Paper Code: BS-110 Paper:
Probability and Statistics for Engineers

List of Practical

1. Fitting of binomial distributions for n and p given.
2. Fitting of binomial distributions, computing mean and variance.
3. Fitting of Poisson distributions for given n and λ.
4. Fitting of negative binomial distribution.
5. Problems related to area property of normal distribution.
6. Fitting of normal distribution when parameters are given/not given.
7. Fitting of gamma distribution.
8. Fitting of exponential distribution.
9. Problems based on measures of dispersion.
10. Problems based on moments, skewness and kurtosis.
11. Karl Pearson correlation coefficient.
12. Correlation coefficient for a bivariate frequency distribution.
13. Lines of regressions.
14. Planes of regressions.
15. Construct box plot of given data
16. Represent the given data using histograms.
17. Performing one way ANOVA and two way ANOVA using block design.
18. Building 2^k factorial design.

Experiment :

Date

Page No.

1. Fitting of binomial distribution for n and p given

Example- A coin is tossed 5 times. what is the probability of getting 0 heads.

Theory- number of trials = 5

Probability of success = 0.5

for binomial distribution

$$P = {}^nC_r \, p^r \, q^{n-r}$$

P = binomial probability

x = no. of times for a specific outcome within n trials

p = probability of success on a single trial

q = probability of failure on a single trial

n = number of trials

$$P(0) = {}^5C_0 \, (1/2)^0 \, (1-1/2)^5$$

$$= \frac{1}{2^5} = 0.03125$$

code -

```
num_trials = 5
prob_success = 0.5
prob = dbinom (x = 0, num_trials,
                        prob_success)
prob
```

```
> num_trials = 5
> prob_success = 0.5
> prob = dbinom(x = 0, num_trials, prob_success)
> prob
[1] 0.03125
>
```

Experiment :

Date

Page No.

2. Fitting of binomial distribution, computing mean and variance.

Example - consider an exam which consist of 25 MCQ. Each question has 5 possible answers. This means that probability of answering a question correctly by chance is 0.2.

Theory -

n = number of trials = 5

p = probability of success on each trial = 0.2

number of observations = 25

mean = n * p (Theoritical)

= 5 × 0.2 = 1

variance = n * p * (1 - p) (Theoritical)

= 5 × 0.2 × 0.8

= 0.8

code -

```
x = rbinom (25, 5, 0.2)
mean(x)
var(x)
```

rbinom function generates required number of random variables of given probability.

DELTA®

```
> x = rbinom(25, 5, 0.2)
> table(x)
x
 0  1  2  3
 5 15  4  1
> mean(x)
[1] 1.04
> var(x)
[1] 0.54
>
```

3. Fitting of poisson distribution for given n and λ.

Example - calls to a customer-service line came at an average rate of 6 every 5 minutes. Stimulate the number of calls in 50 such period.

Theory -

rpois () function in R is used for generating random numbers from a given poisson's distribution.

Here, $n = 50$ (number of values of Return)

$\lambda = 6$

The poisson distribution has a density

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for} \quad x = 0, 1, 2 \ldots$$

code -

```
num_values = 50
lambda = 6
observations = rpois (num_values, lambda)
table (rpois (50, 6))
```

```
> num_values = 50
> lambda = 6
> observations = rpois(num_values, lambda)
> table(rpois(50, 6))

 2  3  4  5  6  7  8  9 11 12
 3  2  4  8 10  8  7  5  1  2
>
```

Experiment :

Date

Page No.

4. Fitting of negative binomial distribution .

Example - A cereal company randomly places 1 of 5 different toys in each box of its cereal. One particular child is trying to collect 6 copies of their favorite toy. stimulate the number of failures before 6th success 100 times.

Theory - Negative binomial distribution describes the number of failed Bernoulli trials that occur before r successes. The probability mass function for such a random variable is

$$f(x) = P(X = x) = \binom{x+r-1}{r-1} p^r q^x$$

where p is probability of success on an individual trial and $q = 1-p$ is the probability of failure.

code -

```
num_obs = 100            # num of observations
size = 6                 # target for no. of successful
                           trials
prob_success = 0.2       # prob. of success in each trial
boxes <- rnbinom (num_obs, size, prob_success)
table (boxes)
```

```
> num_obs = 100
> size = 6
> prob_success = 0.2
> boxes <- rnbinom(num_obs, size, prob_success)
> table(boxes)
boxes
 4  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
 1  1  3  2  2  2  5  3  5  1  5  3  7  5  4  4  6  3  3  6  1  1
29 31 32 33 34 35 36 37 39 41 42 43 46 48 54 56
 5  2  2  3  1  2  1  1  3  1  1  1  1  1  1  1
>
```

Experiment :

Date

Page No.

5. Fitting of normal distributions when parameters are given / not given ?

Example - Flipper lengths of a certain kind of penguin are normally distributed with mean 192.9 mm and standard deviation 7.1 mm. What is the probability that a randomly selected penguin has a flipper less than 200 mm long? more than 200 mm?

Theory - mean = 192.9 mm

sd = 7.1 mm

To calculate the probability, we will use pnorm function which takes 4 arguments

q → limit of interest

mean → μ

sd → σ

lower limit → logical value

code -

```
prob-less200 = pnorm(200, 192.9, 7.1)
prob-less200
prob-greater200 = 1-prob-less200
prob-greater200
```

DELTA

```
> prob_less200 = pnorm(200, 192.9, 7.1)
> prob_less200
[1] 0.8413447
> prob_greater200 = 1 - prob_less200
> prob_greater200
[1] 0.1586553
>
```

Experiment :

Date

Page No.

6. Fitting of gamma distribution.

Examples - calls to a customer-service line come at an average rate of 1 every 3 minutes. what is the probability that more than an hour elapses before 25 calls come in?

Theory - The Gamma distribution is completely specified by two parameters, θ, representing the mean waiting time between events and α, the total number of occurrences needed.

In this case, we will use pgamma function which takes 3 arguments: x, alpha and lambda

here x = 60 (1 hour)
alpha = 25
lambda = 1/3

code -

```
x = 60
alpha = 25
lambda = 1/3
prob = 1 - pgamma(x, alpha, lambda)
prob
```

DELTA

```
> x = 60
> alpha = 25
> lambda = 1/3
> prob = 1 - pgamma(x, alpha, lambda)
> prob
[1] 0.8432274
>
```

Experiment :

Date

Page No.

7. Fitting of exponential distribution.

Example - Calls to a customer-service line come at an average rate of 6 every 5 minutes. What is the probability that atleast 3 minutes elapse without a call?

Theory - The exponential distribution models the waiting time between identical and independent randomly-occuring events like calls to a pizza. Exponential distribution is specified by mean number of occurences in an interval.
In this case, we will use pexp (x, lambda) which returns the probability that the waiting time between two occurences is no more than x.

Code -

```
x = 3
lambda = 6/5
prob = 1 - pexp(x, lambda)
prob
```

DELTA®

```
> x = 3
> lambda = 6/5
> prob = 1 - pexp(x, lambda)
> prob
[1] 0.02732372
>
```

Experiment :

Date

Page No.

8. Lines of regression -

Example - The sales of a company (in million dollars) for each year are shown in table below

X (year)	2005	2006	2007	2008	2009
Y (Sales)	12	19	29	37	45

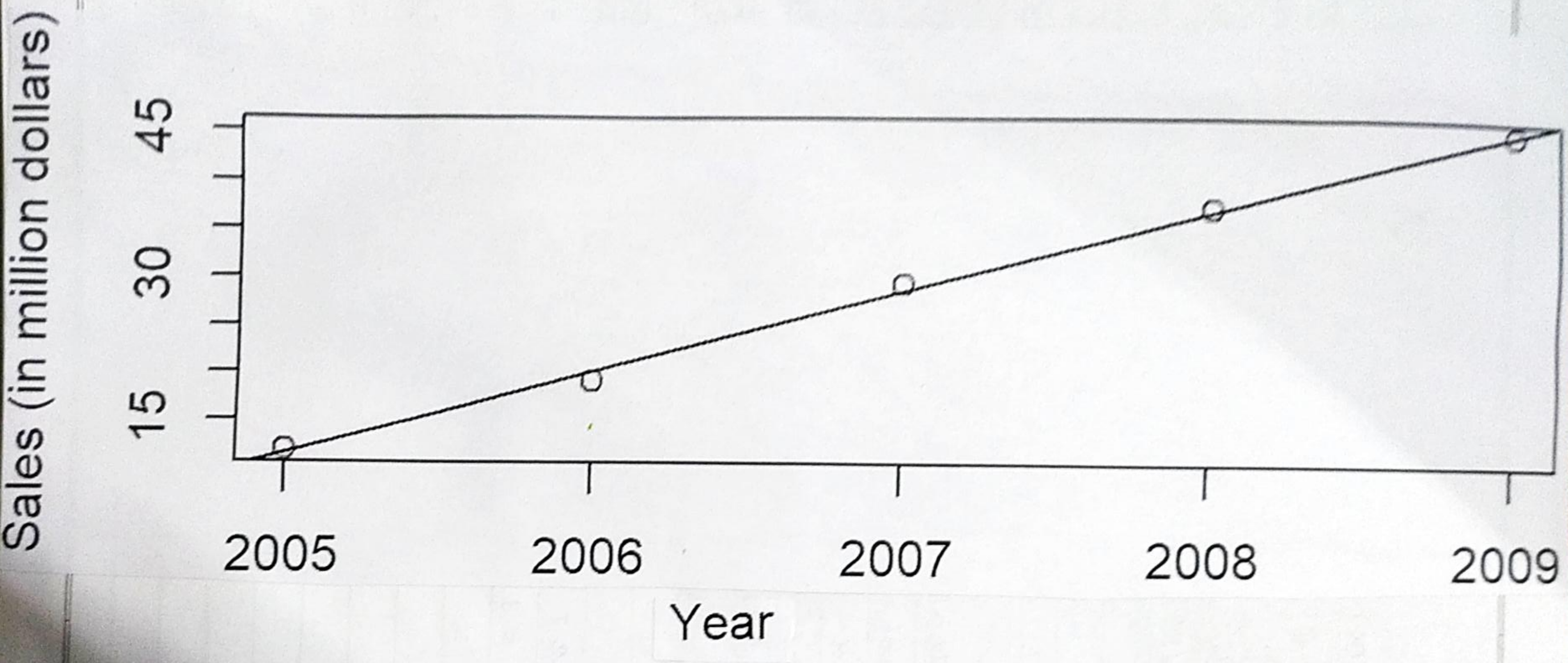
Theory - Regression analysis is a statistical tool used to establish a relationship model between 2 variables. One of these variable is called predictor variable and the other is called response variable. The general mathematical equation for a linear regression is -

$y = ax + b$ where Y = response variable
X = predictor variable

a and b are constants which are called coefficien

code -
```
year = c(2005, 2006, 2007, 2008, 2009)
Sales = c(12, 19, 29, 37, 45)
~~plot~~ table = data.frame (year, sales)
plot ( table $ year, table $ sales , col = "red",
       x lab = 'Year', Y lab = 'Sales' (in million
                                        dollars))
regression = lm(sales ~ year, data = table)
abline ( regression , col = 'blue')
```
Here, we used lm( ) function which is used to fit linear models to dataframes.

DELTA®


Sales (in million dollars)
15
30
45
2005
2006
2007
2008
2009
Year

Experiment :

Date

Page No.

9. Planes of Regression

Example- we have datasets containing 3 variables

(X1, X2, Y) Y: Reaction to Drug B

X1: Reaction to drug A

X2: Heart Rate

Explain Y with variables X1 and X2.

Theory - multiple regression is a statistical technique that can be used to analyze the relationship between a single dependent variable and several independe variables. objective of multiple regression analysis is to use the independent variables whose values are known to predict the value of single dependent val

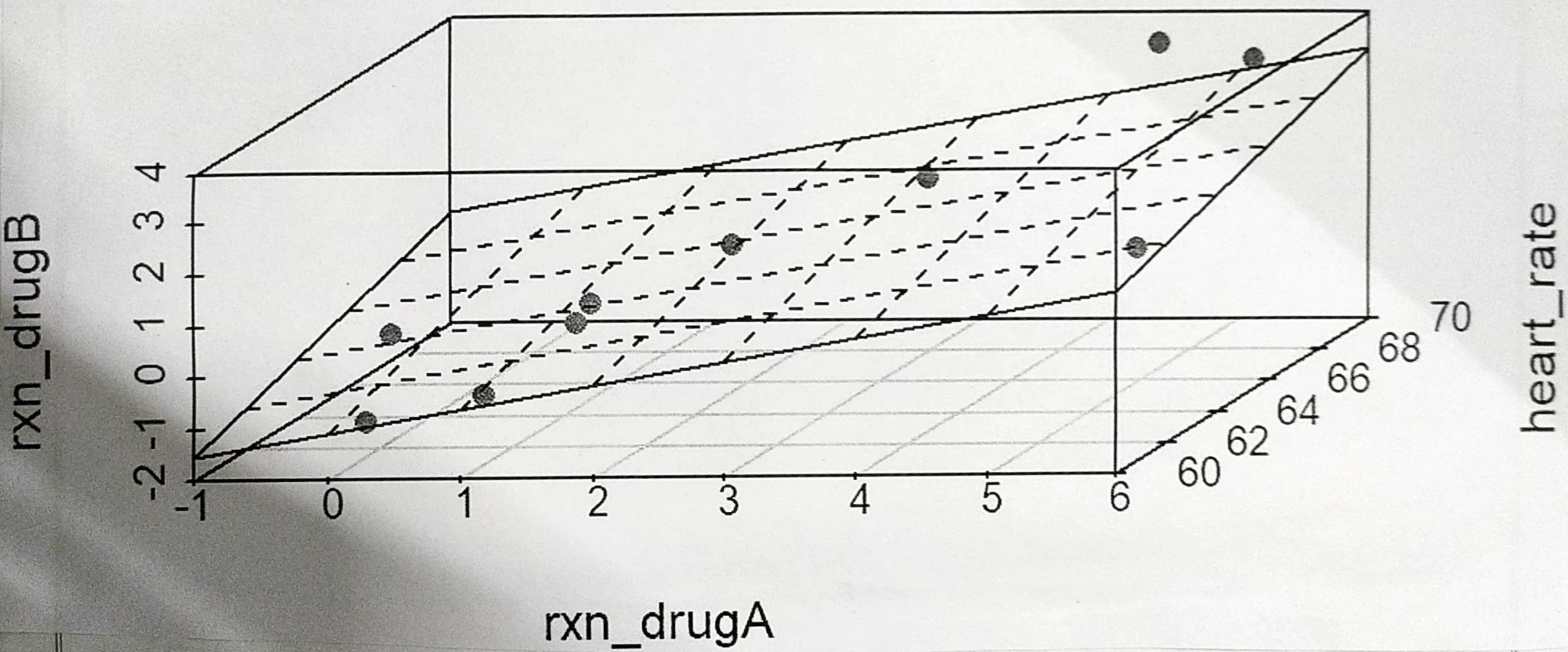
Formula- $Y = a + b_1 X_1 + b_2 X_2 + \dots b_n X_n$

Here Y is independent variable and $X_1, X_2 \dots X_n$ are independent variables.

code-
```
library(scatterplot3d)
rxn-drug A ← c(1.9,0.8,1.1,0.1,-0.1,4.4,4.6,1.6,5.5
heart-rate ← c(66,62,64,61,63,70,68,62,68,66)
rxn-drug B ← c(0.7,-1.0)-0.2,-1.2,-0.1,3.4,0.0,0.8,3.7,
dataset = cbind.data.frame(rxn-drug A,
            heart-rate, rxn-drug B)
plot3d ← scatterplot3d(rxn-drug A, heart-rate,
         rxn-drug B, angle = 55. pch = 15, color = 're
         main = "Regression plane")
```

DELTA®

Regression Plane



Experiment :

Date

Page No.

```
my.lim <- lm(rxn_drug B, ~ rxn_drugA +
             heart_rate , data= dataset)
plot 3d $ plane3d (my.lim , lty.box ="solid")
```

10. Performing one way ANOVA and two way ANOVA.

An anova test is a way to find out if survey or experiment s results are significant. Basically, we are testing groups to see if there is a difference between them.

- one-way or two-way refers to the number of independent variables in our analysis.

One-way Anova:

used to compare two means from two independent groups using F-distribution. The null hypothesis of test is that the two means are equal. Therefore a significant result means 2 means are equal.

Example- suppose we have to compare the lifetimes of four brands of automobile tyre. The lifetimes of these sample observations are measured in mileage run in 1000 miles. For each brand of automobile tyre, 15 observations have been collected.

→ In order to test and compare lifetimes of fourx brands of tyre, we should apply one way ANOVA method as there is only one factor (mileage run) to classify observations-

DELTA®

Experiment : Date ______ Page No. ______

Ho : $\mu_{apollo} = \mu_{bridgestone} = \mu_{CEAT} = \mu_{MRF}$

HA : Not all μs are equal or atleast one is different from others.

code-

```
tyre-data ← read.csv('tyre.csv', header-TRUE
                                  sep=",")
model ← aov(mileage ~Brands, data=tyre-data)
~~sep~~ summary(model)
```

Interpretation - From the results, it is observed that F-statistic value is 17.94 and is highly significant. Thus, it is wise to reject the null hypothesis.

Two-way Anova-

The two way ANOVA test is used to compare the effect of two grouping variables (A and B) on a response variable at the same time.

Example- using the built-in R dataset Tooth Growth which includes information from a study on the effects of vitamin c on tooth growth in Guinea pigs. The trial used 60 pigs who were given one of 3 ~~A~~ vitamin c dose levels (0.5, 1, 2mg/day) via one of two administration routes (orange juice or ascorbic acid) is tooth length affected by supp and dose?

DELTA

```
> tyre_data ← read.csv('tyre.csv',header = TRUE, sep = ",")
> model← aov(Mileage~Brands, data = tyre_data)
> summary(model)
            Df Sum Sq Mean Sq F value   Pr(>F)
Brands       3  256.3   85.43   17.94 2.78e-08 ***
Residuals   56  266.6    4.76
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Experiment :

Date

Page No.

code-

```
data <- ToothGrowth
# R treats dose as a numerical variable
#we will transform it into a factor variable
data$dose <- factor(data$dose,
                    levels = c(0.5, 1, 2),
                    labels = c("D0.5","D1","D2"))
res.aov2 <- aov(len ~ supp + dose, data = data)
summary(res.aov2)
```

- we may deduce from the ANOVA table that both supp and dose are statistically significant. The most important factor variable is dosage. These findings led us to anticipate that modifying the delivery technique (supp) or the vitamin c dose will have a major impact on mean tooth length.

DELTA®

```
> summary(res.aov2)
            Df Sum Sq Mean Sq F value   Pr(>F)    
supp         1  205.4   205.4   14.02 0.000429 ***
dose         2 2426.4  1213.2   82.81  < 2e-16 ***
Residuals   56  820.4    14.7                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```