MATH 4720 / MSSC 5720

Instructor: Mehdi Maadooliat

Chapter 5 (Part C)



Department of Mathematical and Statistical Sciences

CHAPTER 5 (PART C)



- Confidence Interval (CI)
- CI for μ , when σ is known
- Choosing Sample Size for Estimating μ
- A Statistical Test for μ
- Type I and Type II Errors
- Power of a Test
- Choosing Sample Size for Testing μ
- Level of Significance
- P-value
- Inference about μ , when σ is unknown



LEVEL OF SIGNIFICANCE

- The approach to hypothesis testing we discussed so far is called classical approach. Note that it is based on a decision rule for a given α .
- A question arises: What is an appropriate α ?
- Many statisticians object to this approach. They say that an appropriate approach should be that we should indicate how much evidence the data provide in favor of ${\cal H}_0$.
- What to do?
- p value
 - = P(oberving the evidence more extreme than the evidence by the data if H_0 is true)

P-VALUE APPROACH



• Example:

An experiment was done on 35 petri dishes which were first cultured with E.Coli bacteria, and then a solution of the antibacterial soap was added. After 24-hour incubation period, bacterial counts were recorded on each of 35 dishes.

• Sample Information:

$$\bar{y} = 31.2$$
, $s = 8.4$

- $H_0: \mu = 33$
- H_a : μ < 33
- **T.S.** $z = \frac{\bar{y} \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{31.2 33}{8.4/\sqrt{35}} = -1.27$
- Applet: P-value

P-VALUE APPROACH CONT'D



• H_0 : $\mu = 33$ vs H_a : $\mu < 33$.

- **T.S.** z = -1.27
- p value = P(oberving the evidence more extremethan the evidence by the data $| H_0 |$ is true) $= P(Z \le -1.27) = 0.1020$
- Since this evidence for H_0 is not small enough, we do note have sufficient evidence to reject H_0 .
- Suppose $\bar{y} = 30.2$ instead of 31.2

$$z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{30.2 - 33}{8.4/\sqrt{35}} = -1.972$$

- $p value = P(Z \le -1.972) = 0.0244$
- Since this evidence for H_0 is very small, we should reject H_0 .

ANOTHER INTERPRETATION OF P-VALUE:



- *p* − *value*
- = smallest level of α at which H_0 can be rejected.
- In the previous example:
- When $\bar{y}=30.2$, p-value=0.0244. This is the smallest level at which H_0 can be rejected.
- If $\alpha=0.05$, $p-value<\alpha$, so Reject H_0
- If $\alpha=0.01$, $p-value>\alpha$, so Fail to Reject H_0
- In other word:
 - You Reject at any $\alpha > 0.0244$
 - You Fail to Reject at any $\alpha < 0.0244$

GENERAL APPROACH



• If p-value $< \alpha$, Reject H_0 in favor of H_a .

Computing p-value

- $H_0: \mu = \mu_0$
- Compute $z = \frac{\bar{y} \mu_0}{\frac{\sigma}{\sqrt{n}}}$
- H_a : $\mu > \mu_0$ p-value = P(Z > computed z)
- H_a : $\mu < \mu_0$ p-value = P(Z < computed z)
- H_a : $\mu \neq \mu_0$ p-value = 2 * P(Z > |computed z|)

Applet: P-value

THREE EXAMPLES



• Let $\alpha = 0.05$ in the following examples.

1.
$$H_0$$
: $\mu = 5.0$ vs. H_a : $\mu > 5.0$ $n = 16$, $\bar{y} = 6.5$, $\sigma = 2.0$

Given that sample is drawn from normal population.

$$- z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{6.5 - 5.0}{2.0 / \sqrt{16}} = 3.0$$

- p-value =
$$P(Z > 3.0)$$

= $normcdf(3.0, \infty, 0.1)$
= 0.0013

- Conclusion: Since this p-value $< \alpha = 0.05$, we reject H_0 in favor of H_a .

THREE EXAMPLES CONT'D



2.
$$H_0$$
: $\mu = 150$ vs. H_a : $\mu < 150$
 $n = 50$, $\bar{y} = 147.4$, $\sigma = 10$

$$-z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{147.4 - 150}{10/\sqrt{50}} = -1.84$$

- p-value =
$$P(Z < -1.84)$$

= $normcdf(-\infty, -1.84, 0, 1)$
= 0.0329

- Conclusion: Since this p-value $< \alpha = 0.05$, we reject H_0 in favor of H_a .

THREE EXAMPLES CONT'D



3.
$$H_0$$
: $\mu = 35$ vs. H_a : $\mu \neq 35$ $n = 25$, $\bar{y} = 40$, $\sigma = 15$

Given that sample is drawn from normal population.

$$-z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{40 - 35}{15/\sqrt{25}} = 1.67$$

- p-value =
$$2 * P(Z > 1.67)$$

= $2 * normcdf(1.67, \infty, 0, 1)$
= 0.095

- Conclusion: Since p-value < 0.05. We don't have sufficient evidence to reject H_0 in favor of H_a .

INFERENCE ABOUT μ WHEN σ IS UNKNOWN



Hypothesis Testing

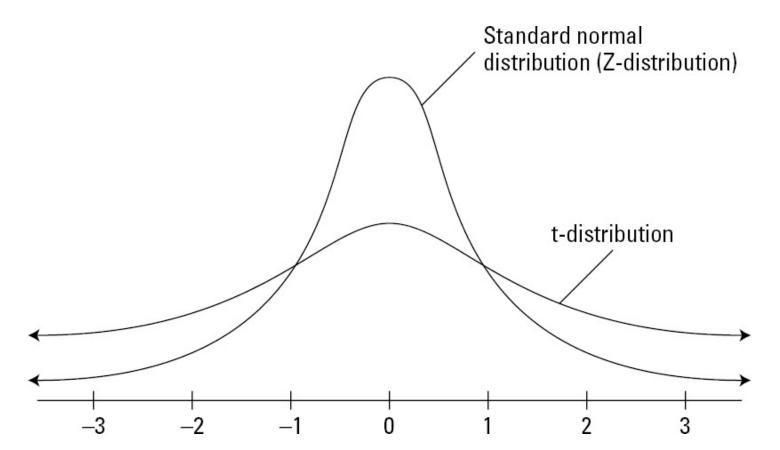
- Sample: $\{y_1, y_2, ..., y_n\}$
- \bar{y} sample mean, s sample st.dev.
- **T.S.** $\hat{z} = \frac{\bar{y} \mu_0}{\frac{s}{\sqrt{n}}}$ (σ replaced by sample st.dev.)
- The distribution of \hat{z} is no longer N(0, 1).

STUDENT T-DISTRIBUTION



•
$$t = \frac{\overline{y} - \mu}{\frac{s}{\sqrt{n}}}$$

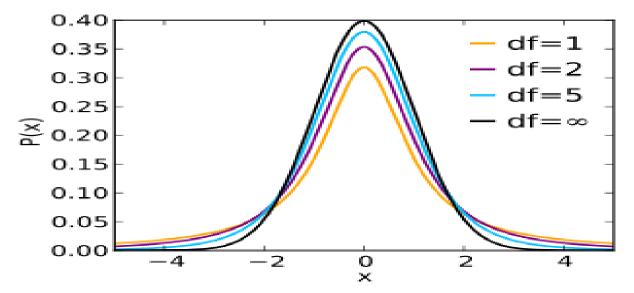
• The distribution of t is called student-t distribution.



REMARKS ON T-DISTRIBUTION



- 1. t-dist. has similar shape as N(0, 1), but is flatter than N(0, 1).
- 2. The t-distribution is symmetric around 0 as N(0,1) is.
- 3. It has a range from $-\infty$ to ∞ as the range of N(0,1).
- 4. Unlike N(0, 1), the t-distribution depends on the degrees of freedom df.
- **5.** As the $df \rightarrow \infty$, t-distribution approaches to N(0, 1).



• Applet: t-distribution vs N(0,1)

INFERENCE ABOUT μ WHEN σ IS UNKNOWN



Hypothesis Testing

$$\begin{array}{cccc} \bullet & H_0\colon \mu=\mu_0 & \text{vs} & H_a\colon \mu>\mu_0 \\ & \text{or} & \mu<\mu_0 \\ & \text{or} & \mu\neq\mu_0 \end{array}$$

• Decision Rule df = n - 1

- H_a : $\mu > \mu_0$: Reject H_0 in favor of H_a if $t > t_\alpha$
- H_a : $\mu < \mu_0$: Reject H_0 in favor of H_a if $t < -t_\alpha$
- H_a : $\mu \neq \mu_0$: Reject H_0 in favor of H_a if $|t| > t_{\alpha/2}$

• Here t_A is a notation for value of t so that the area to the right is A.

- t-table ("D2L > Useful Links > Z, T and Chi^2 Tables")
 - $P(t \ge t_{\alpha})$, where t is a t-distribution with df = n 1.
- t calculator (df)
- Calculator



EXAMPLE

 Consider a population of hypertension group whose average systolic blood pressure (SBP) is 150. You want to determine whether a new treatment is effective in reducing SBP. A clinical trial was conducted on 25 patients of this population. After 6 months of treatment, SBP was recorded on each subject.

•
$$\bar{y} = 147.2, \quad s = 5.5$$

- Is there a sufficient evidence at $\alpha=0.05$ that the new treatment is effective?
- Assume that the distribution of SBP is normal.

EXAMPLE CONT'D



- H_0 : $\mu = 150$ vs. H_a : $\mu < 150$
- **TS:** $t = \frac{\bar{y} \mu_0}{\frac{S}{\sqrt{n}}} = \frac{147.2 150}{\frac{5.5}{\sqrt{25}}} = -2.55$
- **Decision Rule**: Reject H_0 in favor of H_a if $t < -t_\alpha$
- df = n 1 = 25 1 = 24, $\alpha = 0.05$
- $t_{\alpha} = \text{invT}(0.95, 24) = 1.711$
- Reject H_0 in favor of H_a if t < -1.711.
- Conclusion: Is t < -1.711? Yes, since t = -2.55. Thus we reject H_0 , and we have sufficient evidence to conclude that the new treatment is effective.

P-VALUE APPROACH



•
$$H_0$$
: $\mu = 150$ vs. H_a : $\mu < 150$, **TS**: $t = -2.55$

• p-value =
$$P(t < -2.55)$$

= $tcdf(-\infty, -2.55, 24) = 0.0088$

• Since p-value $< \alpha = 0.05$, we reject H_0 in favor of H_a . We have sufficient evidence to conclude that the new treatment is effective.

• P—value Formula:

- H_a : $\mu > \mu_0$, p-value = P(t > computed t)
- H_a : $\mu < \mu_0$, p-value = P(t < computed t)
- H_a : $\mu \neq \mu_0$: p-value = 2 * P(t > |computed t|)

ESTIMATION OF μ USING A CONFIDENCE INTERVAL



• $100(1-\alpha)\%$ Confidence Interval of μ is

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

- Assumption: Either
 - $n \ge 30$
 - the sample is drawn from a normal population.



BOOK EXERCISE 5.39

• A discount tire manufacturer claim that its tires can be driven at least 35,000 miles on the average. A consumer testing agency suspects that this claim is false. A study was conducted on 15 cars in which the testing tires were mounted.

$$n = 15, \ \bar{y} = 31.47, \ s = 5.04$$
 (in thousand miles)

- (a) Estimate the mean miles driven by the tires using 99% confidence interval.
- (b) Is there a sufficient evidence ($\alpha = 0.01$) that the manufacturer's claim is false.

FIRST LET'S DO (B)



•
$$n = 15$$
, $\bar{y} = 31.47$, $s = 5.04$

• **(b)**
$$H_0$$
: $\mu = 35$ vs. H_a : $\mu < 35$

• **TS:**
$$t = \frac{\bar{y} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{31.47 - 35}{\frac{5.04}{\sqrt{15}}} = -2.71$$

- **Decision Rule:** $\alpha = 0.01$. df = 15 1 = 14, $t_{\alpha} = 2.624$
- Reject H_0 in favor of H_a if $t < -t_\alpha = -2.624$
- Conclusion: Is t < -2.624? Yes. Reject H_0 in favor of H_a .
- We have sufficient evidence to conclude that the miles driven by the tires is less than 35,000 miles on the average.

NOW LET'S DO (A)



•
$$n = 15$$
, $\bar{y} = 31.47$, $s = 5.04$

(a) 99% confidence interval of μ .

- Formula: $\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
- Assumption: sample is drawn from a normal population

•
$$\alpha = 0.01$$
, $\alpha/2 = 0.005$, df = 14, $t_{\alpha/2} = 2.977$

• 99% Cl of
$$\mu$$
: $31.47 \pm 2.977 * \frac{5.04}{\sqrt{15}}$ 31.47 ± 3.90

$$27.59 < \mu < 35.3$$

SUMMARY



Case 1: One Variable

	Numerical Variable, σ known	Numerical Variable, σ unknown
Parameter of Interest:	Mean, μ	Mean, μ
Confidence Interval Formula:	$\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\overline{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ with df = $n - 1$
Name of Hypothesis Test, H ₀	One Sample Z Test, $H_0: \mu = \mu_0$	One Sample T Test, $H_0: \mu = \mu_0$
Test Statistic Formula:	$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} \text{with df} = n - 1$
p-value:	$H_a: \mu \neq \mu_0$, p -value = $2P(Z \ge z)$ $H_a: \mu > \mu_0$, p -value = $P(Z \ge z)$ $H_a: \mu < \mu_0$, p -value = $P(Z \le z)$	$H_a: \mu \neq \mu_0$, p -value = $2P(T \geq t)$ $H_a: \mu > \mu_0$, p -value = $P(T \geq t)$ $H_a: \mu < \mu_0$, p -value = $P(T \leq t)$