Hypothesis Testing

CS 700

Acknowledgement

These slides are based on presentations created and copyrighted by Prof. Daniel Menasce (GMU)

Hypothesis Testing

- Purpose: make inferences about a population parameter by analyzing differences between observed sample statistics and the results one expects to obtain if some underlying assumption is true.
- Null hypothesis: $H_0 : \mu = x$
- Alternative hypothesis: $H_1 : \mu \neq x$
- If the null hypothesis is rejected then the alternative hypothesis is accepted

Risks in Decision Making

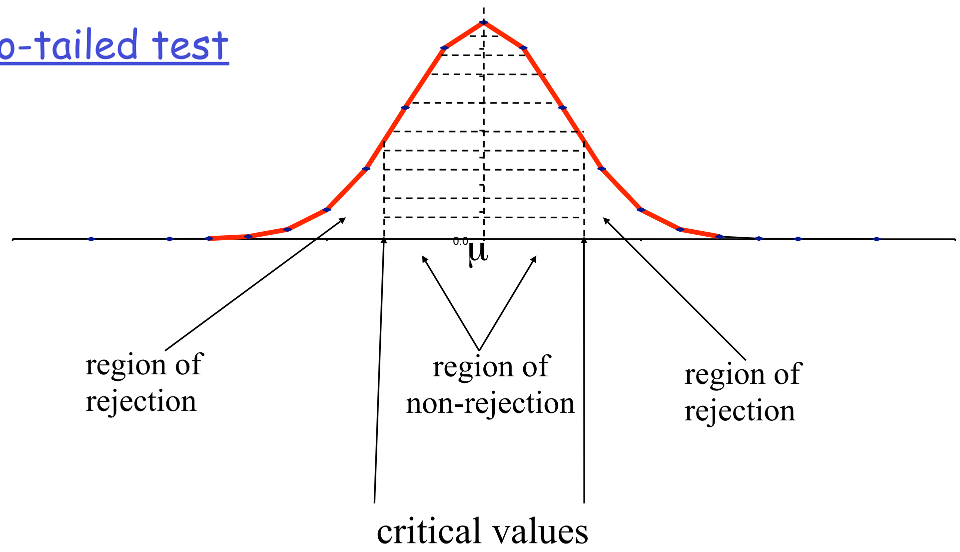
- Type I Error occurs if H_o is rejected when it is true.
  - Pr [H_o is rejected | true] = α
- Type II Error occurs if H_o is not rejected when it is false.
  - Pr[H_o is not rejected | false] = β
- Confidence coefficient:
  - Pr [H_o not rejected | true]= 1- α
- Power of the test:
  - Pr[H_o is rejected |false]= 1-β

	Actual Situation	
	H_o true	H_o false
Accept H_o	Correct decision Confidence=1-α	Type II Error: Pr[Type II]=β
Reject H_o	Type I Error P[Type I]=α	Correct Decision Power=1-β

One-sided and two-sided alternatives

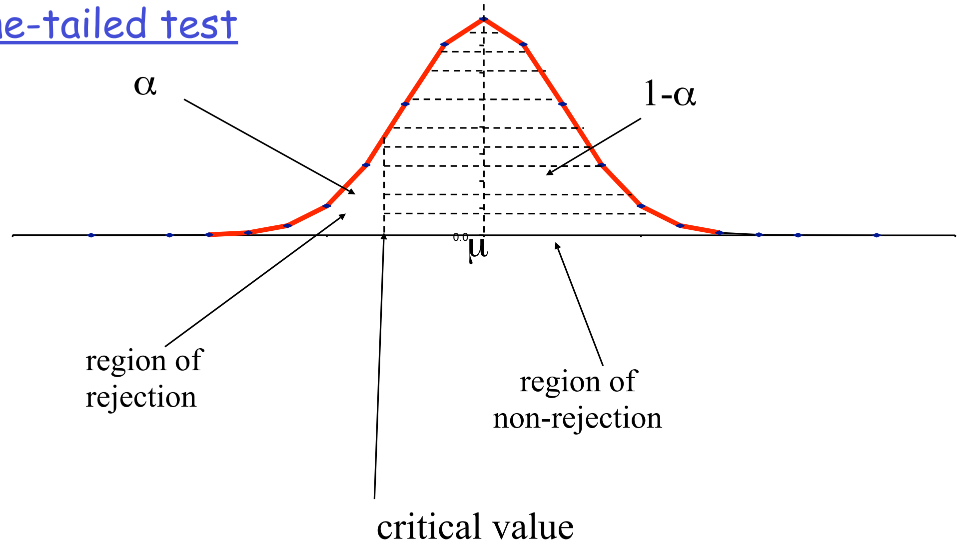
- Traditionally, the null hypothesis is used for a hypothesis set up primarily to see if it can be rejected
  - When the goal of an experiment is to establish an assertion, the negation of the assertion should be taken as the null hypothesis, and the assertion becomes the alternative hypothesis
- Alternative hypotheses usually specify that the population mean (or whatever other parameter is of concern) is not equal to, greater than, or less than the value assumed under the null hypothesis
  - Two-sided alternative $H_1 : \mu \neq x$
  - One-sided alternatives: $H_1 : \mu > x$ or $H_1 : \mu < x$

Two-tailed test



Test statistic: $Z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

One-tailed test



Test statistic: $Z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

Critical regions for two-sided and one-sided alternative hypotheses

Null hypothesis: $\mu = \mu_0$

Alternative hypothesis	Reject null hypothesis if:
$\mu < \mu_0$	$Z < -z_{\alpha}$
$\mu > \mu_0$	$Z > z_{\alpha}$
$\mu \neq \mu_0$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Note that the critical region for accepting the null hypothesis can be used to compute the $(1-\alpha)100\%$ confidence intervals for the population mean μ, i.e. $(\bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}})$

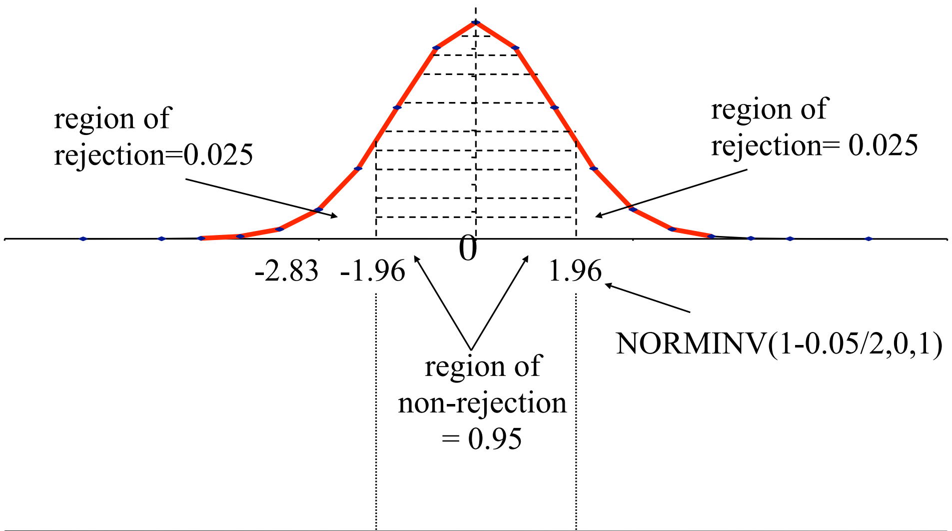
Example of Hypothesis Testing

- A sample of 50 files from a file system is selected. The sample mean is 12.3 Kbytes. The standard deviation is known to be 0.5 Kbytes.

  H_0: $\mu = 12.5$ Kbytes

  H_1: $\mu \neq 12.5$ Kbytes

  Confidence: 0.95



$$Z = \frac{12.3 - 12.5}{\frac{0.5}{\sqrt{50}}} = -2.83$$

Reject H_o

Z Test of Hypothesis for the Mean	
Null Hypothesis $\mu=$	**12.5**
Level of Significance	**0.05**
Population Standard Deviation	**0.5**
Sample Size	**50**
Sample Mean	**12.3**
Standard Error of the Mean	0.070710678
Z Test Statistic	-2.828427125
Two-Tailed Test	
Lower Critical Value	**-1.959961082**
Upper Critical Value	**1.959961082**
***p*-Value**	**0.00467786**
Reject the null hypothesis	

Hypothesis Tests with Unknown σ

- We can estimate the variance by the sample variance
- For large samples, we can use the Z statistic
- For small samples, if the population is assumed to be normally distributed the sampling distribution for the mean follows a t distribution with n-1 degrees of freedom
- t statistic for unknown σ: $t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$

sample standard deviation → s

Example of Hypothesis Testing

- A sample of 5 files from a file system is selected. Assume that file sizes are normally distributed. The sample mean is 12.3 Kbytes. The sample standard deviation is 0.5 Kbytes.

  Ho: $\mu = 12.35$ Kbytes

  H_1: $\mu \neq 12.35$ Kbytes

  Confidence: 0.95

Example

$t = (12.3 - 12.35)/(0.5/\sqrt{5}) = -0.2236$

$\alpha = 0.05$, degrees of freedom = 4

$t_{\alpha/2} = 2.776$ for 4 degrees of freedom

In EXCEL, TINV(0.05,4)

The t test statistic (-0.2236) is between the lower and upper critical values (i.e. -2.776 and 2.776)

So the null hypothesis should not be rejected.

Example of One-Tailed Test

- A sample of 50 files from a file system is selected. The sample mean is 12.35 Kbytes. The standard deviation is known to be 0.5 Kbytes.

  Ho: $\mu = 12.3$ Kbytes

  H_1: $\mu < 12.3$ Kbytes

  Confidence: 0.95

Example of One-Tailed Test

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{12.35 - 12.3}{0.5 / \sqrt{50}} = 0.707 \qquad \text{Statistic}$$

Critical value = NORMINV(0.05,0,1) = -1.645.
Region of non-rejection: Z ≥ −1.645.
So, do not reject H_o. (Z exceeds critical value)

One-tailed Test

Z Test of Hypothesis for the Mean	
Null Hypothesis $\mu=$	**12.3**
Level of Significance	**0.05**
Population Standard Deviation	**0.5**
Sample Size	**50**
Sample Mean	**12.35**
Standard Error of the Mean	0.070710678
Z Test Statistic	0.707106781
Lower-Tail Test	
Lower Critical Value	**-1.644853**
***p*-Value**	**0.760250013**
Do not reject the null hypothesis	

Steps in Hypothesis Testing

1. State the null and alternative hypothesis.
2. Choose the level of significance α.
3. Choose the sample size n. Larger samples allow us to detect even small differences between sample statistics and true population parameters. For a given α, increasing n decreases β.
4. Choose the appropriate statistical technique and test statistic to use (Z or t).

Steps in Hypothesis Testing

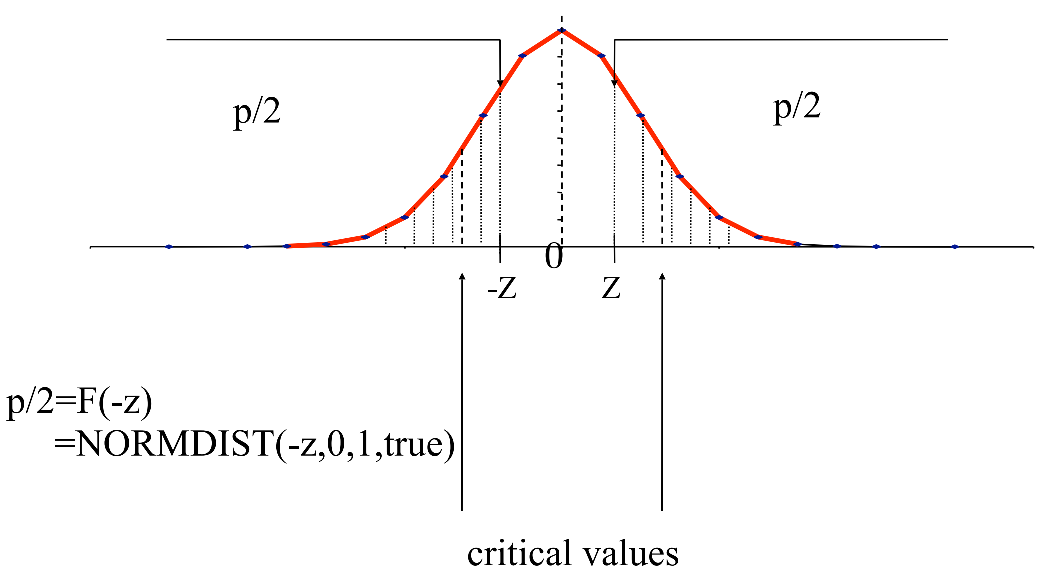
4. Determine the critical values that divide the regions of acceptance and non-acceptance.

5. Collect the data and compute the sample mean and the appropriate test statistic (e.g., Z).

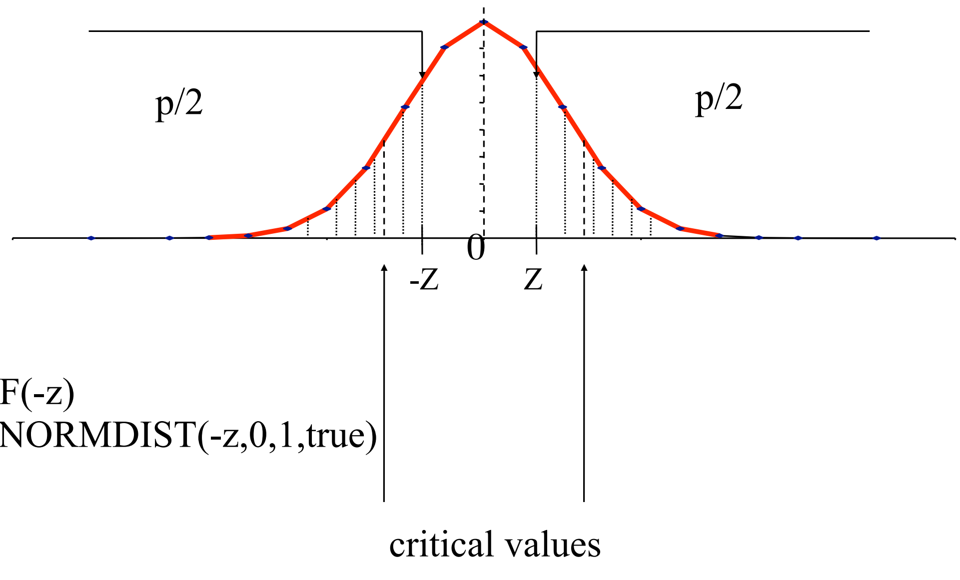
6. If the test statistic falls in the non-reject region, H_o cannot be rejected. Else H_o is rejected.

The p-value Approach

- p-value: observed level of significance. Defined as the probability that the test statistic is equal to or more extreme than the result obtained from the sample data, given that H_o is true.

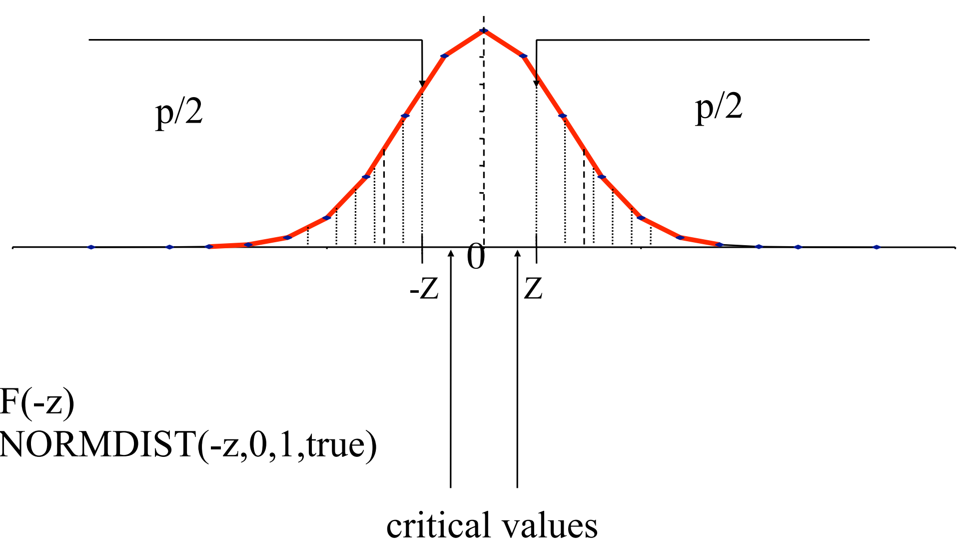


If $p \geq \alpha$ then do not reject H_o, else reject H_o.



p/2=F(-z)
=NORMDIST(-z,0,1,true)

Do not reject H_o



p/2=F(-z)
=NORMDIST(-z,0,1,true)

Reject H_o

Computing p-values

Z Test of Hypothesis for the Mean	
Null Hypothesis μ=	**12.5**
Level of Significance	**0.05**
Population Standard Deviation	**0.5**
Sample Size	**50**
Sample Mean	**12.3**
Standard Error of the Mean	0.070710678
Z Test Statistic	-2.828427125
Two-Tailed Test	
Lower Critical Value	**-1.959961082**
Upper Critical Value	**1.959961082**
***p*-Value**	**0.00467786**
Reject the null hypothesis	

The null hypothesis is rejected because p (0.0047) is less than the level of significance (0.05).

Steps in Determining the p-value.

1. State the null and alternative hypothesis.
2. Choose the level of significance α.
3. Choose the sample size n. Larger samples allow us to detect even small differences between sample statistics and true population parameters. For a given α, increasing n decreases β.
4. Choose the appropriate statistical technique and test statistic to use (Z or t).

Steps in Determining the p-value.

5. Collect the data and compute the sample mean and the appropriate test statistic (e.g., Z)
6. Calculate the p-value based on the test statistic
7. Compare the p-value to α
8. If $p \geq \alpha$ then do not reject H_o, else reject H_o

Rank-sum (Wilcoxon test)

- Non-parameteric test, i.e., does not depend upon distribution of population, for comparing two samples
- Example:
  - Suppose the time between two successive crashes are recorded for two competing computer systems as follows (time in weeks):
    System I: 0.63 0.17 0.35 0.49 0.18 0.43 0.12 0.20 0.47 1.36 0.51 0.45 0.84 0.32 0.40
    System II: 1.13 0.54 0.96 0.26 0.39 0.88 0.92 0.53 1.01 0.48 0.89 1.07 1.11 0.58
  - The problem is to determine if the two populations are the same or if one is likely to produce larger observations than the other

Rank-sum test (cont'd)

- U-test is a non-parameteric alternative to the paired and unpaired t-tests
- First step in the U-test is to rank the data jointly, in increasing order of magnitude

0.12	0.17	0.18	0.20	0.26	0.32	0.35	0.39	0.40	0.43
I	I	I	I	II	I	I	II	I	I
0.45	0.47	0.48	0.49	0.51	0.53	0.54	0.58	0.63	0.84
I	I	II	I	I	II	II	II	I	I
0.88	0.89	0.92	0.96	1.01	1.07	1.11	1.13	1.36	
II	II	II	II	II	II	II	II	I	

- Assign each data item a rank in this order
  - If there are ties among values, the rank assigned to each observation is the mean of the ranks which they jointly occupy

Rank-sum test (cont'd)

- The values in the first sample occupy ranks 1, 2,3,4,6,7,9,10,11,12,14,15,19,20 and 29
- The sum of the ranks for the two samples, $W_1 = 162$ and $W_2 = 273$
- The U-test is based on the statistics

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2}$$

or

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

or on the statistic U which is the smaller of the two

Rank-sum test (cont'd)

- Under the null hypothesis that the two samples come from identical populations, it can be shown that the mean and variance of the sampling distribution of U_1 are

$$\mu_{U_1} = \frac{n_1 n_2}{2}$$

and

$$\sigma_{U_1}^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

- Numerical studies have shown that the sampling distribution of U1 can be approximated closely by the normal distribution when n1 and n2 are both greater than 8

Rank-sum test (cont'd)

- Thus, the test of the null hypothesis that both samples come from identical populations can be based on

$$Z = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}}$$

  which is a random variable having approximately the standard normal distribution

- The alternative hypothesis is either:
  - Population 2 is stochastically larger than Population 1
    - We reject the null hypothesis if $Z < -z_\alpha$
  - Or, Population 1 is stochastically larger than Population 2
    - We reject the null hypothesis if $Z > -z_\alpha$

Example cont'd

- At the 0.01 level of significance, test the null hypothesis that the two samples in our example come from the same population
  - Alternative hypothesis, populations are not identical
  - For $\alpha = 0.01$, we can reject the null hypothesis if $Z < -2.575$ or $Z > 2.575$
    - Calculations: n1 = 15, n2 = 14, W1 = 162
      $U1 = 162 - 15 \times 16/2 = 42$
      $Z = (42 - 15 \times 14/2)/\sqrt{(15 \times 14 \times 30)/12} = -2.75$
  - Since Z is less than -2.575, we reject the null hypothesis; we conclude there is a difference between the two systems