Pengujian Hipotesis

Tim Dosen Pengantar Statistika

UNIVERSITAS AIRLANGGA
Excellence with morality

Statistics

Descriptive Statistics

- Collecting
- Organizing
- Summarizing
- Presenting Data

Inferential Statistics

- Parameter estimate
- **Hypothesis testing**
- Determining relationship
- Making predictions

A **hypothesis** is something that has not yet been proven to be true.

Hypothesis testing is the process of determining whether or not a given hypothesis is true.

Hypothesis testing, examine whether the sample data support or contradict the investigator's conjecture about the true value of the parameter.

Statistical inference deals with **drawing conclusions about population parameters** from an analysis of the sample data.

Netto: 85gr

Customer
Apakah benar, **rata-rata** berat bersih dari indomie goreng ini **setidaknya** 85gr per bungkusnya?
Netto indomie ≥ 85gr

Company
Apakah benar, **rata-rata** berat bersih dari indomie goreng ini **tepat** 85gr per bungkusnya?
Netto indomie = 85gr

Berdasarkan data panjang jalan tol yang dikeluarkan oleh Kementerian PUPR pada tahun 2010, panjang jalan tol di pulau Jawa adalah 472 km. Setelah 10 tahun, seorang pengamat ingin menguji jika panjang jalan tol di pulau Jawa di tahun 2020 lebih panjang daripada panjang jalan tol di tahun 2010.

Formulate Hypothesis

- Null hypothesis (H_0)

A null hypothesis is **an assertion** about **the value of a population parameter**.

It is an assertion that we hold as **true** unless we have sufficient statistical evidence to conclude otherwise.

- Alternative hypothesis (H_1 / Ha)

The alternative hypothesis is the **negation of the null hypothesis**.

H_0	H_1 / Ha
Assumption, status quo	Rejection of H_0
Assumed to be "true", a given	Rejection of an assumption or the given
Negation of the research question	Research question to be "proven"
Always contains an equality (=, ≤, ≥)	Does not contain equality (≠, <, >)

H_0: $\mu =$ $\quad$ H_0: $\mu \leq$ $\quad$ H_0: $\mu \geq$

H_1 : $\mu \neq$ $\quad$ H_1 : $\mu >$ $\quad$ H_1 : $\mu <$

The **decision** made for the hypothesis is **reject H_0** or **fail to reject H_0**, we don't say accept H_0

Formulate Hypothesis

Apakah benar, **rata-rata** berat bersih dari indomie goreng ini **tepat** 85gr per bungkusnya?
$H_0: \mu = 85$ gr
$H_1: \mu \neq 85$ gr

A vendor claims that his company fills any accepted order, on the **average**, in **at most six working days**. You suspect that the average is greater than six working days and want to test the claim of the vendor.
$H_0: \mu \leq 6$ days
$H_1: \mu > 6$ days

A manufacturer of golf balls claims that the **variance** of the weights of the company's golf balls **is controlled to within 0.0028 oz²**.
$H_0: \sigma^2 \leq 0.0028$ oz^2
$H_1: \sigma^2 > 0.0028$ oz^2

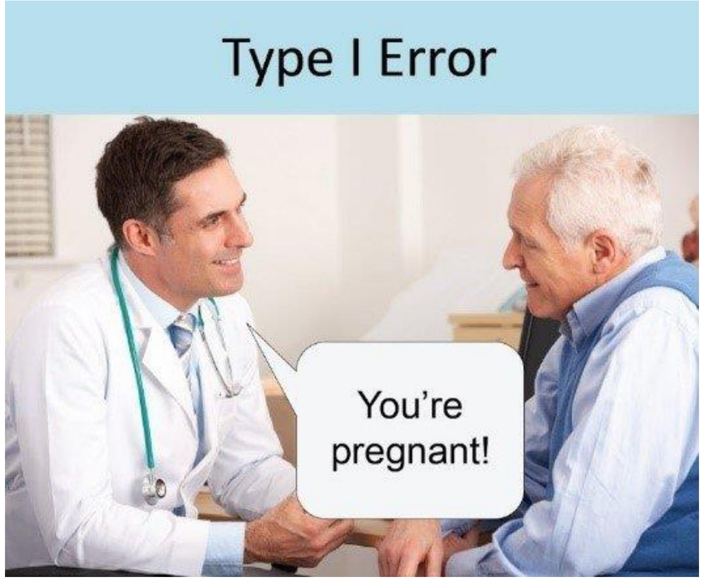
At least 20% of the visitors to a particular commercial Web site where an electronic product is sold are said to end up ordering the product.
$H_0: p \geq 20\%$
$H_1: p < 20\%$

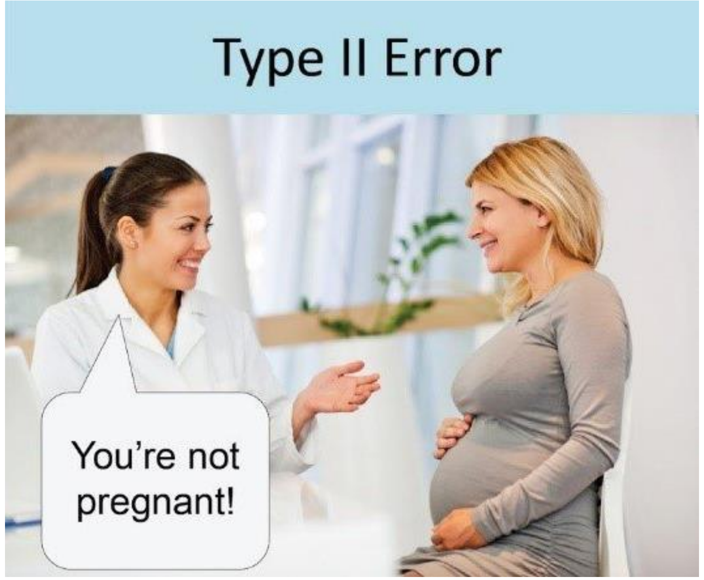
Type I and Type II Error

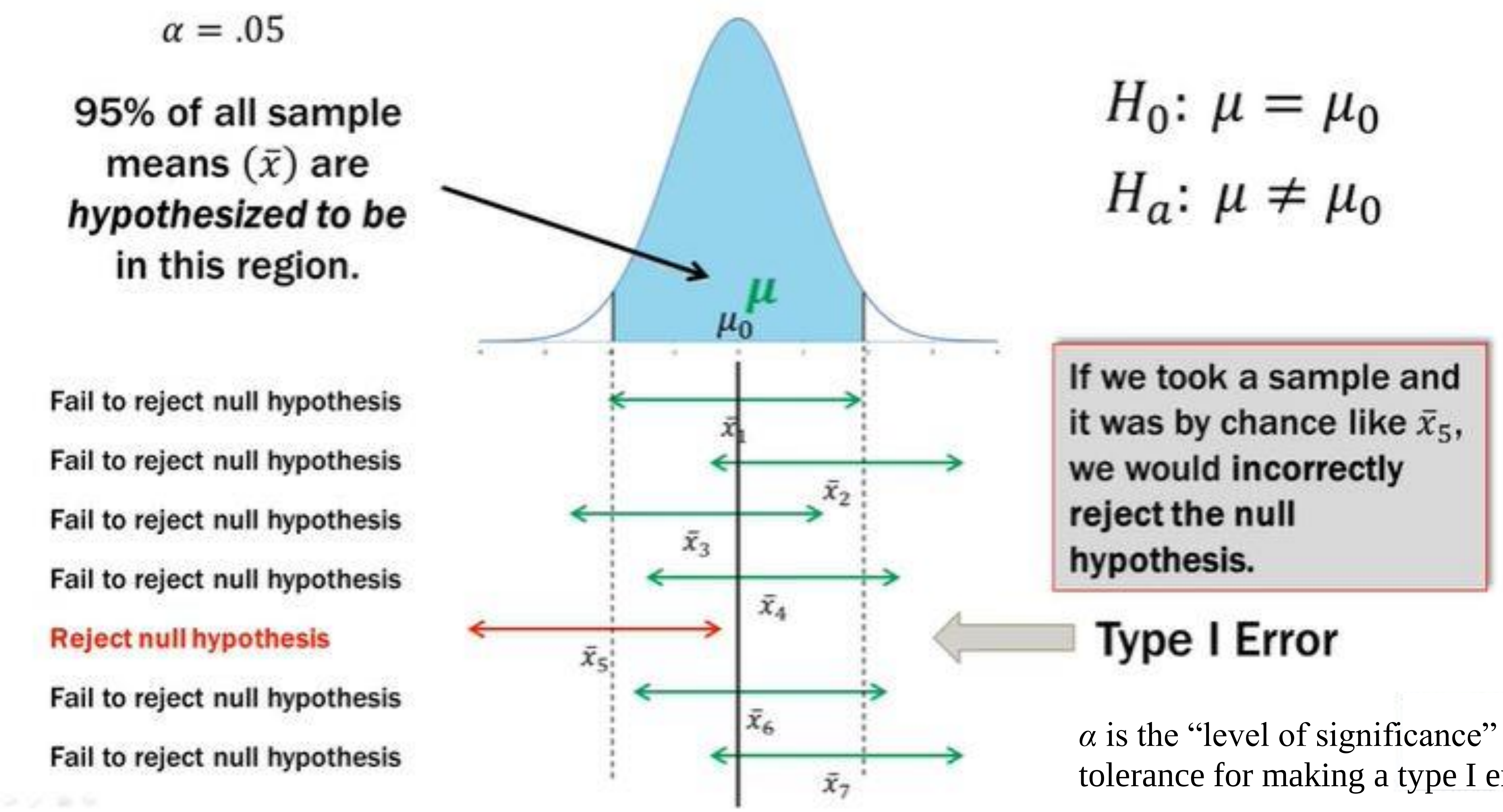
After the null and alternative hypotheses are spelled out, the next step is to **gather evidence**. In all real-world cases, the evidence is gathered from a **random sample** of the population. Thus, there will be chances for **error**.

		Actual condition	
		H_0 TRUE	H_0 FALSE
Conclusion	Fail to reject H_0	Correct (True Positive)	**Type II error** (False Negative)
	Reject H_0	**Type I error** (False Positive)	Correct (True Negative)

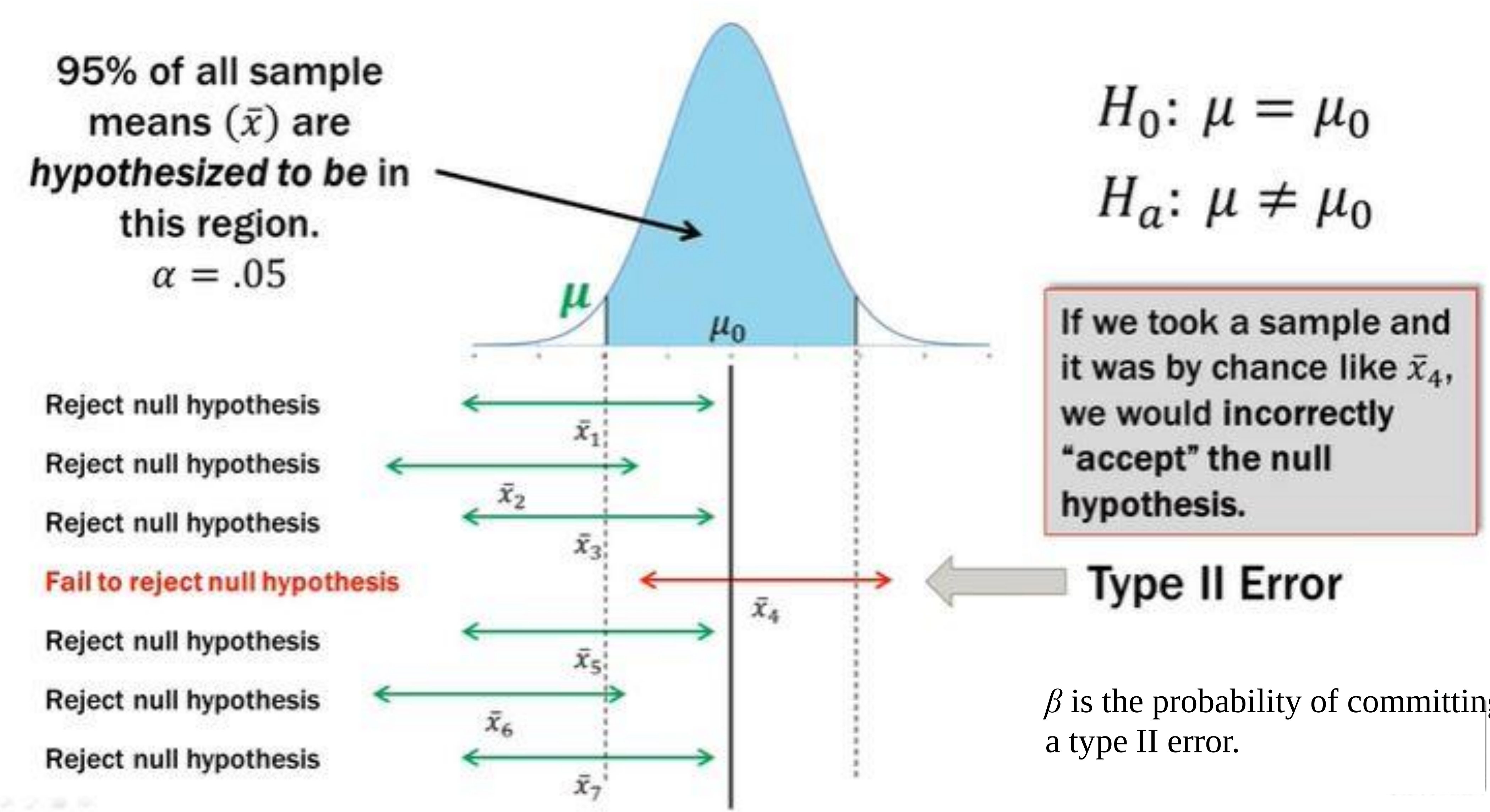
Type I error : rejecting a true null hypothesis
Type II error : accepting a false null hypothesis







α is the "level of significance" of our tolerance for making a type I error



β is the probability of committing a type II error.

The Two-tailed Test Rejection Region

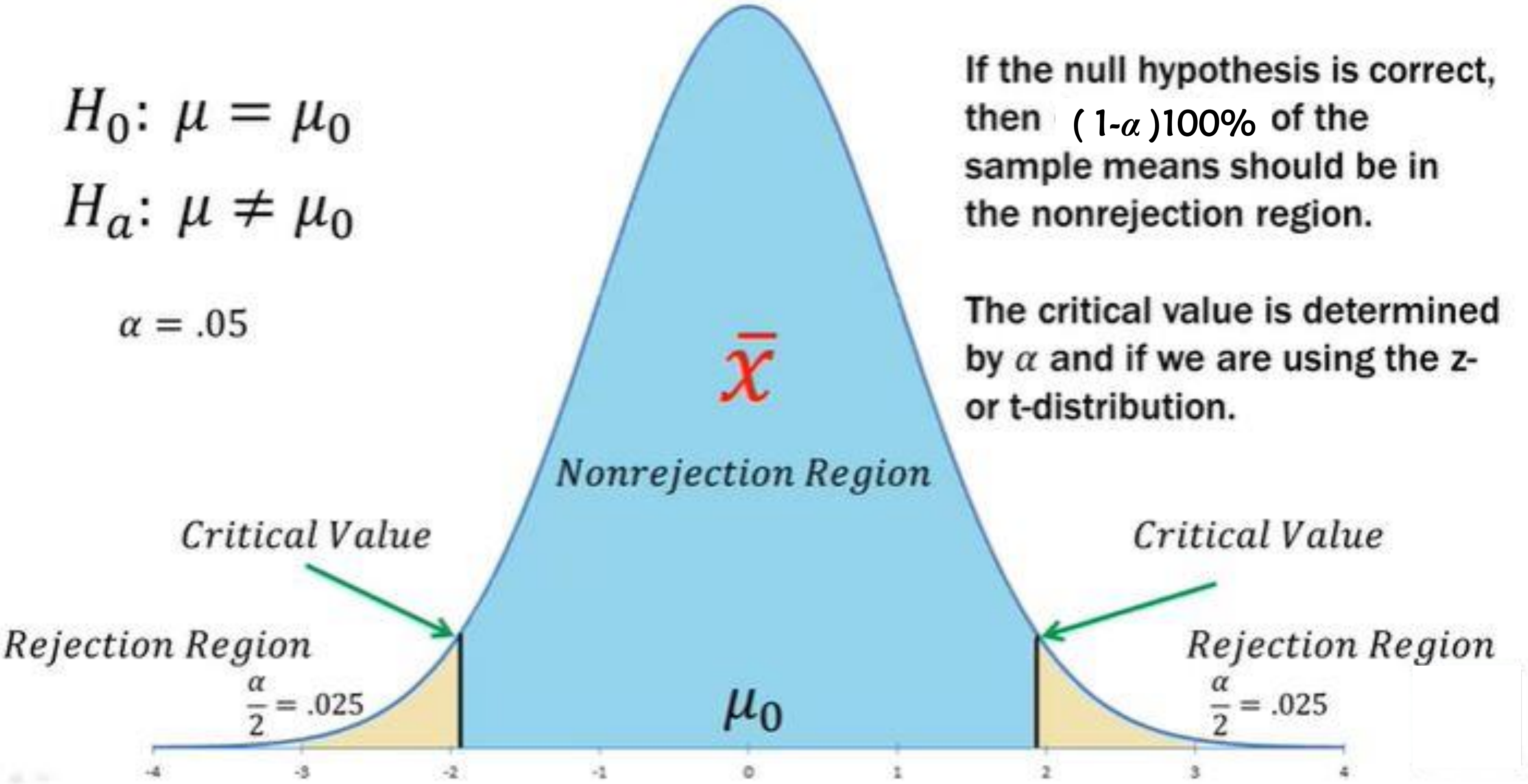
$H_0\colon \mu = \mu_0$

$H_a\colon \mu \neq \mu_0$

$\alpha = .05$

If the null hypothesis is correct, then $(1-\alpha)100\%$ of the sample means should be in the nonrejection region.

The critical value is determined by α and if we are using the z- or t-distribution.



The One-tailed (Lower) Test Rejection Region

In a one-tailed hypothesis test, all of the α is in one tail or the other depending on the alternative hypothesis.

The trick is that H_a "points" to the tail where the critical value and rejection region are.

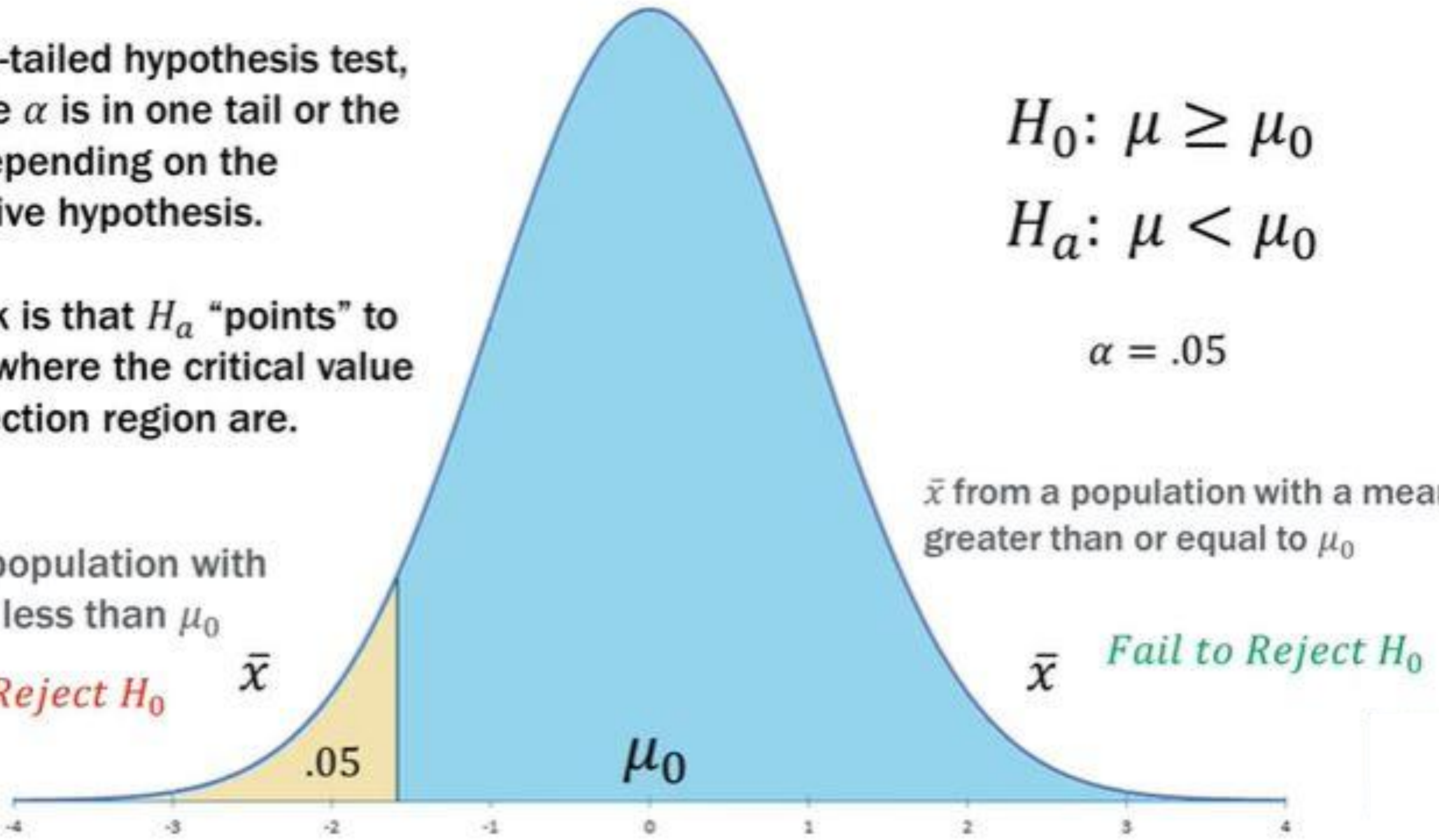
$$H_0\colon \mu \geq \mu_0$$

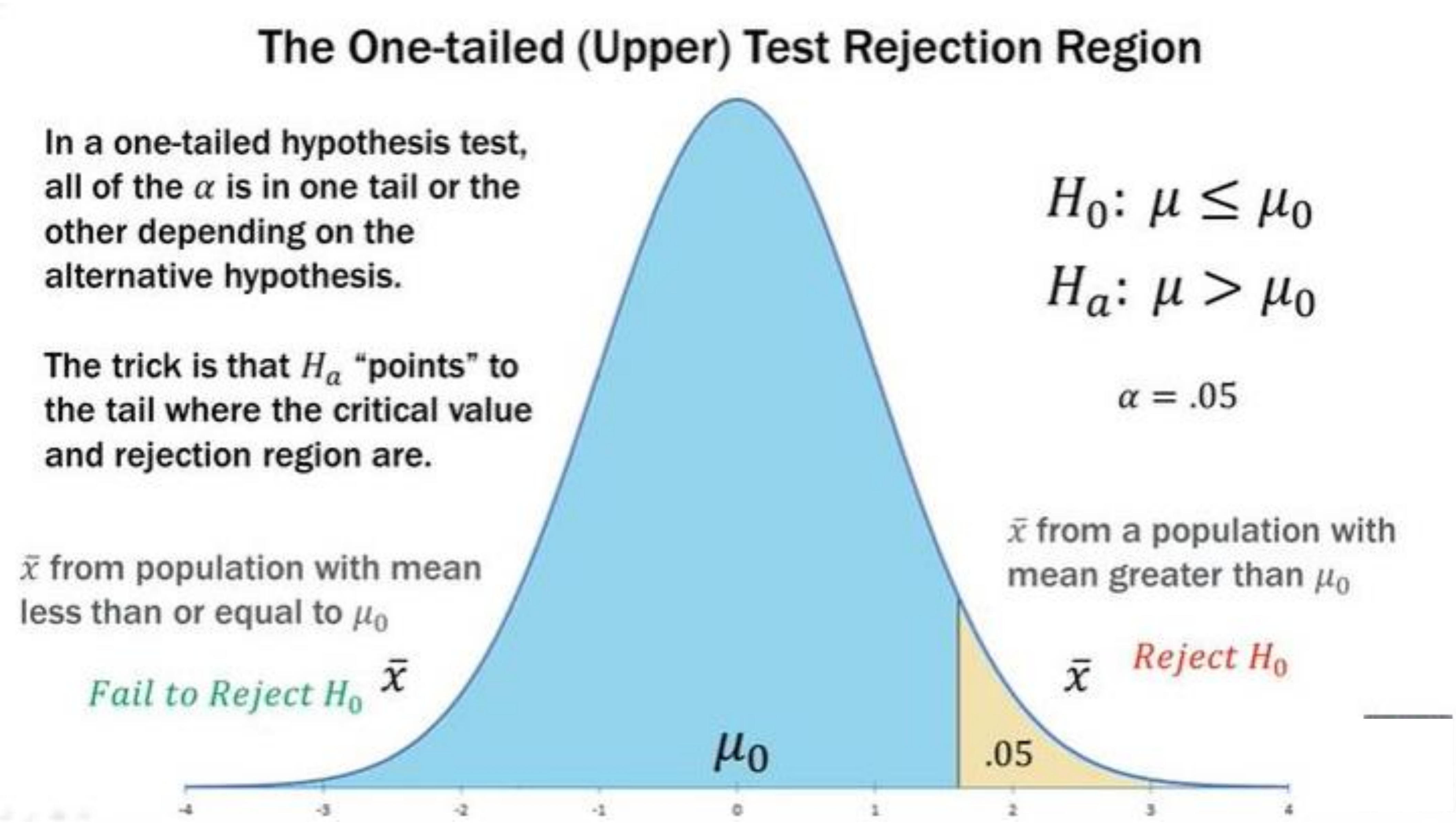
$$H_a\colon \mu < \mu_0$$

$\alpha = .05$

$\bar{x}$ from population with a mean less than μ_0

$\bar{x}$ from a population with a mean greater than or equal to μ_0

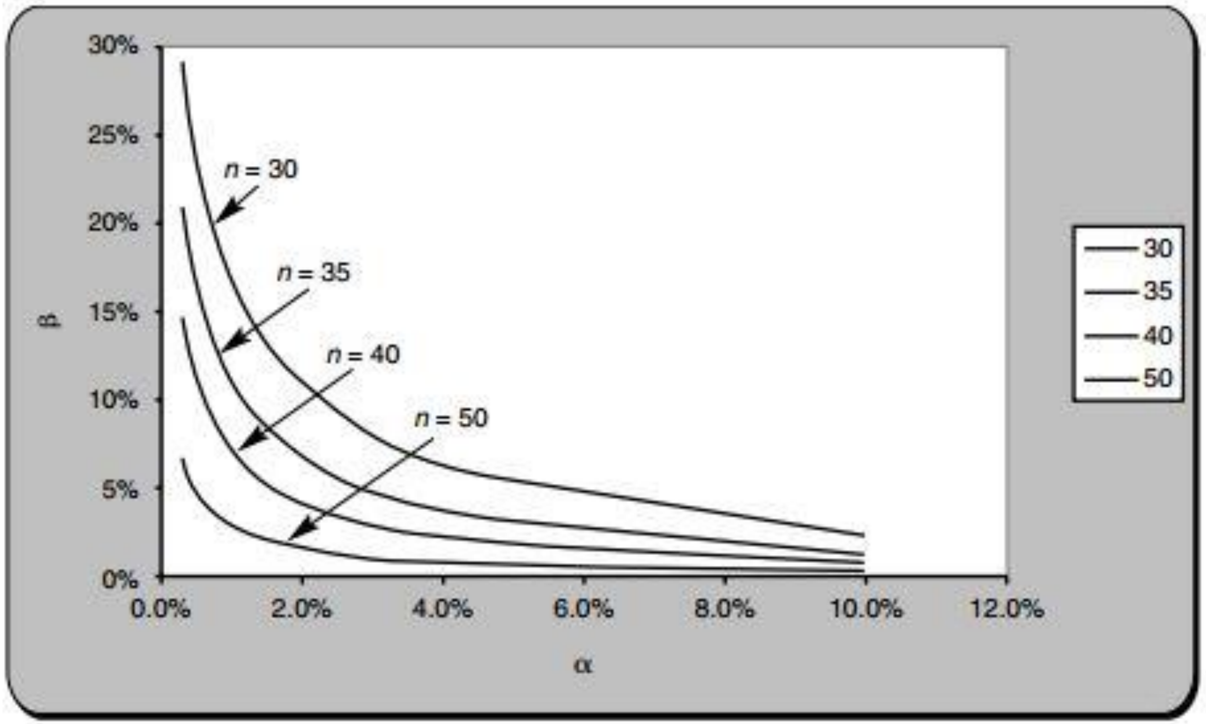



The One-tailed (Upper) Test Rejection Region
In a one-tailed hypothesis test, all of the α is in one tail or the other depending on the alternative hypothesis.
The trick is that H_a "points" to the tail where the critical value and rejection region are.
H_0: $\mu \le \mu_0$
H_a: $\mu > \mu_0$
$\alpha = .05$
$\bar{x}$ from population with mean less than or equal to μ_0
$\bar{x}$ from a population with mean greater than μ_0
Fail to Reject H_0
$\bar{x}$
Reject H_0
$\bar{x}$
μ_0
.05
-4
-3
-2
-1
0
1
2
3
4

P-value

P-value is the **probability** of getting a sample evidence that is equally or more unfavorable to the null hypothesis while the null hypothesis is actually true.

The p-value is a kind of "**credibility rating**" of H_0 in light of the evidence.



Level of Significance

- The most common policy in statistical hypothesis testing is to **establish** a **significance level**, denoted by α, and **to reject H_0 when the p-value falls below it**. When this policy is followed, one can be sure that the maximum probability of **type I error is α.**
- Rule: When the p-value is less than α, reject H_0.
- The standard values for α are 10%, 5%, and 1%.
- The symbol used for the probability of type II error is β.
- Note that β depends on the actual value of the parameter being tested, the sample size, and α.

Want to have low α as well as low β ?
Increase the n

Compute the P-value

H_0: $\mu \geq 1000$

H_1 : $\mu < 1000$

$\bar{X} = 999$

$\sigma = 5$

$n = 100$

$\alpha = 5\%$

$Z_\alpha = Z_{0.05} = 1.645$

Test statistics:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{999 - 1000}{5/\sqrt{100}} = -2.00$$

P-value = $P(Z < -2.00) = 0.0228$

Rejection region: reject H_0 if p-value < α or $Z \leq -Z_\alpha$

Statistical conclusion: Since p-value (0.0228) < α (0.05) and/or $Z(-2.00) \leq -Z_\alpha(-1.645)$, so we **reject** H_0.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Hypothesis Testing

Procedure

1. State the problem
2. Set the hypothesis (H_0 and H_1)
3. Specify the appropriate statistical test
4. Calculate test statistics
5. Set the type I error (α)
6. Declare the decision rule
7. State statistical conclusion
8. Make decision

Types of hypothesis tests:

1. Tests of hypotheses about population means.
2. Tests of hypotheses about population proportions.
3. Tests of hypotheses about population variances.

Hypothesis Testing Population Mean

Large Sample Tests for μ

When the sample size is large, a **Z test** concerning μ is based on the normal test statistic

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

The rejection region is one- or two-sided depending on the alternative hypothesis. Specifically,

$$\begin{array}{lll} H_1: \mu > \mu_0 & \text{requires} & R: Z \geq z_\alpha \\ H_1: \mu < \mu_0 & & R: Z \leq -z_\alpha \\ H_1: \mu \neq \mu_0 & & R: |Z| \geq z_{\alpha/2} \end{array}$$

Hypotheses Tests for μ—Small Samples

To test $H_0: \mu = \mu_0$ concerning the mean of a **normal population,** the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

which has Student's t distribution with $n - 1$ degrees of freedom:

$$\begin{array}{ll} H_1: \mu > \mu_0 & R: T \geq t_\alpha \\ H_1: \mu < \mu_0 & R: T \leq -t_\alpha \\ H_1: \mu \neq \mu_0 & R: |T| \geq t_{\alpha/2} \end{array}$$

The test is called a **Student's *t* test** or simply a ***t* test.**

Testing Population Mean Example

An automatic bottling machine fills cola into 2-liter (2,000-cm³) bottles. A consumer advocate wants to test the null hypothesis that the average amount filled by the machine into a bottle is at least 2,000 cm³. A random sample of 40 bottles coming out of the machine was selected and the exact contents of the selected bottles are recorded. The sample mean was 1,999.6 cm³. The population standard deviation is known from past experience to be 1.30 cm³.

1. Test the null hypothesis at an α of 5%.
2. Assume that the population is normally distributed with the same σ of 1.30 cm³. Assume that the sample size is only 20 but the sample mean is the same 1,999.6 cm³. Conduct the test once again at an α of 5%.
3. If there is a difference in the two test results, explain the reason for the difference.

H_0: $\mu \geq 2000$
H_1 : $\mu < 2000$

$\bar{X}$ = 1999.6; σ = 1.30; n = 40

α = 5%; $Z_{0.05} = 1.645$

Test statistics:

$$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{1999.6-2000}{1.30/\sqrt{40}} = -1.95$$

P-value = $P(Z < -1.95) = 0.0256$

Rejection region: reject H_0 if p-value < α or $Z \leq -Z_\alpha$

Statistical conclusion: Since p-value (0.0256) < α (0.05) and/or $Z(-1.95) \leq -Z_\alpha(-1.645)$, so we **reject** H_0.

n = 20

Test statistics:

$$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{1999.6-2000}{1.30/\sqrt{20}} = -1.38$$

P-value = $P(Z < -1.38) = 0.0838$

Rejection region: reject H_0 if p-value < α or $Z \leq -Z_\alpha$

Statistical conclusion: Since p-value (0.0838) > α (0.05) and/or $Z(-1.38) \geq -Z_\alpha(-1.645)$, so we **fail to reject** H_0.

The greater n, the smaller p-value

Testing Population Proportion

Hypotheses about population proportions can be tested using the binomial distribution or normal approximation to calculate the p-value.

Test statistics :

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

The rejection region is right-sided, left-sided, or two-sided according to

$H_1 : p > p_0$, $\quad Z \geq Z_\alpha$

$H_1 : p < p_0$, $\quad Z \leq -Z_\alpha$

$H_1 : p \neq p_0$, $\quad |Z| \geq Z_{\alpha/2}$

Example: A coin is to be tested for fairness. It is tossed 25 times and only 8 heads are observed. Test if the coin is fair at α = 5%. Let p denote the probability of getting a head, which must be 0.5 for a fair coin.

H_0: $p = 0.5$

H_1 : $p \neq 0.5$

α = 5%; $Z_{0.05/2} = 1.96$

$\hat{p} = \frac{x}{n} = \frac{8}{25} = 0.32$

Test statistics:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.32 - 0.5}{\sqrt{0.5 \times 0.5/25}} = -1.8$$

From binomial distribution, with n = 25, X = 8, we can calculate the p-value:

P-value $= 2 \times P(X \leq 8) = 2 \times 0.054 = 0.108$

Rejection region: reject H_0 if p-value < α or $|Z| \geq Z_{\alpha/2}$

Statistical conclusion: Since p-value (0.108) > α (0.05) and/or $|Z||-1.8| \leq Z_{\alpha/2}(1.96)$, so we **fail to reject** $H_{0.}$

Testing Population Variance

To **test hypotheses about** σ, the test statistic is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Given a level of significance α,

$$\left.\begin{array}{r}\text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma < \sigma_0\end{array}\right\} \text{ if } \quad \frac{(n-1)S^2}{\sigma_0^2} \le \chi^2_{1-\alpha}$$

$$\left.\begin{array}{r}\text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma > \sigma_0\end{array}\right\} \text{ if } \quad \frac{(n-1)S^2}{\sigma_0^2} \ge \chi^2_{\alpha}$$

$$\left.\begin{array}{r}\text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma \ne \sigma_0\end{array}\right\} \begin{array}{l}\text{if } \quad \dfrac{(n-1)S^2}{\sigma_0^2} \le \chi^2_{1-\alpha/2} \\ \text{or } \quad \dfrac{(n-1)S^2}{\sigma_0^2} \ge \chi^2_{\alpha/2}\end{array}$$

Example: A manufacturer of golf balls claims that the company controls the weights of the golf balls accurately so that the variance of the weights is not more than 1 mg^2. A random sample of 31 golf balls yields a sample variance of 1.62 mg^2. Is that sufficient evidence to reject the claim at an α of 5%?

H_0: $\sigma^2 \le 1$
H_1 : $\sigma^2 > 1$

n = 31, S^2= 1.62
Test statistics:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(31-1)1.62}{1} = 48.6$$

Rejection region: reject H_0 if p-value < α or $\chi^2 \ge \chi^2_\alpha$
Statistical conclusion: Since $\chi^2(48.6) \ge \chi^2_\alpha(43.77)$, so we **reject** H_0

Latihan soal

H_0: orang tidak membawa senjata tersembunyi
H_1: orang membawa senjata tersembunyi

Kesalahan tipe I: orang terdeteksi membawa senjata padahal tidak
Kesalahan tipe II: orang terdeteksi tidak membawa senjata padahal membawa

7–1. Sebuah perusahaan farmasi mengklaim bahwa empat dari lima dokter meresepkan obat pereda nyeri yang dihasilkannya. Jika Anda ingin menguji klaim ini, bagaimana Anda menetapkan hipotesis nol dan hipotesis alternatif?

7–3. Ditemukan bahwa peselancar Web akan kehilangan minat pada halaman Web jika pengunduhan membutuhkan waktu lebih dari 12 detik dengan kecepatan baud 28K. Jika Anda ingin menguji keefektifan halaman Web yang baru dirancang sehubungan dengan waktu downloadnya, bagaimana Anda akan menyiapkan hipotesis nol dan hipotesis alternatif?

7–5. Selama kenaikan tajam harga bensin pada musim panas tahun 2006, perusahaan minyak mengklaim bahwa harga rata-rata bensin tanpa timbal dengan nilai oktan minimum 89 di Midwest tidak lebih dari $ 3,75. Jika Anda ingin menguji klaim ini, bagaimana Anda menetapkan hipotesis nol dan hipotesis alternatif?

7–9. Consider the use of metal detectors in airports to test people for concealed weapons. In essence, this is a form of hypothesis testing.

a. What are the null and alternative hypotheses?

b. What are type I and type II errors in this case?

c. Which type of error is more costly?

d. Based on your answer to part (*c*), what value of α would you recommend for this test?

e. If the sensitivity of the metal detector is increased, how would the probabilities of type I and type II errors be affected?

f. If α is to be increased, should the sensitivity of the metal detector be increased or decreased?

7–11. For each one of the following null hypotheses, determine if it is a left-tailed, a right-tailed, or a two-tailed test.

a. $\mu \geq 10$.

b. $p \leq 0.5$.

c. μ is at least 100.

d. $\mu \leq -20$.

e. p is exactly 0.22.

f. μ is at most 50.

g. $\sigma^2 = 140$.

7–12. The calculated z for a hypothesis test is -1.75. What is the p-value if the test is (*a*) left-tailed, (*b*) right-tailed, and (*c*) two-tailed?

Latihan soal

7–16. An automobile manufacturer substitutes a different engine in cars that were known to have an average miles-per-gallon rating of 31.5 on the highway. The manufacturer wants to test whether the new engine changes the miles-per-gallon rating of the automobile model. A random sample of 100 trial runs gives $\bar{x} = 29.8$ miles per gallon and $s = 6.6$ miles per gallon. Using the 0.05 level of significance, is the average miles-per-gallon rating on the highway for cars using the new engine different from the rating for cars using the old engine?

7–48. According to the *New York Times*, the Martha Stewart Living Omnimedia Company concentrates mostly on food.[10] An analyst wants to disprove a claim that 60% of the company's public statements have been related to food products in favor of a left-tailed alternative. A random sample of 60 public statements revealed that only 21 related to food. Conduct the test and provide a *p*-value.

7–67. At Armco's steel plant in Middletown, Ohio, statistical quality-control methods have been used very successfully in controlling slab width on continuous casting units. The company claims that a large reduction in the steel slab width variance resulted from the use of these methods. Suppose that the variance of steel slab widths is expected to be 156 (squared units). A test is carried out to determine whether the variance is above the required level, with the intention to take corrective action if it is concluded that the variance is greater than 156. A random sample of 25 slabs gives a sample variance of 175. Using $\alpha = 0.05$, should corrective action be taken?

Daftar Pustaka

- Aczel, A. and Sounderpandian, J., 2008, *Complete Business Statistics*, 7th *Edition*, McGraw-Hill/Irwin, USA.
- Richard, A.J. and Bhattacharyya, G.K., 2010, *Statistics: Principles and Methods*, 6th *Edition*, John Wiley and Sons, USA.
- https://youtu.be/k80pME7mWRM

Terimakasih

Tim Dosen Pengantar Statistika

UNIVERSITAS AIRLANGGA

Excellence with morality