

Week 5

Introduction to Statistical Tests of Significance

Dr. Esam Mahdi

Department of Mathematics, Statistics and Physics
Qatar University

Learning Objectives

- Identify the type I and type II errors in statistical context.
- Test the hypothesis about the population mean (one and two sided for small and large samples).
- Test the hypothesis about the population proportion (one and two sided for large samples).
- Decide to reject or not reject the null hypothesis using any of the following three methods:
 - Compare the observed test statistics with the cut off point critical one (tabular that can be founded from normal or t tables).
 - Use the Confidence interval around the population parameter.
 - Use the P value (when we have large samples).

Introduction

- Suppose that a pharmaceutical company is concerned that the mean potency μ of an antibiotic meet the minimum government potency standards. They need to decide between two possibilities:
 - *The mean potency μ does not exceed the mean allowable potency.*
 - *The mean potency μ exceeds the mean allowable potency.*
- *This is an example of a test of hypothesis.*
- In general, when a health investigator seeks to understand or explain something, for example the effect of a drug, he/she usually formulates his/her research question in the form of a hypothesis.
- **In the statistical context, a hypothesis is a statement about:**
 - The distribution (e.g., The distribution is normal).
 - The underlying parameter of the distribution (e.g., $\mu = 10$).
 - The relationship between probability distributions (e.g., there is no statistical relationship).
 - The parameters of two or more distributions (e.g., $\mu_1 = \mu_2$).

Criminal court: Jury Decisions

- In criminal court, the accused is “presumed innocent” until “proved guilty beyond all reasonable doubt.”

Jury Decision	Truth	
	Accused is innocent	Accused is Guilty
Guilty	Error	Ok
“Innocent”	Ok	Error



- If the accused is innocent, but the judge decide that he/she is guilty then an error was made (*Type I error*).
- If the accused is guilty, but the judge decide that he/she is innocent then an error was made (*Type II error*).

Hypothesis Tests in Statistical Context

- The hypothesis to be tested is called the **null hypothesis** and will be denoted by H_0 .
 - H_0 assumed to be true until we can prove otherwise.
 - H_0 is usually stated in the null form, indicating no difference or no relationship between distributions or parameters.
- An **alternative hypothesis**, which we denote by H_a is a hypothesis that in some sense contradicts the null hypothesis H_0 .
 - H_a will be accepted as true if we can disprove H_0 .

Type I and Type II errors

Decision	Truth	
	Null hypothesis (H_0) is correct	Null hypothesis (H_0) is false; Equivalently, alternative hypothesis (H_a) is correct
Reject the null hypothesis	Type I error = α (level of significance)	Correct decision ($1 - \beta$) = Power
Accept (fail to reject) the null hypothesis	Correct decision ($1 - \alpha$) = Level of confidence	Type II error = β

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{Type II error}) = P(\text{accept } H_0 \text{ when } H_0 \text{ is false})$$

- There is always a chance of making one of these errors. We'll want to minimize the chance of doing so.
- Minimize α will increase β and vice versa. What we have to do in practice
- In practice, we fix α at some level — say $\alpha = 0.05$ or 0.01 — and β is controlled through the use of sample size where the power is maximized.
- **Power** of a test is the probability of rejecting the null hypothesis when it should be rejected:

$$\text{Power} = 1 - P(\text{Type II error})$$

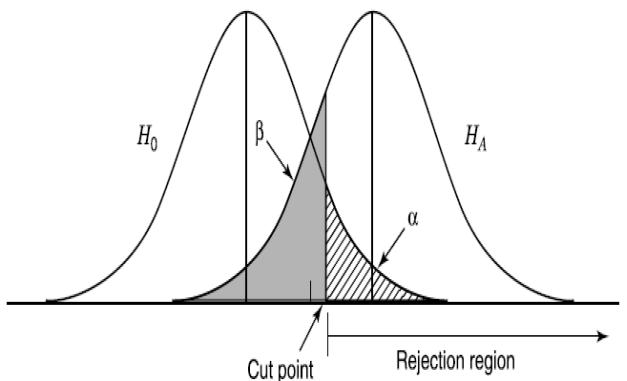


Figure 5.4 Graphical display of a one-sided test.

Example 5.1

Suppose that the national smoking rate among men is 25% and we want to study the smoking rate among men in the New England states.

- Let π be the proportion of New England men who smoke.
- The null hypothesis that the smoking prevalence in New England is the same as the national rate is
$$H_0: \pi = 0.25$$
- Suppose that we plan to take a sample of size $n = 100$ and use this decision making rule:

$$P \leq 0.20, H_0 \text{ is rejected};$$

where P is the proportion obtained from the sample.

Question 1: Find the type I error.

From the central limit theorem we have

$$z = \frac{P - \mu_p}{\sqrt{\sigma_p^2}} \sim N(0,1),$$

$$\text{where } \mu_p = \pi \text{ and } \sigma_p^2 = \frac{\pi(1-\pi)}{n}$$

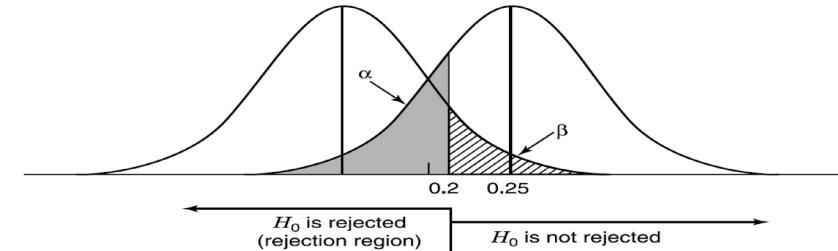


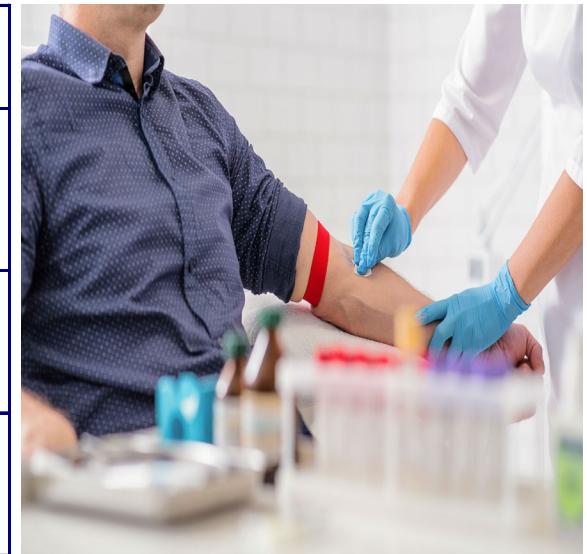
Figure 5.1 Graphical display of type I and type II errors.

Question 2: Suppose that the truth is $H_a: \pi = 0.15$, find the type II error.

Medical Screening Tests: Type I & Type II errors

- Sensitivity = $\frac{\text{number of diseased persons who screen positive}}{\text{total number of diseased person}}$
- Specificity = $\frac{\text{number of healthy persons who screen negative}}{\text{total number of healthy persons}}$

Screening test	Reality	
	- (Healthy person)	+
+	$\alpha = 1 - \text{specificity}$	Sensitivity
-	Specificity	$\beta = 1 - \text{sensitivity}$



Steps for Testing Hypothesis

- Formulate a **null hypothesis** and an **alternative hypothesis**. The statement resulting from the research question forms the alternative hypothesis H_a .
- Fix the **significance level, α** , at some level, the maximum tolerable risk you want to have of making a mistake when H_0 is true but you decide to reject it.
- Select a random sample from the population ($n \geq 25$ is large).
- Compute the sample statistic:
- Determine the rejection region:
 - A rule that tells you for which values of the test statistic, the null hypothesis should be rejected.
- Compare the statistic with the parameter in the null hypothesis.
 - This is called $\text{Test statistic} = \frac{\text{statistic} - \text{hypothesized null value}}{\text{standard error of statistic}}$.
 - The null hypothesis is rejected if there is sufficiently strong evidence from the data to support its alternative. That is; we reject H_0 if the test statistic is located in the rejection region.

One-Sided versus Two-Sided Tests

- There are three types of hypothesis tests:
 - *Left sided test, right sided test, and two-sided test.*
- A one-sided test is indicated for research questions like these:
 - Is a new drug superior to a standard drug?
 - Does the air pollution exceed safe limits?
 - Has the death rate been reduced for those who quit smoking?
- A two-sided test is indicated for research questions like these:
 - Is there a difference between the cholesterol levels of men and women?
 - Does the mean age of a target population differ from that of the general population?

Hypothesis Tests about the Population Mean

- When the population is normally distributed and σ is known (regardless of the sample size) we use the [z-test statistic](#)

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

- In this case, the cut off point (critical tabular value $= \mp Z_\alpha$ or $\mp Z_{\alpha/2}$) will be found from the normal curve associated with the significance level α . (Note that the value μ_0 is the assumed null value of the mean under H_0)
- For any population when the sample size n is large (at least 25) we use the [z-test statistic](#)

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \text{ or } z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \text{ (as } s \text{ can be substituted for } \sigma \text{ for } n \geq 25)$$

- In this case, the cut off point (critical tabular value $= \mp Z_\alpha$ or $\mp Z_{\alpha/2}$) will be found from the normal curve associated with the significance level α .
- When the population is normal and σ is unknown, where the sample size n is small (less than 25), we will use [t-test statistic](#)

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

- In this case, the cut off point (critical tabular value $= \mp t_\alpha$ or $\mp t_{\alpha/2}$) will be found from the t-distribution curve associated with the significance level α and degrees of freedom $n - 1$.

Hypothesis Tests about the Population Mean

- One-sided (right tailed) test

$$H_0: \mu = \mu_0 \text{ versus } H_a: \mu > \mu_0$$

Reject H_0 at the significance of level α if the Z.test $\geq Z_\alpha$ (or t.test $\geq t_{\alpha,n-1}$)

- One-sided (left tailed) test

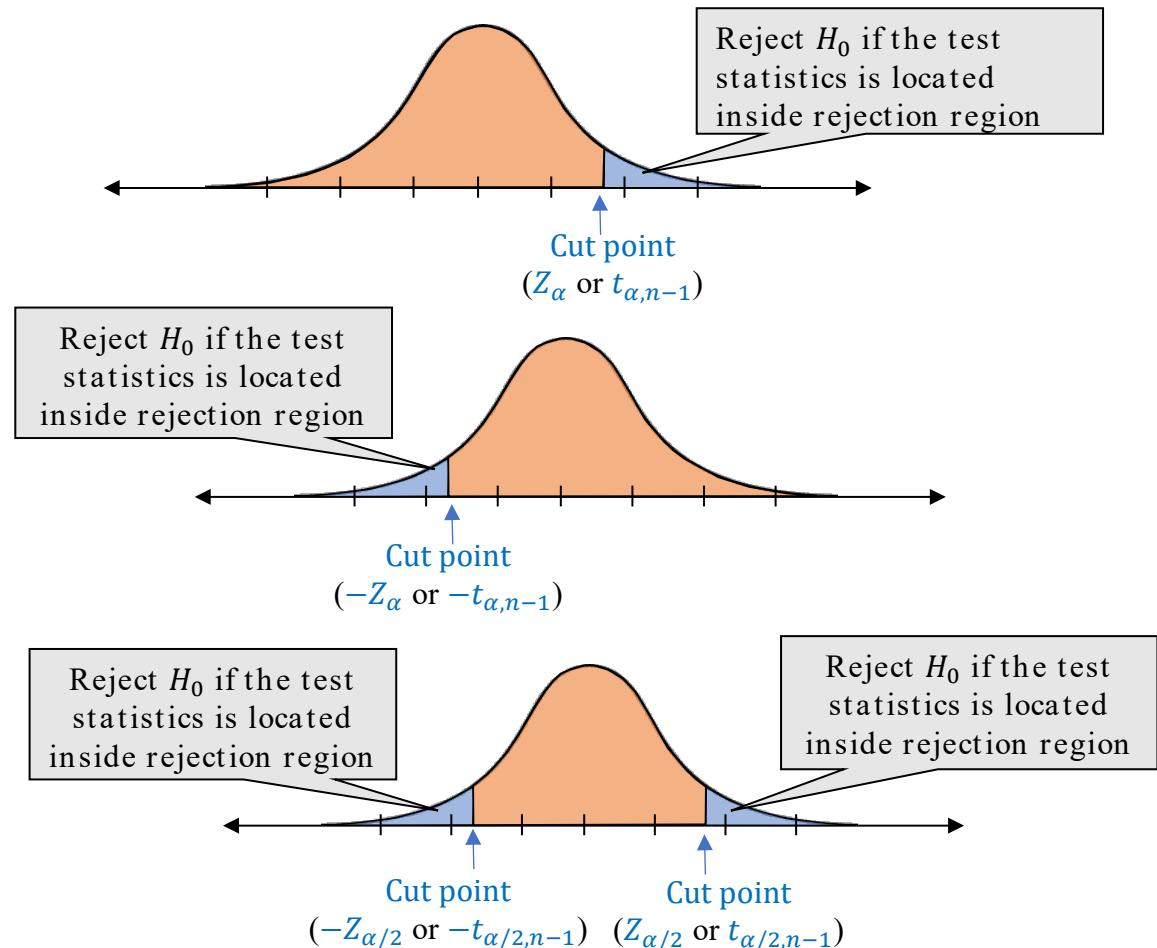
$$H_0: \mu = \mu_0 \text{ versus } H_a: \mu < \mu_0$$

Reject H_0 at the significance of level α if the Z.test $\leq -Z_\alpha$ (or t.test $\leq -t_{\alpha,n-1}$)

- Two-sided (two tailed) test

$$H_0: \mu = \mu_0 \text{ versus } H_a: \mu \neq \mu_0$$

Reject H_0 at the significance of level α if the Z.test $\geq Z_{\alpha/2}$ or if Z.test $\leq -Z_{\alpha/2}$ (or t.test $\geq t_{\alpha/2,n-1}$ or t.test $\leq -t_{\alpha/2,n-1}$)



Hypothesis Tests about the Population Proportion

- A random sample of size $n \geq 25$ from a binomial population to test

$H_0: \pi = \pi_0$ versus $H_a:$ one of three alternatives

- Test statistic derived from the central limit theorem is

$$z \approx \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

- For one-sided (right tailed) test, we *reject H_0 at the significance of level α if*

$$Z_{\text{test}} \geq Z_\alpha$$

- For one-sided (left tailed) test, we *reject H_0 at the significance of level α if*

$$Z_{\text{test}} \leq -Z_\alpha$$

- For two-sided (two tailed) test, we *reject H_0 at the significance of level α if*

$$Z_{\text{test}} \geq Z_{\frac{\alpha}{2}} \text{ or if } Z_{\text{test}} \leq -Z_{\alpha/2}$$

Example 5.2

- Suppose that the national smoking rate among men is 25% and we want to study the smoking rate among men in the New England states. The null hypothesis under investigation is

$$H_0: \pi = 0.25$$

- Of $n = 100$ males sampled, $x = 15$ were found to be smokers. Does the proportion π of smokers in New England states differ from that in the nation?

Solution:

Example 6.1

Example 6.1 A group of investigators wish to explore the relationship between the use of hair dyes and the development of breast cancer in women. A sample of $n = 1000$ female beauticians 40–49 years of age is identified and followed for five years. After five years, $x = 20$ new cases of breast cancer have occurred. It is known that breast cancer incidence over this time period for average American women in this age group is $\pi_0 = 7/1000$. We wish to test the hypothesis that using hair dyes *increases* the risk of breast cancer (a one-sided alternative). We have:



1. A one-sided test with

$$H_A: \pi > \frac{7}{1000}$$

2. Using the conventional choice of $\alpha = 0.05$ leads to the rejection region $z > 1.65$.
3. From the data,

$$\begin{aligned} p &= \frac{20}{1000} \\ &= 0.02 \end{aligned}$$

leading to a “ z score” of:

$$\begin{aligned} z &= \frac{0.02 - 0.007}{\sqrt{(0.007)(0.993)/1000}} \\ &= 4.93 \end{aligned}$$

- (i.e., the observed proportion p is 4.93 standard errors away from the hypothesized value of $\pi_0 = 0.007$).
4. Since the computed z score falls into the rejection region ($4.93 > 1.65$), the null hypothesis is rejected at the 0.05 level chosen. In fact, the difference is very highly significant ($p < 0.001$).

Confidence Intervals and Hypothesis Tests

Suppose that we consider a hypothesis of the form

$$H_0: \mu = \mu_0$$

where μ_0 is a known hypothesized value. A two-sided hypothesis test for H_0 is related to confidence intervals as follows:

1. If μ_0 is not included in the 95% confidence interval for μ , H_0 should be rejected at the 0.05 level. This is represented graphically as shown in Figure 5.7.
2. If μ_0 is included in the 95% confidence interval for μ , H_0 should not be rejected at the 0.05 level (Figure 5.8).

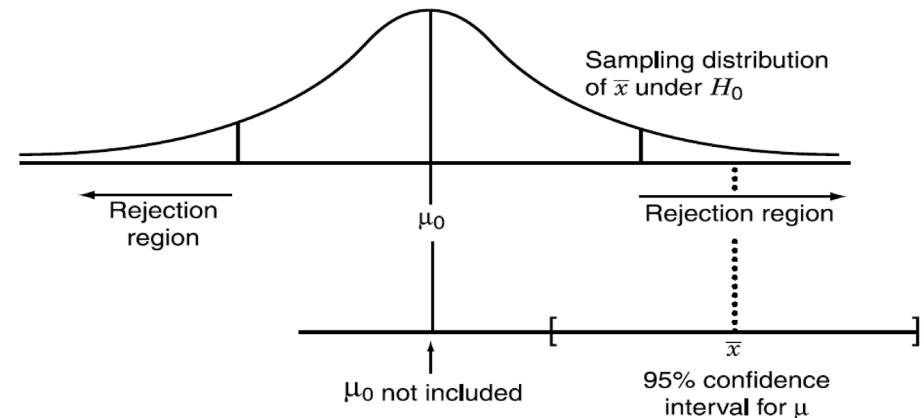


Figure 5.7 μ_0 not included at 95% confidence interval for μ .

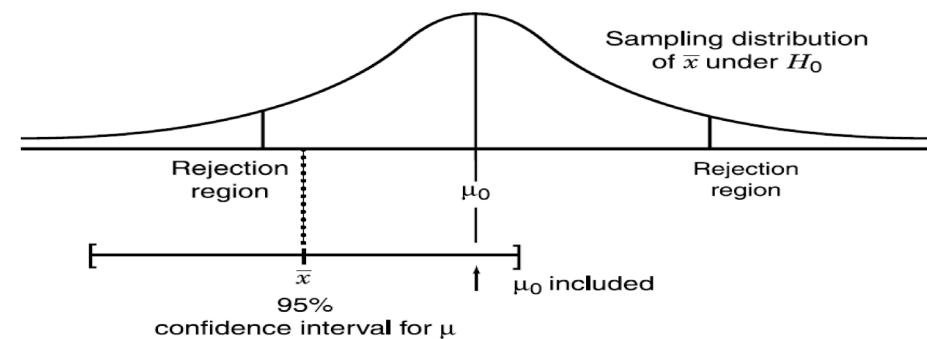


Figure 5.8 μ_0 included in 95% confidence interval for μ .

Example 5.3

- Consider the hypothetical data set in Example 5.2. Our point estimate of smoking prevalence in New England is

$$P = \frac{15}{100} = 0.15$$

Its standard error is

$$SE(P) = \sqrt{\frac{0.15(1 - 0.15)}{100}} = 0.036$$

- Therefore, a 95% confidence interval for the New England states smoking rate π is given by

$$0.15 \mp 1.96 \times 0.036 = (0.079, 0.221)$$

- It is noted that the national rate of 0.25 is not included in that confidence interval. Thus the null hypothesis is rejected at the 5% significant level.

P Values and Hypothesis Tests

- The **p value** (or *probability value*) is the probability of getting values of the test statistic as extreme as, or more extreme than, that observed if the null hypothesis is true.
 1. For a right-tailed test, P-value = (Area in right tail) = $P(Z \geq z_{\text{test}})$.
 2. For a left-tailed test, P-value = (Area in left tail) = $P(Z \leq z_{\text{test}})$.
 3. For a two-tailed test, P-value = 2(Area in tail of test statistic) = $2P(Z \geq |z_{\text{test}}|)$.
- **Compare the P-value with α :**
 - If P-value $\leq \alpha$, then reject H_0 .
 - If P-value $> \alpha$, then fail to reject H_0 .
- **Interpreting a decision based on P-value as in Table 5.2 below:**

TABLE 5.2

<i>p</i> Value	Interpretation
$p > 0.10$	Result is not significant
$0.05 < p < 0.10$	Result is marginally significant
$0.01 < p < 0.05$	Result is significant
$p < 0.01$	Result is highly significant

Example 5.2 (Using p-value)

- Suppose that the national smoking rate among men is 25% and we want to study the smoking rate among men in the New England states. The null hypothesis under investigation is

$$H_0: \pi = 0.25$$

- Of $n = 100$ males sampled, $x = 15$ were found to be smokers. Does the proportion π of smokers in New England states *differ* from that in the nation? **Use p-value!**

Solution: