8.4 *P*-Values

P-Values

- Using the rejection region method to test hypotheses entails first selecting a significance level α.
- Then after computing the value of the test statistic, the null hypothesis H_0 is rejected if the value falls in the rejection region and is otherwise not rejected.
- We now consider another way of reaching a conclusion in a hypothesis testing analysis.
- This alternative approach is based on calculation of a certain probability called a ***P*-value**.
- One advantage is that the *P*-value provides an intuitive measure of the strength of evidence in the data against H_0.

เน้นหัวข้อนี้ มั้ย

8.4 *P*-Values

P-Values

ทดสอบสมมติฐาน

- Using the rejection region method to test hypotheses entails first selecting a significance level α. Type 1 error
- Then after computing the value of the test statistic, the null hypothesis H_0 is rejected if the value falls in the rejection region and is otherwise not rejected.
- We now consider another way of reaching a conclusion เหมือนกัน in a hypothesis testing analysis.
- This alternative approach is based on calculation of a certain probability called a ***P*-value**. ค่า prob ที่มีค่าระหว่าง 0 - 1
- One advantage is that the *P*-value provides an intuitive measure of the strength of evidence in the data against H_0.

Conclution

H_0	True	False
Accept		Type II error
Reject	Type I error	

P-Values

Definition

- The ***P*-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample.

182

P-Values

- This definition is quite a mouthful.
- Here are some key points:

- The P-value is a probability.
- This probability is calculated assuming that the null hypothesis is true.
- Beware: The P-value is not the probability that H_0 is true, nor is it an error probability!
- To determine the P-value, we must first decide which values of the test statistic are at least as contradictory to H_0 as the value obtained from our sample.

183

Example 14

- Urban storm water can be contaminated by many sources, including discarded batteries.
- When ruptured, these batteries release metals of environmental significance.
- The article "Urban Battery Litter" (*J. of Environ. Engr.*, 2009: 46–57) presented summary data for characteristics of a variety of batteries found in urban areas around Cleveland.
- A sample of 51 (n) Panasonic AAA batteries gave a sample mean zinc mass of 2.06g ($\bar{x}$) and a sample standard deviation of 0.141g. (s)

184

Example 14

cont'd

- Does this data provide compelling evidence for concluding that the population mean zinc mass (μ) exceeds 2.0g? — μ_0 – Null value ค่าว่าง
- With μ denoting the true average zinc mass for such batteries, the relevant hypotheses are

$H_0: \mu = \mu_0$

$H_a: \mu > \mu_0$

$n = 51$

$$H_0: \mu = 2.0 \text{ versus}$$
$$H_a: \mu > 2.0.$$

→ $n > 40$ มากเพียงพอ

- The sample size is large enough so that a z test can be used without making any specific assumption about the shape of the population distribution.

185

Example 14

cont'd

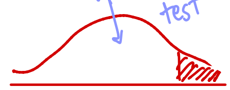
$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

- The test statistic value is

$$z = \frac{\bar{x} - 2.0}{s/\sqrt{n}} = \frac{2.06 - 2.0}{.141/\sqrt{51}} = 3.04$$

z_α

upper tailed test



$z = z_\alpha \geq 3.04$ → Reject H_0

- Now we must decide which values of z are at least as contradictory to H_0.

ขัดแย้ง

- Let's first consider an easier task:

Which values of $\bar{x}$ are at least as contradictory to the null hypothesis (H_0) as 2.06, the mean of the observations in our sample?

ขึ้นไป

Example 14

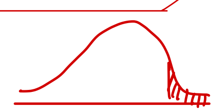
$H_0: \mu = 2.0$

$H_a: \mu > 2.0$

$\bar{x} = 2.06 \downarrow 2.10$

cont'd

- Because > appears in H_a, it should be clear that 2.10 is at least as contradictory to H_0 as is 2.06, and so in fact is *any* $\bar{x}$ value that exceeds 2.06.
- But an $\bar{x}$ value that exceeds 2.06 corresponds to a value of z that exceeds 3.04.
- Thus the P-value is



$Z = 3.04$ →

α = ค่าความน่าจะเป็นของ type 1 error

พท.ใต้เส้นโค้ง P-value = $P(Z \geq 3.04$ when $\mu = 2.0)$

Z ที่ ≥ Z0

- Since the test statistic Z was created by subtracting the null value 2.0 in the numerator, when $\mu = 2.0$—i.e., when H_0 is true—Z has approximately a standard normal distribution.

Example 14

cont'd

$1 - P(Z \leq 3.04)$

$1 - P(3.04)$

$= 1 - \Phi(3.04)$ → standard Normal Distribution

$= 0.0012$

As a consequence,

P-value $= P(Z \geq 3.04$ when $\mu = 2.0)$

$\approx$ area under the z curve to the right of 3.04

$= 1 - F(3.04)$

$= 0.0012$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

P-Values

- We will shortly illustrate how to determine the *P*-value for any z or t test—i.e., any test where the reference distribution is the standard normal distribution (and z curve) or some t distribution (and corresponding t curve).

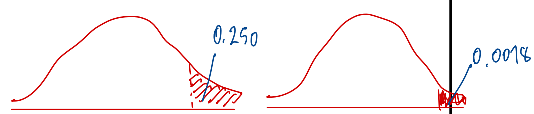
- For the moment, though, let's focus on reaching a conclusion once the *P*-value is available.

- Because it is a ค.น่าจะเป็น probability, the *P*-value 0 – 1 must be between 0 and 1.

P-Values

H_0

- What kinds of *P*-values provide evidence against the null hypothesis?
- Consider two specific instances:



- $\boxed{P\text{-value} = 0.250}$: In this case, fully 25% of all possible test statistic values are at least as contradictory to H_0 as the one that came out of our sample.
- So our data is not all that contradictory to the null hypothesis.
- $\boxed{P\text{-value} = 0.0018}$: Here, only 0.18% (much less than 1%) of all possible test statistic values are at least as contradictory to H_0 as what we obtained. Thus the sample appears to be highly contradictory to the null hypothesis.

190

P-Values

$H_0: \mu = 2.0$

$H_a: \mu > 2.0$

- More generally, *the smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.*

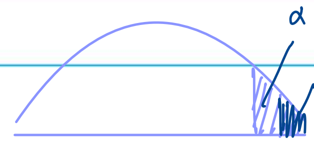
ขัดแย้ง = ปฏิเสธ

- That is, H_0 should be rejected in favor of H_a when the *P*-value is sufficiently small.
- So what constitutes "sufficiently small"?

191

P-Values

Decision rule based on the P-value

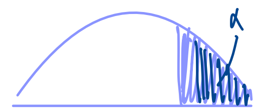


- Select a significance level α (as before, the desired type I error probability).
- Then

reject H_0 if P-value $\le \alpha$

do not reject H_0 if P-value $> \alpha$

ไม่ reject = ยอมรับ



- Thus if the P-value exceeds (มากกว่า) the chosen significance level (α), the null hypothesis cannot be rejected at that level. P-value $> \alpha$ ไม่ปฏิเสธH_0
- But if the P-value is equal to or less than α, then there is enough evidence to justify rejecting H_0.

P-value $\le \alpha$ ปฏิเสธH_0

P-Values

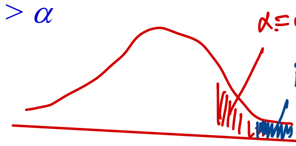
- In Example 14, we calculated P-value = 0.0012.
- Then using a significance level of 0.01 ($\alpha = 0.01$), we would reject the null hypothesis (H_0) in favor of the alternative hypothesis because 0.0012 ≤ 0.01.

P-value $\le \alpha$

ยอมรับ H_a

0.0012 ↓ reject H_0 if P-value $\le$ ↓ 0.01 α

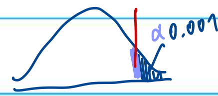
do not reject H_0 if P-value $> \alpha$



H_0: $\mu = 2.0$
H_a: $\mu > 2.0$.

P-Values

P-value = 0.0012



α = 0.001

- However, suppose we select a significance level of only 0.001, which requires more substantial evidence from the data before H_0 can be rejected.
- In this case we would not reject H_0 because 0.0012 > 0.001.
  Accept Ho
- How does the decision rule based on the *P*-value compare to the decision rule employed in the rejection region approach?
- *The two procedures—the rejection region method and the P-value method—are in fact identical.*
  ① α ② α เหมือนกัน

reject H_0 if *P*-value ≤ α

do not reject H_0 if *P*-value > α

194

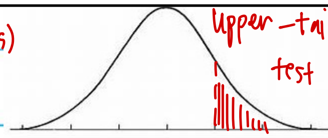
P-Values

- Whatever the conclusion reached by employing the rejection region approach with a particular α, the same conclusion will be reached via the *P*-value approach using that same α.

195

Example 15

P-value = $P(Z \geq 2.10;\ \mu = 1.5)$
$= 1 - P(Z \leq 2.10)$
$= 1 - \phi(2.10)$
$= 1 - 0.9821 = 0.0179$



upper-tailed test

- The nicotine content problem involved testing

$$H_0\text{: } \mu = 1.5 \text{ versus } H_a\text{: } \mu > 1.5$$

$Z = Z_\alpha$
$Z \longrightarrow$ reject H_0
$Z = 2.10$ → Prob → P-value → α

- using a z test (i.e., a test which utilizes the z curve as the reference distribution).
- The inequality in H_a implies that the upper-tailed rejection region $z \geq z_\alpha$ is appropriate.
- Suppose $z = 2.10$.
- Then using exactly the same reasoning as in Example 14 gives P-value = 1 – F(2.10) = 0.0179.

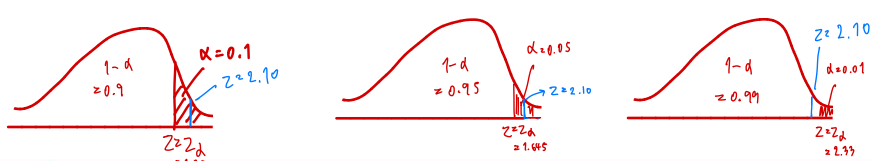
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857

196

Example 15 Suppose $z = 2.10$. Rejection Region

cont'd

- Consider now testing with several different significance levels:
- $\alpha = 0.10 \Rightarrow z_\alpha = z_{.10} = 1.28 \Rightarrow 2.10 \geq 1.28 \Rightarrow$ reject H_0
- $\alpha = 0.05 \Rightarrow z_\alpha = z_{.05} = 1.645 \Rightarrow 2.10 \geq 1.645 \Rightarrow$ reject H_0
- $\alpha = 0.01 \Rightarrow z_\alpha = z_{.01} = 2.33 \Rightarrow 2.10 < 2.33 \Rightarrow$ do not reject H_0 Accept H_0



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916

197

Example 15

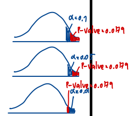
reject H_0 if P-value $\leq \alpha$

do not reject H_0 if P-value $> \alpha$

cont'd

- Because P-*value* = 0.0179 ≤ 0.10 and also 0.0179 ≤ 0.05, using the P-value approach results in rejection of H_0 for the first two significance levels.

P-value	α	Conclution
0.0179	0.1	Reject H_0
0.0179	0.05	Reject H_0
0.0179	0.01	Not Reject (Accept H_0)



- However, for $\alpha = 0.01$, $z = 2.10$ is not in the rejection region and 0.0179 is larger than 0.01.

- More generally, whenever α is smaller than the P-value 0.0179, the critical value z_α will be beyond the calculated value of z and H_0 cannot be rejected by either method.

198

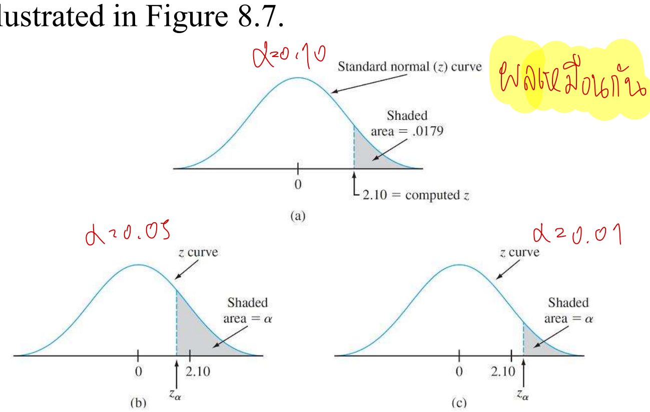
Example 15

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

do not reject H_0 if P-value $> \alpha$

cont'd

This is illustrated in Figure 8.7.



Relationship between α and tail area captured by computed z: (a) tail area captured by computed z; (b) when $\alpha < .0179$, $z_\alpha < 2.10$ and H_0 is rejected; (c) when $\alpha < .0179$, $z_\alpha < 2.10$ and H_0 is not rejected

Figure 8.7

199

P-Values

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

do not reject H_0 if P-value $> \alpha$

P-value
Accept Ho ↓ Reject Ho
↖ α ↗

Proposition

- The P-value is the smallest significance level α at which the null hypothesis can be rejected.

- Because of this, the P-value is alternatively referred to as the **observed significance level** (OSL) for the data.
- It is customary to call the data *significant* (ข้อมูลมีนัยยะสำคัญ) when H_0 is rejected and *not significant* otherwise.

ข้อมูลไม่มีนัยสำคัญ → Accept Ho

ของ α น้อยสุด

- The P-value is then the smallest level at which the data is significant.

P-value = ค่า α น้อยสุดที่ข้อมูลมีนัยสำคัญ (Reject Ho)

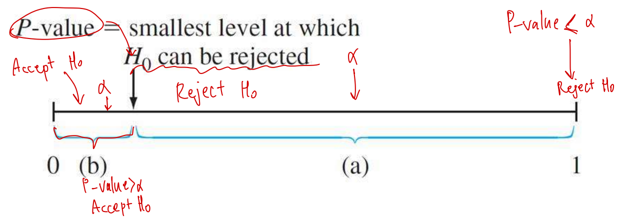
200

P-Values

Reject Ho ถ้า P-value ≤ α

Accept Ho ถ้า P-value > α

- An easy way to visualize the comparison of the P-value with the chosen α is to draw a picture like that of Figure 8.8.



Comparing α and the P-value: (a) reject H_0 when α lies here; (b) do not reject H_0 when α lies here

Figure 8.8

- The calculation of the P-value depends on whether the test is upper-, lower-, or two-tailed.
- However, once it has been calculated, the comparison with α does not depend on which type of test was used.

201

Example 16

- The true average time to initial relief of pain for a best-selling pain reliever is known to be 10 min. $\mu = 10$

- Let μ denote the true average time to relief for a company's newly developed reliever.

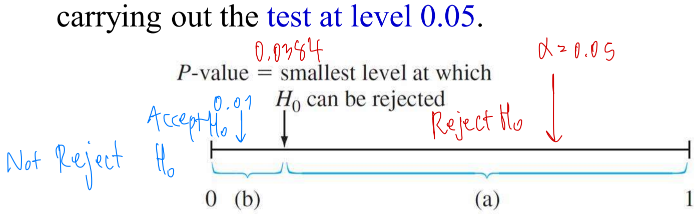
- Suppose that when data from an experiment involving the new pain reliever is analyzed, the *P*-value for testing

H_0: $\mu = 10$ versus
H_a: $\mu < 10$ is calculated as 0.0384.

Example 16

cont'd

- Since $\alpha = 0.05$ is larger than the *P*-value [0.05 lies in the interval (a) of Figure 8.8], H_0 would be rejected by anyone carrying out the test at level 0.05.



Comparing α and the P-value: (a) reject H_0 when α lies here; (b) do not reject H_0 when α lies here

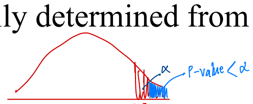
Figure 8.8

- However, at level 0.01, H_0 would not be rejected because 0.01 is smaller than the smallest level (0.0384) at which H_0 can be rejected.

P-Values for *z* Tests

P-Values for *z* Tests

- The *P*-value for a *z* test (one based on a test statistic whose distribution when H_0 is true is at least approximately standard normal) is easily determined from the information in Appendix Table A.3.



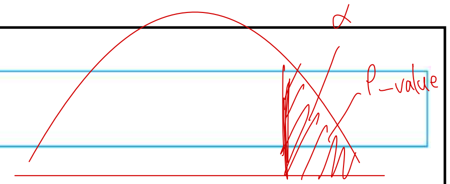
- Consider an upper-tailed test and let z denote the computed value of the test statistic Z.

คำนวณที่ได้

- The null hypothesis is rejected if $z \geq z_\alpha$, and the *P*-value is the smallest a for which this is the case.

- Since z_α increases as α decreases, the *P*-value is the value of a for which $z = z_\alpha$.

P-Values for z Tests



- That is, the P-value is just the area captured by the computed value z in the upper tail of the standard normal curve.

P-value = $P(Z \geq z)$
$= 1 - P(Z \leq z) = 1 - \phi(z)$

- The corresponding cumulative area is $F(z)$, so in this case

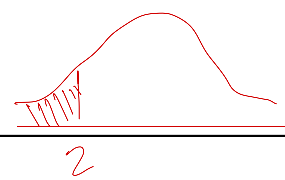
$$\boxed{P\text{-value} = 1 - F(z). \quad = 1 - \phi(z)}$$

cdf

Upper

- An analogous argument for a lower-tailed test shows that the P-value is the area captured by the computed value z in the lower tail of the standard normal curve.

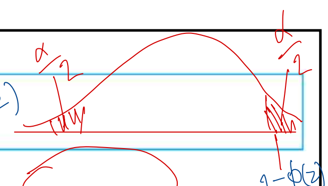
P-value = $P(Z < z)$
$= \phi(z)$



206

P-value = $2(1 - \phi(z))$

P-Values for z Tests



- More care must be exercised in the case of a two-tailed test. Suppose first that z is positive. Then the P-value is the value of α satisfying $z = z_{\alpha/2}$ (i.e., computed z = upper-tail critical value).

- This says that the area captured in the upper tail is half the P-value, so that P-value = $2[1 - F(z)]$.

- If z is negative, the P-value is the α for which $z = -z_{\alpha/2}$, or, equivalently, $-z = z_{\alpha/2}$ so that P-value = $2[1 - F(-z)]$.

207

P-Values for z Tests

- Since $-z = |z|$ when z is negative, P-value $= 2[1 - F(|z|)]$ for either positive or negative z.

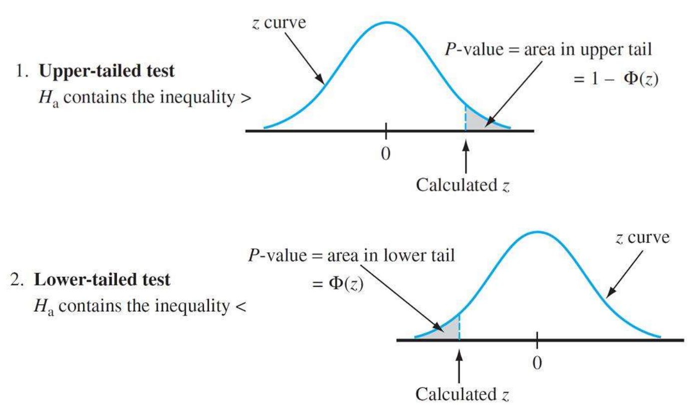
$$P\text{-value: } P = \begin{cases} 1 - \Phi(z) & \text{for an upper-tailed } z \text{ test} \\ \Phi(z) & \text{or an lower-tailed } z \text{ test} \\ 2[1 - \Phi(|z|)] & \text{for a two-tailed } z \text{ test} \end{cases}$$

- Each of these is the probability of getting a value at least as extreme as what was obtained (assuming H_0 true).

208

P-Values for z Tests

The three cases are illustrated in Figure 8.9.



Determination of the P-value for a z test

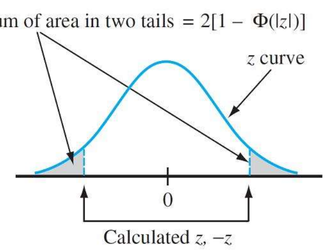
Figure 8.9

209

P-Values for z Tests

cont'd

3. **Two-tailed test**
   H_a contains the inequality $\neq$

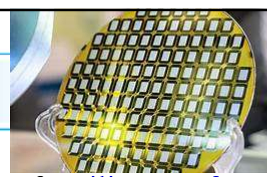


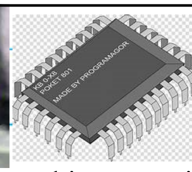
Determination of the P-value for a z test

Figure 8.9

210

Example 17





- The target thickness for silicon wafers used in a certain type of integrated circuit $\mu = 245$ is 245 μm.

$n = 50$

- A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of $\bar{x} =$ 246.18 μm and a sample standard deviation σ of 3.60 μm.

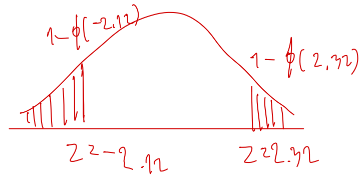
- Does this data suggest that true average wafer thickness is something other than the target value?

211

Example 17

cont'd

1. Parameter of interest: μ = true average wafer thickness
2. Null hypothesis: H_0: $\mu = 245$
3. Alternative hypothesis: H_a: $\mu \neq 245$
4. Formula for test statistic value: $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$ ≈ 2.4618
5. Calculation of test statistic value: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$



212

Reject H_0 P-value ≤ α

Not Reject H_0 P-value > α

Example 17

cont'd

α P-value 0.0204

Not Reject

6. Determination of P-value: Because the test is two-tailed,

   P-value = 2(1 – F(2.32)) = 0.0204

   = 2(1-0.9898)

   α = 0.01

7. Conclusion: Using a significance level of 0.01, H_0 would not be rejected since 0.0204 > 0.01.

   α = 0.01

   At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

   $\mu_0 = 245$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916

213

End of Section 8.4

241