Statistical and Mathematical Methods for Data Analysis

Dr. Faisal Bukhari
Punjab University College of Information Technology (PUCIT)

Textbooks

- **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

- **Elementary Statistics: Picturing the World,** 6^{th} Edition, Ron Larson and Betsy Farber

- **Elementary Statistics**, 13^{th} Edition, Mario F. Triola

Reference books

- **Probability and Statistical Inference, Ninth Edition,** Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman

- **Probability Demystified**, Allan G. Bluman

- **Practical Statistics for Data Scientists: 50 Essential Concepts,** Peter Bruce and Andrew Bruce

- **Schaum's Outline of Probability,** Second Edition, Seymour Lipschutz, Marc Lipson

- **Python for Probability, Statistics, and Machine Learning,** José Unpingco

References

Readings for these lecture notes:

- Probability & Statistics for Engineers & Scientists, Ninth edition, Ronald E. Walpole, Raymond H. Myer

- www.unm.edu/~marley/statppt/fall06/002/day12.ppt

- https://www.technologynetworks.com/informatics/articles/one-way-vs-two-way-anova-definition-differences-assumptions-and-hypotheses-306553#:~:text=A%20one%2Dway%20ANOVA%20only,multiple%20groups%20of%20two%20factors.

These notes contain material from the above resources.

Independent and Dependent Samples.

- Two samples are **independent** if the sample values selected from **one population** are **not related to or somehow paired or matched** with the sample values selected from the other population.

- Two samples are **dependent** (or consist of **matched pairs**) if the members of one sample can be used to determine the members of the other sample. [Samples consisting of **matched pairs** (such as husband wife data) are **dependent**.

- In addition to **matched pairs of sample data**, **dependence** could also occur with samples related **through associations** such as **family members**.]

Confidence Interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observations

If $\bar{d}$ and s_d are the **mean** and **standard deviation**, respectively, of the normally distributed differences of n **random pairs of measurements**, a 100(1 − α)% confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{(\alpha/2, n-1)} \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{(\alpha/2, n-1)} \frac{s_d}{\sqrt{n}}$$

Where,

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n-1}} \text{ OR } s_d = \sqrt{\frac{1}{n(n-1)}\left\{n\sum_{i=1}^{n} d^2{}_i - \left(\sum_{i=1}^{n} d^2{}_i\right)^2\right\}}$$

$$s_d^2 = \frac{\sum(d - \bar{d})^2}{n-1} \text{ OR } s_d^2 = \frac{1}{n(n-1)}\left\{n\sum_{i=1}^{n} d^2{}_i - \left(\sum_{i=1}^{n} d^2{}_i\right)^2\right\}$$

$$d_i = x_{1i} - x_{2j} \text{ OR } d_i = x_{2i} - x_{1j}, \quad \bar{d} = \frac{\sum_{i=1}^{n} di}{n}$$

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

Testing Hypothesis about Paired Observation

a) H_o: $\mu_d = 0$

H_1:$\mu_d < 0$ (One tailed test)

Where $\mu_d = \mu_1 - \mu_2$

b) H_o: $\mu_d = 0$

H_1:$\mu_d > 0$ (One tailed test)

c) H_o: $\mu_d = 0$

H_1:$\mu_d \neq 0$ (Two tailed test)

Test statistic:

$$t_{cal} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}},$$

Where $d_i = x_{1i} - x_{2j}$ OR $d_i = x_{2i} - x_{1j}$

$$\bar{d} = \frac{\sum_{i=1}^{n} di}{n}$$

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}} \text{ OR}$$

$$s_d = \sqrt{\frac{1}{n(n-1)}\{n \sum_{i=1}^{n} d^2{}_i - (\sum_{i=1}^{n} d_i)^2\}}$$

The *t* Test for Dependent Samples: An Example

Eight individuals indicated their attitudes toward socialized medicine before and after listening to a pro–socialized medicine lecture. Attitudes were assessed on a scale from 1 to 7, with higher scores indicating more positive attitudes. The attitudes before and after listening to the lecture were as indicated in the second and third columns of the table. Test for a relationship between the time of assessment and attitudes toward socialized medicine using a correlated groups *t* test.

Individual	Before speech	After speech
1	3	6
2	4	6
3	3	3
4	5	7
5	2	4
6	5	6
7	3	7
8	4	6

Solution

$\mu_D = 0$ (Population mean)

$n = 8$ (Sample size)

$\alpha = 0.05$ (Level of significance)

$\bar{d} = ?$

$s_d = ?$

1. We state our hypothesis as:

H_o: $\mu_d = 0$

H_1: $\mu_d \neq 0$ (Two tailed test)

2. The level of significance is set $\alpha = 0.05$

3. Test statistic to be used is

$$t_{cal} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

4. Calculations:

Before speech	After speech	$d_i = x_{1i} - x_{2j}$	$d^2{}_i$
3	6	-3	9
4	6	-2	4
3	3	0	0
5	7	-2	4
2	4	-2	4
5	6	-1	1
3	7	-4	16
4	6	-2	4
Sum		$\sum_{i=1}^{n} di = -16$	$\sum_{i=1}^{n} d^2{}_i = 42$

$$\bar{d} = \frac{\sum_{i=1}^{n} \mathrm{di}}{n} = -16/2 = \mathbf{-2}$$

$$s_d = \sqrt{\frac{1}{n(n-1)}\{n\sum_{i=1}^{n} \mathrm{d}^2{}_i - (\sum_{i=1}^{n} \mathrm{d}_i)^2\}}$$

$$s_d = \sqrt{\frac{1}{8(8-1)}\{8(42) - (-16)^2\}} = \sqrt{\frac{80}{8(8-1)}} = 1.1952$$

$$t_{cal} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = t_{cal} = \frac{-2 - 0}{\frac{1.1952}{\sqrt{8}}} = \frac{-2}{0.4226}$$

$t_{cal} = -4.7326$

$|t_{cal}| = 4.7326$

5. **Critical region:**

$|t_{cal}| > t_{tab}$, where $t_{tab} = t_{(\alpha/2,\, n-1)}$

Where $t_{tab} = t_{(\alpha/2,\, n-1)} = t_{(0.0250,\, 7)} = 2.365$

6. **Conclusion:** Since calculated value of t_{cal} is greater than t_{tab}, so we reject H_O

Interpret your results.

After the **pro-socialized medicine lecture**, individuals' attitudes toward **socialized medicine** were significantly more positive than before the lecture.

Table A.4 Critical Values of the t-Distribution

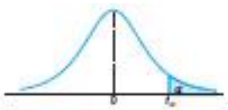

Table A.4 Critical Values of the t-Distribution

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Table A.4 (continued) Critical Values of the t-Distribution

	α						
v	0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290

Analysis of variance (ANOVA)

- **Analysis of variance (ANOVA)** is a method of testing the **equality of three or more population means** by analyzing **sample variances**.

One-Way Analysis of Variance vs Two-Way Analysis of Variance

	One-Way ANOVA	Two-Way ANOVA
Definition	A test that allows one to make comparisons between the means of three or more groups of data.	A test that allows one to make comparisons between the means of three or more groups of data, where two independent variables are considered.
Number of Independent Variables	One.	Two.
What is Being Compared?	The means of three or more groups of an independent variable on a dependent variable.	The effect of multiple groups of two independent variables on a dependent variable and on each other.
Number of Groups of Samples	Three or more.	Each variable should have multiple samples.

One-Way Analysis of Variance: Completely Randomized Design (One-Way ANOVA)

Random samples of **size *n*** are selected from each of ***k*** **populations**. The ***k*** **different populations** are classified on the basis of a **single criterion** such as different **treatments or groups**.

Today the term **treatment** is used generally to refer to the **various classifications**, whether they be different **aggregates**, different **analysts**, different **fertilizers**, or different **regions** of the country.

Assumptions and Hypotheses in One-Way ANOVA

- It is assumed that the ***k*** **populations** are **independent** and **normally distributed** with means $\mu_1, \mu_2, . . . , \mu_k$ and **common variance** σ^2.

- We wish to derive appropriate methods for testing the hypothesis

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

H_1: At least two of the means are not equal

- Let y_{ij} denote the **jth observation** from the **ith treatment** and arrange the data as the table on next slide.

- Here, Y_i is the **total of all observations** in the sample from the **ith treatment**, $\overline{y_{i.}}$ is the mean of **all observations** in the sample from the **ith treatment**,

- $Y_{..}$ is the total of all n_k observations, and $\overline{y_{i..}}$ is the mean of all n_k observations.

k Random Samples

Treatment:	**1**	**2**	$\cdots$	$\boldsymbol{i}$	$\cdots$	$\boldsymbol{k}$	
	y_{11}	y_{21}	$\cdots$	y_{i1}	$\cdots$	y_{k1}	
	y_{12}	y_{22}	$\cdots$	y_{i2}	$\cdots$	y_{k2}	
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
	y_{1n}	y_{2n}	$\cdots$	y_{in}	$\cdots$	y_{kn}	
Total	$Y_{1.}$	$Y_{2.}$	$\cdots$	$Y_{i.}$	$\cdots$	$Y_{k.}$	$Y_{..}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{i.}$	$\cdots$	$\bar{y}_{k.}$	$\bar{y}_{..}$

Sum-of-Squares Identity

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \text{total sum of squares},$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 = \text{treatment sum of squares},$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 = \text{error sum of squares}.$$

SST = SSA + SSE.

Treatment Mean Square vs Error Mean Square (MSE)

- **Treatment Mean Square:**

$$s^2 = \frac{SSA}{k-1}$$

- **Error Mean Square (MSE):**

$$s^2 = \frac{SSE}{k(n-1)}$$

Use of *F*-Test in ANOVA

- The **estimate s^2** is **unbiased** regardless of the truth or falsity of the null hypothesis.

- It is important to note that the **sum-of-squares** identity has **partitioned** not only the **total variability of the data**, but also the **total number of degrees** of freedom.

  That is, $nk - 1 = k - 1 + k(n - 1)$.

F-Ratio for s^2 Testing Equality of Means

- When H_0 is **true**, the ratio $f_{cal} = \frac{s_1^2}{s^2}$ is a value of the random variable F having the ***F*-distribution** with $k-1$ and $k(n-1)$ degrees of freedom.

- Since s_1^2 overestimates σ^2 when H_0 is **false**, we have a one-tailed test with the critical region entirely in the right tail of the distribution.

- The **null hypothesis** H_0 is rejected at the α-level of significance when $f_{cal} > f_\alpha[k-1, k(n-1)]$.

Analysis of Variance for the One-Way ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	SSA	$k-1$	$s_1^2 = \frac{SSA}{k-1}$	$\frac{s_1^2}{s^2}$
Error	SSE	$k(n-1)$	$s^2 = \frac{SSE}{k(n-1)}$	
Total	SST	$kn-1$		

Example 13.1: Test the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_5$ at the 0.05 level of significance for the data of Table 13.1 on absorption of moisture by various types of cement aggregates.

Table 13.1: Absorption of Moisture in Concrete Aggregates

Aggregate:	1	2	3	4	5	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	16,854
Mean	553.33	569.33	610.50	465.17	610.67	561.80

***Solution*:**

1. We state our hypothesis as

$H_0 : \boldsymbol{\mu_1 = \mu_2 = \cdots = \mu_5}$

H_1: At least two of the means are not equal

2. The level of significance is set α = 0.05.

3. Test statistic to be used is

$$\boldsymbol{f_{cal} = \frac{s_1^2}{s^2}}$$

4. Calculations:

Formulae

1. **SST =** $\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i..})^2$ **= total sum of squares**

2. **SSA =** $n \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{i..})^2$ **= treatment sum of squares**

3. **SSE =** $\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$ **= error sum** of squares

Aggregate	1	2	3	4	5	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	**16854**
Mean	553.33	569.33	610.50	465.17	610.67	**561.8**

$\overline{y_{i..}}$ = **16854/30 = 561.80**

Or

$\overline{y_{i..}}$ **= 2809/5 = 561.80**

SSA = $\mathbf{n} \sum_{i=1}^{k} (\overline{y_{i.}} - \overline{y_{i..}})^2$ = treatment sum of squares

= 6(553.33 - **561.80**)^2 + 6(569.33 - **561.80**)^2
+ 6(610.50 - **561.80**)^2 + 6(465.17 - **561.80**)^2
+ 6(610.50 - **561.80**)^2

= 6(-6.47)^2 + 6(7.53)^2+ 6(48.7)^2 + 6(-96.63)^2 + 6(48.7)^2

= 6(14179.2987)= **85075.7922**

SSE = $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$ = error sum of squares

1	2	3	4	5
$(y_{ij} - \bar{y}_{i.})^2$ $=(y_{ij} - 553.33)^2$	$(y_{ij} - \bar{y}_{i.})^2$ $= (y_{ij} - 569.33)^2$	$(y_{ij} - \bar{y}_{i.})^2$ $= (y_{ij} - 610.50)^2$	$(y_{ij} - \bar{y}_{i.})^2$ $= (y_{ij} - 465.17)^2$	$(y_{ij} - \bar{y}_{i.})^2$ $= (y_{ij} - 610.67)^2$
5.4289	658.9489	812.25	2320.349	2272.114
9279.469	113.8489	20.25	261.4689	413.4431
10677.09	3761.369	9900.25	2686.349	7861.784
31566.63	186.8689	1406.25	738.2089	5.444289
2951.749	4053.869	1406.25	2517.029	2055.108
6188.969	2738.429	4422.25	8069.429	4669.44
60669.33	**11513.33**	**17967.5**	**16592.83**	**17277.33**

SSE = $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2$

= 60669.33 + 11513.33 + 17967.5

+ 16592.83 + 17277.33

= **124020.3**

SSE = $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2$ **= error sum** of squares

***SST* =** *SSA* + *SSE*

= 85075.7922 + 124020.3

= **209096.0922**

Method 2

We can first find **SSA**. Then we can find **SST** using formula given below:

SST = $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \overline{y_{i..}})^2$ **= total sum of squares**

After finding **SSA** and **SST**, we will use the following formula to find **SSE**:

SST = SSA + SSE

$\Rightarrow$ ***SSE = SST - SSA***

Analysis of Variance for the One-Way

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	Calculated f
Treatments	SSA = 209377	k – 1 = 5 – 1 = 4	$s_1^2 = \frac{SSA}{k-1}$ $= \frac{209377}{4}$ = 21339.1167	$f_{cal} = \frac{s_1^2}{s^2}$ $= \frac{21339.1167}{4960.8133}$ = 4.30
Error	SSE = 124021	k(n - 1) = 5 (6-1) = 25	$s^2 = \frac{124021}{25}$ = 4960.8133	
Total	SST = 209096.0922	kn -1 = (5)(6) -1 = 29		

5. Critical region:

$f_{cal} > f_{\alpha}[k - 1, k(n - 1)]$

$f_{tab} = f_{\alpha}[k - 1, k(n - 1)] = f_{0.05}[4, 25]$
$= 2.76$

$4.30 > 2.76$

6. Conclusion: Reject H_0

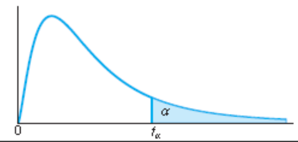

Table A.6 Critical Values of the *F*-Distribution

	$f_{0.05}(v_1, v_2)$								
	v_1								
v_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table A.6 (continued) Critical Values of the F-Distribution

	$f_{0.01}(v_1, v_2)$								
	v_1								
v_2	1	2	3	4	5	6	7	8	9
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table A.6 (continued) Critical Values of the *F*-Distribution

	$f_{0.01}(v_1, v_2)$									
	v_1									
v_2	10	12	15	20	24	30	40	60	120	∞
1	6055.85	6106.32	6157.28	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00