

Week-04-L-06

Select a Statistical Model

Exercises on Classification & Regression

Dr. Amandeep Singh
Imaging Laboratory
Indian Institute of Technology Kanpur



Exercise on Classification



Problem Statement

The following are the details of yield effected by three different pesticides in different fields with same four different crops.

<i>Brand of pesticide</i>	<i>Yield brought to a scale of 10</i>			
	<i>Wheat</i>	<i>Rice</i>	<i>Red Lentil</i>	<i>Sugarcane</i>
<i>x</i>	8	9	5	10
<i>y</i>	7	6	6	9
<i>z</i>	6	6	7	5

Find out whether there is any significant difference in the performance of the three pesticides.

Exercise on Classification

$$\sum x = 8 + 9 + 5 + 10 = 32$$
$$\sum y = 7 + 6 + 6 + 9 = 28$$
$$\sum z = 6 + 6 + 7 + 5 = 24$$

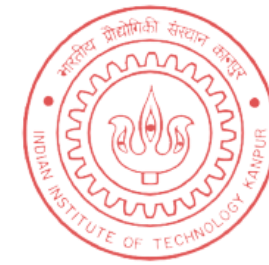
- We first find the sum of all the three varieties:

$$\sum x = 8 + 9 + 5 + 10 = \underline{32}$$

$$\sum y = 7 + 6 + 6 + 9 = \underline{28}$$

$$\sum z = 6 + 6 + 7 + 5 = \underline{24}$$

Exercise on Classification



- *Sample mean for $\bar{x} = \frac{32}{4} = 8$*
- *Sample mean for $\bar{y} = \frac{28}{4} = 7$*
- *Sample mean for $\bar{z} = \frac{24}{4} = 6$*

- *Total number sample items = No. of items for x + No. of items for y + No. of items for z*
= 4+4+4 = 12 (N)

- *Mean of all the samples $\bar{M} = \frac{32+28+24}{12} = \frac{84}{12} = 7$*

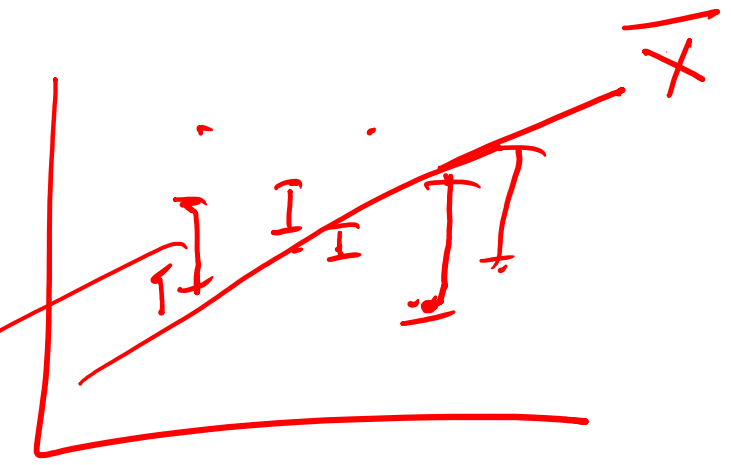
Exercise on Classification



ideas to products
IMAGINEERING
LAB | IIT KANPUR

MedTech
IIT KANPUR

- Sum of Squares of deviation of X:

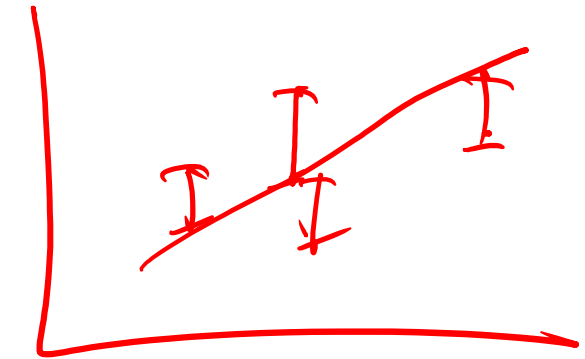


X	$X - \bar{X} = X - 8$	$(X - \bar{X})^2$
8	0	0
9	1	1
5	-3	9
10	2	4
		14

Exercise on Classification



- *Sum of Squares of deviation of Y:*



Y	$Y - \bar{Y} = Y - 7$	$(Y - \bar{Y})^2$
7	0 ✓	0 ✓
6	-1 ✓	1 ✓
6	-1 ✓	1 ✓
9	2 ✓	4 ✓
		6

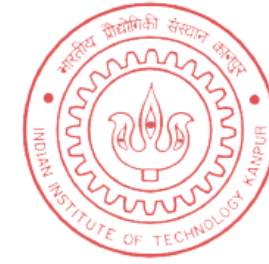
Exercise on Classification



- *Sum of Squares of deviation of Z:*

Z	$Z - \bar{Z} = Z - 6$	$(Z - \bar{Z})^2$
6	0	0
6	0	0
7	-1	1
5	-1	1
		2

Exercise on Classification



- *Sum of squares of deviations within* = $\sum (X - \bar{X})^2 + \sum (Y - \bar{Y})^2 + \sum (Z - \bar{Z})^2$
= $14 + 6 + 2$
= 22

Exercise on Classification



Sum of squares of deviations for total variance:

<u>Pesticide</u>	<u>Crop</u>	<u>Yield</u>	<u>Yield-Mean</u>	<u>(Yield-Mean)²</u>
$x=A$ A	Wheat	8	1 ✓	1 ✓
A	Rice	9	2	4
A	Pulse A	5	-2	4
$y=A$ A	Sugarcane	10	3	9
$x=B$ B	Wheat	7	0	0
B	Rice	6	-1	1
B	Pulse A	6	-1	1
B	Sugarcane	9	2	4
$z=C$ C	Wheat	6	-1	1
C	Rice	6	-1	1
C	Pulse A	7	0	0
C	Sugarcane	5	2	4
				30

Exercise on Classification

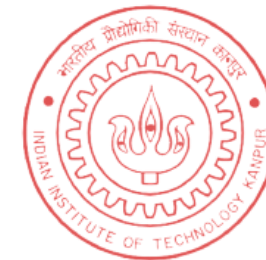


ANOVA TABLE

SS
DoF

<i>Source of Variation</i>	<i>Degrees of Freedom</i>	<i>Sum of Squares of Deviations</i>	<i>Variance</i>
<u><i>Between Varieties</i></u>	<u>$3-1=2$</u>	<u>8</u>	$\frac{8}{2} = 4$
<u><i>Within Varieties</i></u>	<u>$12-3=9$</u>	<u>22</u>	$\frac{22}{9} = 2.44$
<i>Total</i>	<i>12-1</i> <i>11</i>	<i>30</i>	

Exercise on Classification



Calculation of F value:

$$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{4.0}{2.44} = 1.6393$$

Degrees of freedom for greater variance (df_1) = 2

Degrees of freedom for smaller variance (df_2) = 9

Let us take level of significance as 5%

The table value of $F = 4.26$

Exercise on Classification



	DF1	$\alpha = 0.10$								
DF2	1	2	3	4	5	6	7	8	9	10
1	39.863	49.5	53.593	55.833	57.24	58.204	58.906	59.439	59.858	60.195
2	8.5263	9	9.1618	9.2434	9.2926	9.3255	9.3491	9.3668	9.3805	9.3916
3	5.5383	5.4624	5.3908	5.3426	5.3092	5.2847	5.2662	5.2517	5.24	5.2304
4	4.5448	4.3246	4.1909	4.1073	4.0506	4.0098	3.979	3.9549	3.9357	3.9199
5	4.0604	3.7797	3.6195	3.5202	3.453	3.4045	3.3679	3.3393	3.3163	3.2974
6	3.776	3.4633	3.2888	3.1808	3.1075	3.0546	3.0145	2.983	2.9577	2.9369
7	3.5894	3.2574	3.0741	2.9605	2.8833	2.8274	2.7849	2.7516	2.7247	2.7025
8	3.4579	3.1131	2.9238	2.8064	2.7265	2.6683	2.6241	2.5894	2.5612	2.538
9	3.3603	3.0065	2.8129	2.6927	2.6106	2.5509	2.5053	2.4694	2.4403	2.4163
10	3.285	2.9245	2.7277	2.6053	2.5216	2.4606	2.414	2.3772	2.3473	2.3226
11	3.2252	2.8595	2.6602	2.5362	2.4512	2.3891	2.3416	2.304	2.2735	2.2482
12	3.1766	2.8068	2.6055	2.4801	2.394	2.331	2.2828	2.2446	2.2135	2.1878
13	3.1362	2.7632	2.5603	2.4337	2.3467	2.283	2.2341	2.1954	2.1638	2.1376
14	3.1022	2.7265	2.5222	2.3947	2.3069	2.2426	2.1931	2.1539	2.122	2.0954

Exercise on Classification



	DF1	$\alpha = 0.05$									
DF2	1	2	3	4	5	6	7	8	9	10	
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	
2	18.513	19	19.164	19.247	19.296	19.33	19.353	19.371	19.385	19.396	
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.041	5.9988	5.9644	
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.099	4.06	
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.866	3.787	3.7257	3.6767	3.6365	
8	5.3177	4.459	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	
10	4.9646	4.1028	3.7083	3.478	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.948	2.8962	2.8536	
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.671	
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	

Exercise on Classification

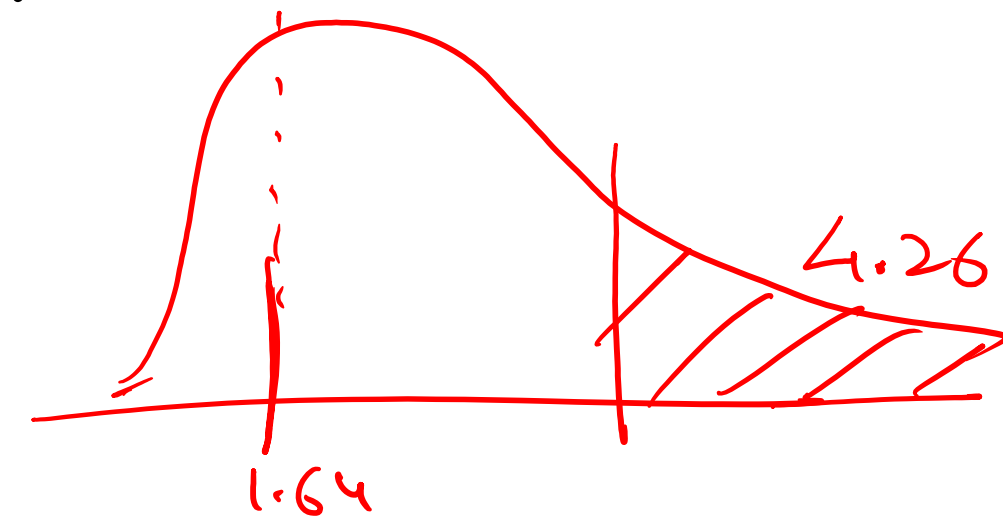


Inference:

On doing the calculation we observe that the table value of F is larger than the calculated value of F . 1.64 4.26

Now, we can say that null hypothesis is accepted.

It's concluded that there is no significant difference in the performance of the different varieties of pesticides used here, at 5% level of significance.



Exercise on Regression



Problem Statement

The data has been provided in the following table:

Factors:

	Fertilizer (g)	Spacing (cm)
Minimum	5	2
Maximum	15	18

Responses:

Height (cm)
Weight (kg)

Exercise on Regression



RSM

$\sum 10, 15 = 3$
 $\sum 10, 18 = 3$

$3+3$

Std	Run	Block	Factor 1 Fertilizer (g)	Factor 2 Spacing (cm)	Response 1 Height (cm)	Response 2 Weight (kg)
1	8	Block 1	5	2	69	1.16
2	10	Block 1	10	2	73	1.18
3	2	Block 1	15	2	65	1.09
4	11	Block 1	5	10	71	1.21
5	6	Block 1	10	10	76	1.29
6	5	Block 1	15	10	67	1.12
7	1	Block 1	5	18	74	1.27
8	9	Block 1	10	18	79	1.31
9	7	Block 1	15	18	68	1.13
10	3	Block 1	10	10	75	1.29
11	4	Block 1	10	10	76	1.28

I.V.

D.V

Exercise on Regression



ANOVA Table for weight

SS / df

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob>F
Model	0.061 ✓	5	0.012 ✓	28.99 ✓	0.0011 ✓
A-Fertilizer	0.015	1	0.015 ✓	35.39 ✓	0.0019 ✓
B-Spacing	0.013	1	0.013 ✓	30.83 ✓	0.0026 ✓
AB	0.001	1	0.001 ✓	02.89	0.1498
A ² ✓	0.025	1	0.025 ✓	61.04	0.0006
B ² ✓	0.001	1	0.001 ✓	2.64	0.1645
Residual	0.002	5	0.001 ✓		

< 0.05

Quadratic

Exercise on Regression



Regression Models

Fert = 108
 Spacing = 18 cm +
 Decision.

$W = \beta_0 + \beta_1 \text{Fert} + \beta_2 \text{Spacing}$

Weight

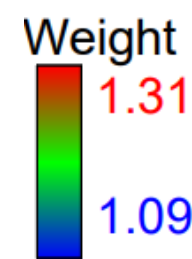
$$= 0.84 \times \text{Fertilizer} + 0.07 \times \text{Spacing} + 0.01 \times \text{Fertilizer} \times \text{Spacing} - 0.0004 \times \text{Fertilizer}^2 - 0.0040 \times \text{Spacing}^2$$

Coded Factors

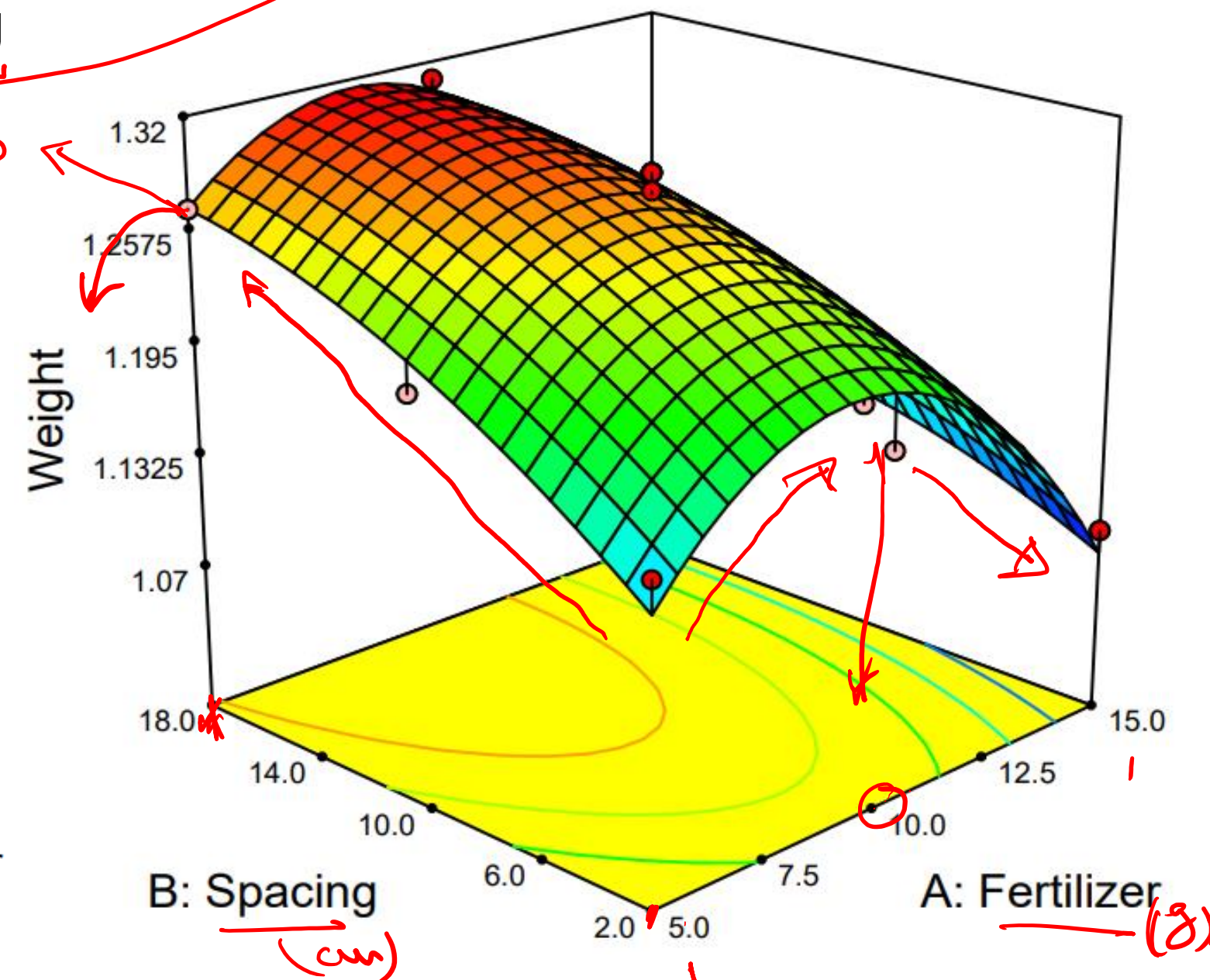
$$= 1.28 - 0.05 \times A + 0.04 \times B - 0.01 \times A \times B - 0.10 \times B^2 - 0.02 \times A^2$$

$R^2 = 0.97$ (Linear)
 $R^2 = 0.91$ Linear

R-Squared	0.96
Adj R-Squared	0.93



X1 = A: Fertilizer
 X2 = B: Spacing



Thank you

