One way
Analysis of Variance (ANOVA)

ANOVA General ANOVA Setting"Slide 43-45)

- Investigator controls one or more factors of interest
  - Each factor contains two or more levels
  - Levels can be numerical or categorical
  - Different levels produce different groups
  - Think of each group as a sample from a different population
  - Observe effects on the dependent variable, Are the groups the same?
- Experimental design: the plan used to collect the data
- Experimental units (subjects) are assigned randomly to groups , Subjects are assumed homogeneous
- Only one factor or independent variable, with two or more levels.
- Analyzed by one-factor analysis of variance (ANOVA)

One-Way Analysis of Variance

- Evaluate the difference among the means of three or more groups

Examples: Number of accidents for 1st, 2nd, and 3rd shift, Expected mileage for five brands of tires.

ANOVA Assumptions "Slide 71"

- Randomness and Independence

  Select random samples from the c groups (or randomly assign the levels)
- Normality

  The sample values for each group are from a normal population
- Homogeneity of Variance

  All populations sampled from have the same variance " Can be tested with Levene’s Test"

(Textbook: P374)

Data frame:

Observations	Groups				
	1	2	…. …. ……	C	Total
1	X_{11}	X_{21}	…..X_{i1} …….	X_{c1}	
2	X_{12}	X_{22}	…..X_{i2}……..	X_{c2}	
….	…	….	…..………..	….	
j	X_{1j}	X_{2j}	…….X_{ij}…….	X_{cj}	
Sum	$X_{1.}$	$X_{2.}$	……X_{j}…..	$X_{c.}$	X (Grand Total)
Samples	n_1	n_2	…..n_j…..	n_c	nThe Total sample
Mean	$\bar{X}_1$	$\bar{X}_2$	…..$\bar{X}_i$……	$\bar{X}_c$	$\bar{\bar{X}}$(Grand Mean)

c : number of groups or levels
n_j : number of values in group j
X_{ij} : i^{th} observation from group j

X : Total mean (Total of all data values)

n: The Total of all samples ($n = n_1 + n_2 + …… + n_j$)
$\bar{\bar{X}}$: Grand mean (mean of all data values)

Analysis of variance is a general method for studying sampled-data relationships. The method enables the difference between two or more sample means to be analyzed, achieved by subdividing the total sum of squares.
Basic idea is to partition total variation of the data into two sources:

Variation Within Groups + Variation Among Groups

Total Variation: the aggregate variation of the individual data values across the various factor levels (SST)

Among-Group Variation: variation among the factor sample means (SSA)

Within-Group Variation: variation that exists among the data values within a particular factor level (SSW)
(Slide 51-52)

The equations used to calculate these totals are : (Slide 53-59)

$$\boxed{SST = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij} - \overline{\overline{X}})^2}$$

$$\boxed{SST = (X_{11} - \overline{\overline{X}})^2 + (X_{12} - \overline{\overline{X}})^2 + \cdots + (X_{cn_c} - \overline{\overline{X}})^2}$$

$$\boxed{SSA = \sum_{j=1}^{c} n_j(\overline{X}_j - \overline{\overline{X}})^2}$$

$$\boxed{SSA = n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 + \cdots + n_c(\overline{X}_c - \overline{\overline{X}})^2}$$

$$\boxed{SSW = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2}$$

$$\boxed{SSW = (X_{11} - \overline{X}_1)^2 + (X_{12} - \overline{X}_2)^2 + \cdots + (X_{cn_c} - \overline{X}_c)^2}$$

Obtaining the Mean Squares (Slide 61)

The Mean Squares are obtained by dividing the various sum of squares by their associated degrees of freedom

Mean Square Among (d.f= c-1) : $\boxed{MSA = \dfrac{SSA}{c-1}}$

Mean Square Within (d.f = n-c) : $\boxed{MSW = \dfrac{SSW}{n-c}}$

Mean Square Total (d.f = n-1) : $\boxed{MST = \dfrac{SST}{n-1}}$

(c-1): The degrees of freedom for the Among group
(n-c): The degrees of freedom for the within group

F Test for differences among more than two means. Slide (63-64)

Step (1): State the null and alternate hypotheses:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots = \mu_c$

H_1: At least two population means are different.

Step (2): Select the level of significance (α)

Step (3): The test statistic: Because we are comparing means of more than two groups, use the F statistic.

$$F_{STAT} = \frac{MSA}{MSW} = \frac{SSA/c-1}{SSW/n-c}$$

Step (4): The critical value:
The degrees of freedom for the numerator are the degrees of freedom for the among group (c-1)
The degrees of freedom for the denominator are the degrees of freedom for the within group (n-c).

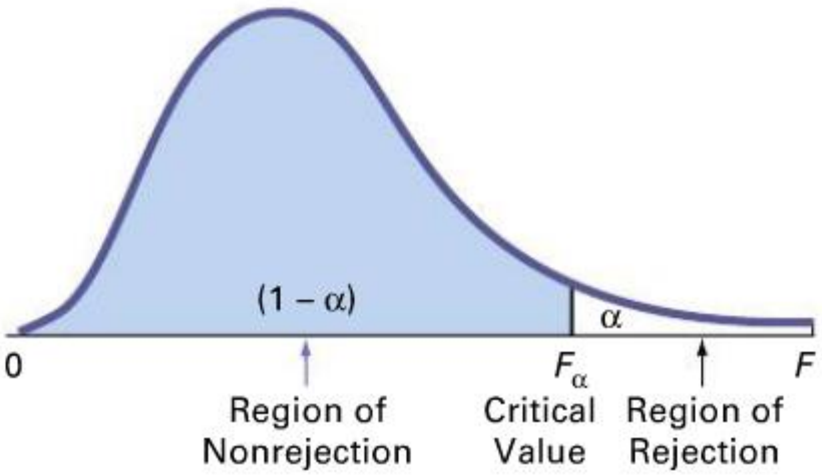
$F_{(\alpha\ ,\ c-1\ ,n-c\ )}$

(Textbook: Table E.5 P548-553)

Step (5) : Formulate the decision Rule and make a decision

Reject H_o

If $F_c > F_{(\alpha\ ,c-1,n-c)}$



ANOVA TABLE

It is convenient to summarize the calculation of the F statistic in (ANOVA Table)
(Slide 62)

Source of variation (S.V)	Degrees of freedom	Sum of Squares (S.S)	Mean Squares (MS)	F- Stat "F-Ratio"
Among groups	c-1	SSA	MSA =SSA/c-1	F_{STAT} = MSA / MSW
Within groups	n-c	SSW	MSW = SSW/n-c	
Total	n-1	SST		

Example (Slide 65-68)

You want to see if three different golf clubs yield different distances. You randomly select five measurements from trials on an automated driving machine for each club. At the 0.05 significance level, is there a difference in mean distance?

	Club 1	Club 2	Club 3
	254	234	200
	263	218	222
	241	235	197
	237	227	206
	251	216	204
Total	1246	1130	1029
Mean	249.2	226	205.8
$\bar{\bar{X}} = 227$			

C=3 $n_1=n_2=n_3= 5$ n=15

SSA = 5 $(249.2 – 227)^2$ + 5 $(226 – 227)^2$ + 5 $(205.8 – 227)^2$ = 4716.4

SSW = $(254 – 249.2)^2 + (263 – 249.2)^2 +\cdots+ (204 – 205.8)^2$ = 1119.6

Step (1): State the null and alternate hypotheses:

H_0 : $\mu_1 = \mu_2 = \mu_3$

H_1 : At least two population means are different.

Step (2): Select the level of significance (α =0.05)

Step (3): The test statistic:

$$F_{STAT} = \frac{MSA}{MSW} = \frac{SSA/c-1}{SSW/n-c} = \frac{4716.4/3-1}{1119.6/15-3} = 25.275$$

Step (4): The critical value:

The degrees of freedom for the numerator (c-1) = 3-1 =2

The degrees of freedom for the denominator (n-c) = 15-3 =12

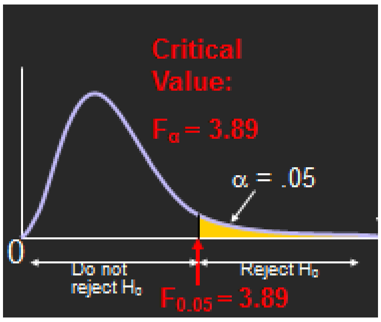
$$F_{(0.05\ ,\ 2,12)} = 3.89$$

Step (5) : Formulate the decision Rule and make a decision

F_{STAT} (25.275) > $F_{(0.05,2,12)}$ (3.89)

Reject Ho at α =0.05

Conclusion: There is evidence that at least one μ_j differs from the rest



P-value = (0.0000)   , $\alpha = 0.05$
P-value < α Reject H_0

ANOVA TABLE

Source of Variation(S.V)	SS	df	MS	F-Ratio
Between Groups	4716.4	2	2358.2	25.275
Within Groups	1119.6	12	93.3	
Total	5836.0	14		

(Textbook P" 378-381 "Starting from paragraph 3, there is an illustrative example)

Additional examples of related samples

Example (1)

Advertisements by Sylph Fitness Center claim that completing its course will result in losing weight. A random sample of eight recent participants showed the following weights before and after completing the course.

- At the 0.01sigenificance level, can we conclude the students lost weight (in pounds?)

-Find the confidence interval for μ_D

Note: 1 kg = 2.20 pounds

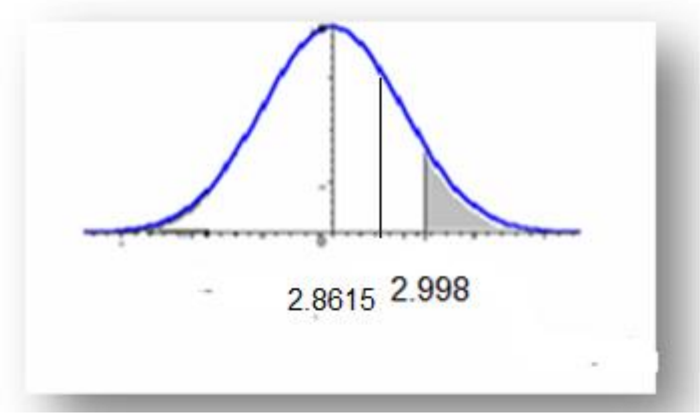
No	Before	After
1	155	154
2	228	207
3	141	147
4	162	157
5	211	196
6	164	150
7	184	170
8	172	165

Solution:

Step (1): State the null and alternate hypotheses

(where $u_1 > u_2$ or $u_1 - u_2 > 0$)

$H_0 : \mu_D \le 0 \; H_1 : \mu_D > 0$



Step (2): Select the level of significance (α=0.01)

Step (3): The critical value

In one –tailed test (Right)

$t_{(\alpha,n-1)} = t_{(0.01,7)}$ =2.998

Reject H_0 if

$t_c > 2.998$

Step (4): The test statistic

No	**Before** **B**	**After** **A**	D **(X1-X2)**	$\left(D_i - \overline{D}\right)$	$\left(D_i - \overline{D}\right)^2$
1	155	154	1	1-8.875=-7.875	62.02
2	228	207	21	21-8.875=12.125	147.02
3	141	147	-6	-6-8.875=-14.87	221.27
4	162	157	5	5-8.875=-3.875	15.02
5	211	196	15	15-8.875=6.125	37.52
6	164	150	14	14-8.875=5.125	26.27
7	184	170	14	14-8.875=5.125	26.27
8	172	165	7	7-8.875=-1.875	3.52
Total			71		538.66

$$\overline{D} = \frac{\sum D}{n} = \frac{71}{8} = 8.875$$

$$\mathrm{S_D} = \sqrt{\frac{\sum(\mathrm{D}_i - \overline{\mathrm{D}})^2}{\mathrm{n}-1}} = \sqrt{\frac{538.66}{7}} = 8.7722$$

$$t_c = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{8.875}{8.7722 / \sqrt{8}} = \frac{8.875}{8.7722 / 2.8284} = \frac{8.875}{3.1015} = 2.8615$$

Step (5): Formulate the Decision Rule

Do not reject H₀. We cannot conclude that the students lost weight

The Paired Difference Confidence Intervalμ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2} \frac{\mathrm{S_D}}{\sqrt{\mathrm{n}}}$$

$$= 8.875 \pm 3.499 \frac{8.7722}{\sqrt{8}}$$

$$= 8.875 \pm 10.8519$$

$$-1.9769 < \hat{\mu}_D < 19.7269$$

Example (2)

Advertisements by Sylph Fitness Center claim that completing its course will result in losing weight. A random sample of eight recent participants showed the following weights before and after completing the course.

- At the 0.01sigenificance level, can we conclude the student's weight is significantly increased? (In pounds)

-Find the confidence interval for μ_D

Note: 1 kg = 2.20 pounds

No	Before	After
1	155	154
2	228	207
3	141	147
4	162	157
5	211	196
6	164	150
7	184	170
8	172	165

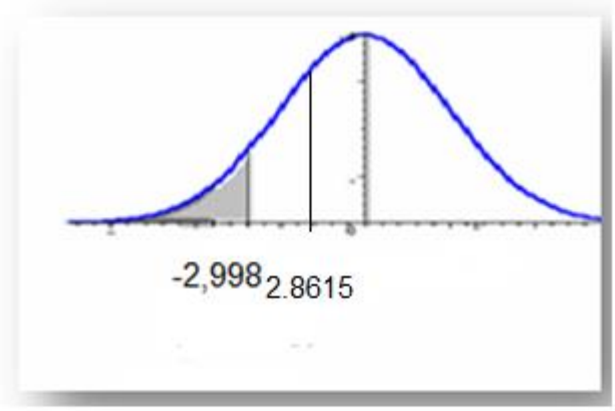
Solution:

Step (1): State the null and alternate hypotheses

(Where $u_1 < u_2$ or $u_1 - u_2 < 0$)

$H_0: \mu_D \geq 0$

$H_1: \mu_D < 0$



Step (2): Select the level of significance (α=0.01)

Step (3): The critical value (One-tailed test (Left)

$-t_{(\alpha,n-1)} = -t_{(0.01,7)}$=-2.998

Reject H_0 if

$t_c < -2.998$

Step (4): The test statistic

No	Before B	After A	D (X_2-X_1)	$\left(D_i - \overline{D}\right)$	$\left(D_i - \overline{D}\right)^2$
1	155	154	-1	-1+8.875=7.875	62.02
2	228	207	-21	-21+8.875=12.125	147.02
3	141	147	6	6+8.875=14.87	221.27
4	162	157	-5	-5+8.875=3.875	15.02
5	211	196	-15	-15+8.875=-6.125	37.52
6	164	150	-14	-14+8.875=-5.125	26.27
7	184	170	-14	-14+8.875=-5.125	26.27
8	172	165	-7	-7+8.875=1.875	3.52
Total			-71		538.66

$$\overline{D} = \frac{\sum D}{n} = \frac{-71}{8} = -8.875$$

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}} = \sqrt{\frac{538.66}{7}} = 8.7722$$

$$t_c = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{-8.875}{8.7722 / \sqrt{8}} = \frac{-8.875}{8.7722 / 2.8284} = \frac{-8.875}{3.1015} = -2.8615$$

Step (5): Formulate the Decision Rule

Do not reject H_o. We cannot conclude that the students lost weight

The Paired Difference Confidence Intervalμ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2} \frac{S_D}{\sqrt{n}}$$

$$= 8.875 \pm 3.499 \frac{8.7722}{\sqrt{8}}$$

$$= 8.875 \pm 10.8519$$

$$-1.9769 < \hat{\mu}_D < 19.7269$$

Example (3)

The management of Discount Furniture, a chain of discount furniture stores in the Northeast, designed an incentive plan for salespeople .To evaluate this innovative plan, 6 salespeople were selected at random, and their weekly income before and after the plan were recorded.

Salespeople	**Before**	**After**
1	$320	340
2	290	285
3	421	475
4	360	365
5	506	525
6	431	431

Was there a significant increase in the typical salesperson's weekly income due to the innovative incentive plan? Use the 0.05 significance level.

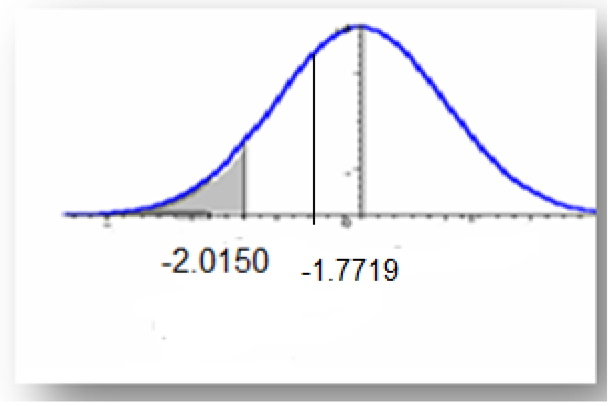
Solution:

Step (1): State the null and alternate hypotheses

(Where $u_1 < u_2$ or $u_1 - u_2 < 0$)

$H_{0:}\ \mu_D \geq 0$

$H_{1:}\ \mu_D < 0$



Step (2): Select the level of significance (α=0.05)

Step (3): The critical value

One-tailed test (Left)

$$-t_{(\alpha,n-1)} = -t_{(0.05,5)} = -2.0150$$

Reject H_0 **if** $t_c < -2.0150$

Step (4): The test statistic

Salespeople	**Before**	**After**	D **(X1-X2)**	$\left(D_i - \overline{D}\right)$	$\left(D_i - \overline{D}\right)^2$
1	$337	340	-3	-3+3=0	0
2	290	285	5	5+3=8	64
3	421	425	-4	-4+3=-1	1
4	360	365	-5	-5+3=-2	4
5	506	513	-7	-7+3=-4	16
6	431	435	-4	-4+3=-1	1
Total			-18		86

$$\overline{D} = \frac{\sum D}{n} = \frac{-18}{6} = -3$$

$$\mathrm{S_D} = \sqrt{\frac{\sum(\mathrm{D}_i - \overline{\mathrm{D}})^2}{\mathrm{n}-1}} = \sqrt{\frac{86}{5}} = 4.1473$$

$$t_c = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{-3}{4.1473 / \sqrt{6}} = \frac{-3}{4.1473 / 2.4495} = \frac{-3}{1..6931} = -1.7719$$

Step (5): Formulate the Decision Rule

Do not reject

The Paired Difference Confidence Interval μ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2} \frac{\mathrm{S_D}}{\sqrt{\mathrm{n}}}$$

$$= -3 \pm 2.5706 \frac{3.7683}{\sqrt{6}}$$

$$= -3 \pm (2.5706)(1.5384) = -3 \pm 3.9546$$

$$-6.9546 < \hat{\mu}_D < 0.9546$$

Example (4)

The management of Discount Furniture, a chain of discount furniture stores in the Northeast, designed an incentive plan for salespeople .To evaluate this innovative plan, 6 salespeople were selected at random, and their weekly income before and after the plan were recorded.

Salespeople	**Before**	**After**
1	$320	340
2	290	285
3	421	475
4	360	365
5	506	525
6	431	431

Was there a significant decrease in the typical salesperson's weekly income due to the innovative incentive plan? Use the 0.05 significance level.

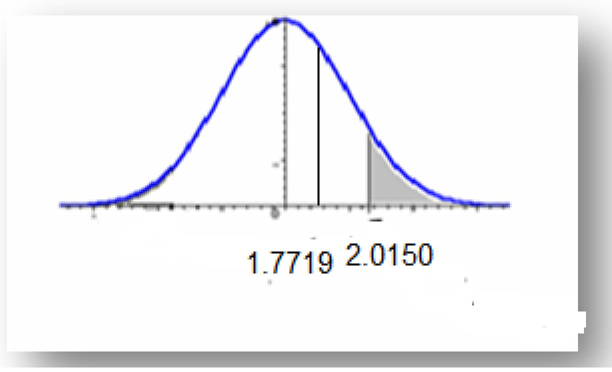
Solution:

Step (1): State the null and alternate hypotheses

(Where $u_2 < u_1$ or $u_2 - u_1 < 0$)

H_0: $\mu_D \leq 0$

H_1: $\mu_D > 0$



Step (2): Select the level of significance (α=0.05)

Step (3): The critical value

One-tailed test (Left)

$t_{(\alpha,n-1)} = t_{(0.05,5)}$ =2.0150

Reject H_0 **if**

$t_c > 2.0150$

Step (4): The test statistic

Salespeople	**Before**	**After**	D **(X2-X1)**	$\left(D_i - \overline{D}\right)$	$\left(D_i - \overline{D}\right)^2$
1	$337	340	3	3-3=0	0
2	290	285	-5	-5-3=-8	64
3	421	425	4	4-3=1	1
4	360	365	5	5-3=2	4
5	506	513	7	7-3=4	16
6	431	435	4	4-3=1	1
Total			18		86

$$\overline{D} = \frac{\sum D}{n} = \frac{18}{6} = 3$$

$$\mathrm{S_D} = \sqrt{\frac{\sum(\mathrm{D}_i - \overline{\mathrm{D}})^2}{\mathrm{n}-1}} = \sqrt{\frac{86}{5}} = 4.1473$$

$$t_c = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{3}{4.1473/\sqrt{6}} = \frac{3}{4.1473/2.4495} = \frac{3}{1..6931} = 1.7719$$

Step (5): Formulate the Decision Rule

Do not reject

The Paired Difference Confidence Intervalμ$_{\mathrm{D}}$ is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2} \frac{\mathrm{S_D}}{\sqrt{\mathrm{n}}}$$

$$= 3 \pm 2.5706 \frac{3.7683}{\sqrt{6}}$$

$$= 3 \pm (2.5706)(1.5384) = 3 \pm 3.9546$$

$$-0.9546 < \hat{\mu}_D < 6.9546$$