Contents

Chapter 1

Linear Models and Regression Analysis

1.1 Analysis of General Linear Model

1. Consider the following model:

   (a) $y_1 = \beta_1 + \beta_2 + \varepsilon_1, \quad y_2 = \beta_1 + \beta_3 + \varepsilon_2, \; y_3 = \beta_1 + \beta_2 + \varepsilon_3.$

   (b) $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \; (i = 1,2,3)$ where $x_1 = -1, \; x_2 = 0, \; x_3 = 1.$

   Write these models in Gauss-Markoff setup. Find the blues of β_is.

2. Consider the following model.

$$y_i = \theta_i + \theta_5 + \varepsilon_i, \; i = 1,2,3$$
$$y_6 = \theta_3 + \theta_7 + \varepsilon_6$$
$$y_4 = \theta_4 + \theta_6 + \varepsilon_4$$
$$y_7 = \theta_2 + \theta_8 + \varepsilon_7$$
$$y_5 = \theta_1 + \theta_7 + \varepsilon_5$$
$$y_8 = \theta_4 + \theta_8 + \varepsilon_8$$

   (a) Write the model in Gauss-Markoff setup and find rank of error space and estimation space.

   (b) Check the estimability of the following parametric functions. i) $\theta_2 - \theta_4$ ii) $\theta_1 + \theta_2 + \theta_3 + 3\theta_5$ iii) $\theta_1 - \theta_2$ iv) $\theta_3 + 2\theta_5 - \theta_1 - 2\theta_8$

   (c) If $\underline{Y} = (60.2 \; 74.39 \; 77.88 \; 94.75 \; 81.47 \; 99.34 \; 111.86 \; 127.68)'$ obtain two different solutions to normal equations and verify that the BLUE of an estimable parametric function is unique even though two different solutions to normal equations are used.

   (d) Obtain an estimate of error variance, BLUE of $X\underline{\beta}$ and variance-covariance of the BLUE of $X\underline{\beta}$

3. Draw a random sample from $N\left(\mu = 50, \sigma^2 = 16\right)$ and obtain the sample percentiles of this sample.

   (a) Based on these percentiles obtain the BLUE of the parameters μ, σ using the general linear model setup $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$.

   (b) Find the variance-covariance of the BLUE of the parameters μ, σ.

   (c) Obtain an unbiased estimate of the error variance in the model. Also give the estimate of variance-covariance of the $\left(\hat{\mu}, \hat{\sigma}\right)$.

   (d) Calculate the SSE using residual vector and present the various sum of squares in an ANOVA. Obtain the 95 % confidence interval for parameterμ.

4. Consider the following model.

$$y_{1j} = \theta_1 + \alpha_j + \varepsilon_{1i},\ j = 1,2,4 \qquad y_{2j} = \theta_2 + \alpha_j + \varepsilon_{2i},\ j = 2,3,4$$

$$y_{3j} = \theta_3 + \alpha_j + \varepsilon_{3i},\ j = 1,2,3 \qquad y_{4j} = \theta_4 + \alpha_j + \varepsilon_{4i},\ j = 1,3,4$$

   (a) Write the model in Gauss-Markoff setup and find rank of error space and estimation space.

   (b) Check the estimability of the following parametric functions. i) $\theta_1 + \alpha_3$ ii) $\theta_1 + \theta_2 + \theta_3 - 3\theta_4$ iii) $3\theta_2 + \alpha_2 + \alpha_3 + \alpha_4$ iv) $\theta_1 - 2\theta_4 + \alpha_3$

   (c) Obtain solution of normal equations and hence obtain the BLUE and variance of BLUE of any one estimable parametric functions in (b) using

$$\left(y_{11}, y_{12}, y_{14}, y_{22}, y_{23}, y_{24}, y_{31}, y_{32}, y_{33}, y_{41}, y_{43}, y_{44}\right) = (73, 74, 71, 75, 67, 72, 73, 75, 68, 75, 72, 75)$$

1.2 Analysis of one way and two way classification models.

1. A study was done to compare Pinot Noir wine made in three different regions. Wine samples from each region were taken and scored by a panel of judges on the wine quality *flavor*. The higher the number, the more favorable the rating.

Region 1	3.1	3.5	4.8	3.1	5.5	5.0	4.8	4.3	4.7	4.3		
	5.1	3.0	4.3	5.5	4.2	3.5	5.7					
Region 2	4.3	3.4	5.0	4.1	4.7	5.0	5.0	2.9	5.0			
Region 3	3.9	4.5	7.0	6.7	5.8	5.6	4.8	5.5	6.6	5.3	5.7	6.0

   (a) Compare the flavor ratings for three different regions graphically using appropriate graph.

   (b) Analyze the data using suitable model. Test whether the flavor ratings for different regions are significantly different.

   (c) Obtain residuals and fitted values for this model.

2. The data in table aside shows the life of electric bulbs from four companies A, B, C and D. Prepare analysis of variance table to find out whether the lives differ from company to company.

Company	Life in Burning hours					
A	1020	1010	1030	1000		
B	1030	1040	1050	1030	1060	
C	990	980	970	960	970	980
D	1040	1050	1030	1070		

3. University of Wisconsin researcher tested the yields of six varieties of *Alfalfa* on each of four separate fields. The data is given in the following table.

Field	Variety					
	A	B	C	D	E	F
1	3.22	3.04	3.06	2.64	3.19	2.49
2	3.31	2.99	3.17	2.75	3.40	2.37
3	3.26	3.27	2.93	2.59	3.11	2.38
4	3.25	3.20	3.09	2.62	3.23	2.37

(a) Using appropriate model test whether the yields of six varieties are significantly different.

(b) Do the yields from different fields differ significantly?

(c) Obtain the BLUEs and estimates of variance of BLUEs of contrast effects in variety and fields.

(d) Present the variation in average yield due to different varieties and fields graphically.

4. The data given represents the results of an experiment performed to determine the effects of 4 different levels of each temperature and Alkali Percent on yield percent of pulp received from a cellulosic raw material.

   (a) Present the variation in average yield due to different temperature and Alkali Percent graphically.

   (b) Prepare analysis of variance table to find out whether the levels of temperature and Alkali percent differ significantly?

Temp → % Alkali ↓	138	143	148	153
12	391	343	312	304
15	373	328	308	282
18	359	319	305	272
21	349	289	301	261

1.3 Simple/multiple Linear Regression and Extra Sum of Squares

1. As part of a study investigating the general health of college students, information was collected from 17 college students. The corresponding data is given in Table 1. It includes age, height, diastolic blood pressure (DBP), systolic blood pressure (SBP) and the volume of air exhaled after one deep breath.

   (a) Observe the relation among variables through the multiple-scatter plot.

   (b) Fit a MLRM to the above data to regress volume of air on other variables recorded.

      i. Test for significance of regression and draw the conclusion. Obtain R^2 for the model. Calculate fitted values, error vector and residuals.

      ii. Obtain the normal probability plot of residuals for the fitted model.

      iii. Obtain the variance covariance matrix of the estimated parameters.

      iv. Test on individual regression coefficients and draw the conclusions.

      v. Obtain 95% confidence interval on regression coefficients.

      vi. Obtain 95% confidence interval on mean response and 95% prediction interval for a new observation on volume(Y) when $x_1 = 21, x_2 = 165, x_3 = 80, x_4 = 115$.

      vii. What is the contribution of height and SBP to the model given that all other regressors are included?

   (c) Split the regression sum of squares in 3 components of extra sum of squares as $SSR(X_4)$, $SSR(X_1/X_4)$ and $SSR(X_3, X_2/X_1, X_4)$.

2. 16 observations on the viscosity of a polymer (y) and two-process variables reaction temperature (X_1) and catalyst feed rate (X_2) are recorded and given in Table 2.

   (a) Fit a multiple regression model to the above data. Estimate the variance-covariance matrix of regression parameters.

   (b) Test for significance of regression. Draw appropriate conclusions.

(c) Obtain the residual vector, variance-covariance matrix of residuals and have the normal probability plot of residuals.

(d) Test the significance of addition of process variable temperature given that the feed rate variable is already there in the regression model i.e. split the regression sum of squares in two components of extra sum of squares as $SSR(X_2)$ and $SSR(X_1/X_2)$

(e) Check whether you need both the regressor variables in the model?

(f) Split the regression sum of squares in two components of extra sum of squares as $SSR(X_1)$ *and* $SSR(X_2/X_1)$.

Table 1						Table 2		
Age	Height	DBP	SBP	Volume		Viscosity	Temp ^0C	Catalyst Feed rate lb/hr
X_1	X_2	X_3	X_4	Y		Y	X_1	X_2
22	175	60	122	3.1		2256	80	8
19	152	70	102	3.4		2340	93	9
23	165	82	118	3		2426	100	10
31	162	90	108	3.2		2293	82	12
21	193	68	120	4.9		2330	90	11
41	137	76	104	2.4		2368	99	8
27	182	76	120	4.5		2250	81	8
26	162	62	116	3.1		2409	96	10
18	172	74	118	4.4		2364	94	12
34	160	76	102	2.9		2379	93	11
31	172	70	112	4.2		2440	97	13
28	160	80	122	3		2364	95	11
24	163	62	118	3		2404	100	8
20	185	78	124	4.7		2317	85	12
21	190	76	120	4.8		2309	86	9
19	168	68	102	4.1		2328	87	12
35	137	60	106	2.3				

3. The following information was obtained from the manager of a city water department for predicting the consumption of water (in gallons) from the size of household:

Household Size(x)	Water Used (y)
2	650
7	1200
9	1300
4	430
12	1400
6	900
9	1800
3	640
3	793
2	925

(a) Is there any relation between X and Y?

(b) Test the hypothesis $H_0 : \rho = 0$.

(c) Fit a simple linear regression model. Give the units and interpret the value of β_1 in the simple linear model.

(d) Find the standard errors of the estimates.

(e) Test the hypothesis $H_0 : \beta_0 = 0$.

(f) Construct the analysis of variance table and test for significance of regression (i.e. is there significance linear relationship between Water Used y and Household Size x). Use both t and F tests to check the significance of regression.

(g) What percentage of total variability in water consumption is explained by this model? Does the model do a good job in this respect?

(h) Find 95% confidence interval on the slope and intercept.

(i) Find the 95% confidence interval on the mean consumption of water.

(j) Predict the consumption of water for size of household is 5.

(k) Examine the normality and independence of errors.

(l) Is the fitted model adequate?

1.4 Lack of Fit test in Simple/Multiple Linear Regression models.

1. The diameter, height, and volume of 31 black cherry trees in Allegheny National Forest are recorded in the file c:\Program Files\mtbwin\data\trees.mtw.

   (a) Observe the relation between the variables through 'multiple scatter plot'.

   (b) Fit the simple linear regression model to regress volume of the tree on diameter of the tree. Perform the lack of fit test.

   (c) Test the hypothesis of no regression of volume on diameter if there is no evidence of lack of fit.

   (d) Variance covariance matrix of BLUE of regression parameters and predict the value of volume when the diameter is 19.

   (e) Check the normality assumption of residuals through normal probability plot.

2. Consider the following data

y	26	24	175	160	163	55	62	100	26	30	70	71
x_1	1	1	1.5	1.5	1.5	1	2	0.5	1	1	1	1
x_2	1	1	4	4	4	2	5	3	2	2	3	3

   Fit the linear regression model of Y on X_1 and X_2. Obtain an estimate of pure error term and hence perform the lack of fit test.

3. The yield of a chemical process is related to the concentration of the reactant and the operating temperature. An experiment has been conducted with the following results.

Yield	81	89	83	91	79	87	84	90
Concentration	1	1	2	2	1	1	2	2
Temperature	150	180	150	180	150	180	150	180

   (a) Fit the linear regression model of Y on X_1 and X_2.

   (b) Obtain an estimate of pure error term and hence perform the lack of fit test.

4. A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machine in an outlet. The two most important variables affecting the delivery time (y) are the number of cases of product stocked (x_1) and the distance walked by the route driver (x_2).

y	16.68	11.50	12.03	14.88	13.75	18.11	8.00	17.83	79.24
x_1	7.00	3.00	3.00	4.00	6.00	7.00	2.00	7.00	30.00
x_2	560.0	226.0	340.0	80.00	150.0	330	110.0	210.0	1460
y	21.50	40.33	21.00	13.5	19.75	24.00	29.00	15.35	19.00
x_1	5.00	16.00	10.00	4.00	6.00	9.00	10.00	6.00	7.00
x_2	605.0	688.0	215.0	255	462	448	776	200	132
y	9.50	35.10	17.90	52.32	18.75	19.83	10.75		
x_1	3.00	17.00	10.00	26.00	9.00	8.00	4.00		
x_2	36	770	140	810	450	635	150		

(a) Fit the linear regression model of Y on X_1 and X_2.

(b) Obtain (i) Standardized residuals (ii) Studentized residuals (iii) PRESS residuals (iv) Standardized PRESS residuals (v) R-student.

(c) Obtain the specified residual plots (i) Normal Probability Plot for residuals (ii) Residuals Vs fitted values (iii) Residuals Vs X_1 (iv) Residuals Vs X_2.

(d) Obtain PRESS and $R^2_{\text{prediction}}$

(e) Construct and interpret a plot of residuals versus time order.

(f) Study the effect of deleting outlier observations on $\hat{\beta}_i, se(\hat{\beta}_i), R^2, MS_{\text{Res}}$.

1.5 Curvilinear/Polynomial Regression

1. Table 3 presents the data concerning the strength of Kraft paper and the % of hardwood in the batch of pulp from which the paper was produced. Show that the quadratic model appears to provide a good fit to the relationship between tensile strength and hardwood concentration.

2. Fit the regression model as specified below for the data given in Table 2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

   (a) Compute t-statistic for each model parameter and conclude on the significance of corresponding regression parameter.

   (b) Use extra sum of squares method to evaluate the contribution of all the quadratic terms x_1^2, x_2^2 and $x_1 x_2$ in regression sum of squares for the regression model fitted in (a). Also test its significance.

   (c) Plot the residuals versus fitted values. Also have the Box plot of residuals to check the existence of any of the outlying observations.

3. The data concerning machine setting and the amount of energy consumed is given in Table 1. Observe the relationship in them by drawing scatter plot. Transform the data using log-transformation and fit the linear model for the transformed variables. Model the relationship between the machine setting and the amount of energy consumed with a quadratic model. Show that the quadratic model appears to provide a good fit to the data.

Table 1				Table 2				Table 3	
Sr. No.	Energy Consumption	Machine Setting		y	x_1	x_2		Hardwood Conc. (x_i) in %	Tensile Strength (y_i) in psi
1	21.6	11.15		26	1	1		1	6.3
2	4.0	15.70		24	1	1		1.5	11.1
3	1.8	18.90		175	1.5	4		2	20
4	1.0	19.40		160	1.5	4		3	24
5	1.0	21.40		163	1.5	4		4	26.1
6	0.8	21.70		55	0.5	2		4.5	30
7	3.8	25.30		62	1.5	5		5	33.8
8	7.4	26.40		100	0.5	3		5.5	34
9	4.3	26.70		26	1	1.5		6	38.1
10	36.2	29.10		30	0.5	1.5		6.5	39.9
				70	1	2.5		7	42
				71	0.5	2.5		8	46.1
								9	53.1
								10	52
								11	52.5
								12	48
								13	42.8
								14	27.8
								15	21.9

1.6 Polynomial Regression using orthogonal Polynomial and Multicollinearity

1. An operations research analyst has developed a computer simulation model of a single of a single item inventory system. He has experimented with the simulation model to investigate the effect of various reorder quantities on the average annual cost of inventory. The data are as shown in the following table.

Reorder Quantity(x)	50	75	100	125	150	175	200	225	250	275
Avg Annual Cost (y)	335	326	316	313	311	314	318	328	337	345

   (a) Observe the scatter plot of the above data. Fit the second order polynomial that expresses the average annual cost as a function of reorder quantity. Use orthogonal polynomials to fit the model. Is there evidence that the quadratic term is statistically significant? Comment on the need for the quadratic term in the model.

   (b) Prepare a scatter plot of observed y versus observed x. Overlay the fitted model on the scatter plot of y versus x.

2. The development engineer is interested in determining if the cotton weight percent in a synthetic fiber affects the tensile strength. He has run the experiment with five levels of cotton weight percent and the average observed tensile strength for the respective levels are recorded. The data are as follows.

Weight percent cotton (x)	15	20	25	30	35
Average tensile strength (y)	49	77	88	108	54

Using orthogonal polynomials fit the quadratic model to regress the average tensile strength of the synthetic fiber on cotton weight percent in the fiber. Also test the significance of each linear, quadratic, cubic and quadratic term in the model. Prepare a scatter plot of observed y versus observed$x$. Overlay the fitted model on the scatter plot of y versus x.

3. Following table presents data concerning the percentage of conversion of n-heptane to acetylene and three explanatory variables. (% conversion in acetylene (P), $\frac{Reactor\ temp-1212.50}{80.623}(T)$, $\frac{Ratio\ of\ H_2-12.44}{5.662}$ (H), $\frac{Contact\ Time-0.0403}{0.03164}$ (C). Use this information for studying multicollinearity, VIF, Mallows Cp statistics, Condition indices, Condition number using subset regression models. (Response: P, Predictors: C, H, T, CH, HT, CT, C^2, H^2, T^2).

% conversion	49	50.2	50.5	48.5	47.5	44.5	28	31.5
Reactor Temp.	1300	1300	1300	1300	1300	1300	1200	1200
H_2	7.5	9	11	13.5	17	23	5.3	7.5
Contact Time	0.012	0.012	0.0115	0.013	0.0135	0.012	0.04	0.038
% conversion	34.5	35	38	38.5	15	17	20.5	29.5
Reactor Temp.	1200	1200	1200	1200	1100	1100	1100	1100
H_2	11	13.5	17	23	5.3	7.5	11	17
Contact Time	0.032	0.026	0.034	0.041	0.084	0.098	0.092	0.086

1.7 Nonlinear regression

1. The initial velocity of an enzymatic reaction in chemical kinetics is related to the substrate concentration. The data for initial rate id a reaction for an enzyme treated with puromycin are as shown in the following table.

Concentration (ppm)	0.02	0.06	0.11	0.22	0.56	1.10
Velocity (counts/min)/min	47	97	123	152	191	200
	76	107	139	159	201	207

   (a) Plot the graph of (i) concentration Vs velocity (ii) 1/velocity Vs 1/concentration.

   (b) Fit the model $y = \theta_1 x/(\theta_2 + x) + \varepsilon$ to the above data by using appropriate transformation.

   (c) Overlay the graph of fitted curve in the original scale on the graph of concentration versus velocity. Also superimpose the fitted straight line for the transformed data on the scatter plot of transformed data.

2. Consider the following observations

x	y	
0.5	0.68	1.58
1.0	0.45	2.66
2.0	2.50	2.04
4.0	6.19	7.85
8.0	56.10	54.20
9.0	89.80	90.20
10.0	147.70	146.30

   (a) Fit the nonlinear regression model $y = \theta_1 e^(\theta_2 x) + \epsilon$ to these data. Discuss how you obtained the starting values.

   (b) Test for significance of regression

(c) Estimate the error variance σ^2.

(d) Test the hypotheses $H_0 : \theta_1 = 0$ and $H_0 : \theta_2 = 0$ Are both model parameters different from zero? If not, refit an appropriate model.

(e) Analyze the residuals from this model. Discuss model adequacy.

1.8 Logistic Regression

1. The table below presents the test-firing results for 25 surface-to-air antiaircraft missiles at targets of varying speed. The result of each test is either a hit ($y = 1$) or a miss ($y = 0$).

Test	Target Speed x (Knots)	y	Test	Target Speed x (Knots)	y
1	400	0	14	330	1
2	220	1	15	280	1
3	490	0	16	210	1
4	210	1	17	300	1
5	500	0	18	470	1
6	270	0	19	230	0
7	200	1	20	430	0
8	470	0	21	460	0
9	480	0	22	220	1
10	310	1	23	250	1
11	240	1	24	200	1
12	490	0	25	390	0
13	420	0			

(a) Fit a logistic regression model to the response variable y. Use a simple linear regression model as the structure for the linear predictor.

(b) Does the model deviance indicate that the logistic regression model from part a is adequate?

(c) Provide an interpretation of the parameter β_1 in this model.

(d) Expand the linear predictor to include a quadratic term in target speed. Is there any evidence that this quadratic term is required in the model?

2. A study was conducted attempting to relate home ownership to family income. Twenty households were selected and family income was estimated, along with information concerning home ownership ($y = 1$ indicates yes and $y = 0$ indicates no). The data are shown below.

Household	Income	Home Ownership Status	Household	Income	Home Ownership Status
1	38000	0	11	38700	1
2	51200	1	12	40100	0
3	39600	0	13	49500	1
4	43400	1	14	38000	0
5	47700	0	15	42000	1
6	53000	0	16	54000	1
7	41500	1	17	51700	1
8	40800	0	18	39400	0
9	45400	1	19	40900	0
10	52400	1	20	52800	1

(a) Fit a logistic regression model to the response variable y. Use a simple linear regression model as the structure for the linear predictor.

(b) Does the model deviance indicate that the logistic regression model from part a is adequate?

(c) Provide an interpretation of the parameter β_1 in this model.

(d) Expand the linear predictor to include a quadratic term in income. Is there any evidence that this quadratic term is required in the model?

3. A study was performed to investigate new automobile purchases. A sample of 20 families was selected. Each family was surveyed to determine the age of their oldest vehicle and their total family income. A follow - up survey was conducted 6 months later to determine if they had actually purchased a new vehicle during that time period (y = 1 indicates yes and y = 0 indicates no). The data from this study are shown in the following table.

Income x_1	Age x_2	y	Income x_1	Age x_2	y
45000	2	0	37000	5	1
40000	4	0	31000	7	1
60000	3	1	40000	4	1
50000	2	1	75000	2	0
55000	2	0	43000	9	1
50000	5	1	49000	2	0
35000	7	1	37500	4	1
65000	2	1	71000	1	0
53000	2	0	34000	5	0
48000	1	0	27000	6	0

(a) Fit a logistic regression model to the data.

(b) Does the model deviance indicate that the logistic regression model from part i) is adequate?

(c) Interpret the model coefficients β_1 and β_2.

(d) What is the estimated probability that a family with an income of $45,000 and a car that is 5 years old will purchase a new vehicle in the next 6 months?

(e) Expand the linear predictor to include an interaction term. Is there any evidence that this term is required in the model?

(f) For the model in part i), find statistics for each individual model parameter.

(g) Find approximate 95% confidence intervals on the model parameters for the logistic regression model from part a).

Chapter 2

Parametric Inference

2.1 Sampling distribution of Statistics/estimator

1. Based on the simulated data from $N(\mu = 10, \sigma^2 = 1)$. Demonstrate the distribution and it's descriptive statistics of following statistics
   (a) $T_1 = \overline{X}_n$
   (b) $T_2 = S_n^2$ where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$
2. Demonstrate the distribution and its descriptive statistics of following statistics which is based on the simulated data from$Poisson(\lambda = 15)$
   (a) $T_1 = \overline{X}_n$
   (b) $T_2 = S_n^2$where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$
3. Based on the simulated data from $U(0, \theta = 10)$.Demonstrate the distribution and its descriptive statistics of estimator.
   (a) $T_1 = 2\overline{X}_n$
   (b) $T_2 = X_{(n)}$
4. Based on the simulated data from$Gamma(\alpha = 8, \beta = 10)$.Demonstrate the distribution and its descriptive statistics of estimator $T_1 = \tilde{\alpha} = \frac{m_1^2}{m_2 - m_1^2}$,$T_{21} = \tilde{\beta} = \frac{m_2 - m_1^2}{m_1}$where $m_1 = E(X) = \overline{X}$and $m_2 = E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

2.2 Estimation by Method of Moments

1. Generate a random sample of size $n = 100$ from $Bino\left(n = 6, p = 0.86\right)$.
   Obtain method of moment estimates of n and p based on your sample.
2. Generate random sample of size $n = 100$ from Gamma distribution, with pdf is given by

$$f(x, \alpha, \beta) = \begin{cases} \frac{1}{\alpha\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}, x, \alpha, \beta > 0 \\ \\ 0, O.W. \end{cases}$$

   Take $\alpha = 7 and \beta = 4$ Obtain method of moment estimates of $\alpha and \beta$ based on your sample.

3. Consider a random sample of size $n = 20$ given below from Beta distribution of first kind with pdf is
$$f(x,a,b) = \begin{cases} \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & O.W. \end{cases}$$
0.133904, 0.224011, 0.062282, 0.122690, 0.051213, 0.620577, 0.000650, 0.271087, 0.827109, 0.929479, 0.561905, 0.160865, 0.362565, 0.001039, 0.883135, 0.393477, 0.990179, 0.045737, 0.243155, 0.080471
Obtain estimates of a, b based on your sample using method of moments.

4. Generate a random sample of size $n = 20$ from truncated Poisson distribution, which is truncated at point zero and obtained method of moment estimates of parameter θ based on your sample.

2.3 Plotting Likelihood Function

1. Generate a random sample of size $n = 50$ from $Poisson(\lambda = 5)$distribution. Plot a likelihood function against the different values of λ and check whether likelihood function attains maxima at i)$\lambda = \overline{X}$(sample mean) ii) $\lambda = 5$

2. Generate a random sample of size $n = 50$ from $Bernoulli(P = 0.8)$ distribution. Plot a likelihood function against the different values of P and check whether likelihood function attains maxima at i)$P = \overline{X}$(sample mean) ii) $P = 0.8$

3. Generate a random sample of size $n = 20$ from $Normal\left(\mu = 10, \sigma^2 = 1\right)$distribution. Plot a likelihood function for different values of μ when $\sigma = 1$ is known and check whether likelihood function attains maxima at i)$\mu = \overline{X}$(sample mean) ii) $\mu = 10$

4. Draw a random of size $n = 10$ from *Exponential* distribution with mean θ.Plot a likelihood function and check whether likelihood function attains maxima at i)$\theta = \overline{X}$ ii) $\theta = \overline{X} - 0.5$

5. Consider a random sample of size $n = 10$ from $U(0,\theta)$ distribution. 2.06874 5.82498 2.65698 3.58650 2.78196 1.36005 4.49083 5.74037 3.41361 1.36519 Plot a likelihood function and check whether likelihood function attains maxima at i)$\theta = 2\overline{X}$ ii) $\theta = X_{(n)}$

2.4 Unbiased estimator and MLE of probabilities

1. Take a random sample of size m=1000 from $B(n = 6, p = 0.45)$. Obtain the unbiased estimator and maximum likelihood estimator of the following.

2. $P(X = i), i = 0,1,2,3,4,5,6.$ Plot estimated probability mass function (pmf) and exact probability mass function (pmf) using graph.

   (a) $P(X \le b)\ where\ b = 3$
   (b) $p(a < X)\ \ where\ a = 4$
   (c) $p(a < X < b)\ where\ b = 5\ and\ a = 2$

3. Consider an experiment of throwing 2 dice simultaneously. Simulate this experiment 1000 times. Based on this simulation estimate the following probabilities of events

   (a) $A = \{X + Y \le 7\}$
   (b) $B = \{X = Y\}$

   Where X: Outcome from first dice. Y: Outcome from second dice.

2.5 Unbiased estimator

1. Demonstrate the distribution of following statistics which is function of random sample from $Poisson(\lambda = 6)$

   (a) $T_1 = \overline{X}$

   (b) $T_2 = S^2$

   Conclude that which one of the statistics is unbiased for parameterλ. Also compare both the statistics in terms of variance and suggest good one.

2. Demonstrate the distribution of statistics which is based on random sample take from $B(n = 7, p = 0.45)$.Check $T_1 = \overline{X}/n$ is unbiased or not.

2.6 Plotting power function

1. It is desired to test $H_0 : \mu = 10$ versus $H_1 : \mu > 10$on the basis of random sample of size $n = 25$ from the normal population with unknown mean and variance $\sigma^2 = 4$.If the probability of type-I error is to be $\alpha = 0.025$ and the test function used is,

$$\phi(x) = \begin{cases} 1 & if \overline{x} > c \\ 0 & o.w. \end{cases}$$

   Where $\underline{X} = (X_1, X_2, ..., X_{25})$ denote the random sample.

   (a) Find the value of c.

   (b) Find $\beta_\phi(\mu)$, Power function for $\mu = 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15+$

   (c) Plot the curve μ versus $\beta_\phi(\mu)$.

2. Suppose we want to test $H_0 : \theta = 1$ versus $H_1 : \theta > 1$ based on one observation from Poisson distribution with mean θ, Use the test function

$$\phi(x) = \begin{cases} 1 & if\ x > 2 \\ 0.8 & if\ x = 2 \\ 0 & if\ x < 2 \end{cases}$$

   Find the power function$\beta_\phi(\theta) = E(\phi(x))$and plot the θ versus$\beta_\phi(\theta)$

2.7 Most Powerful Test

1. Develop MP test for testing $H_0 : \lambda = \lambda_0$ versus$H_0 : \lambda = \lambda_1$ $(\lambda_1 > \lambda_0)$based on a random sample of size $n = 30$ from a Poisson(λ) with $\alpha = 0.05$.Find Power of your test. Generate a random sample of size $n = 30$ from Poisson(λ) with$\lambda = 2$ and based on this sample give your decision about acceptance or rejection of H_0 for testing $H_0 : \lambda = 2$ versus $H_0 : \lambda = 4$ with $\alpha = 0.05$using test given by you.

2. Develop MP test for testing $H_0 : P = P_0$ versus$H_0 : P = P_1$ $(P_1 > P_0)$based on a random sample of size $n = 25$ from a Bernoulli(P) with $\alpha = 0.1$.Find Power of your test. A random sample of size $n = 25$from Bernoulli(P) are given and based on this sample give your decision about acceptance or rejection of H_0 for testing $H_0 : P = 0.5$ versus$H_0 : P = 0.7$ with $\alpha = 0.1$ using test given by you. 1 0 1 0 0 0 0 0 0 1 0 0 1 1 0 1 1 0 0 0 1 1 0 0 1

2.8 Uniformly Most Powerful Test

1. Develop UMP test for testing $H_0 : \theta = 4$ versus $H_1 : \theta > 4$ $(\theta_1 > \theta_0)$ based on a random sample of size $n = 10$ from a $N(\theta, 1)$ with$\alpha = 0.05$. Consider a sample data 4.78, 4.49, 4.20, 3.93, 5.12, 3.92, 2.57, 4.08, 5.38, 2.97 and base on a sample give your decision about acceptance or rejection of H_0 using test given by you. Find power of the test given by you if $\theta = 4.5$.

2. Let$X_1, X_2, X_3..., X_{25}$ be a random sample of size 25 from $Binomial(5, p)$ distribution.

   (a) Obtain UMP test of size $\alpha = 0.04$ for testing $H_0 : p = 0.4$ versus $H_1 : p > 0.4$

   (b) Obtain power of the test.

   (c) For the following sample data what is your decision about the acceptance or rejection of H_0 3, 2, 4, 2, 1, 2, 3, 1, 1, 4, 3, 5, 2, 4, 3, 1, 1, 2, 2, 3, 2, 4, 2, 2, 0

3. A sample of size $n = 5$ is drawn from distribution with pdf is given below.

$$f(x,\theta) = \begin{cases} \theta\, e^{-x\theta}, & x > 0, \theta > 0 \\ 0 & ow \end{cases}$$

   Develop UMP test for testing $H_0 : \theta = 1$ versus $H_0 : \theta > 1$ $(\theta_1 > \theta_0)$ based on a random sample of size $n = 5$ from $Exp(\theta)$ with $\alpha = 0.1$ Consider a sample data 0.01100, 0.42132, 0.03996, 1.32345, 0.66958 and based on this sample give your decision about acceptance or rejection of H_0 using test given by you. Find the power of the test given by you if $\theta = 1.75$