# CENG 222
## Statistical Methods for Computer Engineering

## Week 10

Chapter 9

9.4 Hypothesis Testing

# Testing Hypotheses

- Hypothesis $H_0$ and the alternative $H_A$ are two mutually exclusive statements about some unknown parameter $\theta$.
- Testing steps:
  - Collect data
  - Compute a test statistic
  - State if there is sufficient evidence to reject $H_0$ in favor of $H_A$
- Examples: 9.22, 9.23, and 9.24

# Type I and Type II errors and level of significance

| | Result of the test | |
|---|---|---|
| | Reject $H_0$ | Accept $H_0$ |
| $H_0$ is true | *Type I error* | correct |
| $H_0$ is false | correct | *Type II error* |

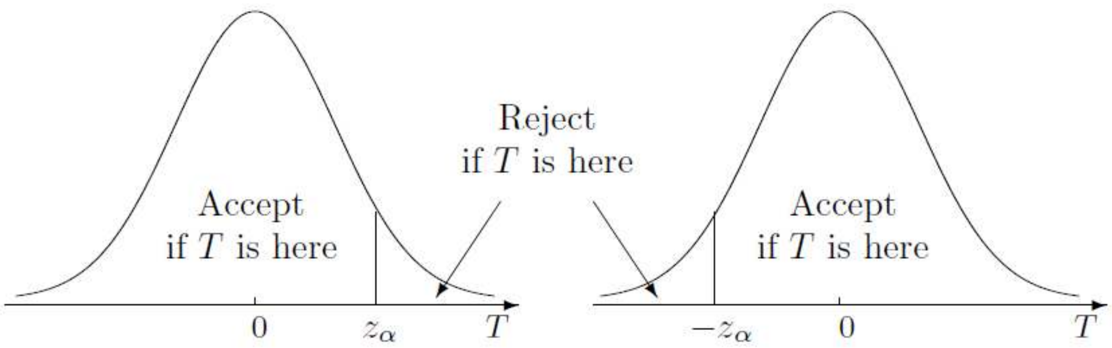
- We generally want to limit Type I errors (also called false positives) while minimizing Type II errors (also called false negatives)
- Significance level is the probability to observe Type I errors.

# Type I and Type II errors and level of significance

- We generally want to limit Type I errors (also called false positives) while minimizing Type II errors (also called false negatives)
- $\alpha = \boldsymbol{P}(\text{reject } H_0 | H_0 \text{ is true})$
- Power of the test is the probability to avoid Type II error
  - Power = sensitivity = recall = True Positive Rate
  - $p(\theta) = \boldsymbol{P}(\text{reject } H_0 | \theta; H_A \text{ is true})$
- See:
  - https://en.wikipedia.org/wiki/Sensitivity_and_specificity

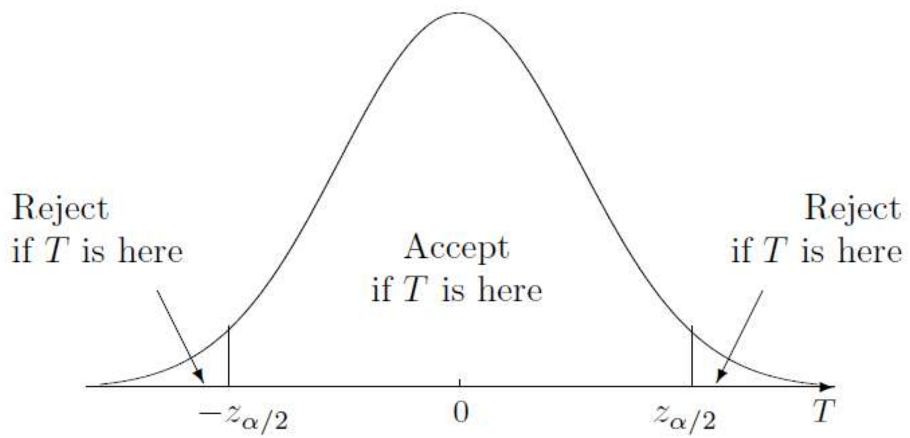
# Types of alternatives

- Two sided, one-sided left tail, and one-sided right tail alternatives




*(a) Right-tail Z-test*

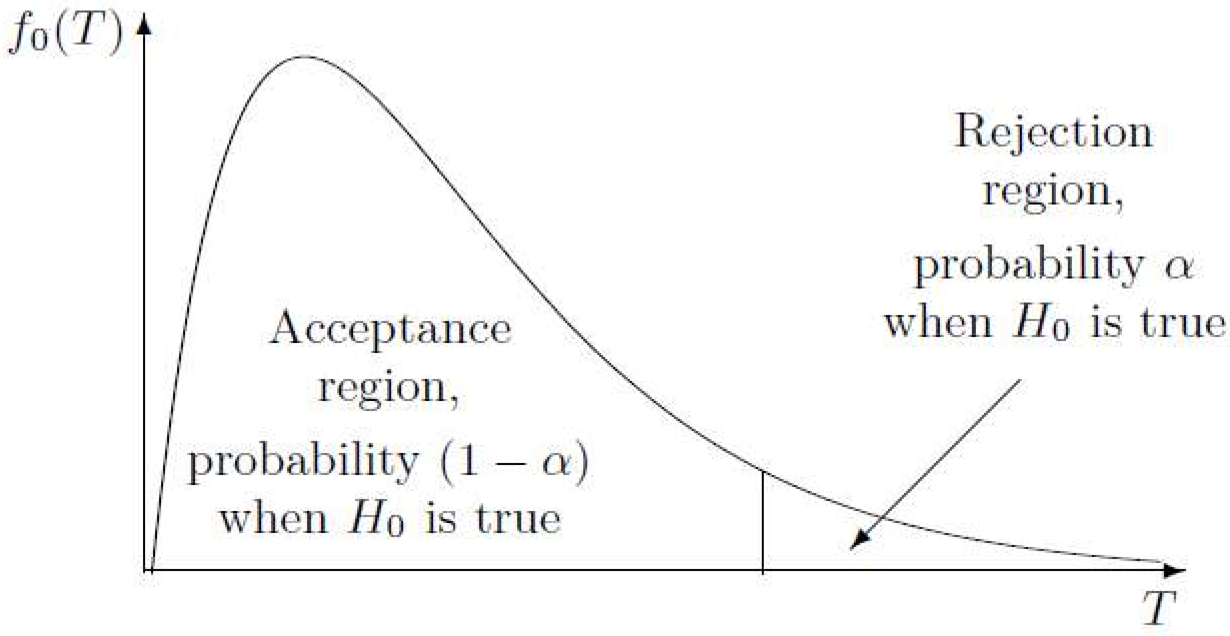
*(b) Left-tail Z-test*




*(c) Two-sided Z-test*

# Null distribution and acceptance/rejection regions

- The distribution of the test statistic $T$ is called the null distribution.



# Z-test

- If the null distribution of the test statistic is Standard Normal, the tests are called Z-tests.
- Z-tests are used when we know population variance.
- T-tests are used for unknown population variance
- Tests can be performed for one sample, two samples (e.g., when comparing two populations)
- Common hypotheses are about population means, proportions, and differences.

# Two-tail Z-test

- Data: $X_1, \dots, X_n$ from Normal$(\mu, \sigma)$ with unknown $\mu$ and known $\sigma$
- Test $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$
  1. Find $\pm z_{\alpha/2}$. Acceptance region is $[-z_{\alpha/2}, z_{\alpha/2}]$
  2. Compute the test statistic
  $$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
  3. If $Z$ belongs to the acceptance region do not reject $H_0$
  Otherwise, reject $H_0$

# One-sided right tail Z-test

- Data: $X_1, \ldots, X_n$ from Normal$(\mu, \sigma)$ with unknown $\mu$ and known $\sigma$
- Test $H_0 : \mu = \mu_0$ versus $H_A : \mu > \mu_0$
  1. Find $z_\alpha$. Acceptance region is $(-\infty, z_\alpha]$
  2. Compute the test statistic

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

  3. If $Z$ belongs to the acceptance region do not reject $H_0$
     Otherwise, reject $H_0$

# One-sided left tail Z-test

- Data: $X_1, \dots, X_n$ from Normal$(\mu, \sigma)$ with unknown $\mu$ and known $\sigma$
- Test $H_0 : \mu = \mu_0$ versus $H_A : \mu < \mu_0$
  1. Find $z_\alpha$. Acceptance region is $[-z_\alpha, +\infty)$
  2. Compute the test statistic
  $$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
  3. If $Z$ belongs to the acceptance region do not reject $H_0$
  Otherwise, reject $H_0$

# Unknown variance: T-test

- We use the estimator for variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- and use the $t$-distribution with $n-1$ degrees of freedom

# Z-test summary

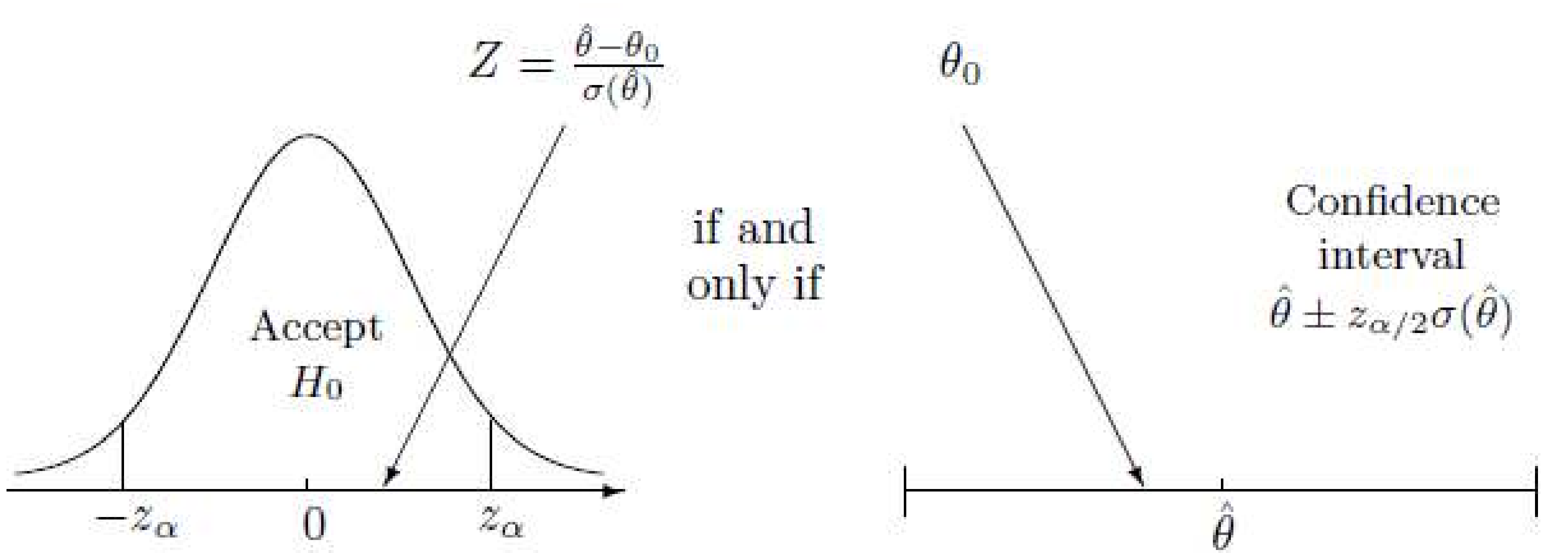
| Null hypothesis $H_0$ | Parameter, estimator $\theta,\ \hat\theta$ | If $H_0$ is true: $\mathbf{E}(\hat\theta)$ | If $H_0$ is true: $\mathrm{Var}(\hat\theta)$ | Test statistic $Z = \dfrac{\hat\theta - \theta_0}{\sqrt{\mathrm{Var}(\hat\theta)}}$ |
|---|---|---|---|---|
| One-sample Z-tests for means and proportions, based on a sample of size $n$ | | | | |
| $\mu = \mu_0$ | $\mu,\ \bar X$ | $\mu_0$ | $\dfrac{\sigma^2}{n}$ | $\dfrac{\bar X - \mu_0}{\sigma/\sqrt{n}}$ |
| $p = p_0$ | $p,\ \hat p$ | $p_0$ | $\dfrac{p_0(1-p_0)}{n}$ | $\dfrac{\hat p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ |
| Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size $n$ and $m$ | | | | |
| $\mu_X - \mu_Y = D$ | $\mu_X - \mu_Y$, $\bar X - \bar Y$ | $D$ | $\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}$ | $\dfrac{\bar X - \bar Y - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ |
| $p_1 - p_2 = D$ | $p_1 - p_2$, $\hat p_1 - \hat p_2$ | $D$ | $\dfrac{p_1(1-p_1)}{n} + \dfrac{p_2(1-p_2)}{m}$ | $\dfrac{\hat p_1 - \hat p_2 - D}{\sqrt{\frac{\hat p_1(1-\hat p_1)}{n} + \frac{\hat p_2(1-\hat p_2)}{m}}}$ |
| $p_1 = p_2$ | $p_1 - p_2$, $\hat p_1 - \hat p_2$ | $0$ | $p(1-p)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)$, where $p = p_1 = p_2$ | $\dfrac{\hat p_1 - \hat p_2}{\sqrt{\hat p(1-\hat p)\left(\frac{1}{n} + \frac{1}{m}\right)}}$ where $\hat p = \dfrac{n\hat p_1 + m\hat p_2}{n+m}$ |

# T-test summary

| Hypothesis $H_0$ | Conditions | Test statistic $t$ | Degrees of freedom |
|---|---|---|---|
| $\mu = \mu_0$ | Sample size $n$; unknown $\sigma$ | $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$ | $n - 1$ |
| $\mu_X - \mu_Y = D$ | Sample sizes $n$, $m$; unknown but equal standard deviations, $\sigma_X = \sigma_Y$ | $t = \dfrac{\bar{X} - \bar{Y} - D}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$ | $n + m - 2$ |
| $\mu_X - \mu_Y = D$ | Sample sizes $n$, $m$; unknown, unequal standard deviations, $\sigma_X \neq \sigma_Y$ | $t = \dfrac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$ | Satterthwaite approximation, formula (9.12) |

# Confidence intervals versus two-sided level-$\alpha$ tests

- A level-$\alpha$ two sided test is the same as testing whether a given test statistic is in the $(1 - \alpha)100\%$ confidence interval $[a,b]$




- Examples: 9.31, 9.35 (one-sided test)

# P-value

- Instead of a fixed significance level $\alpha$, we can compute the boundary level for acceptance/rejection for the computed test statistic
  - This computed value is called the p-value
  - It is the probability that another sample results in a more extreme test static.