M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc. II **Subject: STS-653-MJP**

Practical No. 1: Estimation and elimination of trend component. Variate difference Method

---

Q.1 Consider the following AirPassengers data related to Monthly Airline Passenger Numbers 1949 - 1960.

i) Plot the time series data and describe the trend visually.

ii) Estimate the trend using a 3-point and 5-point moving average method.

iii) Subtract the estimated trend from the original data and plot the detrended series

iv) Comment on the nature of the trend based on the detrended series.

v) Using the same data:
Calculate the first-order differenced series $\Delta X_t = X_t - X_t - 1,\ t=2,3,\cdots,144.$ where X_t represents the monthly Airpassengers Airline passenger. Plot the graph of the differenced series and discuss its characteristics (e.g., randomness, trend removal).

Q.2 Consider the following sunspots data related to Monthly Sunspot Numbers, 1749-1983.

i) Plot the time series data and describe the trend visually.

ii) Estimate the trend using a appropriate smoothing method.

iii) Subtract the estimated trend from the original data and plot the detrended series.

iv) Comment on the nature of the trend based on the detrended series.

v) Using the same dataset:
a) Apply the variate difference method to calculate the first-order differenced series.
b) Plot the differenced series and explain its relevance in stabilizing the trend.
c) Compare the detrended and differenced series for trend elimination effectiveness.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc. II **Subject: STS-653-MJP**

Practical No. 2: Estimation and elimination of seasonal component.

486	474	434	441	435	401	414	414	386	405
411	389	414	426	410	441	459	449	486	510
506	549	579	581	630	666	674	729	771	785

Q.1 Consider the following time series observations (30 time points in rows)

1.

i) Plot the 30 time points and visually identify the nature of the trend and seasonal component.

ii) Apply the filter [a-2, a-1, a0, a1, a2] = [-1, 4, 3,4, -1] * 1/9 and discuss the results.

iii) Plot the smoothed series and discuss how the filter has modified the data.

iv) Calculate the Mean Absolute Deviation (MAD) and Mean Square Deviation (MSD) for the fitted model.

Q.2 Consider the AirPassengers data related to Monthly Airline Passenger Numbers 1949-1960

1. Estimate seasonal component of monthly airline Passenger using ratio to moving average method.

2. Remove seasonal component to deseasonalize the time series and plot the deseasonalized data.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc. II **Subject: STS-653-MJP**

Practical No. 3: Examining Stationarity. Sample ACF and PACF

Q1. Use 'LakeHuron' dataset to estimate the mean . Examine the stationarity of time series using Sample ACF and PACF

Q2. Use 'BJSales' dataset to estimate mean . Examine the stationarity of time series using sample ACF and PACF

Q3. Use 'Johnson Johnson' dataset from R to estimate mean. Examine the stationarity of time series using Sample ACF and PACF.

Q4. Use 'AirPassengers' dataset from R estimate mean. Examine the stationarity of time series using Sample ACF and PACF.

M. V. P. Samaj's
K. T. H. M. College, Nashik
Department of Statistics
Practical Sheet

Class: M.Sc. II **Subject: STS-653-MJP**

Practical No. 4: Identification of moving average (MA) and Auto regressive (AR) process and its order selection

---

Q1) Beijing Air Quality Data (2010-2014):

Read the file pollution.csv. This dataset includes hourly PM2.5 concentration readings from the US Embassy in Beijing, recorded between 2010 and 2014. Then solve the following questions.

i) Plot the time series data and describe the observed trends and patterns.

ii) Compute and plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

iii)Based on the ACF and PACF plots, determine whether the data follows an AR(p) or MA(q) process, and suggest the appropriate order.

iv) Fit an AR(p) or MA(q). Report the estimated parameters.

v) Use the estimated AR model to forecast the next 12 months**.** Plot the original time series with forecasted values**.**

Q2) UK Tourist Visits Data (1980–2020):

Read the file UKTouristsVisits.csv. This dataset contains quarterly data of tourists visiting the UK, which can be used for seasonal time series analysis. Then solve the following questions.

i) Plot the time series data and describe the observed trends and patterns

ii) Compute and plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

iii) Based on the ACF and PACF plots, determine whether the data follows an AR(p) or MA(q) process, and suggest the appropriate order.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statiatics
Practical Sheet

Class: M. Sc.-II **Subject: STS-653-MJP**

Practical No. 5: Yule-Walker estimation for AR(p) model.

Q.1) use the "AirPassengers" dataset (monthly airline passengers from 1949–1960).

i) Plot the time series data and observe its trend, seasonality, and patterns.

ii) Based on the ACF and PACF plots, suggest a suitable AR(p) order for the model.

ii) Use the Yule-Walker method to estimate the AR(p) model for the dataset.

iv) Write the estimated mathematical equation of the AR(p) model based on the computed coefficients.

v) Fit AR models with different orders (p = 1, 2, 3, …) using the Yule-Walker method.

vi) Use the estimated AR model to forecast the next 12 months. Plot the original time series with forecasted values.

Q.2) **Beijing Air Quality Data (2010-2014):**

Read the file pollution.csv. This dataset includes hourly PM2.5 concentration readings from the US Embassy in Beijing, recorded between 2010 and 2014. Then solve the following questions.

i) Plot the time series data and observe its trend, seasonality, and patterns.

ii) Based on the ACF and PACF plots, suggest a suitable AR(p) order for the model.

iii) Use the Yule-Walker method to estimate the AR(p) model for the dataset.

iv) Write the estimated mathematical equation of the AR(p) model based on the computed coefficients.

v) Fit AR models with different orders (p = 1, 2, 3, …) using the Yule-Walker method.

vi) Use the estimated AR model to forecast the next 12 months. Plot the original time series with forecasted values.

M. V. P. Samaj's
K. T. H. M. College, Nashik
Department of statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No.6: Fitting MA model using Least squares regression.

Q) Load the monthly sunspots dataset in R. Check for stationarity and apply differencing if needed. Fit an MA(2) model using the least squares method.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M. Sc.-II **Subject: STS-653-MJP**

Practical No.-07 Residual Analysis and Diagnostic checking.

Q1 Load the AirPassengers dataset.

a. Visualize the dataset to inspect trends, seasonality, or patterns.

b. Fit an appropriate time series model based on visual inspection and prior analysis.

c. Extract the residuals from the fitted model.

d. Perform residual analysis by:

- Plotting the residual time series.
- Creating an ACF plot of residuals to check for autocorrelation.

e. Conduct a Ljung-Box test on the residuals to test for independence (white noise property).

f. Interpret the results to assess whether the model is appropriate for the dataset.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M. Sc.-II **Subject: STS-653-MJP**

Practical No. 8: Fitting ARMA model.

---

1. Read the Amazon.csv file and fit the ARMA(p, q) model also find the values of AICC and BIC.

2. Read the dataset gold.csv and perform the following tasks:
   i) Plot the original time series and check for stationarity.
   ii) If the series is non-stationary, apply appropriate transformations to make it stationary.
   iii) Identify suitable ARMA (p, q) model orders using ACF and PACF plots.
   iv) Fit the selected ARMA model and compute the **AICC** and **BIC** values.

M. V. P. Samaj,s
K. T. H. M. College, Nashik
Department of Statistics
Practical Sheet

Class: M.Sc.-II **Subject: STS-653-MJP**
Practical No. 9: Dickey Fuller Unit Root Test.

Q1. Load the monthly-housing.csv dataset.

a) Visualize the time series data to inspect trends, seasonality, or patterns.
b) Perform the Augmented Dickey-Fuller (ADF) test on the series to check for stationarity.
c) Interpret the ADF test results: State the null hypothesis and alternative hypothesis.
d) Report the ADF test statistic and p-

value. Q2. Load the Gold.csv dataset.

a) Visualize the time series data to inspect trends, seasonality, or patterns.
b) Perform the Augmented Dickey-Fuller (ADF) test on the series to check for stationarity.
c) Interpret the ADF test results: State the null hypothesis and alternative hypothesis.
d) Report the ADF test statistic and p_value.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc.-II **Subject: STS-653-MJP**

Practical No. 10: Identification of ARIMA(p d q) process and order selection .

Q1: Load the Gold.csv dataset and perform the following steps:

1. Visualize the Gold price data to identify potential trends, seasonality, or patterns.
2. Conduct an ADF test to check for stationarity.
3. If the series is non-stationary, apply appropriate differencing to achieve stationarity.
4. Plot the ACF and PACF of the stationary series.
5. Based on the ACF/PACF plots, identify suitable values for: p,d,q

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc.-II **Subject: STS-653-MJP**

Practical No. 11: Select a series and obtain Mean, Variance and auto covariance Autocorrelation upto lag 5.

Q.1) Consider the following series of 70 conservative from a both of chemical process. Estimate the Mean, Variance, autocovariance and autocorrelation upto lag 5.

.

(1-15)	(16-30)	(31-45)	(46-60)	(61-70)
47	44	50	62	68
64	80	71	44	38
23	55	56	64	50
71	37	74	43	60
38	74	58	52	39
64	51	58	38	59
55	57	45	59	40
41	50	54	55	57
59	60	36	41	54
48	45	54	53	23
71	57	48	49	
35	50	55	34	
57	45	45	35	
40	25	57	54	
58	59	50	45	

Q.2) Consider the data of Gold.csv .

a) Obtain the time series plot of this data

b) Estimate the mean and variance

c) Calculate the ACVF and ACF for 50 lags.

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M.Sc.- II **Subject: STS-653-MJP**

Practical No. 12: Compute and plot the empirical autocovariance function and the Empirical autocorrelation.

Q. Load the given time series dataset (AirPassengers) Visualize the time series data to identify trends, seasonality, or patterns. Compute and plot the Empirical Autocovariance Function (ACVF).Compute and plot the Empirical Autocorrelation Function (ACF). Based on the ACF plot, analyze the following:

- Identify any significant spikes and their possible implications for AR or MA processes.
- Comment on the presence of seasonality or non-stationarity (if observed).

M. V. P. Samaj's
K. T. H. M. College, Nashik. Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No: 14 Stratified random sampling (Ratio and Regression method of estimation)

---

1. Following table give the yield of paddy and area in acres of 5 villages, selected by simple random sampling without replacement from 12 villages in certain region. We have mean area per villages ($\bar{x}$ = 988.75).

Village No.	Area (in acres)X	Yield of paddy (Y)
1	1054	10316
2	973	7025
3	1089	10512
4	1054	8963
5	894	8783

a) Estimate the yield of paddy per village ($\bar{y}$) by ratio method and regression method .Also obtain the S.E. of your estimates.

b) Estimate efficiency of regression method over ratio method .Also estimate efficiency of regression method over SRSWOR.

2. An experienced farmer makes an eye estimate of the weight of peaches xi on each tree in an orchard of N=200 trees. He finds X_{total}= 11600 lb .The peaches are picked and weighed on the simple random sample of 10 trees with the following results.

		Tree number										
		1	2	3	4	5	6	7	8	9	10	Total
Act. Wt	Yi	61	42	50	58	67	45	39	57	71	53	543
Est. Wt	Xi	59	47	52	60	67	48	44	58	76	58	569

Compute the estimate by using ration and regression method, also find their SEs .Estimate efficiency of regression over ratio method.

M. V. P. Samaj's
K. T. H. M. College, Nashik Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No: 15 Circular Systematic Sampling

1. The following are the data in population related to the number of seedlings in every individual foot of sawn bed for 78 feet length.

26	16	27	37	04	36	20	21
28	9	20	14	05	20	21	26
11	22	25	14	11	43	15	16
16	26	39	24	09	27	14	18
07	17	24	18	25	20	13	11
22	39	25	17	16	21	09	19
44	21	18	14	13	18	25	27
26	14	44	38	22	19	17	29
31	40	55	36	18	24	07	
26	30	39	29	06	30	30	

A circular systematic sample of size 8 is to be drawn from the above population. Compute the variance of the sample mean for circular systematic sampling. Compare this variance with that of mean of sample of size 8 drawn using SRSWOR.

M. V. P. Samaj's
K. T. H. M .College, Nashik. Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No: 16 Cluster sampling with equal and unequal sample size.

1. For studying the cultivation practices and yield of apple ,a pilot sample survey was conducted in district of Himachal Pradesh (India).The yield (in Kilogram's) of 15 clusters of 4 trees each, selected at random out of 308 bearing trees in a village , are given below

Cluster No.	Tree No.			
	1	2	3	4
1	5.53	4.84	0.69	15.79
2	26.11	10.93	10.08	11.18
3	11.08	0.65	4.21	7.56
4	12.66	32.52	16.92	37.02
5	0.87	3.56	4.81	27.54
6	6.40	11.68	40.05	5.12
7	54.21	34.63	52.55	37.20
8	1.24	35.97	29.54	25.28
9	37.94	47.07	19.64	28.11
10	54.92	17.69	26.24	6.77
11	25.52	38.10	24.74	1.90
12	45.98	5.17	1.17	6.53
13	7.13	34.35	12.18	9.86
14	14.23	16.89	28.93	21.70
15	3.53	40.76	5.15	1.25

Estimate
(i) The average yield per tree as well as the production of apple in the village and their standard errors.
(ii) The intra-cluster correlation coefficient between trees within clusters.
(iii) The efficiency of cluster sampling as compared to simple random sampling.

2. The Excel file "worker.xls" gives information about workers of India.
In all N=600 It gives district wise information about number of worker (Female / male, main/ marginal, urban/ rural) in different states of India.

Consider the each state as a cluster and female main worker in district of that state as a sampling unit. Draw a sample 6 cluster by simple random sampling without replacement out of 35 states of India.

State	**Sr.No**	**no of observations in state**	**State**	**Sr.No**	**no of observations in state**
JK	1	14	west bengal	19	18
Himachal pradesh	2	12	jharkhand	20	18
punjab	3	17	orissa	21	30
chandigarh	4	1	chandigarh	22	16
uttaranchal	5	13	madhya pradesh	23	45
harayana	6	19	gujrat	24	25
delhi	7	9	damman and diu	25	2

harayana	8	32	dadra and nagar haveli	26	1
up	9	70	maharashtra	27	35
bihar	10	37	andra pradesh	28	23
sikkim	11	4	karnatka	29	27
arunachal	12	13	goa	30	2
pradesh					
nagaland	13	8	lakshadweep	31	1
manipur	14	9	kerela	32	14
mizoram	15	8	tamil nadu	33	38
tripura	16	4	pondichery	34	3
meghalaya	17	7	andman and nukobar	35	2
asam	18	23			N=600

Random no.	**Selected states**	**mi**
14	Manipur	9
25	Damman and Diu	2
13	Nagaland	8
9	UP	70
31	Lakshadweep	1
27	Maharashtra	35

Samples of size 6 drawn from 35 states female main workers

x1	**x2**	**x3**	**×4**	**x4**	**x5**	**x6**
62768	2401	45104	37872	68767	1407	175521
20350	6168	70949	84071	48301		187988
25018		33864	45549	40506		456529
15620		22922	114319	38811		375155
37924		21622	25176	75082		185312
41645		16302	41054	60394		158107
31067		46889	67387	103385		284450
22308		26567	30703	132760		143119
15580			71423	73088		333442
			64431	209478		140069
			135744	94169		158098
			75861	91160		139489
			20967	76504		247503
			80453	112736		367419
			48372	60213		350991
			30254	43190		179373
			36830	101416		209441
			16091	72525		244860
			33288	96463		369258
			46691	79663		683739
			13925	47190		564288
			18898	95183		485828
			48610	95412		182690

			56338	94009		208691
			48493	65108		820515
			70875	135542		650726
			97166	84397		348241
			103722	72051		242637
			18260	153829		209496
			29835	124330		540407
			15031	51597		395088
			24615	109299		239462
			32681	30235		105573
			94461	77786		489663
			36065	67184		312459

i)Estimate the average number of female main worker per district together with its relative standard error .
iii)Find efficiency of your estimator with respective to corresponding estimator used in SRSWOR.

M. V. P. Samaj's

K. T. H. M. College, Nashik Department of Statistics

Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No 17: Jackknife and Bootstrap methods of estimation (For Ratio and Regression coefficient, Coefficient of variation, Correlation coefficient)

---

1. Estimate the coefficient of variation of the following observation using Bootstrap estimate in R. Generate sample of 1000 with replacement from given data. Find their mean and variance. Plot a histogram of these values. Compute the upper 95.5% and lower 2.5% sample quantiles. Estimate the bias .Find the bootstrap-corrected estimate of CV.

   8 , 26 , 6.33 , 10.4 , 5.27 , 5.35 , 5.61 , 6.12 , 6.19 , 5.2 , 7.01 , 8.74 , 7.78 , 7.01 , 6 , 6.5 , 8 , 5.12 , 7.41 , 6.52 , 6.21 , 12.28 , 5.6 , 5.38, 6.6, 8.74.

2. Obtain the jackknife estimator of population correlation coefficient given the following sample. Draw 10 sample each of size 10.

X	24	26	32	36	43	52	62	56	52	21
Y	22	28	5	18	14	14	8	8	10	24

   Also estimate the bias and standard error of estimator.

3. Obtain the estimator of coefficient of symmetry based on moments using boot strap method given the following sample. Sample of 8 each of size 8.

Sr.No.	1	2	3	4	5	6	7	8
Xi	22	26	58	54	30	35	12	28

   Also estimate the bias and standard error of estimator.

4. A contractor has kept the data regarding a delay of his work and the penalty which he had to pay on 6 of his earlier project obtain the estimator of mean delay time and mean penalty using

   i) Boot strap method

   ii) Jack knife method

Project no.	1	2	3	4	5	6
Delay time in second	32	4	16	7	12	27
Penalty in thousands Rs.	2300	30	1500	150	700	1800

M. V. P. Samaj's
K. T. H. M. College, Nashik.
Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No: 18 Two Stage Sampling

---

1. From a certain book giving biographic of eminent living persons 10, pages were selected at random from every selected page, two persons were selected at random and their ages are noted. The book has 872 pages with biographers of about 14 to 21 persons on each page estimate the average age and it's standard deviation using this data when average number of cluster is 17

Sr. No.	M_i	Y_{il}	Y_{2i}
1	15	47	30
2	19	38	51
3	19	43	35
4	16	55	41
5	16	59	45
6	18	39	38
7	20	71	64
8	18	35	46
9	16	63	47
10	16	63	47

2. At an experimental station there were 100 fields sown with wheat .Each field was divided in to 16 plots of equal size. Out of 100fields,10 were selected by simple random sampling without replacement. From each selected field, 4 plots were chosen by SRSWOR. The yields in kg/ pot are given below.

field	plot			
	1	2	3	4
1	4.31	4.78	3.86	4.02
2	4.61	4.12	3.16	4.12
3	3.72	4.11	4.17	5.70
4	3.75	4.58	3.62	3.78
5	3.12	4.68	3.92	4.32
6	4.08	4.24	4.04	5
7	4.28	4.66	4.04	3.84
8	4.20	4.72	4.96	3.08
9	4.40	4.66	3	4.04
10	4.16	4.24	4.32	4.02

Estimate the average yield per plot and find standard error of your estimator.

M. V. P. Samaj's
K. T. H. M. College, Nashik. Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No: 19 Probability Proportional to Size (PPS) sampling.

---

1. A village has 10 orchards. Select a sample of 4 orchards, with replacement and with probability proportional to the number of trees in it, using cumulative total method.

Orchard No.	1	2	3	4	5	6	7	8	9	10
No. of tress	150	50	80	100	200	160	40	220	60	140

2. A sample of 10 villages drawn from a Tehsil , with PPSWR is given below .The population in 1951census (X) is used a size. Y is the cultivated area. Total population of the Tehsil in 1951 census is 415149 and N=800. Estimate the total cultivated area its standard error by using sampling.

X	5511	865	2535	3523	8368	7357	5131	4654	1146	1165
Y	4824	924	1948	3013	7378	5506	4051	4060	809	1013

3. A population of 15 states is given below. Draw a PPSWR sample of size of 5 ,using Non- real estate farm loans as the size by Lahiri's method. Estimate the population total and the population mean for the real estate farm loan from this sample. Also estimate the variance of these estimates.

States	Non-real estate farm loans	Real estate farm loans
AI	348.334	409
AK	3.433	2.605
AZ	431.439	54.633
AR	848.317	907.7
CA	3928.732	1343.461
CO	906.281	315.809
CT	4.373	7.13
DE	43.229	42.808
FL	464.516	825.748
GA	540.696	939.46
HI	38.067	40.775
ID	1006.036	53.753
IL	2610.572	2131.048
IN	1022.782	1213.024
IA	3909.738	2327.025

M. V. P. Samaj's
K. T. H. M. College, Nashik
Department of Statistics
Practical Sheet

Class: M. Sc. II **Subject: STS-653-MJP**

Practical No 15: Stratified random sampling (Various types of allocation method)

Table shows the number of inhabitants in thousands of 64 large cities in the United States in 1920. The data were obtained by taking the cities which ranked fifty to sixty-eighth in the United States in total number of inhabitants in 1920. The cities are arranged in two strata, the first containing the 16 largest cities and the second the remaining 48 cities. The total number of inhabitants in all 64 cities in 1920 is to be estimated from sample of size 24.Find standard error of the estimated total for:

a) A simple random sample.

b) A stratified random sample with proportional allocation.

c) A stratified random sample with Neyman allocation.

h = 1		h = 2					
797	457	314	172	121	235	138	113
773	438	298	172	120	235	138	110
748	415	296	163	119	216	138	110
734	401	258	162	118	208	138	108
588	387	256	161	118	201	136	106
577	381	243	159	116	192	132	104
507	324	238	153	116	180	130	101
507	315	237	144	113	179	126	100

Table: Number of inhabitants (in thousands)