Unit 4

4.1 Karl Pearson's Correlation Co-efficient

Correlation is the study of relationship between two independent variables.

Karl pearson's correlation co-efficient is

$$r = r(x, y) = r_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

where,

$$\begin{aligned}
cov(x, y) &= \frac{\sum xy}{n} - \overline{x}\,\overline{y} \\
\sigma_x &= \sqrt{\frac{\sum x^2}{n} - (\overline{x})^2} \\
\sigma_y &= \sqrt{\frac{\sum y^2}{n} - (\overline{y})^2}
\end{aligned}$$

n is the number of data

$$\overline{x} = \frac{\sum x}{n}$$

$$\overline{y} = \frac{\sum y}{n}$$

Note:

1. Correlation co-efficient between -1 and 1. i.e., $-1 \leq r \leq 1$
2. $r = \dfrac{N \sum XY - (\sum X).(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2}\sqrt{N \sum Y^2 - (\sum Y)^2}}$

Problem 1 *Calculate the Karl pearson's co-efficient of correlation to the following data.*

x	65	66	67	67	68	69	70	72
y	*67*	*68*	*65*	*68*	*72*	*72*	*69*	*71*

.

Solution:

X	Y	X^2	Y^2	XY
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
544	552	37028	38132	37560

.

$$\begin{aligned} r &= \frac{N\sum XY - (\sum X).(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}} \\ &= \frac{(8\times 37560) - (544).(552)}{\sqrt{(8\times 37028) - (544)^2}\sqrt{(8\times 38132) - (552)^2}} = 0.6047 \end{aligned}$$

4.2 Rank correlation

Spearsman's rank correlation coefficient

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Where, $d_i = x_i - y_i$

Note: If ranks are repeated, then

$$\rho = 1 - \frac{6\left[\sum d_i^2 + C.F_1 + C.F_2 + \cdots\right]}{n(n^2-1)}$$

Where, $d_i = x_i - y_i$

C.F's are correction factor and it can be calculated by $C.F = \dfrac{m(m^2-1)}{12}$ Here m is the number of times, the data has been repeated.

Problem 1 *Calculate the spearsman's rank correlation to the following data.*

x	68	64	75	50	64	80	75	40	55	64
y	*62*	*58*	*68*	*45*	*81*	*60*	*68*	*48*	*50*	*70*

.

Solution:

X	Y	Rank of X	Rank of Y	$d_i = x_i - y_i$	d_i^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	-5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\sum d_i^2 = 72$

.

In value of X,

75 is repeated 2 times and which having the rank as 2 and 3. $\therefore$ the rank of 75 $= \dfrac{2+3}{2} = 2.5$ and

$C.F_1 = \dfrac{m(m^2-1)}{12} = \dfrac{2(2^2-1)}{12} = 0.5$

64 is repeated 3 times and which having the rank as $5, 6$ and 7. $\therefore$ the rank of 64 $= \dfrac{5+6+7}{3} = 6$ and

$C.F_2 = \dfrac{m(m^2-1)}{12} = \dfrac{3(3^2-1)}{12} = 2$

In value of Y,

68 is repeated 2 times and which having the rank as 3 and 4. $\therefore$ the rank of 68 $= \dfrac{3+4}{2} = 3.5$ and

$C.F_3 = \dfrac{m(m^2-1)}{12} = \dfrac{2(2^2-1)}{12} = 0.5$

$$
\begin{aligned}
\therefore\ \rho &= 1 - \frac{6\left[\sum d_i^2 + C.F_1 + C.F_2 + C.F_3\right]}{n(n^2-1)} \\
&= 1 - \frac{6\left[72 + 0.5 + 2 + 0.5\right]}{10(10^2-1)} \\
&= 1 - 0.4545 \\
&= 0.5454
\end{aligned}
$$

Exercise

Problem 1 *10 competitors in a musical contest were ranked by 3 judges x, y and z. Find out which pair of judges having the same likings of music.*

x	1	2	3	4	5	6	7	8	9	10
y	10	6	7	9	5	4	3	2	1	8
z	8	10	9	7	6	5	4	3	2	1

Ans.: $\because$ ρ_{zx} is greater than the ρ_{xy} and ρ_{yz} x and z having the same likings of music.

4.3 Regression

Regression is the mathematical study of average relationship between the independent variables x and y.

Lines of regression of x on y

$$(x - \overline{x}) = b_{xy}(y - \overline{y})$$

Lines of regression of y on x

$$(y - \overline{y}) = b_{yx}(x - \overline{x})$$

where b_{xy} and b_{yx} are regression co-efficients. It is given by

$$b_{xy} = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sum(y - \overline{y})^2} \text{ and } b_{yx} = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sum(x - \overline{x})^2}$$

Note:

$$r = \sqrt{b_{xy}\, b_{yx}}$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

The point of intersection of the lines of regression of y on x and x on y is the mean value of x and y.

Problem 1 *From the following data find*

1. *Two lines of regressions*
2. *Coefficient of correlation between the marks of economics and statistics*
3. *The most likely marks in statistics when the marks in economics is 30.*

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution: Let x be marks in Economics and y be marks in Statistics

$$\overline{x} = \frac{\sum x}{n} = \frac{320}{10} = 32 \text{ and } \overline{y} = \frac{\sum y}{n} = \frac{380}{10} = 38$$

x	y	$(x-\overline{x})$	$(y-\overline{y})$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380	0	0	140	398	-93

.

$$b_{xy} = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(y-\overline{y})^2}$$
$$= \frac{-93}{398} = -0.2336$$

and

$$b_{yx} = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$$
$$= \frac{-93}{140} = -0.6642$$

correlation co-efficient is $= \sqrt{b_{xy}\, b_{yx}} = \sqrt{-0.2336 \times -0.6642} = 0.393$

Line of regression of x on y is $(x-\overline{x}) = b_{xy}(y-\overline{y})$

$$(x-32) = -0.2336(y-38)$$
$$x - 32 = -0.2336y + 8.8768$$
$$x = -0.2336y + 8.8768 + 32$$
$$x = -0.2336y + 40.8768 - - - - (1)$$

Line of regression of y on x is$(y-\overline{y}) = b_{yx}(x-\overline{x})$

$$(y-38) = -0.6642(x-32)$$
$$y - 38 = -0.6642x + 21.2544$$
$$y = -0.6642x + 21.2544 + 38$$
$$y = -0.6642x + 59.2544 - - - - (2)$$

Now, to find y when $x = 30$

$$eqn.(2) \Rightarrow y = -0.6642(30) + 59.2544 = 39.3284$$

$$\therefore \text{ Marks in Statistics} = 39.32$$

Problem 2 *Two variables x and y have the regression lines* $3x + 2y - 26 = 0,\ 6x + y - 31 = 0$ *find the*

1. *mean value of x and y*

2. *correlation co-efficient between x and y*

3. *the variance of y when the variance of x is 25*

Solution:

$$\text{Given} \quad 3x + 2y - 26 = 0 \tag{1}$$

$$6x + y - 31 = 0 \tag{2}$$

1. mean value of x and y

   Solving (1) and (2), we get $x = 4$ and $y = 7$

   $\therefore \overline{x} = 4$ and $\overline{y} = 7$

2. correlation co-efficient between x and y

   Let $3x + 2y - 26 = 0$ be line of regression of x on y

   Then

   $$3x + 2y - 26 = 0 \Rightarrow 3x = -2y + 26 \Rightarrow x = -\frac{2}{3}y + 12$$

   $$\therefore b_{xy} = -\frac{2}{3}$$

   Let $6x + y - 31 = 0$ be line of regression of y on x

   Then

   $$6x + y - 31 = 0 \Rightarrow y = -6x + 31 \Rightarrow y = -6x + 31$$

   $$\therefore b_{yx} = -6$$

   $$r = \sqrt{b_{xy}\, b_{yx}} = \sqrt{-\frac{2}{3} \times -6} > 2$$

   Since the correlation coefficient should not exceed 1, $3x + 2y - 26 = 0$ can not be a line of regression of x on y and $6x + y - 31 = 0$ can not be a line of regression of y on x. $\therefore$ we have to consider

$3x + 2y - 26 = 0$ be line of regression of y on x

$$3x + 2y - 26 = 0 \Rightarrow 2y = -3x + 26 \Rightarrow y = -\frac{3}{2}y + 13$$

$$\therefore b_{yx} = -\frac{3}{2}$$

and consider $6x + y - 31 = 0$ be line of regression of x on y

$$6x + y - 31 = 0 \Rightarrow 6x = -y + 31 \Rightarrow x = -\frac{1}{6}y + \frac{31}{6}$$

$$\therefore b_{xy} = -\frac{1}{6}$$

$$r = \sqrt{b_{xy}\, b_{yx}} = \sqrt{-\frac{3}{2} \times -\frac{1}{6}} = 0.5 < 1$$

3. the variance of y when the variance of x is 25 ($\sigma_x^2 = 25$)

   i.e., $\sigma_x = 5$, we have to find σ_y

$$\begin{aligned} b_{xy} &= r\frac{\sigma_x}{\sigma_y} \\ \sigma_y &= r\frac{\sigma_x}{b_{xy}} \\ &= 0.5\frac{5}{-\frac{1}{6}} = -15 \\ \sigma_y^2 &= 225 \end{aligned}$$

$- - - - - - - -$

4.4 One way classification – CRD

Working Procedure:

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

H_1 :Not all equal.

x_1	x_2	$\cdots$	x_c	x_1^2	x_2^2	$\cdots$	x_c^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sum x_1$	$\sum x_2$	$\cdots$	$\sum x_c$	$\sum x_1^2$	$\sum x_2^2$	$\cdots$	$\sum x_c^2$

Step1:

$N =$ **Total No. of observations** $= r \times c$.   "r and s are no. of rows and columns in the given data"

Step2:

$$T = \sum x_1 + \sum x_2 + \cdots + \sum x_c$$

Step3:

$$C.F. = \frac{T^2}{N}$$

Step4:

$$TSS = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_c^2 - C.F.$$

Step5:

$$SSC = \frac{\left(\sum x_1\right)^2}{n_1} + \frac{\left(\sum x_2\right)^2}{n_2} + \cdots + \frac{\left(\sum x_c\right)^2}{n_c} - C.F.$$

where $n_1, n_2, \cdots, n_c$ are no. of entries in each columns.

Step6:

$$SSE = TSS - SSC$$

Step7:

one way

ANOVA table

Source of variation	Sum of squares	d.f.	Mean square	$F-$ ratio	$F-$ table value
Between columns	SSC	k $c - 1$	$MSC = \dfrac{SSC}{c-1}$	$F = \begin{cases} \dfrac{MSC}{MSE} & \text{if } MSC > MSE \\ \dfrac{MSE}{MSC} & \text{if } MSE > MSC \end{cases}$	$F_{0.05}(Nr, Dr)$
Error	SSE	$N - c$ k	$MSE = \dfrac{SSE}{N-c}$		

Here Nr =corresponding degrees of freedom of the Numerator in $F-$ ratio

and Dr =corresponding degrees of freedom of the Denominator in $F-$ ratio.

Step8:

Inference:

If Cal.F <tab. F, we accept the H_0.

If Cal.F >tab. F, we reject the H_0.

	B_1	B_2	B_3	B_4
V_1				
V_2				
V_3				

4.5 Two way classification – RBD

Working Procedure:

H_0 : There is no significant difference between row factor and column factor.

H_1 : There is a significant difference between row factor and column factor.

	x_1	x_2	$\cdots$	x_c	Total	x_1^2	x_2^2	$\cdots$	x_c^2
y_1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\sum y_1$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\sum y_2$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_r	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\sum y_r$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\sum x_1$	$\sum x_2$	$\cdots$	$\sum x_c$	T	$\sum x_1^2$	$\sum x_2^2$	$\cdots$	$\sum x_c^2$

Step1:

N = **Total No. of observations** $= r \times c$. "r and s are no. of rows and columns in the given data"

Step2:

$T = \sum x_1 + \sum x_2 + \cdots + \sum x_c$

Step3:

$C.F. = \dfrac{T^2}{N}$

Step4:

$TSS = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_c^2 - C.F.$

Step5:

$SSC = \dfrac{(\sum x_1)^2}{n_1} + \dfrac{(\sum x_2)^2}{n_2} + \cdots + \dfrac{(\sum x_c)^2}{n_c} - C.F.$

where $n_1, n_2, \cdots, n_c$ are no. of entries in each columns.

Step6:

$SSR = \dfrac{(\sum y_1)^2}{m_1} + \dfrac{(\sum y_2)^2}{m_2} + \cdots + \dfrac{(\sum y_r)^2}{m_r} - C.F.$

where $m_1, m_2, \cdots, m_r$ are no. of entries in each rows.

Step6:

$SSE = TSS - SSC - SSR$

Step7:

two-way **ANOVA table**

Source of variation	Sum of squares	d.f.	Mean square	$F-$ ratio	$F-$ table value
Between columns	SSC	$c-1$	$MSC = \frac{SSC}{c-1}$	$F_C = \begin{cases} \frac{MSC}{MSE} & \textbf{if } MSC > MSE \\ \frac{MSE}{MSC} & \textbf{if } MSE > MSC \end{cases}$	$F_{0.05}(Nr, Dr)$
Between rows	SSR	$r-1$	$MSR = \frac{SSR}{r-1}$	$F_R = \begin{cases} \frac{MSR}{MSE} & \textbf{if } MSR > MSE \\ \frac{MSE}{MSR} & \textbf{if } MSE > MSR \end{cases}$	$F_{0.05}(Nr, Dr)$
Error	SSE	$(r-1)(c-1)$	$MSE = \frac{SSE}{(r-1)(c-1)}$		
Total	TSS	$(rc-1)$			

Here Nr =corresponding degrees of freedom of the Numerator in $F-$ ratio

and Dr =corresponding degrees of freedom of the Denominator in $F-$ ratio.

Step8:

Inference:

For Between columns:

If Cal.F_C <tab. F, we accept the H_0.

If Cal.F_C >tab. F, we reject the H_0.

For Between rows:

If Cal.F_R <tab. F, we accept the H_0.

If Cal.F_R >tab. F, we reject the H_0.

Problems

Problem 1 *Four machines A, B, C, D are used to produce a certain kind of cotton fabric. Samples of size 4 with each unit as 100 square meters are selected from the outputs of the machines at random and the number of flaws in each 100 square meters are counted, with the following result.*

A	*B*	*C*	*D*
8	*6*	*14*	*20*
9	*8*	*12*	*22*
11	*10*	*18*	*25*
12	*4*	*9*	*23*

Solution: Here only one factor is involved, namely performance. We want to test with 4 samples for each.

So, we use one-way classification.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ i.e., the machines do not differ significantly in their performance.

H_1 :Not all are equal in performance.

x_1	x_2	x_3	x_4	x_1^2	x_2^2	x_3^2	x_4^2
8	6	14	20	64	36	196	400
9	8	12	22	81	64	144	484
11	10	18	25	121	100	324	625
12	4	9	23	144	16	81	529
40	28	53	90	410	216	745	2038

Step1:

$N =$ **Total No. of observations** $= r \times c = 4 \times 4 = 16.$

Step2:

$T = 40 + 28 + 53 + 90 = 211$

Step3:

$$C.F. = \frac{T^2}{N} = \frac{211^2}{16} = 2782.56$$

Step4:

$TSS = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_c^2 - C.F. = 410 + 216 + 745 + 2038 - 2782.56 = 626.44$

Step5:

$$\begin{aligned} SSC &= \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3} + \frac{(\sum x_4)^2}{n_4} - C.F. \\ &= \frac{(40)^2}{4} + \frac{(28)^2}{4} + \frac{(53)^2}{4} + \frac{(90)^2}{4} - 2782.56 \\ &= 400 + 196 + 702.25 + 2025 - 2782.56 = 540.69 \end{aligned}$$

Step6:

$SSE = TSS - SSC = 626.44 - 540.69 = 85.75$

Step7:

ANOVA table

Source of variation	**Sum of squares**	**d.f.**	**Mean square**	$F-$ **ratio**	$F-$ **table value**
Between columns	540.69	$c-1=3$	$MSC = \frac{540.69}{3} = 180.23$	$F = \frac{180.23}{7.15}$	$F_{0.05}(3,12) = 3.49$
Error	85.75	$16-4=12$	$MSE = \frac{85.75}{12} = 7.15$		

Step8:

Inference:

Since Cal.$F >$tab. F, we reject the H_0. $\therefore$ the 4 machines differ in their performance significantly.

Problem 2 *The sales of 4 salesmen in 3 seasons are tabulated here. Carry out an analysis of variance.*

	Salesmen			
Seasons	*A*	*B*	*C*	*D*
Summer	*45*	*40*	*38*	*37*
Winter	*43*	*41*	*45*	*38*
Monsoon	*39*	*39*	*41*	*41*

Solution:

In this problem the data is given according to two factors season and salesmen. So, we do a two-way analysis of variance.

In order to simplify computations we shall code the data by subtracting 40 from each value.

coding value = 40

H_0 : There is no significant difference between salesmen and between seasons.

H_1 : There is a significant difference between salesmen and between seasons.

	x_1	x_2	x_3	x_4	Total	x_1^2	x_2^2	x_3^2	x_4^2
y_1	5	0	-2	-3	0 Σy_1	25	0	4	9
y_2	3	1	5	-2	7 Σy_2	9	1	25	4
y_3	-1	-1	1	1	0 Σy_3	1	1	1	1
Total	7	0	4	-4	7	35	2	30	14

Σx_1 Σx_2 Σx_3 Σx_4

$T = 7$

Step1:

$N = r \times c = 4 \times 3 = 12.$

Step2:

$T = 7 + 0 + 4 - 4 = 7$

Step3:

$$C.F. = \frac{T^2}{N} = \frac{7^2}{12} = 4.083$$

Step4:

$$TSS = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_c^2 - C.F. = 35 + 2 + 30 + 14 - 4.083 = 76.917$$

$(5^2 + 0^2 + -2^2 + -3^2 + \cdots + 1^2) - 4.083 = 76.917 = TSS$

$81 - 4.083 = 76.917$

Step5:

$$\begin{aligned} SSC &= \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3} + \frac{(\sum x_4)^2}{n_4} - C.F. \\ &= \frac{(49)^2}{3} + 0 + \frac{(4)^2}{3} + \frac{(-4)^2}{3} - 4.083 \\ &= \frac{81}{3} - 4.083 = 22.917 \end{aligned}$$

Step6:

$$\begin{aligned} SSR &= \frac{(\sum y_1)^2}{n_1} + \frac{(\sum y_2)^2}{n_2} + \frac{(\sum y_3)^2}{n_3} - C.F. \\ &= 0 + \frac{(7)^2}{4} + 0 - 4.083 \\ &= 12.25 - 4.083 = 8.167 \end{aligned}$$

Step6:

$SSE = TSS - SSC - SSR = 76.917 - 22.917 - 8.167 = 45.833$

$SSE = 45.833$

Step7:

ANOVA table

Source of variation	**Sum of squares**	**d.f.**	**Mean square**	$F-$ **ratio**	$F-$ **table value**
Between columns	SSC=22.917	$c-1=4-1=3$	$MSC=\frac{22.917}{3}=7.639$	$F_C=\frac{7.639}{7.6388}=1$	$F_{0.05}(3,6)=4.76$
Between rows	SSR=8.167	$r-1=3-1=2$	$MSR=\frac{8.167}{2}=4.0835$	$F_R=\frac{7.6388}{4.0835}=1.87$	$F_{0.05}(6,2)=19.33$
Error	SSE=45.833	$(r-1)(c-1)=6$	$MSE=\frac{45.833}{6}=7.6388$		
Total	TSS=76.917	$(rc-1)=11$			

Step8:

Inference:

For Between columns:

If Cal.F_C <tab. F, we accept the H_0.

For Between rows:

If Cal.F_R <tab. F, we accept the H_0.

$\therefore$ There is no significant difference between the salesmen and between the seasons so far as sales is concerned.

Exercise

1) A completely randomized design experiment with 10 plots and 3 treatments gave the following results:

Plot No.	1	2	3	4	5	6	7	8	9	10
Treatment	A	B	C	A	C	C	A	B	A	B
Yield	5	4	3	7	5	1	3	4	1	7

Analyze the results for treatment effects.

2) The following are the number of mistakes made in 5 successive days by 4 technicians working for a photographic laboratory test at a level of significance $\alpha=0.01$. Test whether the difference among the four sample means can be attributed to chance.

	Technician		
I	II	III	IV
6	14	10	9
14	9	12	12
10	12	7	8
8	10	15	10

3) The following data represent the number of units of production per day turned out by different workers using 4 different types of machines

		Machine type			
		A	B	C	D
Workers	1	44	38	47	36
	2	46	40	52	43
	3	34	36	44	32
	4	43	38	46	33
	5	38	42	49	39

(a) Test whether the five men differ with respect to mean productivity and

(b) Test whether the mean productivity is the same for the four different machine types.

4) The sales of 4 salesmen in 3 seasons are tabulated here. Carry out an analysis of variance.

	Salesmen			
Seasons	A	B	C	D
Summer	36	36	21	35
Winter	28	29	31	32
Monsoon	26	28	29	29

Ex: Perform two-way ANOVA for the following data:

		Treatments		
plots of land	A	B	C	D
I	38	40	41	39
II	45	42	49	36
III	40	38	42	42