

**DELHI TECHNOLOGICAL
UNIVERSITY**

**PROBABILITY AND STATISTICS (MC-
205)**

PRACTICAL FILE



SUBMITTED TO:
PROF. JAMKHONGAM TOUTHANG

SUBMITTED BY:
NITYA MITTAL
(2K19/MC/089)

INDEX

S. No.	Name of Experiment	Page	Date	Signature
1	IMPORTING DATA INTO SPSS FROM TEXT AND MICROSOFT EXCEL FILES			
2	(a) MERGING OF FILES (b) SPLITTING OF FILES			
3	PICTORIAL REPRESENTATION OF DATA			
4	DISTRIBUTION OF CURVES			
5	DESCRIPTIVE STATISTICS FOR DATA IN SPSS			
6	CORRELATION AND REGRESSION			
7	HYPOTHESIS TESTING			
8	<i>t</i> -TEST			
9	CHI-SQUARE TEST			
10	ANOVA			

EXPERIMENT 6

Correlation and Regression

SOURCE CODE:

```
avghours<- data.frame(sports=c("Basketball","Badminton","Table  
Tennis","Lawn Tennis","Football"), avg_hours= c(2,3,2,1.5,3.5))  
numofstudents<-data.frame(num_of_students= c(10,23,65,35,16))  
sports<- cbind(avghours,numofstudents)  
new<- cbind(sports=c("Cricket"), avg_hours=c(4), num_of_students=c(34))  
total<-rbind(sports,new)
```

```
x<-as.numeric(total$avg_hours)  
y<-as.numeric(total$num_of_students)  
z<-cbind(x,y)
```

1. Correlations:

```
cor(z, use="complete.obs", method="kendall")  
cov(z, use="complete.obs")  
cor(z, use="complete.obs", method="spearman")  
cor(z, use="complete.obs", method="pearson")  
cor(x,y)
```

2. Correlations (Pictorial Representation):

```
plot(x, y, main="Scatterplot Example", xlab="Average Hours ", ylab="Num  
of Students ", pch=19)  
abline(lm(y~x), col="red")  
lines(lowess(x,y), col="blue")
```

```
pairs(~x+y,data=z, main="Simple Scatterplot Matrix")
```

3. Regression:

```
w=c(12,23,31,16,22,43)
```

```
z<-cbind(x,y,w)
```

```
numofequip<-data.frame(num_of_equip=c(12,23,31,16,22,43))
```

```
total<-cbind(total,numofequip)
```

```
fit<-lm(total$num_of_students~total$avg_hours+total$num_of_equip,  
data=total)
```

```
summary(fit)
```

```
coefficients(fit)
```

```
vcov(fit)
```

```
influence(fit)
```

```
fit1 <- lm(total$num_of_students ~ total$avg_hours + total$num_of_equip,  
data=total)
```

```
fit2 <- lm(total$num_of_students ~ total$avg_hours, data=total)
```

```
anova(fit1, fit2)
```

4. Regression (Pictorial Representation) :

```
layout(matrix(c(1,2,3,4),2,2))
```

```
plot(fit)
```

OUTPUT:

1. Correlations

```

> z
      x  y
[1,] 2.0 10
[2,] 3.0 23
[3,] 2.0 65
[4,] 1.5 35
[5,] 3.5 16
[6,] 4.0 34

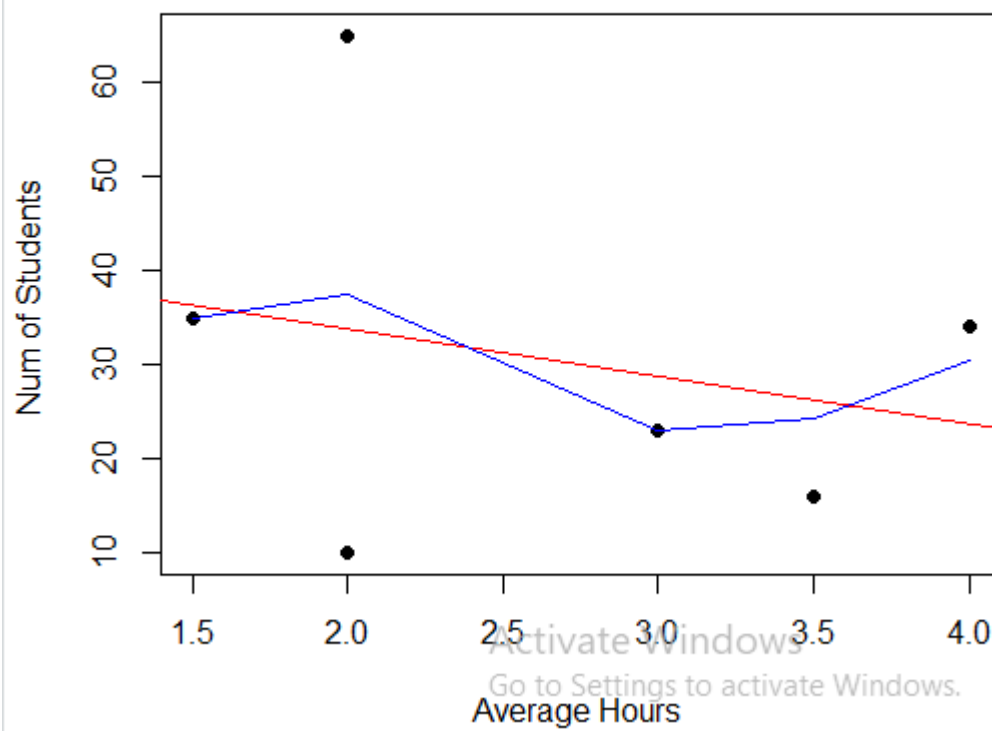
> cor(z, use="complete.obs", method="kendall")
      x      y
x 1.0000000 -0.1380131
y -0.1380131 1.0000000
> cov(z, use="complete.obs")
      x      y
x 0.9666667 -4.9
y -4.9000000 381.9
> cor(z, use="complete.obs", method="spearman")
      x      y
x 1.0000000 -0.2898855
y -0.2898855 1.0000000
> cor(z, use="complete.obs", method="pearson")
      x      y
x 1.0000000 -0.2550251
y -0.2550251 1.0000000
> |

> cor(x,y)
[1] -0.2550251
> |

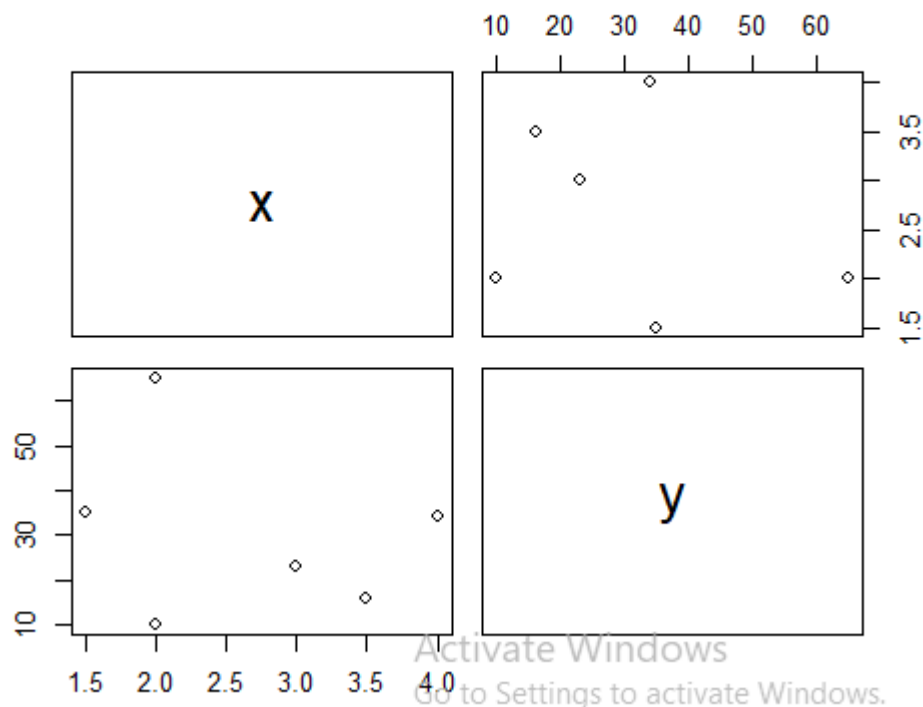
```

2. Correlations (Pictorial Representation):

Scatterplot Example



Simple Scatterplot Matrix



3. Regression:

```
> summary(fit)

Call:
lm(formula = total$num_of_students ~ total$avg_hours + total$num_of_equip,
    data = total)

Residuals:
ALL 6 residuals are 0: no residual degrees of freedom!

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -11.316          NA      NA     NA
total$avg_hours2 -13.421          NA      NA     NA
total$avg_hours3 -32.263          NA      NA     NA
total$avg_hours3.5 -36.368        NA      NA     NA
total$avg_hours4 -79.158          NA      NA     NA
total$num_of_equip  2.895          NA      NA     NA

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared:  1,    Adjusted R-squared:  NaN
F-statistic:  NaN on 5 and 0 DF,  p-value: NA

~ |

> coefficients(fit)
      (Intercept) total$avg_hours2 total$avg_hours3 total$avg_hours3.5 total$avg_hours4 total$num_of_equip
      -11.315789    -13.421053    -32.263158    -36.368421    -79.157895         2.894737

> |

> vcov(fit)
      (Intercept) total$avg_hours2 total$avg_hours3 total$avg_hours3.5 total$avg_hours4 total$num_of_equip
(Intercept)      NaN              NaN              NaN              NaN              NaN              NaN
total$avg_hours2  NaN              NaN              NaN              NaN              NaN              NaN
total$avg_hours3  NaN              NaN              NaN              NaN              NaN              NaN
total$avg_hours3.5 NaN              NaN              NaN              NaN              NaN              NaN
total$avg_hours4  NaN              NaN              NaN              NaN              NaN              NaN
total$num_of_equip NaN              NaN              NaN              NaN              NaN              NaN
> influence(fit)
$hat
 1 2 3 4 5 6
1 1 1 1 1 1

$coefficients
      (Intercept) total$avg_hours2 total$avg_hours3 total$avg_hours3.5 total$avg_hours4 total$num_of_equip
1              0              0              0              0              0              0
2              0              0              0              0              0              0
3              0              0              0              0              0              0
4              0              0              0              0              0              0
5              0              0              0              0              0              0
6              0              0              0              0              0              0

$sigma
 1 2 3 4 5 6
0 0 0 0 0 0

> total
  sports avg_hours num_of_students num_of_equip
1 Basketball      2             10           12
2 Badminton       3             23           23
3 Table Tennis    2             65           31
4 Lawn Tennis    1.5            35           16
5 Football       3.5            16           22
6 Cricket        4             34           43

> |
```

Analysis of Variance Table

Model 1: total\$num_of_students ~ total\$avg_hours + total\$num_of equip

Model 2: total\$num_of_students ~ total\$avg_hours

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
--	--------	-----	----	-----------	---	--------

1	0	0.0				
---	---	-----	--	--	--	--

2	1	1512.5	-1	-1512.5		
---	---	--------	----	---------	--	--

> |

4. Regression (Pictorial Representation) :

