9

Tests of Hypotheses for a Single Sample

LO

- Formulate the two hypotheses; identify critical values, test statistic and compute P-value for a test of hypotheses
  - on population mean.
  - population proportion.
  - population variance.
- Identify type of errors in testing hypotheses.

Introduction

A longtime authorized user of the account makes 0.2 seconds between keystrokes. One day, the following times between keystrokes were recorded when a user typed correctly username and password: 0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27.

Is he/she an authorized use of a computer account assuming Normal distribution of these times?

Statistical hypothesis

- Definition. A statistical hypothesis is a statement about the parameters of one or more populations.
- Example.

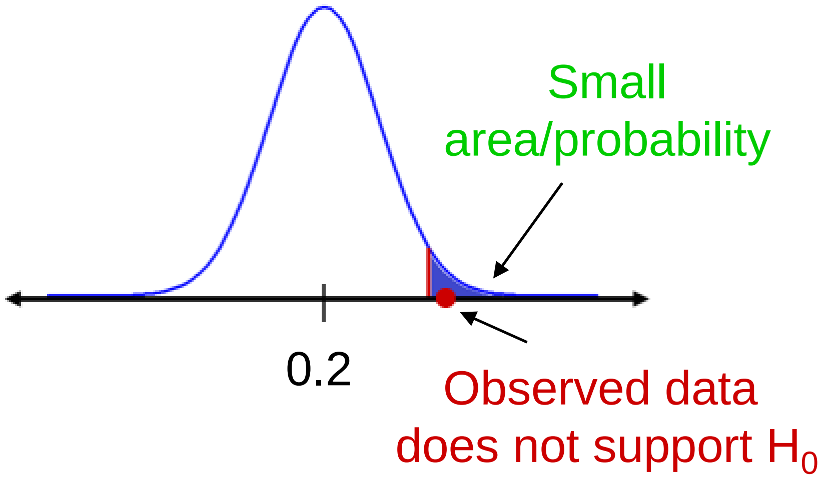
Null hypothesis H_0: $\mu = 0.2$ seconds (authorized use)

Alternative hypothesis H_1: $\mu \neq 0.2$ (unauthorized use)

Observed data: 0.24, 0.22, 0.26, 0.34, 0.35
0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15,
0.17, 0.28, 0.38, 0.40, 0.37, 0.27.

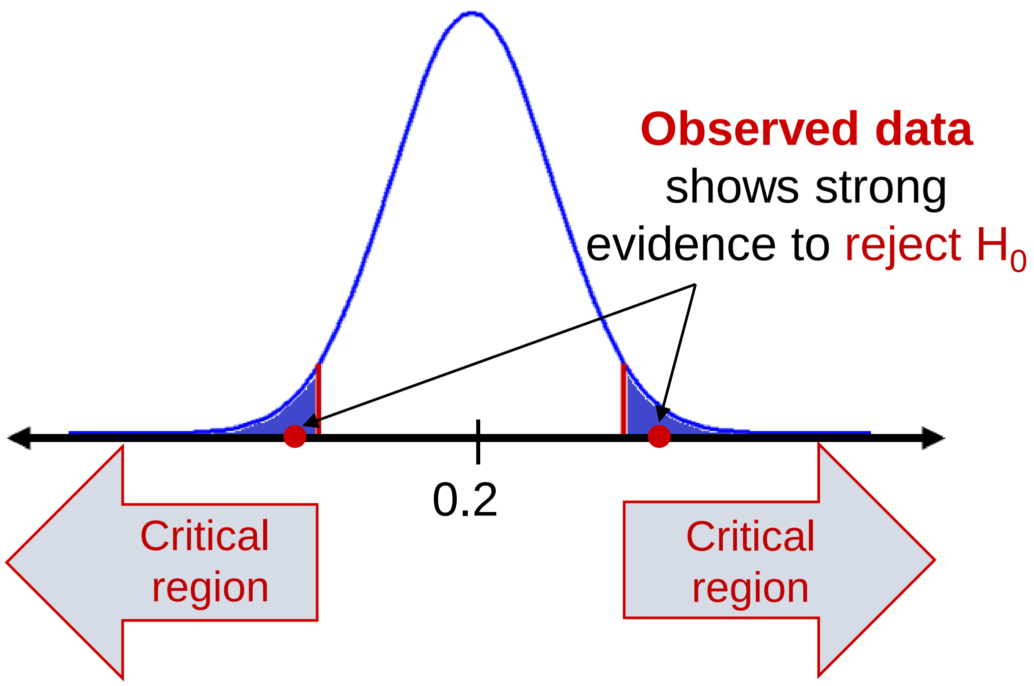
- $\bar{x}$ closes to 0.2 ➔ Does not conflict with H_0.
- $\bar{x}$ is considerably different from 0.2

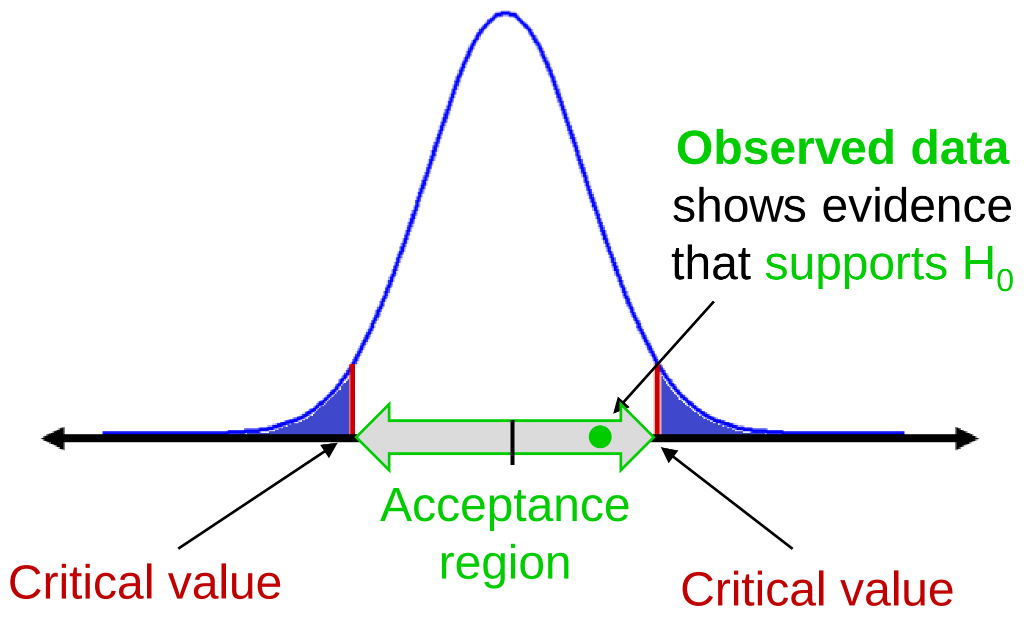
➔ An evidence in support of H_1.



H_0: $\mu = 0.2$

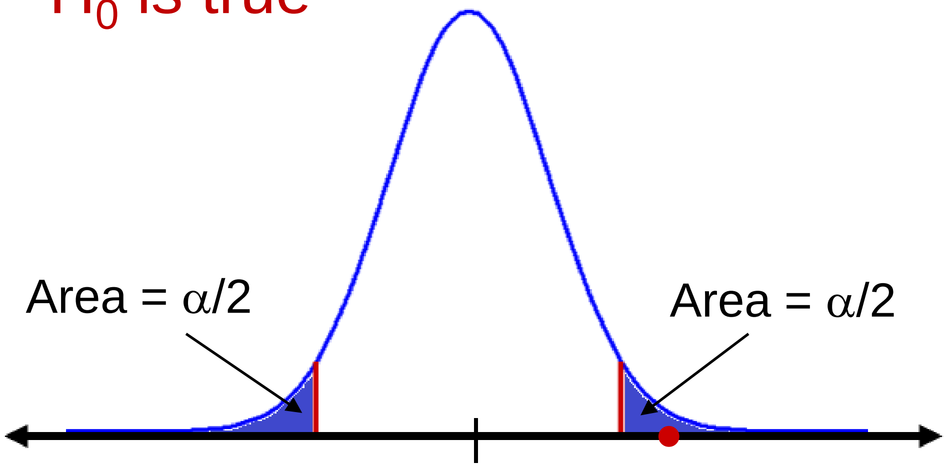
H_1: $\mu \neq 0.2$





Significance level α and Types of errors

H_0 is true



If observed data falls in the critical region (probability = $\alpha/2 + \alpha/2 = \alpha$)
➔ We reject H_0

We can make errors!

Type I error: Reject H_0 when H_0 is actually true
➔ $\alpha = 0.05$ is probability of type I error

α: significance level, or the α-error, or the size of the test (usual α is in [0.01, 0.1])

Note. Type II error: Fail to reject H_0 when H_0 is actually false.

Example

A longtime authorized user of the account makes 0.2 seconds between keystrokes (assuming Normal distribution of these times) and standard deviation σ = 0.07. One day, the following times between keystrokes were recorded when a user typed correctly username and password: 0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27.

➔ n = 18, the sample mean $\bar{x}$ = 0.29 sec.

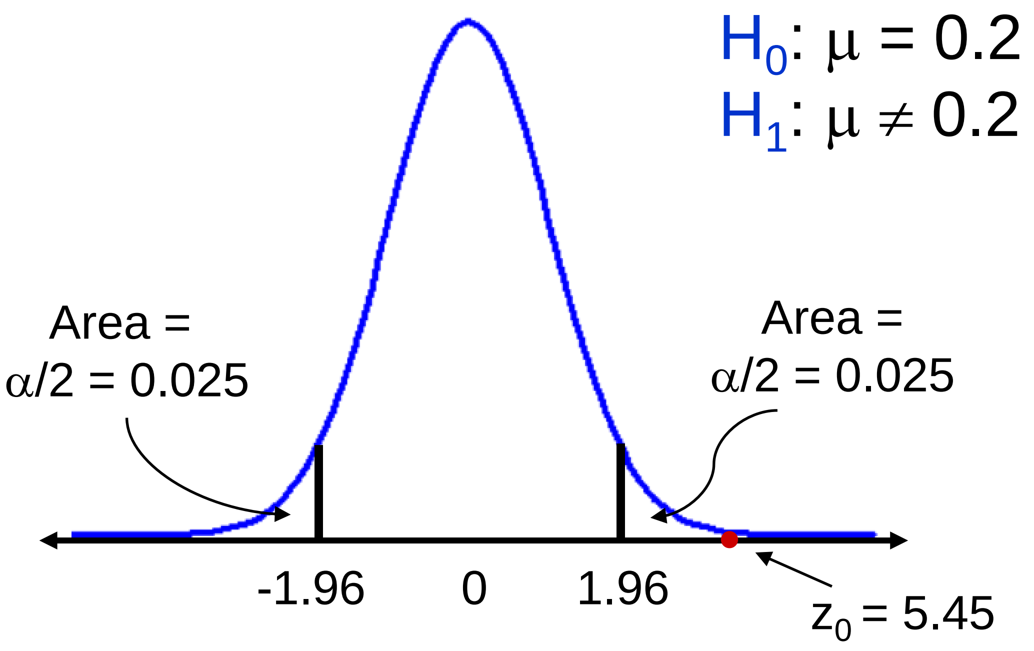
Suppose H_0 is true, that is μ = 0.2, based on the CLT,

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ➔ $z_0 = \frac{0.29-0.2}{0.07/\sqrt{18}} = 5.45$

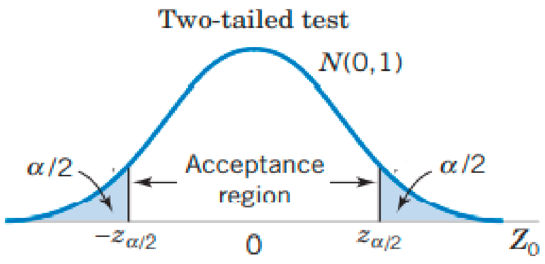
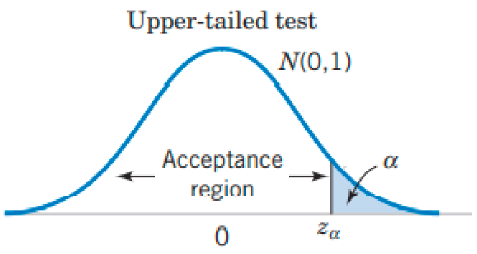
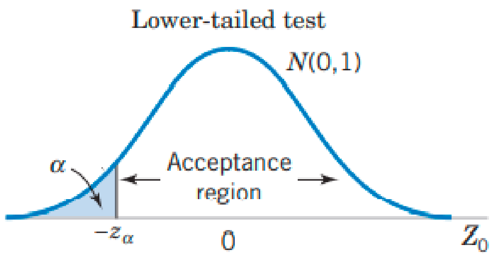
At a significance level of 0.05 (α = 0.05), $z_{\alpha/2} = z_{0.025} = 1.96$

Because $z_0 > z_{\alpha/2}$ ➔ We have strong evidence to reject H_0

Example (cont.)

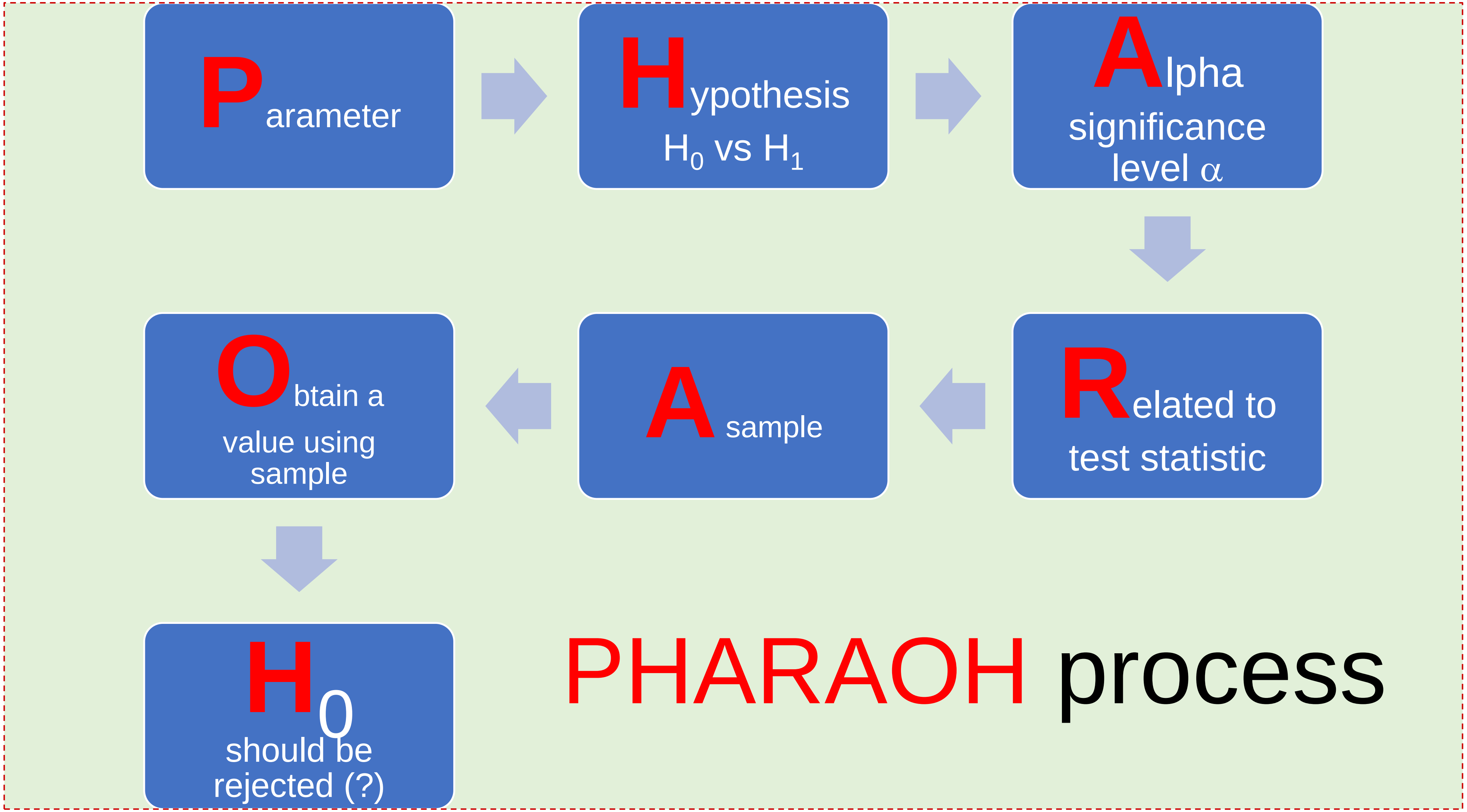


One-Sided and Two-Sided Tests

Two-tailed **Hypotheses**	One-tailed **Hypotheses**	
H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$
		

Example (one-tailed hypotheses). To verify that the average connection speed is 54 Mbps, we test the null hypothesis $H_0 : \mu = 54$ against the alternative hypothesis $H_1 : \mu < 54$, because we worry about a low connection speed.

Hypothesis Tests process



Example

The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800. After an equipment upgrade, the average number of users at 100 randomly selected moments of time is 5200. Does it indicate, at a 5% level of significance, that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.

Parameter: the population mean μ

Hypotheses: H_0: $\mu = 5000$, H_1: $\mu > 5000$

Alpha: $\alpha = 0.05$

Related to test statistic: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

A sample: $\bar{x} = 5200$, $n = 100$

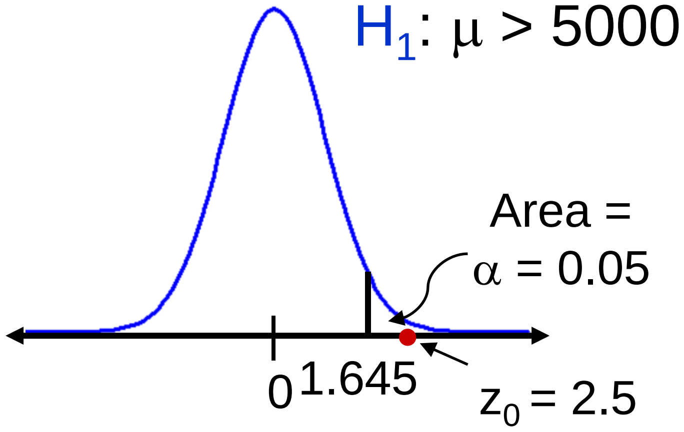
Obtain a value from sample data and test statistic:

$$z_0 = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{5200-5000}{800/\sqrt{100}} = 2.5$$

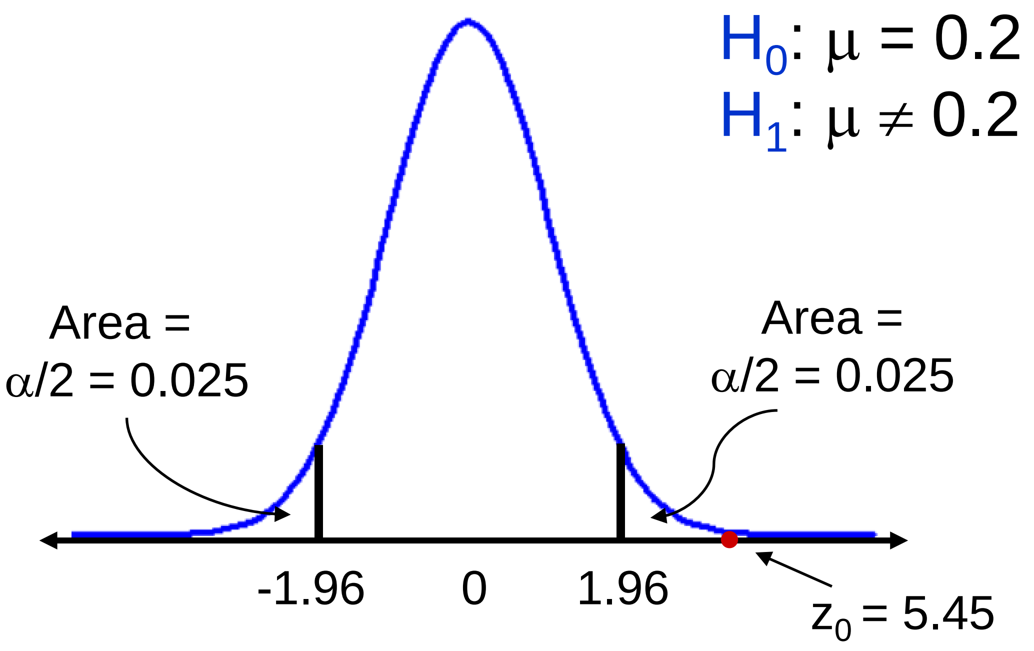
H_0 should be rejected since $z_0 = 2.5 > z_\alpha = 1.645$

H_0: $\mu = 5000$

H_1: $\mu > 5000$

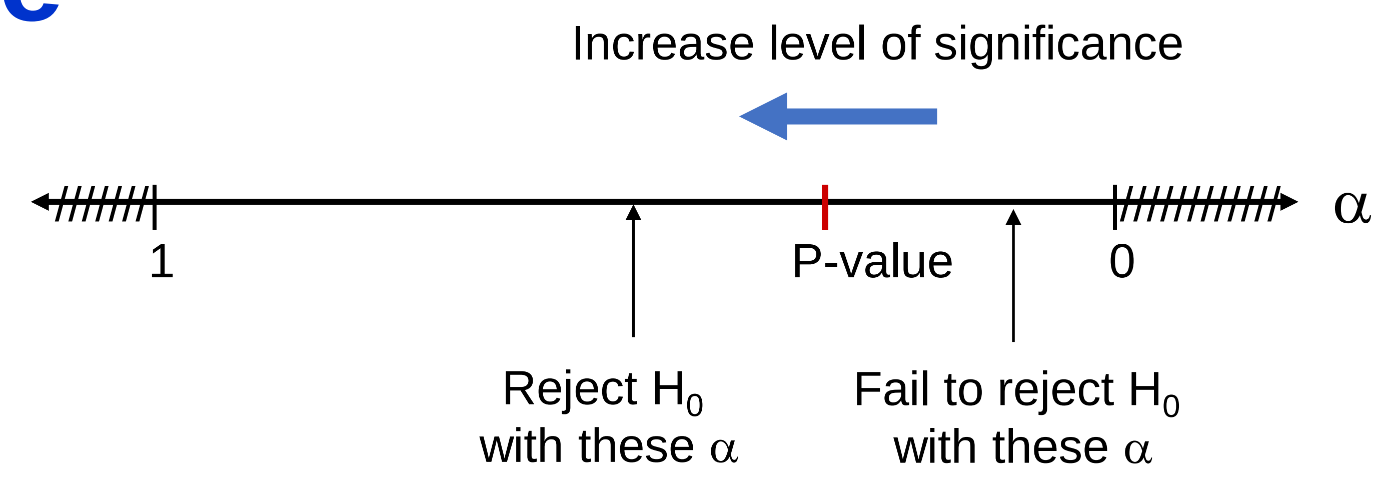


P-value



H_0: $\mu = 0.2$
H_1: $\mu \neq 0.2$
Area =
$\alpha/2 = 0.025$
Area =
$\alpha/2 = 0.025$
-1.96
0
1.96
$z_0 = 5.45$

P-value



Definition. The *P-value* (computed from observed data) is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

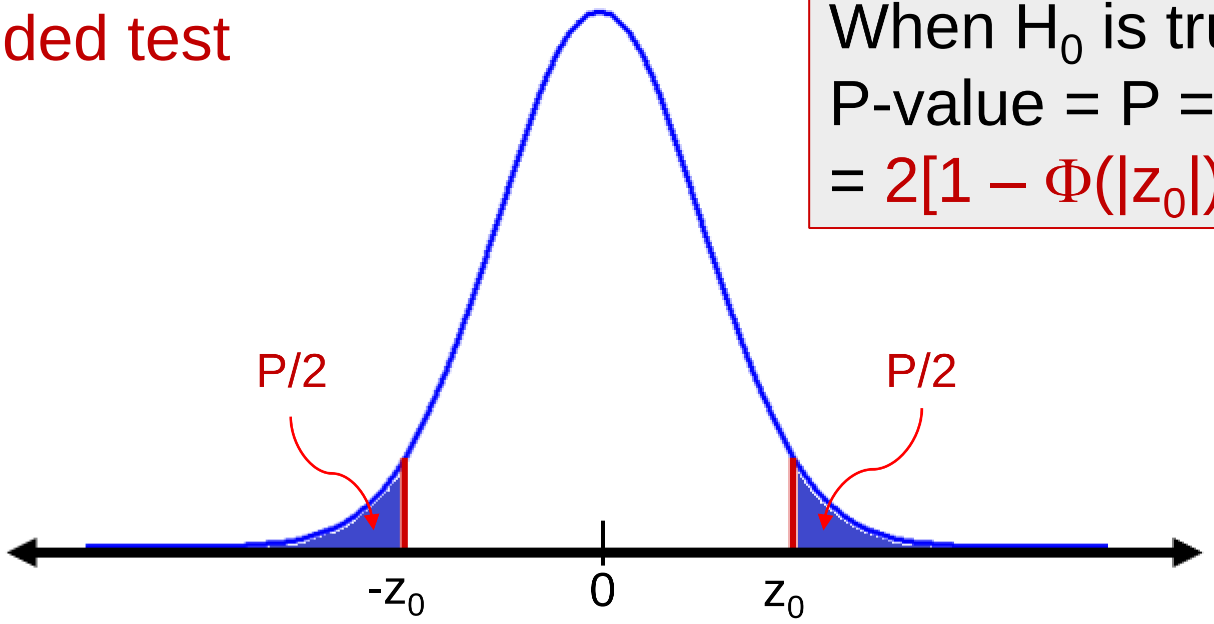
α < P-value > ➔ Fail to reject the null hypothesis H_0.

$\alpha \geqslant$ P-value ➔ H_0 should be rejected

(Usual significance levels $\alpha \in [0.01, 0.1]$)

Computing P-value

Two-sided test



When H_0 is true,
P-value = P = $P\{Z > z_0\} + P\{Z < -z_0\}$
$= 2[1 - \Phi(|z_0|)]$

P-value - Example

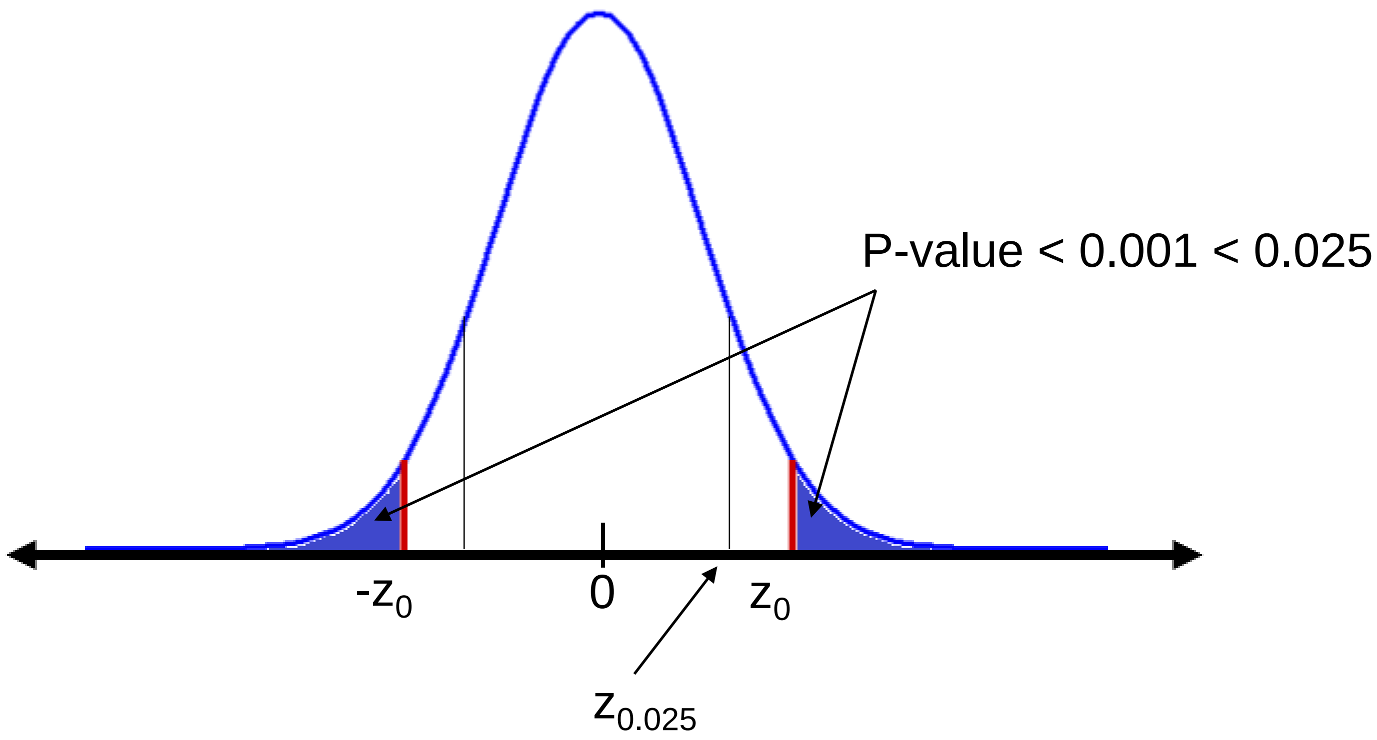
A longtime authorized user of the account makes 0.2 seconds between keystrokes (assuming Normal distribution of these times) and standard deviation of times between keystrokes is 0.07. One day, the following times between keystrokes were recorded when a user typed correctly username and password: 0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27.

➔ $n = 18$, the sample mean $\bar{x} = 0.29$ sec.

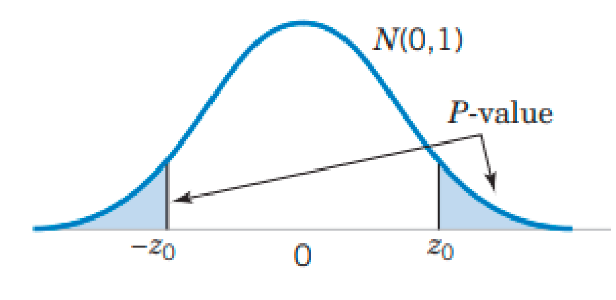
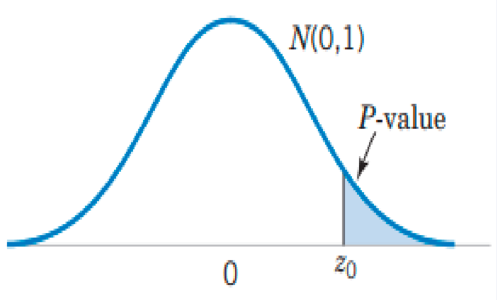
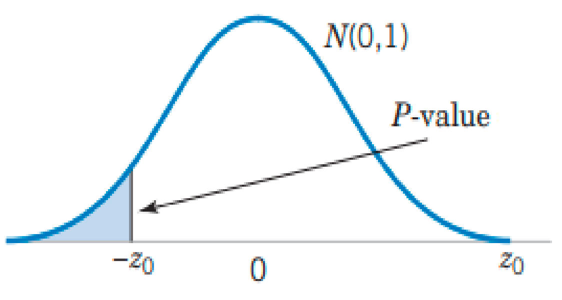
Suppose H_0 is true, that is $\mu = 0.2$, based on the CLT,

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ➔ $z_{obs} = \frac{0.29-0.2}{0.07/\sqrt{18}} = 5.45$

- The P-value is $P = P\{Z \geq z_{obs}\} + P\{Z < -z_{obs}\} = 2[1 - \Phi(4.013)] < 0.0001$
- Given $\alpha = 0.05$, for example, we reject H_0.


P-value < 0.001 < 0.025
$-z_0$
0
z_0
$z_{0.025}$

One-Sided and Two-Sided Tests

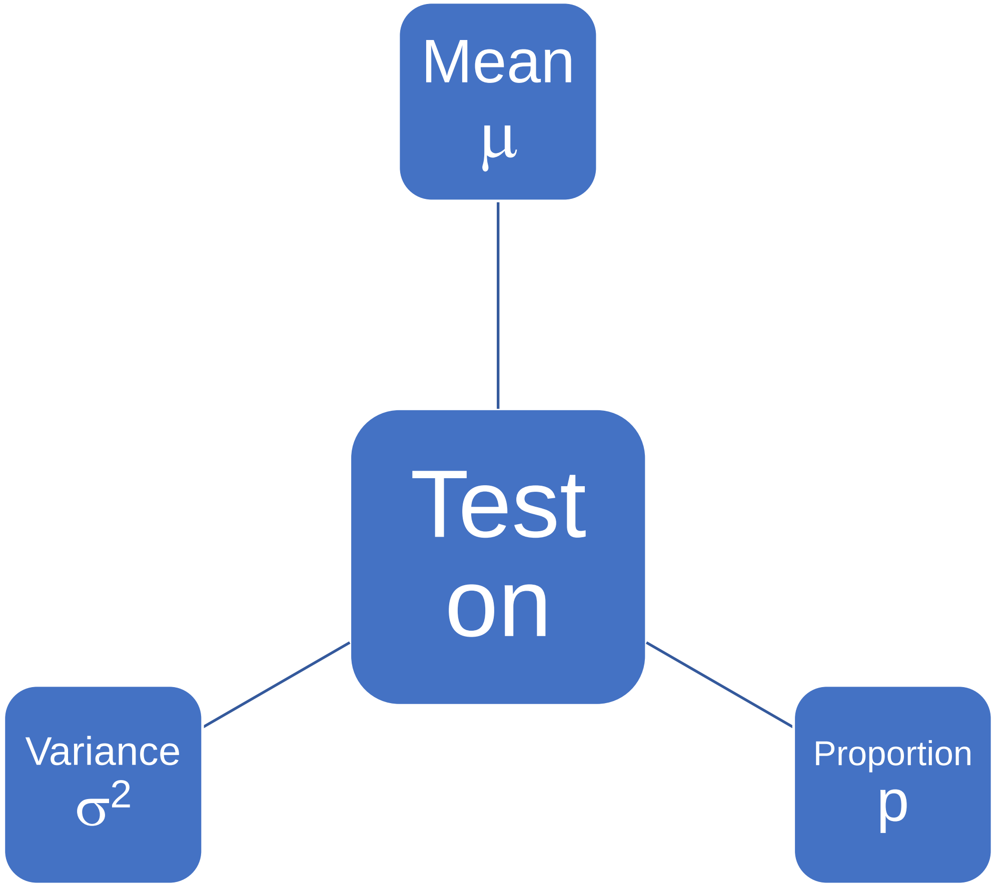
<table>
<tr><th>Two-sided Hypotheses</th><th colspan="2">One-sided Hypotheses</th></tr>
<tr><td>H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$</td><td>H_0: $\mu = \mu_0$
H_1: $\mu > \mu_0$</td><td>H_0: $\mu = \mu_0$
H_1: $\mu < \mu_0$</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

Hypothesis Tests - General Procedure

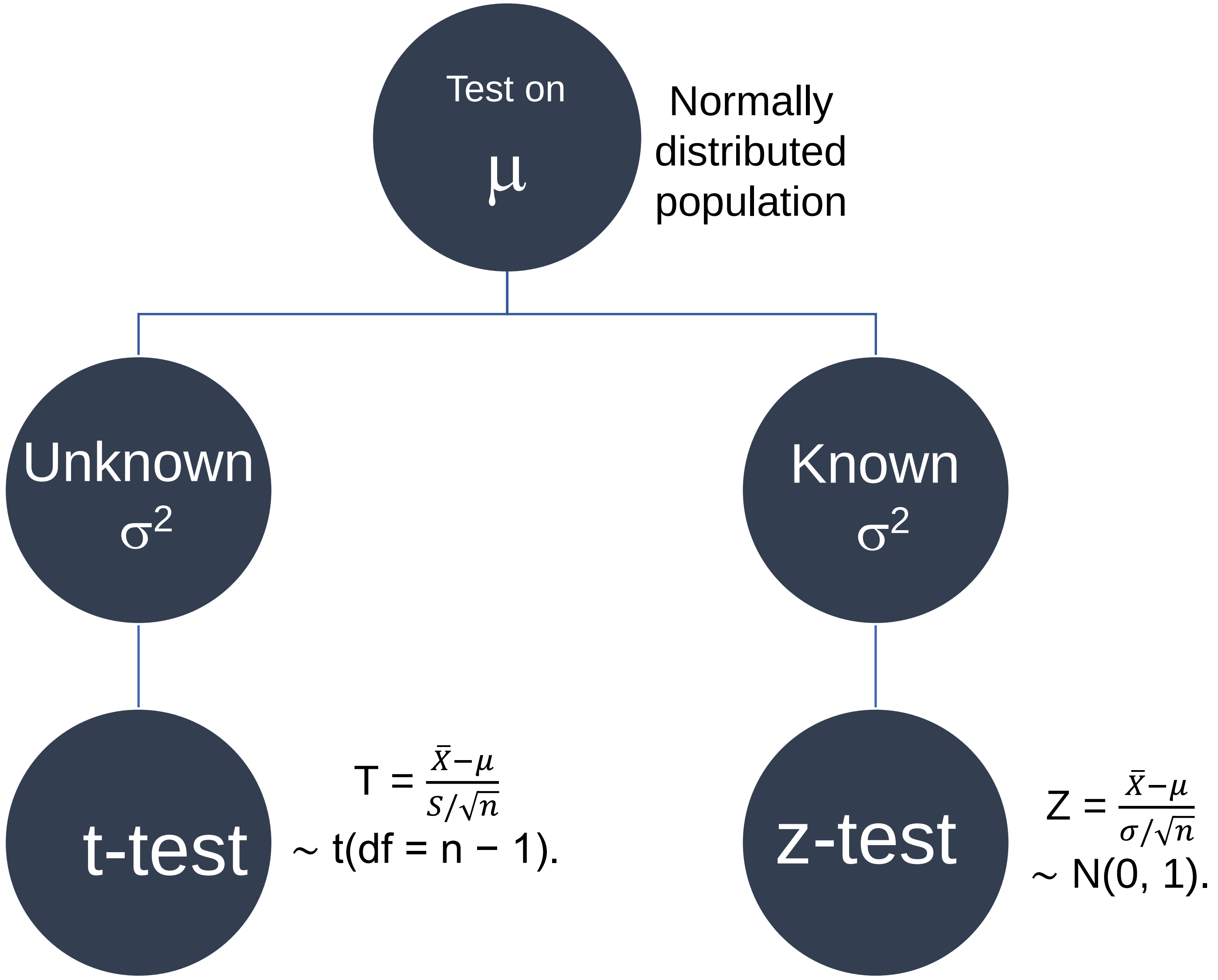
Parameter → **H_0** null hypothesis → **A**lternative hypothesis H_1 → **R**elated to test statistic → **A** sample quantities from sample → **O**btain P-value → **H_0** should be rejected (?)

PHARAOH process

Common Parameters in Hypothesis Testing



Hypothesis Tests on the Mean



Hypothesis Tests on the Mean (z-test)

Parameter: mean μ

Hypotheses: H_0: $\mu = \mu_0$

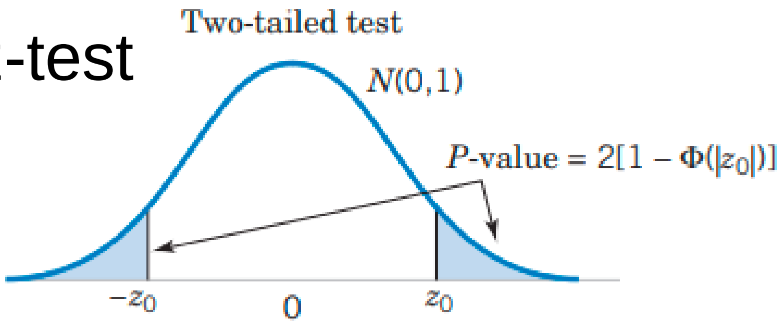
Alternative hypothesis: H_1: $\mu \neq \mu_0$

Relate to a test statistic (z-test)

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Any sample mean $\bar{x}$ for a computation using z-test

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

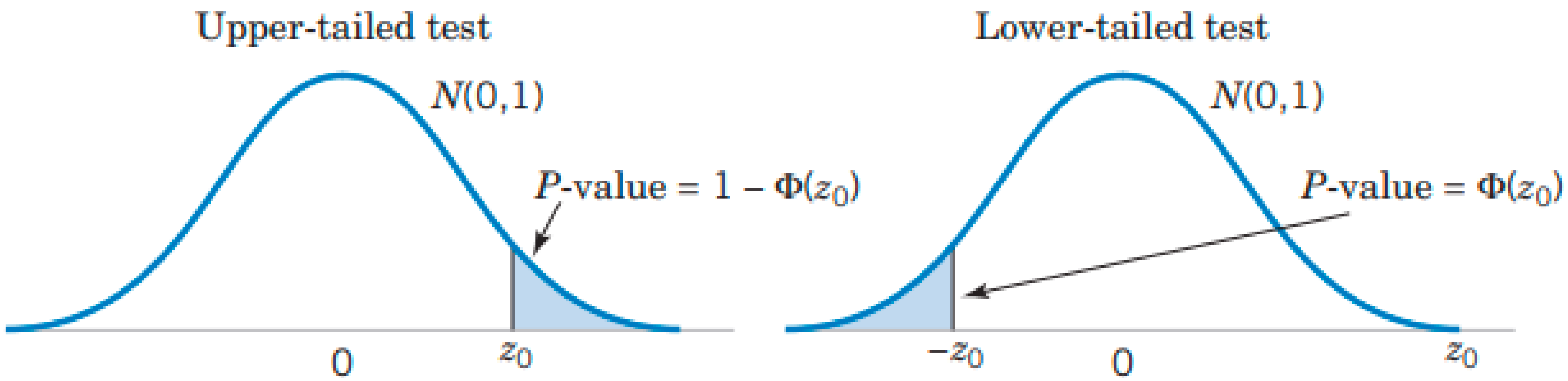


Obtain the P-value (two-tailed test):

$$P = 2[1 - \Phi(|z_0|)]$$

H_0 should be rejected (P-value < α) or not (P-value > α)

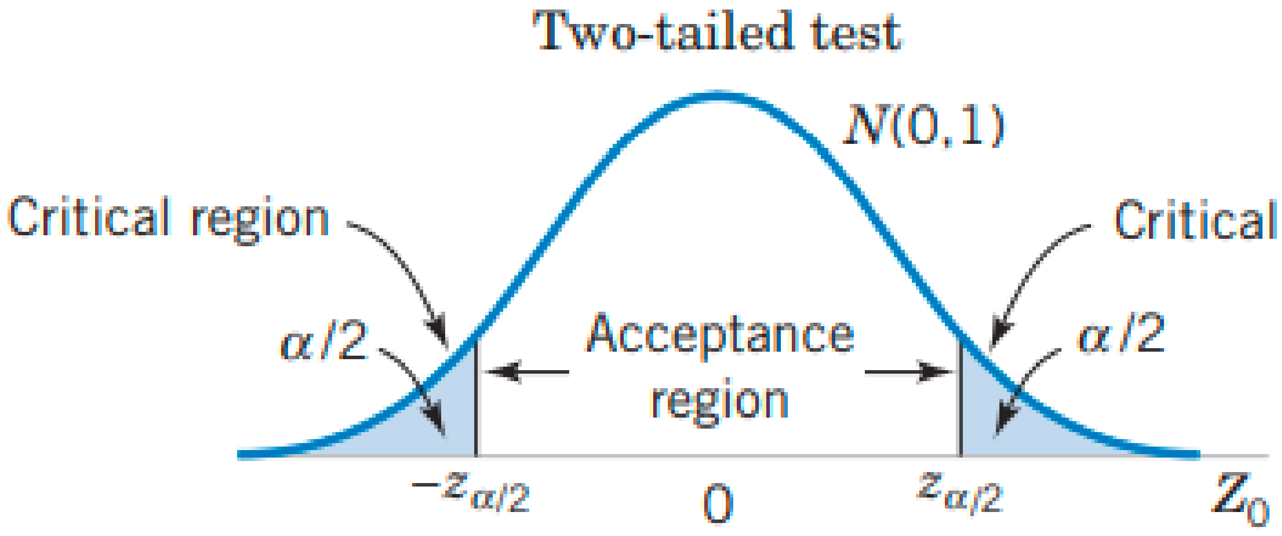
One-sided tests on the mean (z-test)



$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- Upper-tailed test: P-value = $1 - \Phi(z_0)$
- Lower-tailed test: P-value = $\Phi(z_0)$

z-test using significance level (two-sided test)



H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$

z-test: $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Fail to reject H_0 if

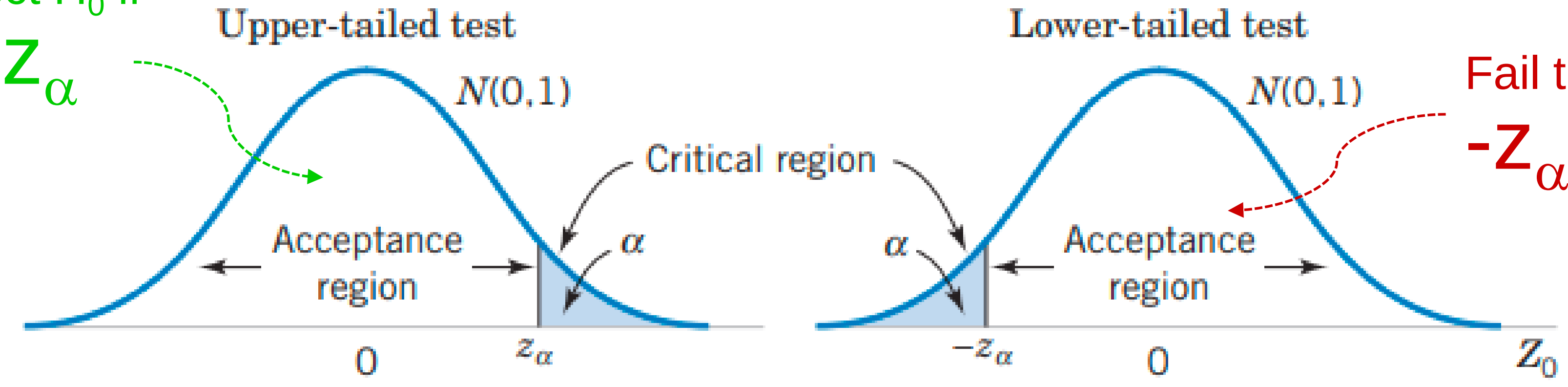
$$-Z_{\alpha/2} \leqslant Z_0 \leqslant Z_{\alpha/2}$$

z-test using significance level (one-sided tests)

$$\text{z-test: } Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Fail to reject H_0 if

$Z_0 \leqslant Z_\alpha$



Fail to reject H_0 if

$-Z_\alpha \leqslant Z_0$

$H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$

$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$

Example

Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $n = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

Example (cont.)

Parameter: The parameter of interest is μ, the mean burning rate

Hypotheses: H_0: $\mu = 50$

Alternative hypothesis: H_1: $\mu \neq 50$

Relate to a test statistic (z-test)

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Any sample mean $\bar{x}$ for a computation using z-test

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

Obtain the P-value (two-tailed test):

$$P = 2[1 - \Phi(3.25)] = 0.0012 < 0.05$$

H_0 should be rejected at the 0.05 level of significance

Practical Interpretation: We conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

Hypothesis Tests on the Mean (t-test)

Parameter: mean μ

Hypotheses: H_0: $\mu = \mu_0$

Alternative hypothesis: H_1: $\mu \neq \mu_0$

Relate to a test statistic (z-test)

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

A sample mean $\bar{x}$ and sample std s for a computation using t-test

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Obtain the P-value (two-tailed test):

$$P = \text{Probability above } |t_0| + \text{probability below } -|t_0|$$

H_0 should be rejected (P-value < α) or not (P-value > α)

T-test

Testing Hypotheses on the Mean of a Normal Distribution, Variance Unknown

Null hypothesis: $H_0\!: \mu = \mu_0$

Test statistic: $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

Alternative Hypotheses	***P*-Value**	**Rejection Criterion for Fixed-Level Tests**
$H_1\!: \mu \neq \mu_0$	Probability above $\lvert t_0 \rvert$ and probability below $-\lvert t_0 \rvert$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1\!: \mu > \mu_0$	Probability above t_0	$t_0 > t_{\alpha,n-1}$
$H_1\!: \mu < \mu_0$	Probability below t_0	$t_0 < -t_{\alpha,n-1}$

The calculations of the *P*-values and the locations of the critical regions for these situations are shown in Figs. 9-10 and 9-12, respectively.

t-test (unknown σ^2 case) - Example

A longtime authorized user of the account makes 0.2 seconds between keystrokes (assuming Normal distribution of these times). One day, the following times between keystrokes were recorded when a user typed correctly username and password: 0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27.

➔ n = 18, the sample mean $\bar{x}$ = 0.29 sec, s = 0.074.

Test statistic T = $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

➔ $t_0 = \frac{0.29-0.2}{0.074/\sqrt{18}} = 5.16$

➔P-value = $P(T > t_0) + P(T < t_0) =$

Note that $t_{0.025} =$

χ^2-test on σ^2

Test on σ^2

χ^2 - test

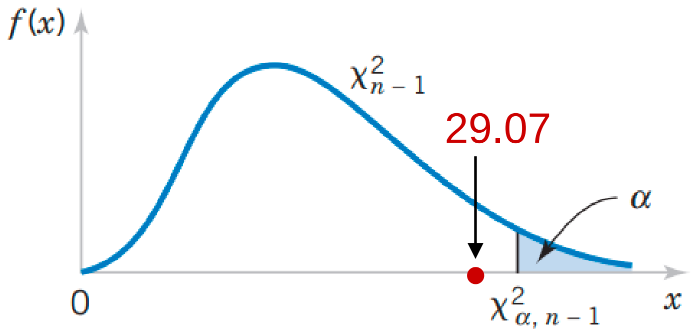
Null hypothesis: $H_0\colon \sigma^2 = \sigma_0^2$

Test statistic: $$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

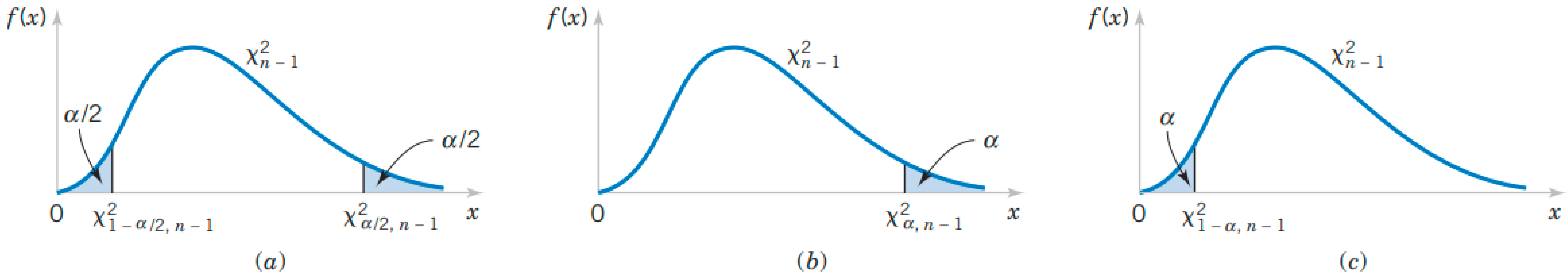
Alternative hypothesis	Rejection criteria
$H_1\colon \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or $\chi_0^2 < -\chi^2_{\alpha/2,n-1}$
$H_1\colon \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha,n-1}$
$H_1\colon \sigma^2 < \sigma_0^2$	$\chi_0^2 < -\chi^2_{\alpha,n-1}$

χ^2-test - example

- An automated filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.
- Parameter: σ^2
- Hypotheses H_0: $\sigma^2 = 0.01$, H_1: $\sigma^2 > 0.01$
- Alpha: $\alpha = 0.05$
- Related to a test statistic: $\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$
- A sample: $n = 20$, $s^2 = 0.0153$
- Obtain a value from sample and test statistic: $\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{19(0.0153)}{0.01} = 29.07$
- $\chi_0^2 = 29.07 < \chi^2_{0.05,19} = 30.14$ ➔ Fail to reject H_0.



χ^2-test



(a) H_1: $\sigma^2 \neq \sigma_0^2$ (b) H_1: $\sigma^2 > \sigma_0^2$ (c) H_1: $\sigma^2 < \sigma_0^2$

Example

Creators of a new software claim that it measures the remaining notebook battery life with a standard deviation as low as 5 minutes. To test this claim, a fully charged notebook was disconnected from the power supply and continued on its battery. The experiment was repeated 50 times, and every time the predicted battery life was recorded. The sample standard deviation of these 50 normally distributed measurements was equal 5.6 minutes.

At the 1% level of significance, do these data provide evidence that the actual standard deviation is greater than 5 min?

Test $H_0 : \sigma = 5$ against a right-tail $H_1 : \sigma > 5$.

Tests on a Proportion

Test on proportion p

z - test

- We wish to test:
- H_0: $p = p_0$
- H_1: $p \neq p_0$
- X: X be the number of observations in a random sample of size n that belongs to the class associated with p

➔ $X \sim N[np, np(1 - p)]$,

Suppose H0 is true, we use the test statistic:

$$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

Summary of Approximate Tests on a Binomial Proportion

Testing Hypotheses on a Binomial Proportion

Null hypotheses: $H_0\colon p = p_0$

Test statistic: $$Z_0 = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1\colon p \neq p_0$	Probability above $\lvert z_0 \rvert$ and probability below $-\lvert z_0 \rvert$ $P = 2[1 - \Phi(z_0)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1\colon p > p_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1\colon p < p_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

Example

A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?

- Parameter: proportion p
- Hypothesis H_0: $p = 0.05$,
- Alternative hypothesis H_1: $p < 0.05$
- Related to a test statistic $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$
- A sample: x = 4, n = 200
- Obtain the P-value computed from sample and test statistic: $z_0 = = \frac{4 - 200(0.05)}{\sqrt{200(0.05)(1-0.05)}} = -1.95$

And P-value = $\Phi(-1.95) = 0.0256 < 0.05$

➔ Reject H_0.

Summary

PHARAOH process
Parameter
Hypotheses: H_0 vs H_1
Related to a test statistic
A random sample
Obtain the P-value
H_0: Reject or fail to reject