Hypothesis Testing

Munmun Akter

Lecturer

NFE, DIU

Objectives

- Able to formulate statistical hypothesis
- Discuss between two types of error in hypothesis testing
- Able to decide accepting or rejecting a statistical hypothesis at a specific level of significance
- Choose appropriate tests statistics for a particular set of data.

Symbols

1) H_o – *Null Hypothesis*
2) H_1 – *Alternative Hypothesis*
3) β – *Greek Letter Beta which is the probability of committing a Type 2 Error*
4) α – *Greek letter Alpha which denotes a probability of committing a Type 1 Error and is known as the Level of Significance*
5) z
6) σ – *Greek letter Sigma which means the Variance*
7) σ_x - *the standard deviation of the sampling distribution of the mean*
8) μ - *Greek letter 'mu' which is the mean of the normal population*
9) n – *Sample size*
10) $\bar{x}$ - *Sample mean*
11) t – *t distribution; a case where the population standard deviation is unknown*
12) s – *standard deviation*

Definition

- Hypothesis is defined as a **statement about one or more populations**.
- The hypothesis is frequently **concerned** with the **parameters** of the **population** about which statement is made

- **Example**: A hospital administrator may hypothesize that the average length of stay of patients admitted to the hospital is five days.

Types of Hypotheses

- Researchers are concerned with **two types** of hypotheses
- 1**. Research hypothesis**: It is the conjecture or supposition that motives the research.
- Research hypothesis lead to directly **statistical hypothesis**.

- 2. **Statistical hypothesis**: this are stated in such a way that they may be **evaluated** by appropriate **statistical techniques**.

Types of statistical hypothesis

- **Two (2) types**:

1. **Null hypothesis** (H_0): The hypothesis **to be tested** (**State hypothesis value** before **sampling**).
2. **Alternative hypothesis** (H_A/H_1): Negation of null hypothesis.

Example:

H0: There is no significant difference in CKD prevalence between male and female ($\mu_M = \mu_{F)}$.

H1: There is significant difference in CKD prevalence between male and female ($\mu_M \neq \mu_{F)}$.

Types of Errors

- **Two (2) types**

1. **Type 1 Error:** It refers when we reject null hypothesis while it is true (Wrongly reject H_0).

E.g: H_0: There is no difference between the effect of two new establish drugs.

Type 1 error occurs if we conclude, there is significance difference between the effect of two drugs when there isn't actually no differences.

Prob (Type 1 error) = Significance level = α

2. Type 2 Error: It refers when we accept null hypothesis while it is false.

E.g: H_0: There is no significance difference in CKD prevalence between male and female.

Type 2 error occurs if we conclude that the prevalence of CKD in male and female is equal when there is actually differences.

Prob (Type 2 error) = β

Types of Errors

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct

Elements of Hypothesis testing

- Null hypothesis
- Alternative hypothesis
- Test Statistic: A **sample** statistic **used** to decide whether **to reject the null hypothesis.**
- Rejection region
- Calculation of test statistic
- Conclusion (Numerical values falls in the rejection region or not).

Level of significance (α)

- To specify **the probability of committing a Type 1 error** which is known **as level of significance**.

- We can determine the Critical values which define

✓**Region of Rejection (Critical region)**

✓**Region of Acceptance**

- The value of α = 5% or 0.05 means the region of rejection is 0.05 and the acceptance region is 0.95.

One-Tailed and Two Tailed Tests

- Where H_1 is Directional, ***One-Tailed Test***
- Where H_1 is Non-Directional, ***Two-Tailed Test***

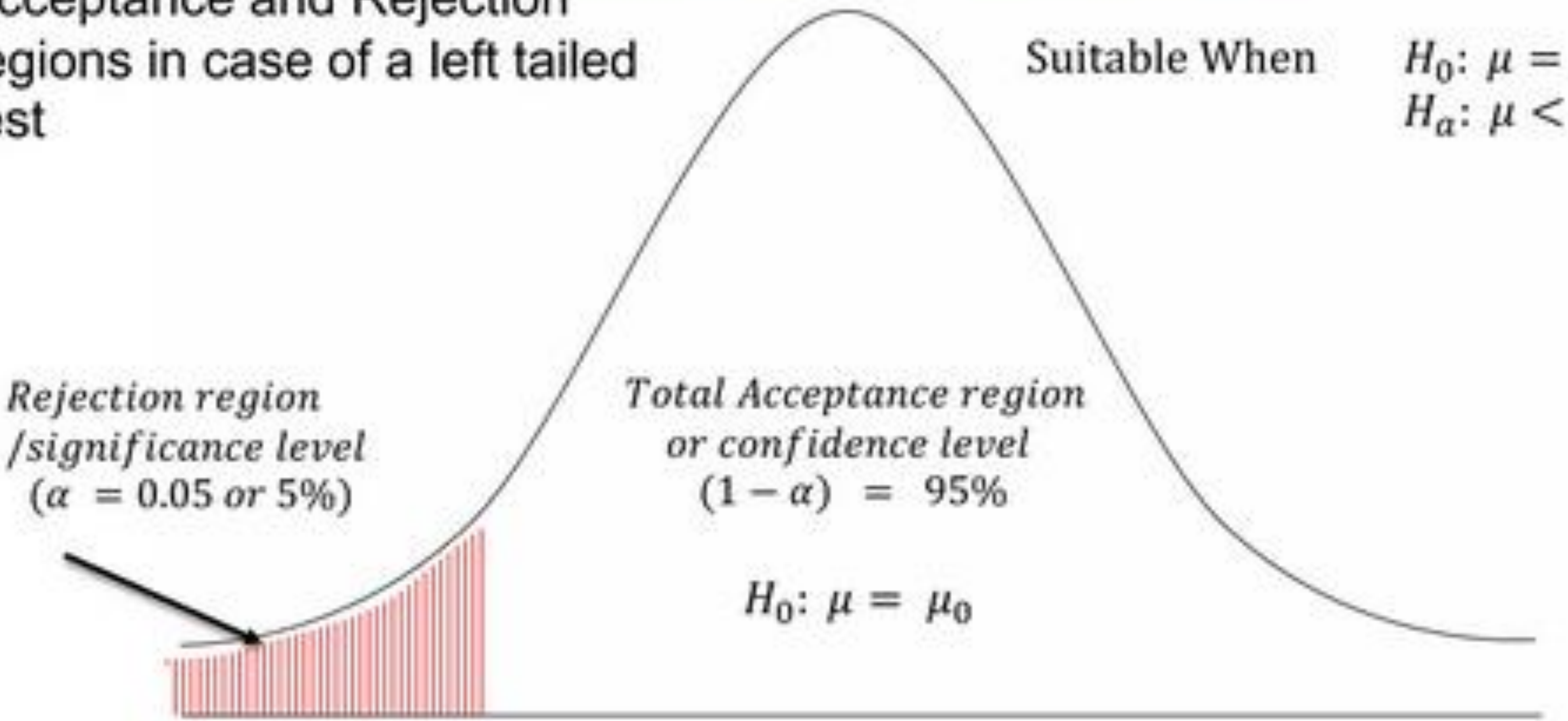
TYPE OF TESTS	DIFFERENCE
One-Tailed Test	Region of Rejection lies entirely in one end of the distribution. Hypothesizing a Range of Values
Two Tailed Test	Involves a Critical Region which is split into two equal parts placed in each tail of the distribution. A value of the parameter is being hypothesized.

Mathematical Formulation of H_1	Region of Rejection
Greater Than (>)	Area of Rejection is placed entirely in the Right Tail of the Distribution
Less Than (<)	Region of Rejection is in the Left Tail
Not Equal To (≠)	Both Tails contain Equal areas serving as Critical Regions

Left tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a left tailed test

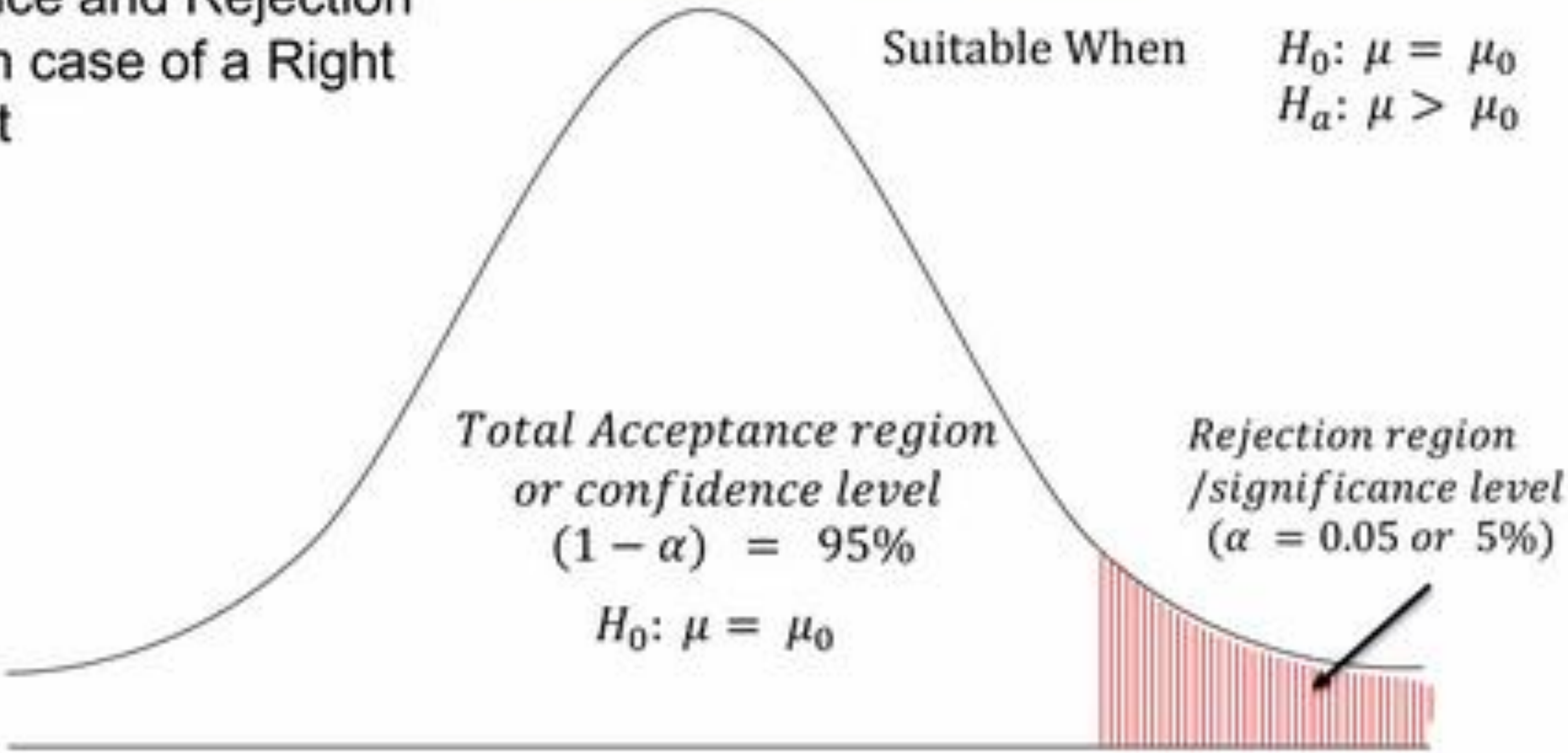
Suitable When $H_0: \mu = \mu_0$
$H_a: \mu < \mu_0$



Right tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Right tailed test

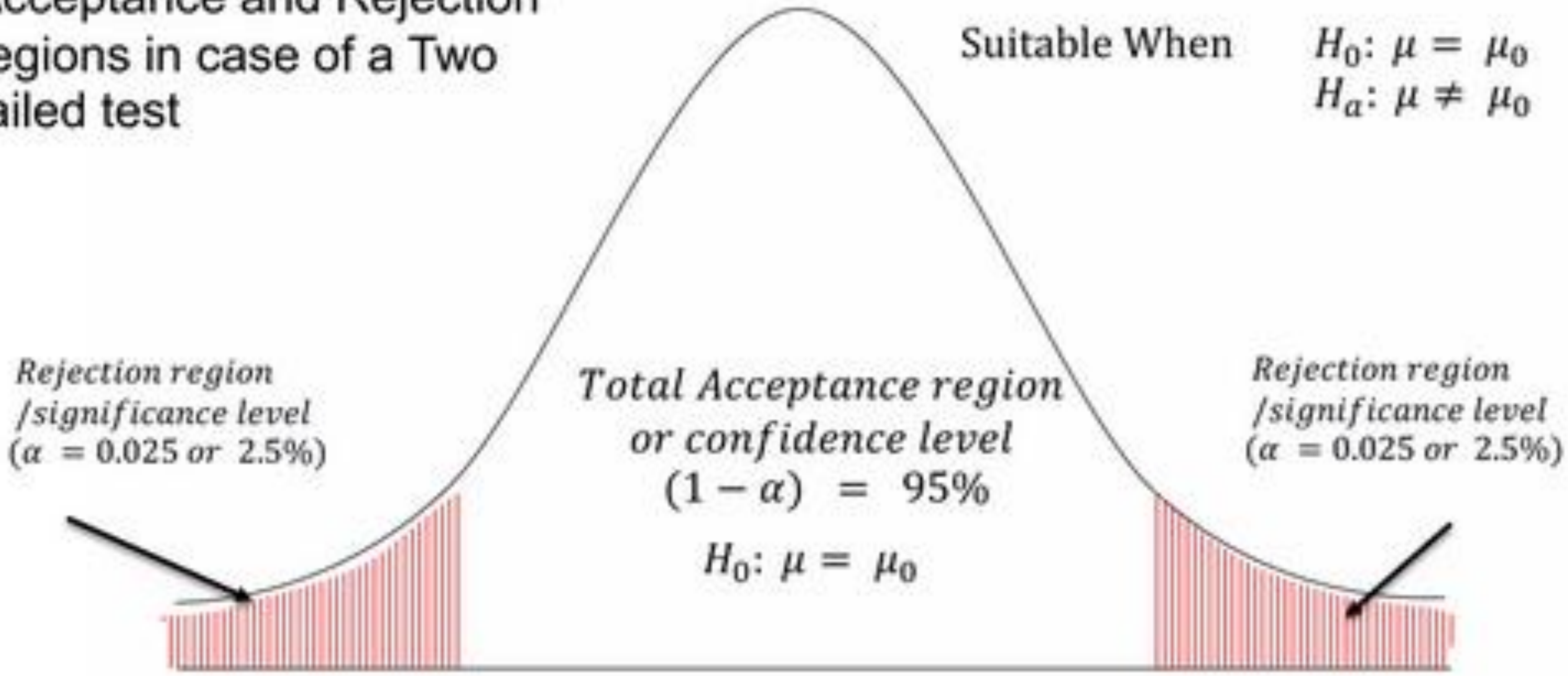
Suitable When $H_0: \mu = \mu_0$
$H_a: \mu > \mu_0$



Two tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Two tailed test

Suitable When $H_0: \mu = \mu_0$
$H_a: \mu \neq \mu_0$



Hypothesis step testing

- Hypothesis testing will be presented as a nine (9) step procedure:

1. Data
2. Assumptions
3. Hypothesis
4. Test statistics
5. Distribution of test statistics
6. Decision rule
7. Calculation of the Test statistics
8. Statistical decision
9. Conclusion

What is test statistics?

- A test statistic is a **sample statistic computed from sample data**. The value of the test statistic is **used in determining whether or not** we may **reject the null hypothesis.**
- The **decision rule** of a statistical hypothesis test is a rule that **specifies the conditions** under which the **null hypothesis may be rejected**.
- There are **two possible states of nature**
- 1. H_0 is true
- 2. H_0 is false
- There are **two possible decisions**:
- 1. Fail or reject H_0 as true
- Reject H_0 as false.

P-value

- The p-value is the **probability** of **obtaining a value of the test statistics** as **extreme or more extreme than the actual value** obtained, when **the null hypothesis is true**.
- The p-value is the **smallest level of significance (alpha/ α)** at which the null hypothesis may be rejected using the obtained value of the test statistic
- When the **p-value is less than α, reject H_0.**

P-value Table

The P-value table shows the hypothesis interpretations:

P-value	Decision
P-value > 0.05	The result is not statistically significant and hence don't reject the null hypothesis.
P-value < 0.05	The result is statistically significant. Generally, reject the null hypothesis in favour of the alternative hypothesis.
P-value < 0.01	The result is highly statistically significant, and thus rejects the null hypothesis in favour of the alternative hypothesis.

z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

$\bar{x}$ = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

A confidence interval refers to the **probability that a population parameter will fall between two set values**.
uses a percentage level, often **95 percent, to indicate the degree of uncertainty of its construction**.

Example

- A company that delivers packgs within a large city area claims that it takes ana average of 28 minutes for a package to be delivered from your door to the destination. Test this claim this hypothesis.
- **Answer:**

1**.Set the null and alternative hypothesis:**

H_0 : average time = 28

H1: average time ≠ 28

2. Collect sample data (will be given in question)

N=100

Mean time = 31.5

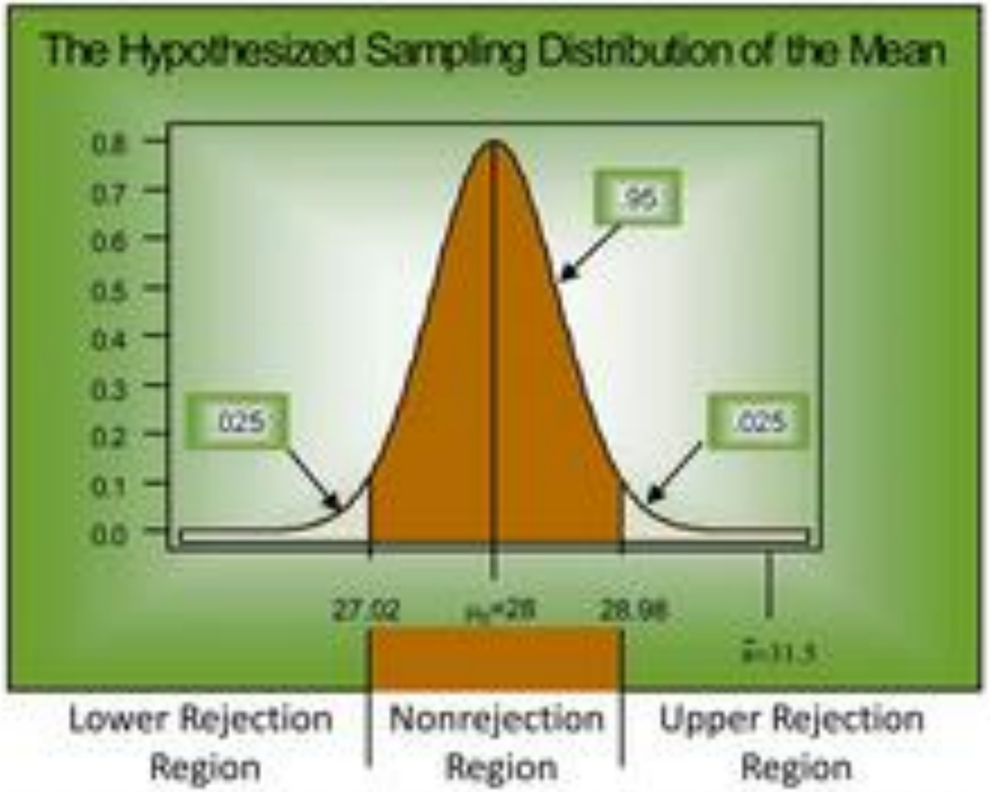
S = 5

3. **Confidence interval at 95% = $\bar{x} \pm Z_{0.05}(s/\sqrt{n})$** = 31.5 ± 1.96 (5/10) =31.5 ± 0.98

= [30.52, 32.48]

We can be 95% sure that the average time for package delivery is between 30.52 and 32.48 minutes. As the asserted value 28 is not in this 95% CI, we may reject null hypothesis.

The Decision Rule



Construct a (1-□) nonrejection region around the hypothesized population mean.

- ✓ Do not reject H_0 if the sample mean falls within the nonrejection region (between the critical points).
- ✓ Reject H_0 if the sample mean falls outside the nonrejection region.

Hypothesis testing

- Three different types of hypothesis test, namely

1. Test of hypothesis about **population means**
2. Test of hypothesis about **population proportions**
3. Test of hypothesis about **population variance**

Testing Population Means

- **Cases in which the test statistic is Z:**

s is known and the **sample size is at least 30**. (Population need not be normal)

- **Cases in which the test statistic is t:**

s is unknown but the **sample standard deviation is known** and the population is normal.

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

$\bar{x}$	= sample mean
μ	= population mean
σ	=population standard deviation
n	=sample size

One-Sample T-Test

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

$\bar{x}$ = obersved mean of the sample
μ = assumed mean
s = standard deviation
n = sample size

Two-Sample T-Test

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = observed mean of 1st sample
$\bar{x}_2$ = observed mean of 2nd sample
s_1 = standard deviation of 1st sample
s_2 = standard deviation of 2nd sample
n_1 = sample size of 1st sample
n_2 = sample size of 2nd sample

Example

- A machine fills cola into two liter (2000cc) bottles. A consumer wants to test the null hypothesis that the average amount filled by the machine into a bottle is at least 2000cc. A random sample of 40 bottles coming out of the machine was selected and the exact content of the bottles recorded. The sample mean was 1999.6cc. The population SD is known from past experience to be 1.30cc. Test the null hypothesis at 5% significance level.
- **Answer:**

H_0: $\mu = 2000$
H_1: $\mu \neq 2000$
n = 40
For 5% significance level, z = -1.645

Test statistics, $Z = \bar{x} - \mu / (\sigma/\sqrt{n})$
Don't reject null hypothesis if z = -1.645
Reject null if $Z \neq -1.645$

Calculation:
N = 40
$\bar{x} = 1999.6$
$\mu=2000$

Z= -1.95; Reject null hypothesis

Population variances

- For testing hypothesis about population variances, the test statistics (Chi-square) is used.
- Where chi-square (X^2) is the claimed value of the population variance in the null hypothesis. The degree of freedom for chi-square random variable is (n-1).

Chi Squared (Variance, not Proportions)	$$\chi^2 = \frac{(n-1)s^2}{\sigma_o^2}$$	n = the sample size s^2 = the sample vari-ance σ^2 = the population variance

Practices

- 1. In a sample of 49 adolescents who served as the subjects in an immunologic study, a variable of interest was the diameter of skin test reaction to an antigen. The sample mean and standard deviation were 21 and 11 mm erythema respectively. Can it be concluded from these data that the population mean is less then 30? Let significance level at 5%.
- 2. A research assume that systolic pressure in a certain population of males is normally distributed with a standard deviation of 16. A simple random sample of 64 males from the population had a mean systolic pressure reading of 133. At the 0.05 level of significance, do these data provide sufficient evidence for us to conclude that the population mean is greater than 130.

Hypothesis testing: The differences between two population mean

Hypothesis testing involving the difference between two population means is most frequently employed to determine whether or not it is reasonable to conclude that the two population means are unequal. In such cases, one or the other of the following hypotheses may be formulated:

1. $H_0: \mu_1 - \mu_2 = 0, \qquad H_A: \mu_1 - \mu_2 \neq 0$
2. $H_0: \mu_1 - \mu_2 \geq 0, \qquad H_A: \mu_1 - \mu_2 < 0$
3. $H_0: \mu_1 - \mu_2 \leq 0, \qquad H_A: \mu_1 - \mu_2 > 0$

Condition:

- 1. When **two independent** simple random samples has been drawn from a normally distributed population with a **known variance**, the test statistic for testing null hypothesis:
- $Z = (\overline{x}_1 - \overline{x}_2) – (\mu_1 - \mu_2)/\sqrt{((\sigma_1^2/n_1) + (\sigma_2^2/n_2))}$
- 2. If **variance is unknown**;
- $t = (\overline{x}_1 - \overline{x}_2) – (\mu_1 - \mu_2)/\sqrt{((s_p^2/n_1) + (s_p^2/n_2))}$

Example 1: Z-test

Researchers wish to know if the data they have collected provide sufficient evidence to indicate a difference in mean serum uric acid levels between normal individuals and individuals with Down's syndrome. The data consist of serum uric acid readings on 12 individuals with Down's syndrome and 15 normal individuals. The means are $\bar{x}_1 = 4.5$ mg/100 ml and $\bar{x}_2 = 3.4$ mg/ml.

- Assume variance for normal population is 1 and variance for Down's syndrome is 1.5
- **Calculation**:
- 1. Hypothesis: H_0: $\mu_1 = \mu_2$; H_A: $\mu_1 \neq \mu_2$.
- Decision rule: at 0.05 significance level; critical value of Z =± 1.96. Accept H_0 if -1.96 < Z estimated < 1.96.
- Z = 2.57, Reject null hypothesis as Z>1.96.

Example 02: t-test

The purpose of a study by Eidelman et al. (A-6) was to investigate the nature of lung destruction in the lungs of cigarette smokers before the development of marked emphysema. Three lung destructive index measurements were made on the lungs of lifelong nonsmokers and smokers who died suddenly outside the hospital of nonrespiratory causes. A larger score indicates greater lung damage. For one of the indexes the scores yielded by the lungs of a sample of nine nonsmokers and a sample of 16 smokers are shown in Table 7.3.1. We wish to know if we may conclude, on the basis of these data, that smokers, in general, have greater lung damage as measured by this destructive index than do nonsmokers.

TABLE 7.3.1 Lung Destructive Index Scores for Example 7.3.2

Nonsmokers:	18.1,	6.0,	10.8,	11.0,	7.7,	17.9,	8.5,	13.0,	18.9
Smokers:	16.6,	13.9,	11.3,	26.5,	17.4,	15.3,	15.8,	12.3,	18.6,
	12.0,	24.1,	16.5,	21.8,	16.3,	23.4,	18.8		

Calculation:

1. *Data* See statement of problem.
2. *Assumptions* The data constitute two independent simple random samples of lungs, one sample from a population of nonsmokers (NS) and the other sample from a population of smokers (S). The lung destructive index scores in both populations are approximately normally distributed. The population variances are unknown but are assumed to be equal.
3. *Hypotheses* H_0: $\mu_S \leq \mu_{NS}$, H_A: $\mu_S > \mu_{NS}$.
4. *Test Statistic* The test statistic is given by Equation 7.3.2.
5. *Distribution of Test Statistic* When the null hypothesis is true, the test statistic follows Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.
6. *Decision Rule* Let $\alpha = .05$. The critical values of t are ± 2.0687. Reject H_0 unless $-2.0687 < t_{\text{computed}} < 2.0687$.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

7. *Calculation of Test Statistic* From the sample data we compute

$$\bar{x}_S = 17.5, \quad s_S = 4.4711, \quad \bar{x}_{NS} = 12.4, \quad s_{NS} = 4.8492$$

Next we pool the sample variances to obtain

$$s_p^2 = \frac{15(4.4711)^2 + 8(4.8492)^2}{15 + 8} = 21.2165$$

We now compute

$$t = \frac{(17.5 - 12.4) - 0}{\sqrt{\dfrac{21.2165}{16} + \dfrac{21.2165}{9}}} = 2.6573$$

8. *Statistical Decision* We reject H_0, since $2.6573 > 2.0687$; that is, 2.6573 falls in the rejection region.

9. *Conclusion* On the basis of these data we conclude that the two population means are different; that is, we conclude that, as measured by the index used in the study, smokers have greater lung damage than nonsmokers. For this test $.01 > p > .005$, since $2.500 < 2.6573 < 2.8073$.

Z table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−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
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148