sufficient evidence to support the research hypothesis that the average body temperature is less than 98.6°F.

14

What if we wish to test whether the average body temperature is different from 98.6°F or not using *t*-test with the same data?

The *p*-value is equal to **twice** the *p*-value of the left-sided test which will be less than .02.



15

Decision Rule

***p*-value approach**: Compute *p*-value,

if $H_a: \mu \neq \mu_0$, p-value = $2 \cdot P(T \geq |t|)$
if $H_a: \mu > \mu_0$, p-value = $P(T \geq t)$
if $H_a: \mu < \mu_0$, p-value = $P(T \leq t)$

reject H_0 if p-value < α

16

Decision Rule

Critical value approach: Determine critical value(s) using α, **reject H_0** against

i) $H_a: \mu \neq \mu_0$, if $t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$
(or $|t| > t_{\alpha/2}$)
ii) $H_a: \mu > \mu_0$, if $t > t_\alpha$
iii) $H_a: \mu < \mu_0$, if $t < -t_\alpha$

17

Remarks

- If the sample size is **relatively large** (>30) both z and t tests can be used for testing hypothesis. The number 30 is just a reference for general situations and for practicing problems. *In fact, if the sample is from a very skewed distribution, we need to increase the sample size or use nonparametric alternatives such Sign Test or Signed-Rank Test.*
- Many commercial packages only provide *t*-test since standard deviation of the population is often unknown.

18

Example

A random sample of **ten** 400-gram soil specimens were sampled in location A and analyzed for certain contaminant. The sample data are the followings:

65, 54, 66, 70, 72, 68, 64, 50, 81, 49

The contaminant levels are normally distributed. Test the hypothesis, at the level of significance **0.05**, that the true mean contaminant level in this location **exceeds 50 mg/kg**.

19

Step 1

What is the hypothesis to be tested?

Ho: $\mu = 50$

H_a: $\mu > 50$

20

Step 2

Which test can be used for testing the hypothesis above? (Check assumptions.)

One sample t-test. Why? Because the random sample was from a **normal population** and unknown variance.

Compute Test Statistic:

Test statistic: $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \dfrac{63.9 - 50}{\frac{10.17}{\sqrt{10}}} = 4.32$

The value of the test statistic is 4.32 with a p-value between .005 and .0005 from table. P-value from SPSS is .00096.

21

Step 3

Decision Rule:

Specify a level of significance, α, for the test. α = .05

Critical value approach:

Reject Ho if $t > t_{.05} = 1.833$

p-value approach:

Reject Ho if p-value < 0.05

22

Step 4

Conclusion:

Since t=4.32 > 1.833,

(or p-value = .00096 < 0.05)

we reject the null hypothesis. The data provide sufficient evidence to support the alternative hypothesis that the average contaminant level in this location exceeds 50 mg/kg.

23

Statistical Significance

A statistical report shows that the average blood pressure for women in certain population is significantly *different from* a recommended level, with a *p*-value of 0.002 and the *t*-statistic of – 6.2. It generally means that the difference between the actual average and the recommended level is *statistically* significant. And, it is a two-sided test.

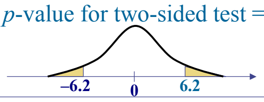
(Practical Significance?)

24

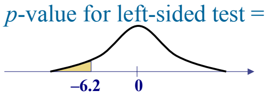
Statistical Report

$t = -6.2$

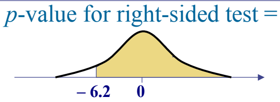
p-value for two-sided test = **.002**



p-value for left-sided test = **.001**



p-value for right-sided test = **.999**

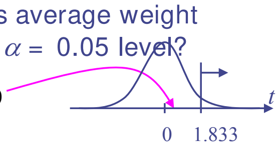


25

Average Weight for Female Ten Years Old Children In US

Info. from a random sample: $n = 10$, $\bar{x} = 80$ lb, $s = 18.05$ lb. Is average weight greater than 78 lb at $\alpha = 0.05$ level?

Test Statistic: $t = \dfrac{80 - 78}{\dfrac{18.05}{\sqrt{10}}} = 0.350$



$t_\alpha = t_{.05}$, d.f. = 10 – 1 = 9, $t_{0.5,\,9} = 1.833$

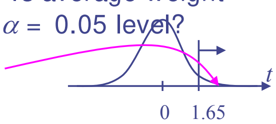
Reject H_0, if $t = 0.35 < 1.833$. **Failed to reject H_0!**

26

Average Weight for Female Ten Years Old Children In US

Info. from a random sample: $n = 400$, $\bar{x} = 80$ lb, $s = 18.05$ lb. Is average weight greater than 78 lb at $\alpha = 0.05$ level?

Test Statistic: $t = \dfrac{80 - 78}{\dfrac{18.05}{\sqrt{400}}} = 2.22$



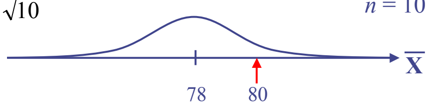
$t_\alpha = t_{.05}$, d.f. = 400 – 1 = 399, $t_{0.5,\,399} = 1.65$

Reject H_0, if $t = 2.22 > 1.65$. **Reject H_0!**

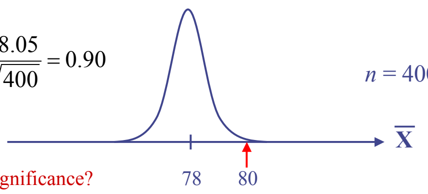
27

Sampling Distribution

S.E. $= \dfrac{18.05}{\sqrt{10}} = 5.71$



S.E. $= \dfrac{18.05}{\sqrt{400}} = 0.90$



Practical Significance?

28