Lecture 08: Hypothesis Testing

Applied Statistics - STAN - 5.37 & 5.38

12 & 14 December 2020

Lecturer: Erika Siregar, SST, MS

General Idea

- Populasi → sample → estimasi parameter → ukur seberapa baik estimasi yang dilakukan
  - Confidence interval → apakah estimator berada di dalam selang kepercayaan?

    $0.677 < p < 0.723$

    $164.49 < \mu < 180.61$

  - Hypothesis testing → membuat 2 asumsi terkait estimator: null dan alternatif. Uji hipotesis mana yang lebih meyakinkan untuk diterima.

Untuk CI Perlu:

- Tetapkan **significance level** a.k.a **peluang error yang bersedia kita tolerir** a.k.a **α**
- Berdasarkan **α,** tentukan confidence level dari estimator yang dihasilkan → 1-**α**
- Tentukan error yang dapat terjadi.
- Bangun CI-nya.

Untuk hypothesis testing, perlu:

- Menghitung test statistics
- Hitung probability terjadinya (p-value)
- Bandingkan p-value dengan **α**

Today's Plan

Last week: estimate a parameter and its CI.

Now: testing our claim about the estimator

1. Construct a hypothesis: H0 & H1
2. Testing a Claim about a Proportion
3. Testing a Claim About a Mean: σ Known
4. Testing a Claim About a Mean: σ Not Known
5. Testing a Claim About a Standard Deviation or Variance

Hypothesis and Hypothesis Testing

- a claim or statement about a population parameter.
- Example of a hypothesis:
  - $\mu < 1.000 \rightarrow$ the mean of daily covid19 cases in Indonesia is less than 1.000
  - $\sigma \neq 10 \rightarrow$ the standard deviation of midterm scores is not equal to 10.
  - $p > 0.8 \rightarrow$ the proportion of people who pay taxes on time is greater than 0.8.
- A **hypothesis test** (or **test of significance**): a procedure for testing a claim about a population parameter.

How The Hypothesis Testing Work

- Define our claim and state it as the Alternative Hypothesis (H1)
  - a statement that a certain population parameter **<u>is unequal</u>** to some claimed value.
  - Use inequality symbol: $\neq$, <, >
- Define another hypothesis which is the null/default condition → null hypothesis (H0)
  - Use equality symbol: =
- Example:
  - H0: , H1:
  - H0: , H1:
- Focus on testing the H0, the result could be:
  - reject H0 or fail to reject H0.

Practicing on constructing H0 & H1

1. Sebuah maskapai mengklaim bahwa rata-rata berat badan penumpang pesawat adalah kurang dari 70 kg.
   H0:
   H1:

2. Sebuah survey mengklaim bahwa suhu tubuh orang-orang yang tertular covid19 memiliki standar deviasi lebih dari 0.85℃.
   H0:
   H1:

3. Sebuah klaim menyatakan bahwa 15% mahasiswa STAN tidak memiliki social media.
   H0:
   H1:

Steps

1. Read the case carefully and highlight all the keywords → n, α, μ, σ, etc.
2. Define the claimed hypothesis → alternative hypothesis (H1) → contains inequality symbol: ‘<’, ‘>’, ≠
3. Define null hypothesis (H0) → the opposite of H1 → contains equality ‘=’. Assume that it’s True
4. Identify the type of hypothesis (right-tailed, left-tailed, two-tailed).
5. Chose a significance level (α) based on the level of consequences.
   a. Common values: 1%, 5%, 10%
   b. α is
      i. peluang error → seberapa besar yang bersedia kita tolerir → mempengaruhi critical region.
      ii. From α, we can obtain critical value.
6. Compute the test statistics → check comparison of z, t, & χ^2
   a. Proportion

$$\boxed{z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}}$$

   b. variance

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

   c. Mean

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} \qquad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Use z if σ is known
Use t if σ is unknown

Steps (Cont.)

8. Make a decision/conclusion → 3 methods:
   a. p-value
      i. Probability of the test statistics.
      ii. p-value vs α
      iii. Reject H0 if p-value ≤ α
   b. Traditional (critical value)
      i. Diketahui α → cari critical value-nya.
      ii. Reject H0 if test statistics falls within critical region
   c. Confidence interval
      i. Buat CI berdasarkan kasus → refer to lecture 07.
      ii. reject H0 if the claimed value does not include in the CI.
9. Make interpretation:
   Enough (or not enough) evidence to support the claim that ...

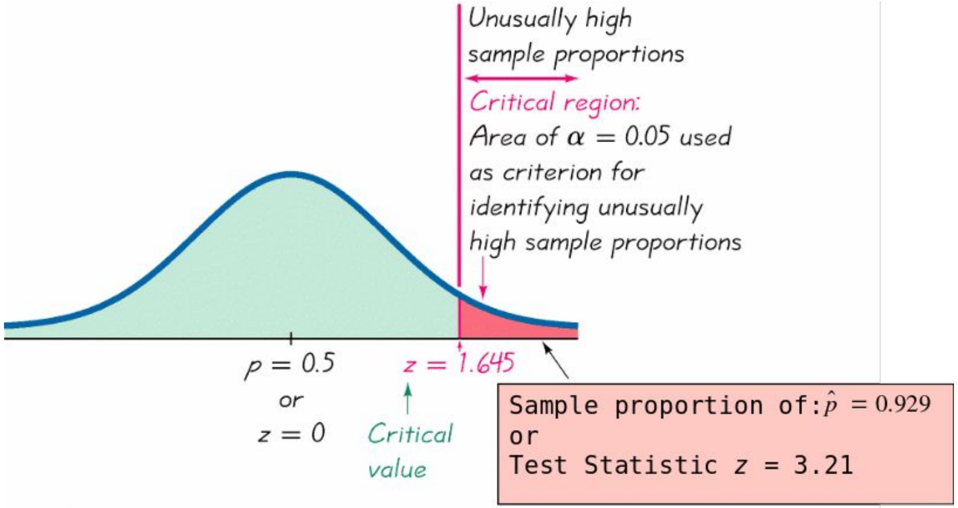
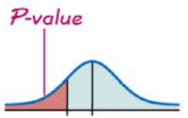
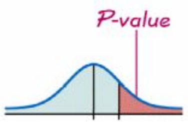
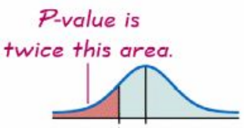
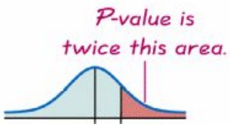
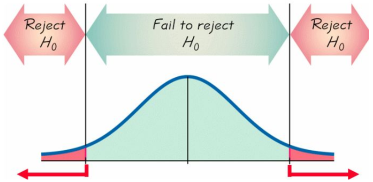


Table 8-2 Confidence Level for Confidence Interval

		Two-Tailed Test	One-Tailed Test
Significance Level for Hypothesis Test	0.01	99%	98%
	0.05	95%	90%
	0.10	90%	80%

Type of Hypothesis Test, Critical Region & p-value

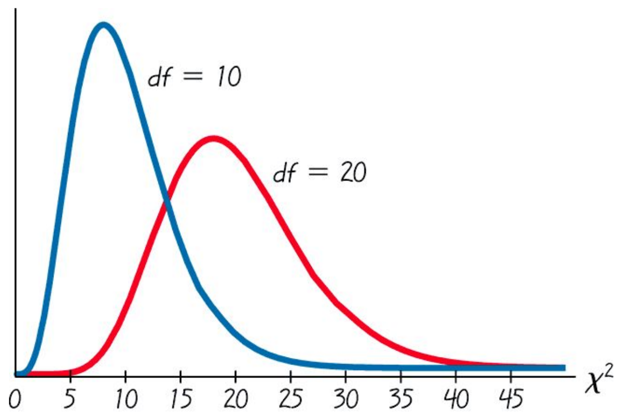
- p-value or probability value is the probability of H0 to be true. → obtain from test statistics
- If p-value is very low → < chosen significance level → reject H0.

No	Hypothesis Test Type & Critical Region	p-value
1	left-tailed (H0: =, H1: **<**)	Critical region lies to the left of Critical value 
2	right tail (H1 **>**)	Critical region lies to the right of Critical value 
3	Critical region in two tails (H1 ≠)	Half of the critical region on the left and the other half on the right.   

Back to steps

z vs t vs $\mathscr{X}^2$

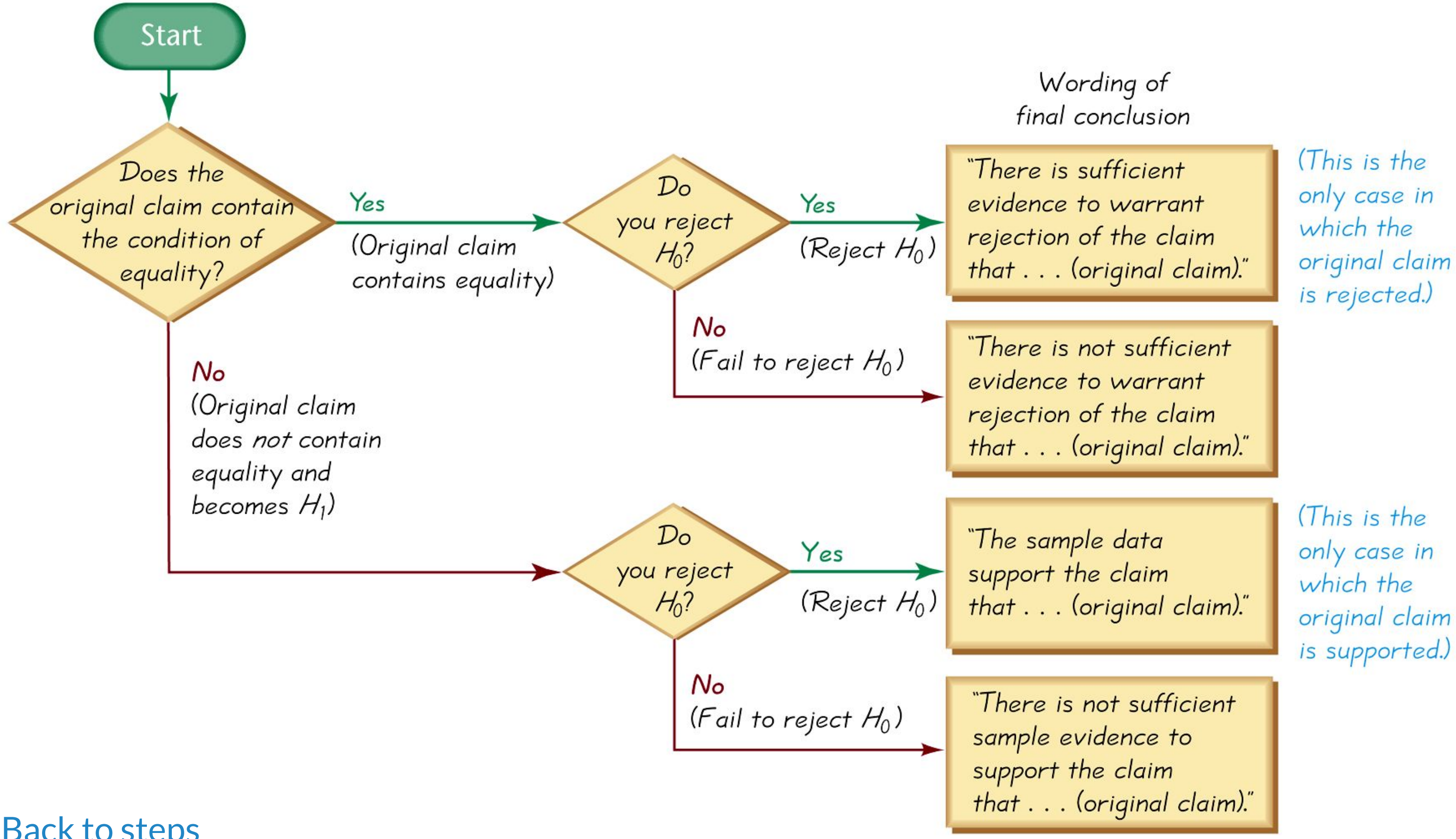
No	Criteria	z	t	$\mathscr{X}^2$
1	curve	Symmetric, Bell shaped	Symmetric, Bell shaped	asymmetric
2	Negative value	yes	yes	no
2	μ	0	0	varies with sample size n
3.	σ	1	> 1, varies with sample size *n*	varies with sample size n
4	df	no	yes	yes
5	Table reading	Area to the left of z	Area to the left of t	Area to the right of $\mathscr{X}^2$
5	R syntax	pnorm(z), qnorm(p)	pt(), qt()	pchisq(), qchisq()



Chi-square has different distribution for different df

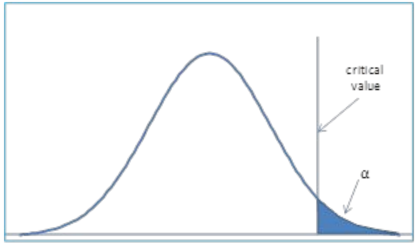
Back to steps

Decision and Interpretation



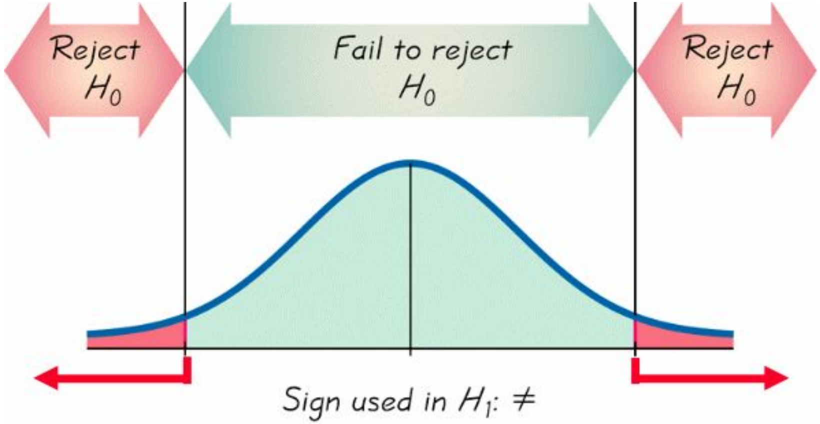
[Back to steps]

Case 1

No	Claim	Hypothesis	Critical Value	P-value & decision
1.	Sebuah laboratorium mengklaim bahwa metode reproduksi yang mereka kembangkan mampu **meningkatkan** peluang memiliki bayi perempuan. 	H0: $p = 0.5$ H1: $p > 0.5$	The sample proportion of 13 girls in 14 births results in p = 13/14 = 0.929. Using p = 0.5 and n = 14, we find the value of the test statistic as follows: $z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.929 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{14}}} = 3.21$	$P(z < 3.21) =$ 0.9993363. Thus, p-value = area to the right of z = 1 - 0.9993 = **0.0006636749**
2	Sebuah laboratorium mengklaim bahwa metode reproduksi yang mereka kembangkan mampu membuat perbedaan dalam peluang memiliki bayi perempuan.	H0: $p = 0.5$ H1: $p \neq 0.5$	3.21	P-value = 2* area to the right of z = 2***0.0006636749 = 0.00132735**

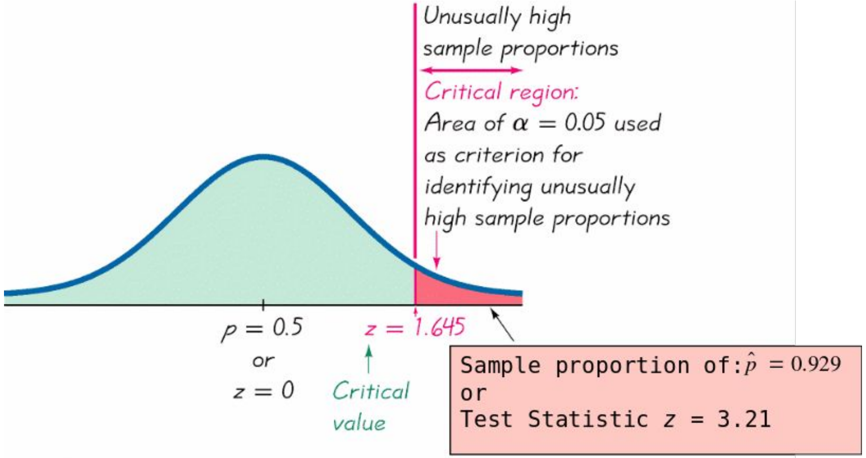
Interpretation

Meaning:
The small P-value shows that there is a **very small chance of getting the sample results that led to a test statistic of z = 3.21**. This suggests that with the XSORT method of gender selection, the **likelihood of having a baby girl is different from 0.5. → reject H0.**



A z score of 3.21 is "unusual" (because it is greater than 2, remember that μ = 0 and σ = 1).

The critical/rejection region (red-shaded area) is the set of all values of the test statistic **that cause us to reject the null** hypothesis.



“

Rewrite steps yang telah kita bahas ke dalam bentuk mindmap

(5 minutes)

The Error of Hypothesis Test

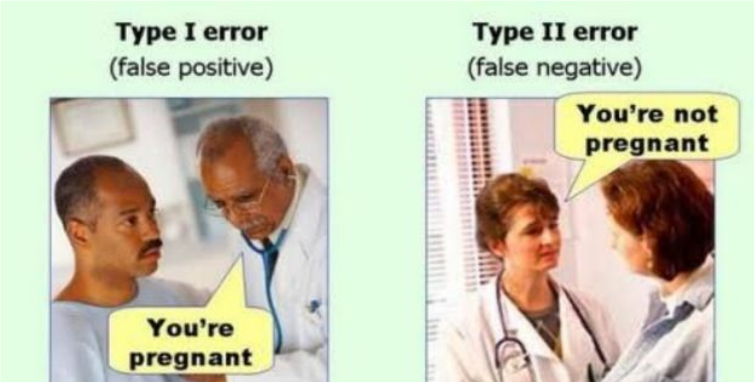
		Reality	
		Positive (H0 False)	Negative (H0 True)
Testing result & decision	Positive (reject H0)	**True Positive** **Power of the test (1-□)**	False Positive Type I error (α)
	Negative (fail to reject H0)	False Negative Type II error (□)	**True Negative**

- **Type I error represents with α.**
- **Type II error represents with β**

goals

Salah menolak

Salah menerima



The power of hypothesis test

- the probability of rejecting a false H0.
- (1 – □) → common requirement is **at least 0.80**.
- the probability of supporting H1 that is true.

Example:

1. Assume that we are conducting a hypothesis test of the claim that a method of gender selection increases the likelihood of a baby girl, so that the probability of a baby girls is $p > 0.5$.
   a. Define the H0 & H1
   b. Identify a type I error.
   c. Identify a type II error.

Answer:

a. **type I error**: Conclude that there is sufficient evidence to support $p > 0.5$, when in reality $p = 0.5$.
b. **Type II error**: Fail to reject $p = 0.5$ (and therefore fail to support $p > 0.5$) when in reality $p > 0.5$.

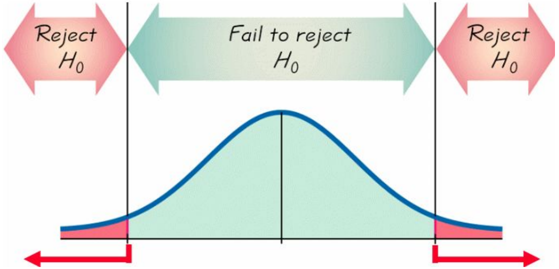
Controlling the Type I & II Errors

- If α is fixed, an increase in sample size *n* will cause a decrease in □.
- If sample size n is fixed, a decrease in α will cause an increase in □, and vice versa.
- To **decrease both α and □, increase the sample size *n***.

Case 1

First step: highlight and write down all the keywords

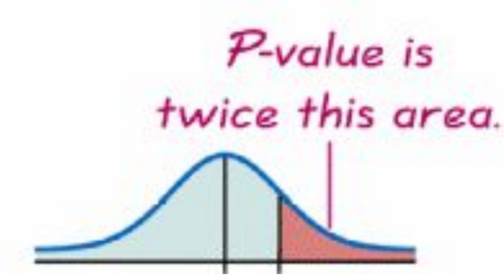
A study in which 57 out of 104 pregnant women correctly guessed the sex of their babies. Use these sample data to test the claim that the success rate of such guesses is no different from the 50% success rate expected with random chance guesses. Use a 0.05 significance level.



Answer:

- Parameter being tested: p → p = 0.5
- H0: p = 0.5 , H1: p ≠ 0.5
- α = 0.05
- $\hat{p}$ = 57/104
- Compute the test statistics

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{57}{104} - 0.50}{\sqrt{\frac{(0.50)(0.50)}{104}}} = 0.98$$



Case 1 (Cont.)

- Test statistics = 0.98 → P(z < 0.98) = pnorm(0.98) → **0.8364569**
  - Area to the right of test statistics = 1 - pnorm(0.98) = **0.1635431**
- two-tailed test, p-value is **twice** the area to the right of test statistic = 2 * 0.1635431 = **0.3270861**.
- Compare p-value with α: 0.3270861 > 0.05 → fail to reject H0.
- Conclusion:
  There is no sufficient evidence to warrant rejection of the claim that women who guess the sex of their babies have a success rate equal to 50%.
  → *tidak cukup bukti untuk menolak klaim bahwa 50% wanita mampu menebak dengan benar jenis kelamin bayinya.*

Case 2

People have died in boat accidents because an obsolete estimate of the mean weight of men was used. We obtain these sample statistics: n = 40 and $\bar{x}$ = 172.55 lb. Research from several other sources suggests that the population of weights of men has a standard deviation given by = 26 lb. Use these results to test the claim that men have a mean weight greater than 166.3 lb, which was the weight in the National Transportation and Safety Board's recommendation M-04-04. Use a 0.05 significance level.

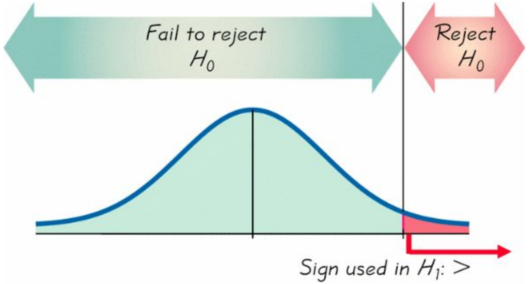
Answer:

n = 40, $\bar{x}$ = 172.55 lb, σ = 26, α = 0.05.

H0: μ = 166.3

H1: μ > 166.3

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{172.55 - 166.3}{\frac{26}{\sqrt{40}}} = 1.52$$



right-tailed test, so P-value is the area is to the right of z = 1.52 → 1 - pnorm(1.52) = 0.06425549 ≈ 0.0643.

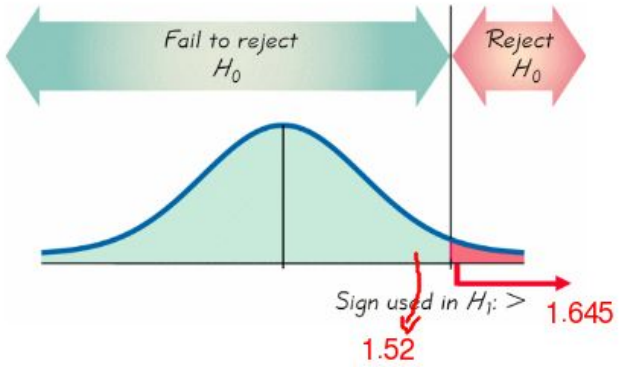
-

- traditional method: z 1.52 falls outside the critical region → fail to reject H0.

Case 2 (Cont.)

a. **p-value method**:
   a. test statistics vs α → 0.0643 > 0.05 → fail to reject H0.

b. **Traditional method**:
   a. check if test statistics falls within critical region
   b. With α = 0.05 → critical value = 1.645
   c. Meaning $\mathbf{z_{1.52}}$ falls outside the critical region → fail to reject H0.



Case 2 (Cont.)

c. **Confidence Interval method**:

a. Use a one-tailed test with α = 0.05. → based on table 8-2 → create 90% CI

b. $E = z_{\alpha/2} \, \sigma/\sqrt{n}$ → 1.959964 * 26/√40 = 6.761929.

c. 90% CI:

Using R:

```
> er <- qnorm(1 - (0.05))*26/sqrt(40)
> xbar <- 172.55
> leftCi <- xbar - er
> rightCi <- xbar + er
> print(paste(leftCi, " < μ < ", rightCi))
[1] "165.788070957618  < μ <  179.311929042382"
> print(paste(round(leftCi,3), " < μ < ", round(rightCi,3)))
[1] "165.788  < μ <  179.312"
```

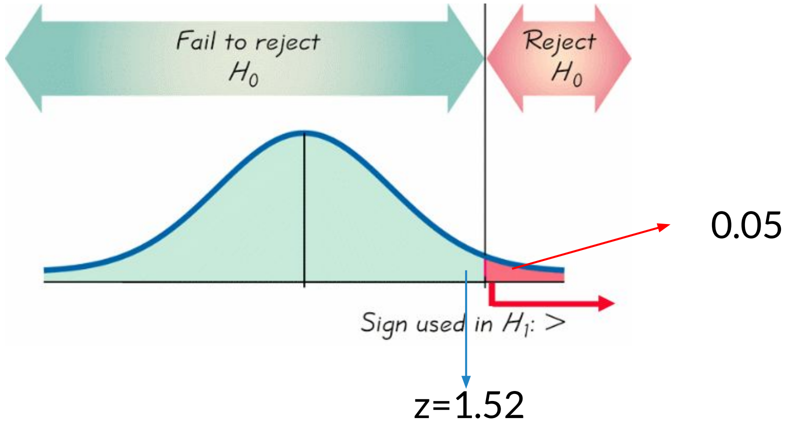
Table 8-2 Confidence Level for Confidence Interval

		Two-Tailed Test	One-Tailed Test
Significance Level for Hypothesis Test	0.01	99%	98%
	0.05	95%	90%
	0.10	90%	80%

d. 166.3 lb is contained in the CI → we cannot support a claim that **μ** is greater than 166.3 → fail to reject the null hypothesis.

Case 2 (Cont.)

Interpretation:
There is not sufficient evidence to support a conclusion that the population mean is greater than 166.3 lb. A sample mean such as 172.55 lb could easily occur by chance since it is a likely value.



Case 3

People have died in boat accidents because an obsolete estimate of the mean weight of men was used. We obtain these sample statistics: n = 40 and $\overline{x}$= 172.55 lb, and s = 26.33 lb. Do not assume that the value of **σ** is known. Use these results to test the claim that men have a mean weight greater than 166.3 lb, which was the weight in the National Transportation and Safety Board's recommendation M-04-04. Use a 0.05 significance level, and the traditional method outlined in Figure 8-9.

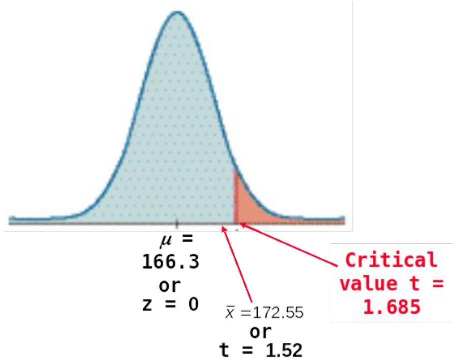
Answer:

n = 40 → df = 39 ; $\overline{x}$= 172.55 ; s = 26.33 ; α = 0.05

H0: μ = 166.3

H1: μ > 166.3

$$t = \frac{\overline{x} - \mu_{\overline{x}}}{\frac{s}{\sqrt{n}}} = \frac{172.55 - 166.3}{\frac{26.33}{\sqrt{40}}} = 1.501$$



a. Dengan metode p-value:
> pt(1.501, 39, lower.tail = FALSE)
[1] 0.07070451

0.07070451 > 0.05 → fail to reject H0

Dengan traditional method:
> qt(0.05, 39, lower.tail = FALSE)
[1] 1.684875 → (dengan tabel = 1.685)

t = 1.501 does not fall in the critical region bounded by t = 1.685 → fail to reject the null hypothesis.

Case 3 (Cont.)

Because we fail to reject the null hypothesis, we conclude that there is not sufficient evidence to support a conclusion that the population mean is greater than 166.3 lb, as in the National Transportation and Safety Board's recommendation.

Case 4

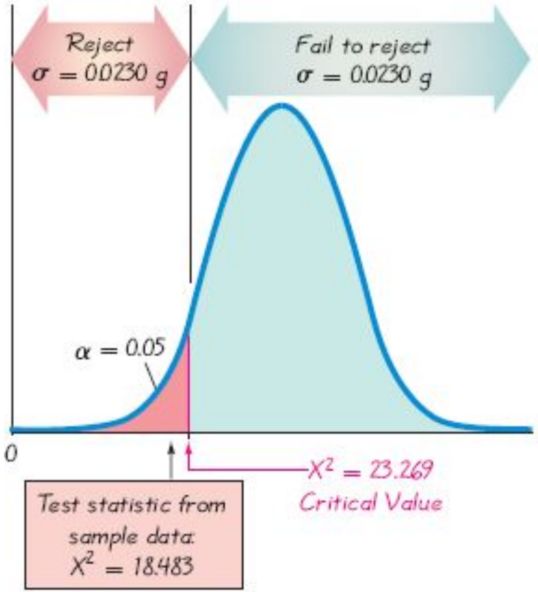
37 penny coins have a mean of weight of 2.49910 g and a standard deviation of 0.01648 g. U.S. Mint specifications require that pennies be manufactured so that the mean weight is 2.500 g. A hypothesis test will verify that the sample appears to come from a population with a mean of 2.500 g as required, but use a 0.05 significance level to test the claim that the population of weights has a standard deviation less than the specification of 0.0230 g.

Answer:
n=37 ; $\overline{x}$= 2.49910 ; s = 0.01648 ;
α=0.05 → critical value: **23.26861**.

H0: σ = 0.0230
H1: σ < 0.0230

Distribution: χ^2

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(37-1)(0.01648)^2}{0.0230^2} = 18.483$$



Decide using traditional method:
test statistic is in the critical region → reject H0.

Interpretation: There is sufficient evidence to support the claim that the standard deviation of weights is less than 0.0230 g

Theoritical Questions

1. What's hypothesis?
2. Jelaskan tentang 2 jenis hipotesis yang diperlukan dalam uji hipotesis.
3. Apa itu significance level?
4. Sebutkan 3 jenis uji hipotesis.
5. Apa itu test-statistics?
6. What's p-value? And how do we obtain it?
7. What's critical value?
8. Critical region
9. Dari data 50 sampel dan α=5%, tentukan p-value dari test statistics berikut (gunakan tabel dan R). Note that uji adalah 1 tail **>** $\rightarrow$ area to the right of test statistics.
    a. $z = 0.27$ $\rightarrow$ R: pnorm(z)
    b. $t = 2.021$ $\rightarrow$ R: pt(t, df, lower.tail=FALSE)
    c. $\mathcal{X}^2 = 12.40$ $\rightarrow$ R: pchisq($\mathcal{X}^2$, df, lower.tail=FALSE)
10. Jelaskan distribusi peluang yang digunakan dalam uji hipotesis:
    a. Proporsi
    b. Mean (σ diketahui)
    c. Mean (σ tidak diketahui)
    d. Varians

Theoritical Questions

11. Ceritakan tentang 2 jenis error yang dapat terjadi dalam uji hipotesis
12. Apa itu power of test?

Exercise 1:

Berikut adalah tinggi dari beberapa siswa yang di-sample secara random dari kelas 5-37 (cm): 168, 167, 166, 164, 165, 168, 165, 168, 168, 167, 170, 166, 170, 168, 170, 166. Dengan menggunakan significance level 0.01, ujilah klaim bahwa siswa kelas 5-37 memiliki tinggi dengan standar deviasi kurang dari 7.5 cm:

a. Dengan p-value method
b. Dengan traditional method
c. Dengan CI method

Exercise 2

Dari pengukuran suhu badan terhadap 106 orang yang dipilih secara acak, diketahui bahwa rata-rata suhu tubuh mereka adalah 98.2℉ dengan standar deviasi sebesar 0.62℉. Jika α=5%, ujilah the *common belief* bahwa mean dari populasi adalah 98.6℉.

a. Dengan p-value method
b. Dengan traditional method
c. Dengan CI method

Thank You