Critical values for z-test:

	LOS 1%	LOS 5%
One-tailed	2.33	1.645
Two-tailed	2.58	1.96

Hypothesis Tests About a Population Mean

Case I: Test for mean with σ Known

(For large sample (sample size $\geq$ 30), we can use s in the place of σ)

Test statistic: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ → Normal distribution

Example 1 : A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130°. A sample of n= 9 systems, when tested, yields a sample average activation temperature of 131.08°F. If the distribution of activation times is normal with standard deviation 1.5°F, does the data contradict the manufacturer's claim at significance level .01?

Example 2 : In 64 randomly selected hours of production, the mean and the standard deviation of the number of acceptable pieces produced by a automatic stamping machine are $\bar{x}$ = 1,038 and s = 146. At the 0.05 level of significance, does this enable us to reject the null hypothesis μ = 1,000 against the alternative hypothesis μ > 1,000?

Example 3 : A random sample of 6 steel beams has a mean compressive strength of 58,392 psi (pounds per square inch) with a standard deviation of 648 psi. Use this information and the level of significance α = 0.05 to test whether the true average compressive strength of the steel from which this sample came is 58,000 psi. Assume normality.

Case II: Test for mean with σ Unknown

Test statistic: $T = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ → t distribution with degrees of freedom: $n-1$

Example 1: New York is known as "the city that never sleeps". A survey was conducted to test the hypothesis that New Yorkers sleep less than 8 hours a night on average. A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are: $n = 25$, $\bar{X} = 7.73$, $s = 0.77$. Test the claim at 5% LOS.

Example 2: Georgianna claims that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons. We have a random sample of 20 children from the city, with a mean of 4.6 years of piano lessons and a standard deviation of 2.2 years.

(a) Evaluate Georgianna's claim (or that the opposite might be true) using a hypothesis test.
(b) Construct a 95% confidence interval for the number of years students in this city take piano lessons, and interpret it in context of the data.

Example 3: Light bulbs of a certain type are advertised as having an average lifetime of 750 hours. The price of these bulbs is very favorable, so a potential customer has decided to go ahead with a purchase arrangement unless it can be conclusively demonstrated that the true average lifetime is smaller than what is advertised. A random sample of 50 bulbs gave an average lifetime of 738.44 with a standard deviation of 38.20.

What conclusion would be appropriate for a significance level of .05? A significance level of .01? What significance level and conclusion would you recommend?

Example 4: A manufacturer claims that the average tar content of a certain kind of cigarette is μ = 14.0. In an attempt to show that it differs from this value, five measurements are made of the tar content (mg per cigarette): 14.5, 14.2, 14.4, 14.3, 14.6. Show that the difference between the mean of this sample and the average tar claimed by the manufacturer, μ = 14.0, is significant at α = 0.05. Assume normality.

Hypothesis Testing with Two Independent Samples

A motivational example: Suppose that there are two types of food available for milking cows. A farmer wishes to test which type of food helps cows to produce more yield of milk.

An experiment: The farmer can select two independent groups of cows who produce similar milk yields. One group is given the food A and the other group is given the food B. After one week, the farmer calculates the means and standard deviations of milk yields for each group and then use his knowledge to decide the type of food which gives better yield.

Further examples:
1. Compare the average age at first marriage of females in two ethnic groups.
2. Compare the average efficiency of two brands of fertilisers.

Test assumptions
1. Two populations are **independent** and ***normally distributed*** **po**pulations with ***equal variances*****.**
2. Two independent samples are drawn (one from each population).

Null hypothesis: $H_0 : \mu_1 = \mu_2$ or $H_0 : \mu_1 - \mu_2 = 0.$

Alternate hypothesis: $H_1 : \mu_1 > \mu_2$ or $H_1 : \mu_1 - \mu_2 > 0$ (one-sided);
$H_1 : \mu_1 < \mu_2$ or $H_1 : \mu_1 - \mu_2 < 0$ (one-sided);
$H_1 : \mu_1 \neq \mu_2$ or $H_1 : \mu_1 - \mu_2 \neq 0$ (two-sided).

Z-test for comparing two independent large samples

Test statistic: $Z = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$ if population distributions are normal with variances known

Test statistic: $Z = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$ if population distributions are normal with variances unknown

Example 1: Analysis of a random sample consisting of 20 specimens of cold-rolled steel to determine yield strengths resulted in a sample average strength of $\overline{X_1} =$ 29.8 ksi. A second random sample of 25 two-sided galvanized steel specimens gave a sample average strength of $\overline{X_2} =$ 34.7 ksi. Assuming that the two yield-strength distributions are normal with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, does the data indicate that the corresponding true average yield strengths 1 and 2 are different? Carry out a test at significance level 0.01.

Example 2: To test the claim that the resistance of electric wire can be reduced by more than 0.050 ohm by alloying, 32 values obtained for standard wire yielded $\overline{X_1}$ = 0.136 ohm and s_1 = 0.004 ohm, and 32 values obtained for alloyed wire yielded $\overline{X_2}$ = 0.083 ohm and s_2 = 0.005 ohm. At the 0.05 level of significance, does this support the claim?

t-test comparing two independent small samples

Assumption: both population distributions are normal and that $\sigma_1 = \sigma_2 = \sigma$ (variances are equal but unknown.

Test statistic: $T = \dfrac{\overline{X_1} - \overline{X_2}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ → degrees of freedom: $n_1 + n_2 - 2$ where ${S_p}^2 = \dfrac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{(n_1 + n_2 - 2)}$

The combined variance s_p^2 is called pooled variance.
It is the weighted average of of the two individual sample variances, weighted by their df.

Example 1: An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Example 2: A feeding test is conducted on a herd of 25 dairy cows to compare two diets, A and B. A sample of 13 cows randomly selected from the herd are fed diet A and the remaining cows are fed with diet B. From observations made over a three-week period, the average daily milk production (in L) is recorded for each cow:

Diet A (x1):	44	44	56	46	47	38	58	53	49	35	46	30	41
Diet B (x2):	35	47	55	29	40	39	32	41	42	57	51	39	--

Matched Pairs Comparisons (Paired t-test)

A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample.

Motivating example: A medical researcher wants to compare the effects of two treatments. The best way to get satisfactory results from this experiment is to give both treatments to the same group of people with a long time gap (e.g. Monday treatment I and Tuesday treatment II), and repeat this for a number of individuals. In this way any effects other than treatment, e.g. personal characteristic, can be eliminated.

Further examples of where this might occur are:

• Before-and-after observations on the same subjects (e.g. students' aptitude test results before and after a particular coaching).

Suppose a sample of n students were given a diagnostic test before studying a coaching and then again after completing coaching. We want to find out if, in general, coaching leads to improvements in students' performance. For $i = 1, .., n$,

x_i: score before coaching

y_i = score after coaching

$d_i = x_i - y_i$ and $\bar{d} = mean(d_i)$

Test statistic: $T = \frac{\bar{d} - 0}{S_d/\sqrt{n}}$ with degrees of freedom: $n - 1$

Example 1: The following are the average weekly losses of worker-hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation:

Before:	45	73	46	124	33	57	83	34	26	17
After:	36	60	44	119	35	51	77	29	24	11

Use the 0.05 level of significance to test whether the safety program is effective.

Example 2: A large pharmaceutical company wants to compare the effect of a new drug (A) for chronic insomnia against an existing drug (B). The best way to do this test is as follows: Week 1: Select a group of individuals suffering from chronic insomnia. Give the old drug and record the number of hours they sleep. Week 2: Give the new drug and record the number of hours they sleep. Suppose that the above study consists of 6 people with chronic insomnia and their hours of sleep each night are recorded as given below. Test whether the new drug is more effective at 5% LOS.

New Drug A	4.8	4.1	5.8	4.9	5.3	7.4

Old Drug B	3.9	4.2	5.0	4.9	5.4	7.1

Example 3: In a study of the effectiveness of physical exercise in weight (in pounds) reduction, a group of 12 persons engaged in a prescribed program of physical exercise for one month showed the following results. Use the 0.01 level of significance to test whether the prescribed program of exercise is effective.

Weight before	209	178	169	212	180	192	158	180	170	153	183	165
Weight after	196	171	170	207	177	190	159	180	164	152	179	162

Grand Total: $T_1 + T_2 + \cdots + T_k = T$

Total no. of observations: $n_1 + n_2 + \cdots + n_k = N$

Total sum of squares : $\boldsymbol{SST = \sum\sum x_{ij}^2 - \frac{T^2}{N}}$; Degrees of freedom: N-1

Between-sample sum of squares: $\boldsymbol{SSB = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{N}}$ Degrees of freedom: k-1

(or treatment sum of squares)

Error sum of squares : $\boldsymbol{SSE = SST - SSB}$ Degrees of freedom: N-k
(or within-sample sum of squares)

Null hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$
Alternate hypothesis $H_1: Not\ all\ \mu_i's\ are\ equal$

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	Critical value F_α with df (k-1, N-k)
Between (Treatments)	SSB	k-1	$v_1 = \frac{SSB}{k-1}$	$F = \frac{v_1}{v_2}$	$F_{(k-1,\ N-k)}$
Error (Within)	SSE	N-k	$v_2 = \frac{SSE}{N-k}$		
Total	SST	N-1			

Compare F and F_α. If $F > F_\alpha$, reject H_0, else do not reject H_0

Randomized Block Design (2-way ANOVA)

- Divide the group of experimental units into **b** homogeneous groups of size **k**. These homogeneous groups are called blocks.
- The treatments are then randomly assigned to the experimental units in each block - one treatment to a unit in each block.

A typical layout for the randomized complete block design using 3 treatments in 4 blocks is as follows:

Block 1	Block 2	Block 3	Block 4
t_2	t_1	t_3	t_2
t_1	t_3	t_2	t_1
t_3	t_2	t_1	t_3

Blocks

		1	2	..	j	..	b	
	1	x_{11}	x_{12}	..	x_{1j}	..	x_{1b}	$T_{1\bullet}$
	2	x_{21}	x_{22}	..	x_{2j}	..	x_{2b}	$T_{2\bullet}$
Treatments	:	:	:		:		:	
	i	x_{i1}	x_{i2}	..	x_{ij}	..	x_{ib}	$T_{i\bullet}$
	:	:	:		:		:	
	k	x_{k1}	x_{k2}	..	x_{kj}	..	x_{kb}	$T_{k\bullet}$
		$T_{\bullet 1}$	$T_{\bullet 2}$		$T_{\bullet j}$		$T_{\bullet k}$	T

Total sum of squares : $SST = \sum\sum x_{ij}^2 - \frac{T^2}{N}$; Degrees of freedom: kb-1

Treatment sum of squares : $SSTr = \frac{1}{b}\sum_{i=1}^{k} T_{i\bullet}^2 - \frac{T^2}{N}$ Degrees of freedom: k-1

Block sum of squares : $SSBl = \frac{1}{k}\sum_{j=1}^{b} T_{\bullet j}^2 - \frac{T^2}{N}$ Degrees of freedom: b-1

Error sum of squares : $SSE = SST - SSTr - SSBl$ Degrees of freedom: (k-1)(b-1)

Null hypotheses H_{0T}: No difference among the treatments H_{0B}: No difference among the blocks
Alternate hypotheses H_{1T}: Not all treatments are same H_{1B}: Not all blocks are same

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	Critical value F_α with df (k-1, N-k)
Treatments	SSTr	k-1	$v_1 = \frac{SSTr}{k-1}$	$F_1 = \frac{v_1}{v_3}$	$F_{(k-1,\ (k-1)(b-1))}$
Blocks	SSBl	b-1	$v_2 = \frac{SSE}{b-1}$	$F_2 = \frac{v_2}{v_3}$	$F_{(b-1,\ (k-1)(b-1))}$
Error	SSE	(k-1)(b-1)	$v_3 = \frac{SSE}{(k-1)(b-1)}$		
Total	SST	ab-1			

Compare calculated values and critical values write the conclusion.

Example 1: A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee as in Table.1.

Program 1	Program 2	Program 3	Program 4
9	10	12	9
12	6	14	8
14	9	11	11
11	9	13	7
13	10	11	8

Table.1

Example 2: The data in Table.2 represent the number of hours of relief provided by five different brands of headache tablets administered to 25 subjects experiencing fevers of 100◦F or more. Perform the analysis of variance and test the hypothesis at the 0.05 level of significance that the mean number of hours of relief provided by the tablets is the same for all five brands. Discuss the results.

A	B	C	D	E
5.2	9.1	3.2	2.4	7.1
4.7	7.1	5.8	3.4	6.6
8.1	8.2	2.2	4.1	9.3
6.2	6.0	3.1	1.0	4.2
3.0	9.1	7.2	4.0	7.6

Table.2

Example 3: An investigator organized a study of mileages obtainable from three different brands of petrol. Using 15 identical cars set to run at the same speed, the investigator randomly assigned each brand of petrol to 5 of the cars. Each of the cars was then run on 10 litres of petrol, with total mileages obtained as in Table.3. Perform ANOVA to test the hypothesis that the average mileage obtained is the same for all three types of petrol. Use 5% LOS.

Petrol 1	Petrol 2	Petrol 3
220	244	252
251	235	272
226	232	250
246	242	238
260	225	256

Table.3

Example 4: In an experiment to see whether the amount of coverage of light-blue interior latex paint depends either on the brand of paint or on the brand of roller used, 1 gallon of each of four brands of paint was applied using each of three brands of roller, resulting in the data (number of square feet covered) in Table.4. State and test the hypothesis appropriate for deciding whether paint brand has any effect on coverage at 5% LOS.

	Roller 1	Roller 2	Roller 3
Paint 1	454	446	451
Paint 2	446	444	447
Paint 3	439	442	444
Paint 4	444	437	443

Table.4

Example 5: Table.5 shows the yields per acre of four different plant crops grown on lots treated with three different types of fertilizer. Using analysis of variance, test at the 0.01 level of significance whether

(1) there is a significant difference in yield per acre due to fertilizers;
(2) there is a significant difference in yield per acre.

	Crop 1	Crop 2	Crop 3	Crop 4
Fertilizer A	4.5	6.4	7.2	6.7
Fertilizer B	8.8	7.8	9.6	7.0
Fertilizer C	5.9	6.8	5.7	5.2

Table.5

Example 6: Three different washing machines were employed to test four different detergents. Table.6 give a coded score of the effectiveness of each washing. At 5% LOS, perform ANOVA to test the hypotheses that i) the detergent used does not affect the score; ii) the machine used does not affect the score.

	Detergent 1	Detergent 2	Detergent 3	Detergent 4
Machine 1	53	54	56	50
Machine 2	50	54	58	45
Machine 3	59	60	62	57

Table.6