Curriculum Vitae

August 2020

KHALIL AHMAD BAIG

Assistant Professor of Statistics, (Consultant Biostatistician and Data Analyst)

- **I will provide you the output results with concluding statistical data analysis report in a publishable form.**
- **100% surety of data confidentiality with integrity.**

Cell	WhatsApp: +923076778815
E-mails	ka.datascientist@gmail.com, khalilahmad@stat.qau.edu.pk
Links	https://orcid.org/0000-0002-6137-7616; https://www.linkedin.com/in/khalildatascientist/

Qualification

2019-to-date	PhD (Statistics)	Quaid-i-Azam University, Islamabad (Pakistan)	(N.Y.A)
2016-2018	MPhil (Statistics)	Quaid-i-Azam University, Islamabad (Pakistan)	(74.55%)
2005-2007	MSc. (Statistics)	University of Agriculture, Faisalabad (Pakistan)	(71.58%)
2003-2005	BSc.	University of the Punjab, Lahore (Pakistan)	(54.50%)
2000-2002	ICS.	B.I.S.E, Faisalabad (Pakistan)	(63.63%)
1998-2000	Matric	B.I.S.E, Faisalabad (Pakistan)	(70.94%)

Work Experience

20-03-2007 to Date	**Consultant Data Analyst**	Data Analysis Solutions, Pakistan
20-03-2009 to Date	**Assistant Professor**	Higher Education Department, Pakistan
20-03-2009 to 04-01-2018	**Lecturer**	Higher Education Department, Pakistan
12-06-2008 to 19-03-2009	**Lecturer**	The University of Faisalabad, Faisalabad

Membership of Associations

- **President** of QAU Library Society, Quaid-i-Azam University, Islamabad, Pakistan.
- **Co-editor** of Quaidian Magazine, Quaid-i-Azam University, Islamabad, Pakistan.
- **PhD Advisor** of Quaidian Statistical Society, Quaid-i-Azam University, Islamabad, Pakistan.
- **Member**, Pakistan Statistical Association

Computer Skills

- Have sufficient background of computing and analytical capabilities
  - Computer oriented and has fairly good command on the execution and application of Statistical packages like **SPSS, STATA, R**, **Python, STATISTICA, Minitab, and EViews**.
- Experienced user of:
  - Operating systems (Windows, DOS)
  - Document processing softwares (e.g., Microsoft Word and Word Perfect)
  - Spreadsheet software (e.g., Microsoft Excel)
  - Presentation softwares (e.g., Microsoft PowerPoint)
  - Database softwares (e.g., Microsoft Access and Oracle)
    Computer Languages (C, C++, Visual Basic)

Teaching Interest

- Biostatistics for Epidemiology, Categorical Data Analysis

Scholarships/Awards

- University Merit Scholarship for MSc in Statistics Part-I at University of Agriculture, Faisalabad
- Indigenous PhD Scholarship of Higher Education Commission, Islamabad

Courses Taught

MATH-114	Business Mathematics, 2008, 2009
MS-207	Probability and Stochastic Process, 2008, 2009
STAT-221	Quantitative Decision Making, 2008, 2009
STAT-223	Business Statistics, 2008, 2009
STAT-224	Statistical Inference, 2008, 2009
MA-356	Statistical Methods in Textile Engineering, 2008, 2009
	Statistics for Intermediate students, 2009 to Date

Statistics for Bachelor students, 2009 to Date

Courses Studied in Ph.D (Credit Hours, 18)

Time Series Analysis, Statistical Pattern Recognition, Estimation Theory, Algebraic Coding Theory, Game Theory, Econometric Forecasting.

Courses Studied in M.Phil (Credit Hours, 24)

Advanced Probability Theory, Surveys Sampling-I, Surveys Sampling-II, Randomized Response, Linear Models, Numerical analysis and Stochastic Simulation, Stochastic Process, Advanced Spatial Data Analysis.

Courses Studied in MSc (Credit Hours, 55+4)

Probability and Distribution Theory-I, Statistical Methods, Sample Surveys-I, Theory of Matrices and Numerical Analysis, Estimation and Testing of Hypotheses, Experimental Design-I, Sample Surveys-II, High Level Programming-I (C++), Computer Programming, Theory and Application of Linear Models, Experimental Design-II, Special Problem, Multivariate Analysis, Database Management Systems (Oracle), Econometrics-I, General Statistical Concepts, Quality Control, Seminar, Survey/Research, Operations Research, Econometrics-II.

Publications

Refereed research paper

1. Hassan, I., M. Sohail, J. Piracha, and **K. Ahmad** (2013). Implementation Status of TQM Practices in Textile and Apparel Industrial Organization: A Case Study from Faisalabad, Pakistan. British Journal of Economics, Management & Trade 3(3): 201-223.
2. S.H. Raza, M. Riaz, H.M. Zakria, M. Sarwar and **K. Ahmad** (2013). The Effect of Farm Size and Locality on Dairy Economic Traits in Small and Medium Dairy Farmers in District Gujranwala, Pakistan. http://en.engormix.com/MA-dairy-cattle/dairy-industry/articles/the-effect-farm-size-t3026/472-p0.htm
3. M. Atiq, W. Ahmad, M. Rafique, S.T. Sahi, A. Rehman, M. Younis, M. Shafiq, **K. Ahmad,** T.M. Ahmad, U. Nawaz (2014). Genetic Potential Of Cotton Germplasm For Management Of Bacterial Blight Disease. Pakistan Journal of Phytopathology, Vol. 26 (01) 2014.107-110
4. M. Atiq, S. Asad, M. Rafique, N.A. Khan, A. Rehman, M. Younis, M. Shafiq, **K. Ahmad**, N. Bashir and W.A. Khan (2014). Identification Of Source Of Resistance In Mung Bean Germplasm Against Charcoal Rot Disease. Pakistan Journal of Phytopathology, Vol. 26 (01) 2014.131-134

MPhil Thesis

1. **Ahmad, K. and Shabbir, J.** (2018). “Use of Fuzzy Tools in Estimation of Population Parameters”. Department of Statistics, Quaid-i-Azam Univ. Islamabad, Pakistan

MSc Research Report

1. **Ahmad, K**. (2007). “Exploring and Forecasting the Inflation in Pakistan from 1947 to 2007 using ARIMA Methodology”. IS thesis. Department of Mathematics and Statistics, Univ. of Agriculture, Faisalabad, Pakistan

Trainings and Workshops Offered

1. Tree Week GAT Preparation Programme for MPhil and PhD scholars, held from 02-03-2009 to 21-03-2009 in Department of Arabic and Islamic Studies, University of Faisalabad, Pakistan

Trainings/ Conferences and Workshops Attended

1. Focusing Statistical Education at College Level, under the scheme of Learning Innovation Department of Higher Education Commission, Pakistan, held on August 19-20, 2008 at University of Agriculture, Faisalabad, Pakistan.
2. ISO 9001:2000 QMS Awareness and Application of SQC Tools in Education**,** held on November 18-19, 2008 at The University of Faisalabad, Pakistan.
3. Annual Conference “6th Annual Three Days Nazaria-i-Pakistan Conference” 2014, 20-22 February. Organized by Nazaria-i-Pakistan Trust, Lahore, Pakistan.
4. International workshop “Biochar for climate – friendly Agriculture shifting paradigms towards higher precision and efficiencies” 2014. 24-27 March. Organized by Agro-climatology lab, Department of Agronomy, University of Agriculture Faisalabad, Pakistan.
5. International symposium on “Strategies for overcoming food security problems through utilization of rain-fed areas” 2014. 26-28 March. Organized by Department of Agronomy, University of Sargodha, Pakistan.

6. 14th International Conference on "Emerging Technologies" 2018. 21-22 November. Organized by Department of Computer Sciences, Quaid-i-Azam University, Islamabad, Pakistan.

Abstracts in the Scientific Proceedings and Seminars etc.

	Title	Year	Particulars of proceedings/ Seminars in which presented
1	Atiq, M., M. R. Bashir., M. A. Zeshan., M.W. Ashraf and **K. Ahmad** and M. Sajid.2014. "Biochar ; as management tool for fusarium wilt of chillies".	2014	International workshop on biochar "Biochar for climate–friendly Agriculture shifting paradigms towards higher precision and efficiencies" (24-27, March). University of Agriculture Faisalabad, Pakistan. PP.41.
2	Atiq, M., M. R. Bashir., M. A. Zeshan., M.W. Ashraf and **K. Ahmad** and M. Sajid.2014. "Efficiency of organic amendments in the soil for the management of fusarium wilt of chillies"	2014	International workshop on biochar "Biochar for climate–friendly Agriculture shifting paradigms towards higher precision and efficiencies" (24-27, March). University of Agriculture Faisalabad, Pakistan. PP.80
3	M. Atiq, A. Karamat, A.R. Khalid, M. Younas, M.Shafiq, **K. Ahmad**, and H. Rizwan. 2014. "Antifungal potential of plant extracts and chemicals for the management of black scurf disease of potato"	2014	International symposium on "Strategies for overcoming food security problems through utilization of rain-fed areas" (26-28 March). University of Sargodha. PP-33.

Some Recently Completed Projects (Sample of results)

(1) Project-I on Covid-19 Patients

Table 5 Sensitivity Analysis of Covid-19 Patients

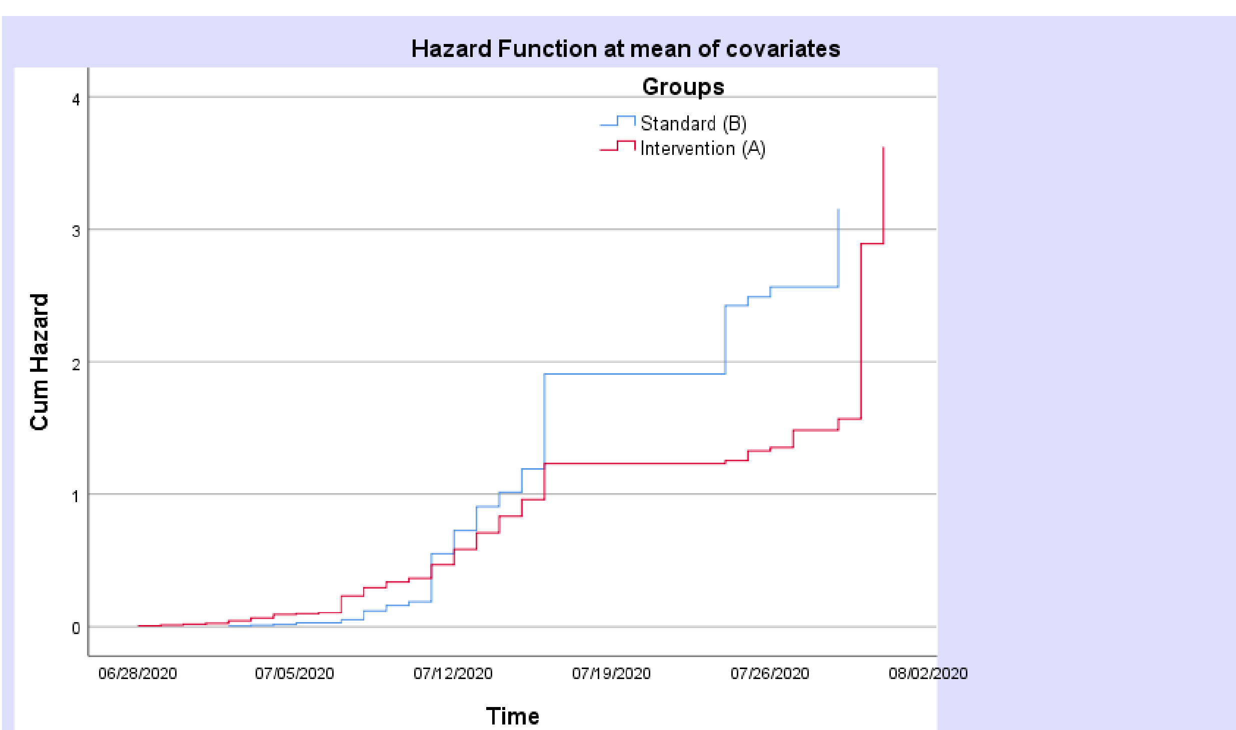
	Effect of Intervention (A) versus Standard Univariate analysis HR or coefficient (95% CI)	Effect of Intervention (A) versus Standard Multivariate analysis HR or coefficient (95% CI)
ICU admission	0.995 (0.960, 1.031)	0.942 (0.885, 1.003
Mechanical ventilation	0.859 (0.545, 1.354)	0.581 (0.281, 1.201)
Length of hospital stay	1.002 (1.003, 1.041)	1.030 (1.008, 1.053)

The table 5 showed the hazard ratios (HR) computed by both the univariate and multivariate Cox regression analysis, which coefficients predict the hazard for the terminal event as a function of the covariates in the model. The hazard ratio of ICU admission is 0.995 < 1 indicated that it decreased the need mechanical ventilation for the intervention group which means that the intervention group has more survival time as compared to the standard group. The hazard ratio of mechanical ventilation is 0.859 < 1 indicated that it decreased the need mechanical ventilation for the intervention group which means that the intervention group has more survival time as compared to the standard group. The hazard ratio of length of stay at hospital is 1.002 > 1 indicated that its increased length of stay at hospital for the intervention group which means that the patients of intervention group stayed more at hospital as compare to standard group for their survival.

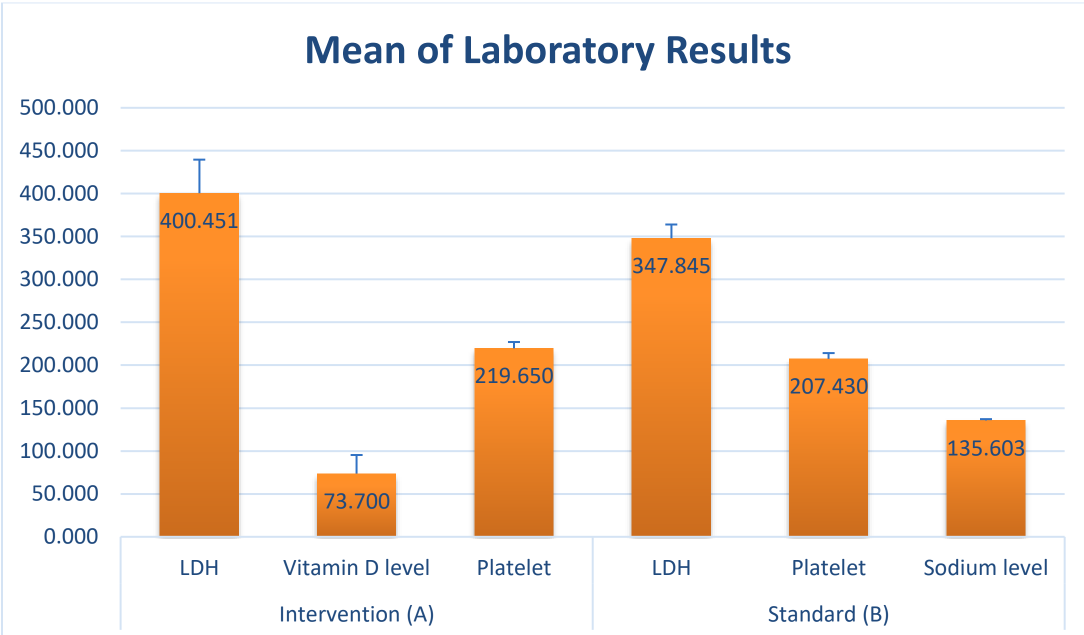
Cox regression model

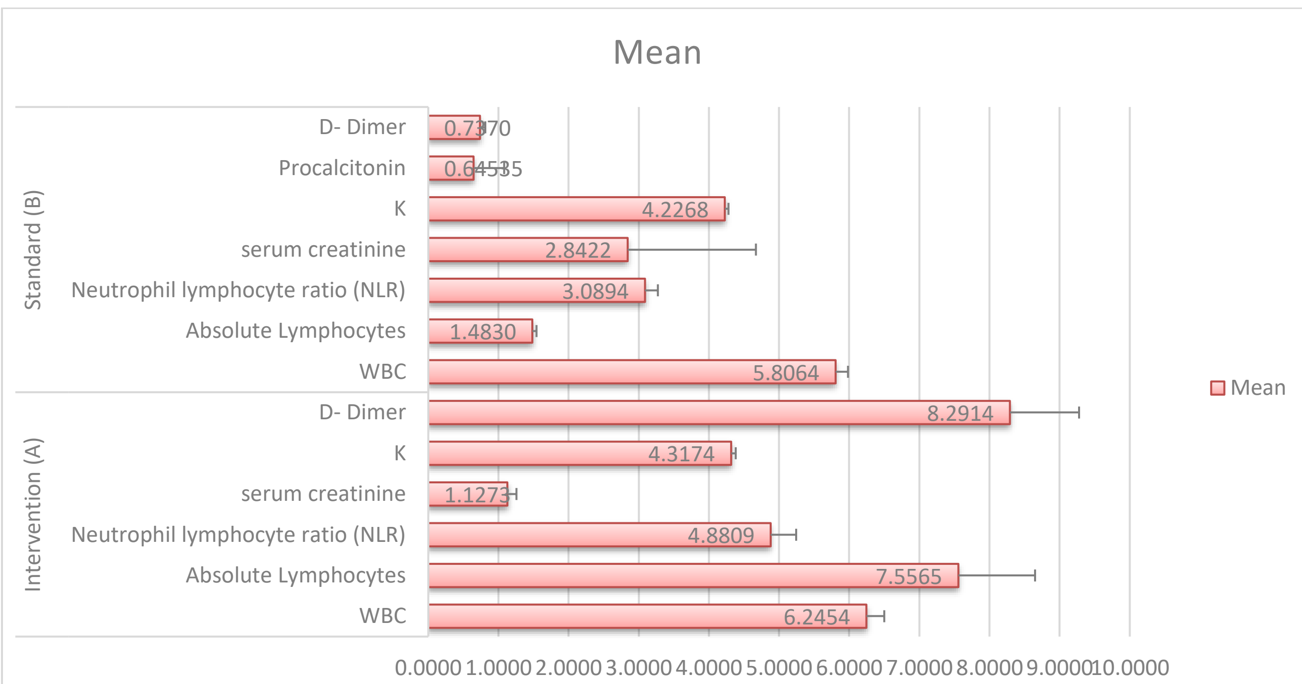
Variables	B	SE	Sig.	Exp(B)	95.0% CI for Exp(B) Lower	95.0% CI for Exp(B) Upper	P-value
Days in Hospital	0.030	0.011	0.007	1.030	1.008	1.053	0.036
Mechanical Ventilation	-0.544	0.371	0.143	0.581	0.281	1.201	
ICU Admission	-0.060	0.032	0.064	0.942	0.885	1.003	

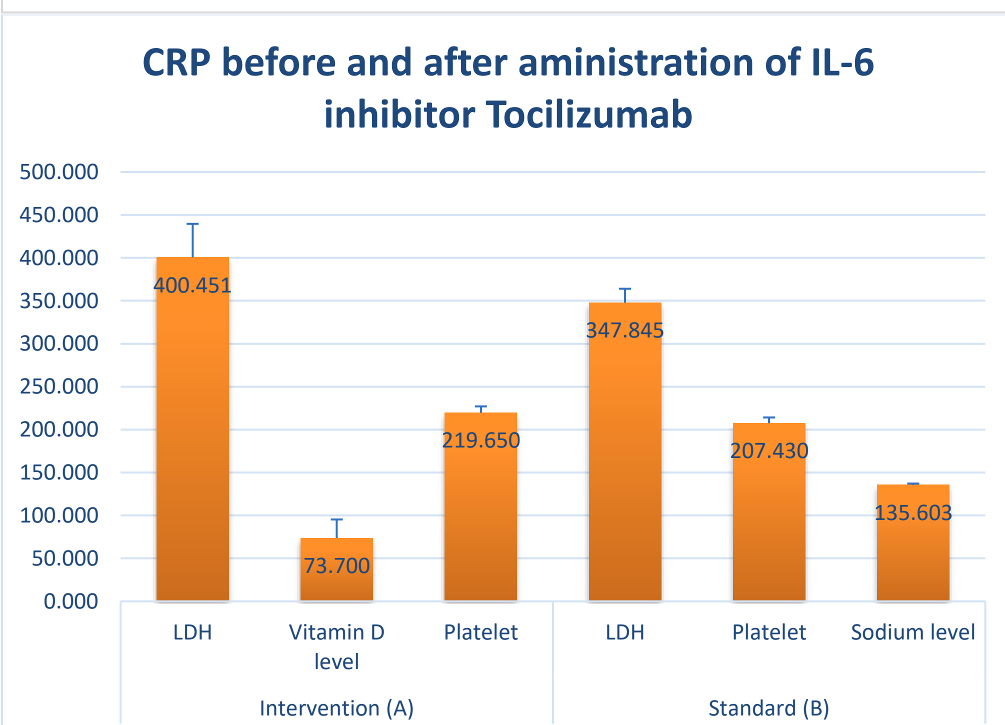
The regression coefficients predict the hazard for the terminal event as a function of the covariates in the model. A positive coefficient indicates a positive relationship between the covariate and the hazard for the mortality. This means that higher values on the covariate is associated with less survival time. A negative coefficient indicates a negative relationship between the covariate and the hazard for the terminal event. Higher values on the covariate are associated with longer survival time. The highly significant positive coefficient of days in hospital B1= 0.03, p-value =0.007<0.



From the survival graph as the line of intervention group is above than the line of standard group which showed that the patients of intervention group have less probability of mortality as compare to standard group. Similarly, from the cumulative hazard graph it is represented that the patients from the intervention group have less hazard as compared to standard group.







(2) Project-II on Covid-19 data

Complete Analysis precise output for observational study

Table No. 1 Association of Epidemiologic Characteristics with Hospital Course

Attributes	Categories	**Hospital Course**				**Outcome**		
		No ICU & No vent.	Yes ICU & No vent.	Yes ICU & Yes vent.	Chi-square P-value	Discharged	Passed away	Chi-square P-value
Age Group	<18	9 (100%)	0 (0%)	0 (0%)	0.006^{**}	9 (100%)	0 (0%)	0.013^{*}
	18-45	192 (94.1%)	12 (5.9%)	0 (0%)		204 (100%)	0 (0%)	
	46-65	125 (86.2%)	8 (5.5%)	12 (8.3%)		138 (95.2%)	7 (4.8%%)	
	>65	33 (89.2%)	2 (5.4%)	2 (5.4%)		35 (94.6%)	2 (5.4%)	
Gender	Male	250 (89.9%)	17 (6.1%)	11 (4%)	0.592^{NS}	272 (97.8%)	6 (2.2%)	0.728^{NS}
	Female	109 (93.2%)	5 (4.3%)	3 (2.6%)		114 (97.4%)	3 (2.6%)	
BMI	<24	34 (100%)	0 (0%)	0 (0%)	0.249^{NS}	34 (100%)	0 (0%)	0.043^{*}
	24-30	101 (87.8%)	10 (8.7%)	4 (3.5%)		111 (96.5%)	4 (3.5%)	
	Above 30-40	67 (87.0%)	2 (2.6%)	8 (10.4%)		73 (94.8%)	4 (5.2%)	
	>40	10 (76.9%)	2 (15.4%)	1 (7.7%)		12 (92.3%)	1 (7.7%)	
	Not available	147 (94.2%)	8 (5.1%)	1 (0.6%)		156 (100%)	0 (0%)	

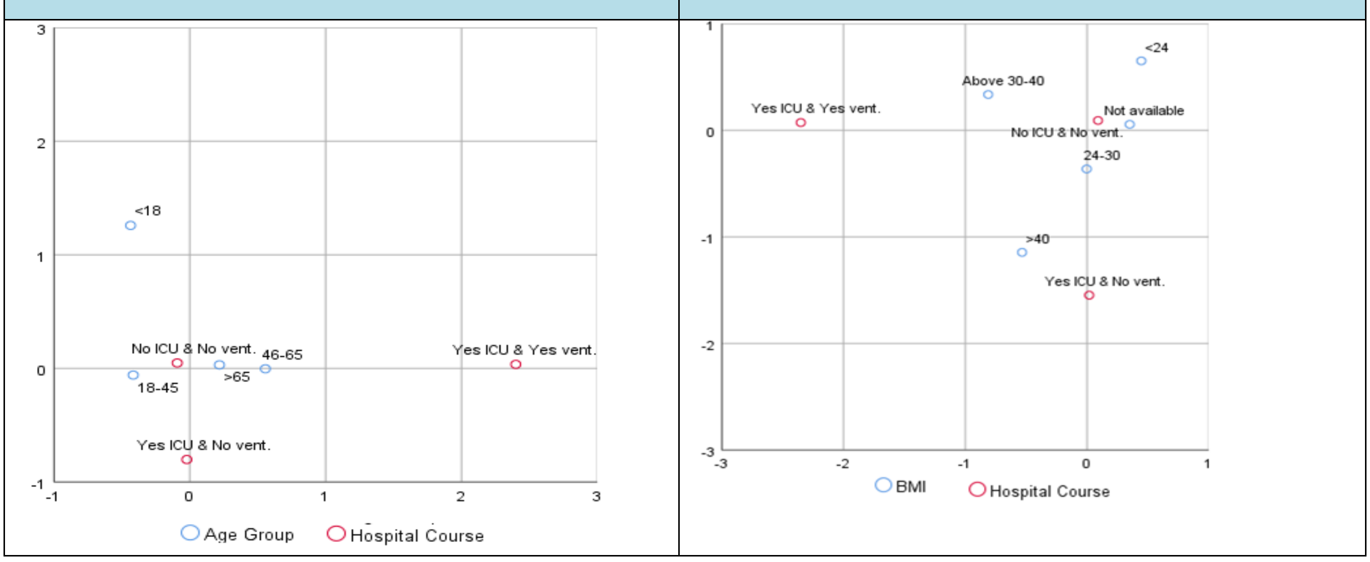
** Highly significant as p-value < 0.01, * Significant as p-value < 0.05, NS ** Nonsignificant as p-value > 0.05

Table 1, represented the epidemiologic characteristics associated with both hospital course (No admission to ICU and no need for mechanical ventilation, Yes admission to ICU and no need for mechanical ventilation, Yes admission to ICU and also yes need for mechanical ventilation, No admission to ICU and yes need for mechanical ventilation) and outcome (discharged, passed away) attributes. There is not a single observation is found about no admission to ICU but yes for need of mechanical ventilation. It is showed that the age group is highly significantly associated with hospital course as the p-value < 0.01 and significantly associated with outcome, p-value < 0.05. For the age <18 years, all patients 9 (100%) did not require admission to ICU and no need for mechanical ventilation and 9 (100%) were discharged. For age 18-45 year 192 (94.1%) did not require admission to ICU and no need for mechanical ventilation, 12 (5.9%) require admission to ICU and no need for mechanical ventilation and 204 (100%) were discharged. For age 46-65 year 125(86.2%) did not require admission to ICU and no need of mechanical ventilation, 8 (5.5%) were required admission to ICU without mechanical ventilation, 12 (8.3%) were required admission to ICU with mechanical ventilation, and 138 (95.2%) were discharged, 7 (4.8%) were passed away. For age more than 65 year 33(89.2%) did not require admission to ICU and no need of mechanical ventilation, 2 (5.4%) were required admission to ICU without mechanical ventilation, 2 (5.4%) were required admission to ICU with mechanical ventilation, and 35 (94.6%) were discharged, 2 (5.4%) were passed away. We noticed that when we move left to right along the categories of hospital course and downward along the categories of age group, the percentage on average have increasing trend, which showed that the both attributes are moving in the same direction that is attributes are positively associated. The Chi-Square test of association between age group and hospital course is highly significant as p-value is 0.006<0.01, also the Chi-Square test of association between age group and outcome is significant as p-value is 0.013<0.05. As the association between gender and hospital course, outcome is nonsignificant, but it is observed that male patients are more than double of female patients in each category of both hospital course and outcome. As more males have to go out from home as compare to females. So, from this information it can be inferred that stay at home is best option to be save.

Figure 1. Correspondence Analysis for the Association of Epidemiologic Characteristics with ICU admission and need for Mechanical Ventilations

(a) Age with ICU admission and Need of Ventilation

(b) BMI with ICU admission and Need OF Ventilation

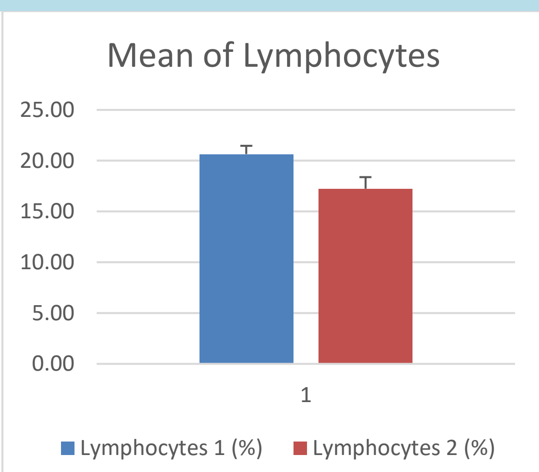


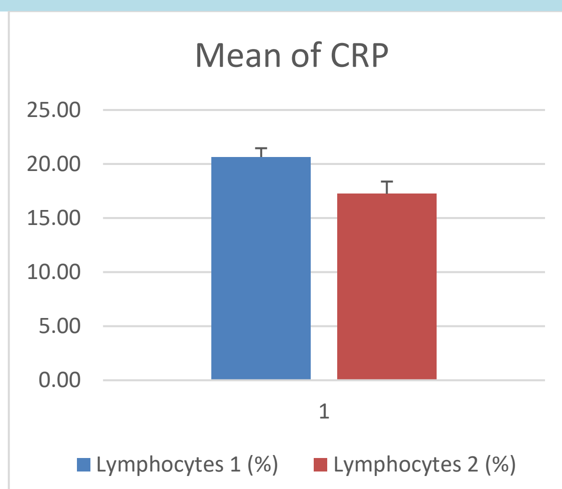
From figure 1(a) the age groups 18-45, >65 are more strongly associated with no admission to ICU and no need for mechanical ventilation, as there is less distance between this age group and No admission to ICU and no need for mechanical ventilation. Also these age group have less distance form required admission to ICU and no need for mechanical ventilation. Needed admission to ICU along with mechanical ventilation more associated with 46-65 and 65 year as compare to both 18-45 and <18. The patients with less than 18 years are not required both admission to ICU and mechanical ventilation.

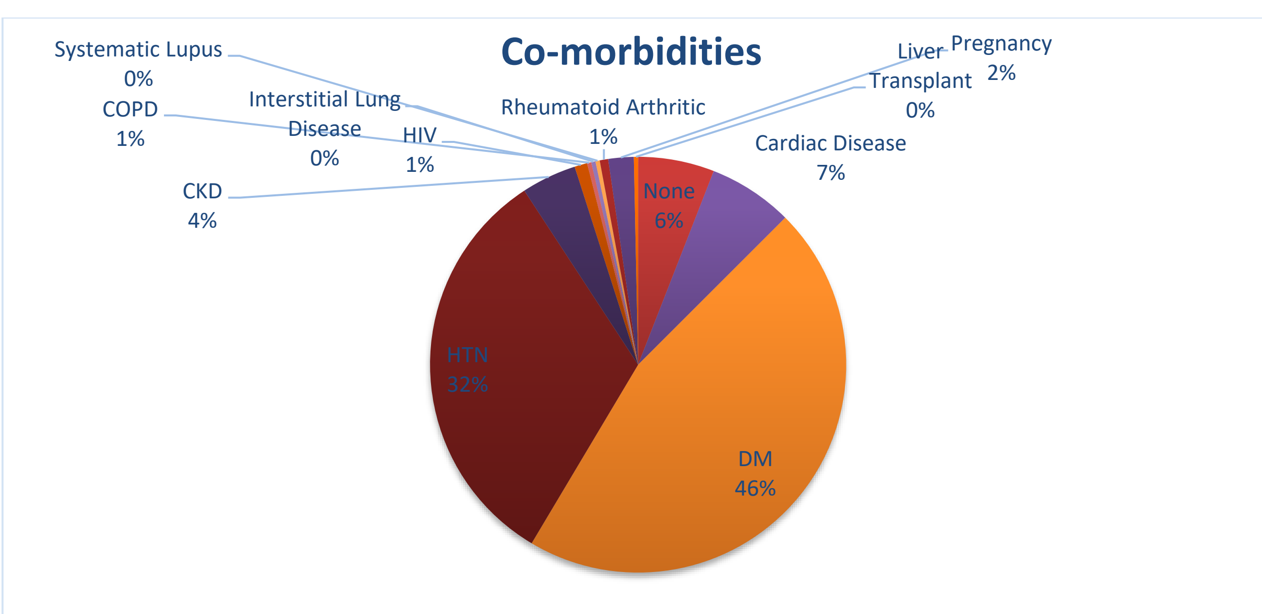
Association of Clinical Characteristics with Hospital Course					
		Hospital Course			
	Symptoms	No ICU & No vent.	Yes ICU & No vent.	Yes ICU & Yes vent.	Chi-sqare P-value
Symptoms	Fever	258 (88.1%)	21 (7.2%)	14 (4.8%)	0.542
	Cough	232 (89.6%)	18 (%)	9 (3.5%)	
	SOB	1430 (81.8%)	17 (10.7%)	12 (7.5%)	
	Diarrhea	53 (88.3%)	6 (10%)	1 (1.7%)	
ST Chest findings	Bilateral veiling	1 (100%)	0 (%)	0 (0%)	0.029*
	Bilateral infiltrates	1 (100%)	0 (%)	0 (0%)	
	Bilateral opacities	2 (100%)	0 (%)	0 (0%)	
	Bilateral Consolidations	55 (77.5%)	7 (9.9%)	9 (12.7%)	
	Bilateral Ground GO	71 (89.9%)	5 (6.3%)	3 (3.8%)	
	Bilateral infiltrates	7 (87.5%)	0 (0%)	1 (12.5%)	
	Unilateral Consolidation	39 (90.7%)	4 (9.3%)	0 (0%)	
	Unilateral infiltrates	1 (100%)	0 (0%)	0 (0%)	
	Unilateral Ground GO	5 (100%)	0 (0%)	0 (0%)	
	Unilateral opacities	2 (66.7%)	1 (33.3%)	0 (0%)	
	Unilateral veling	1 (100%)	0 (0%)	0 (0%)	
	normal	170 (96.6%)	5 (2.8%)	1 (0.6%)	
	Not done	4 (100%)	0 (0%)	0 (0%)	

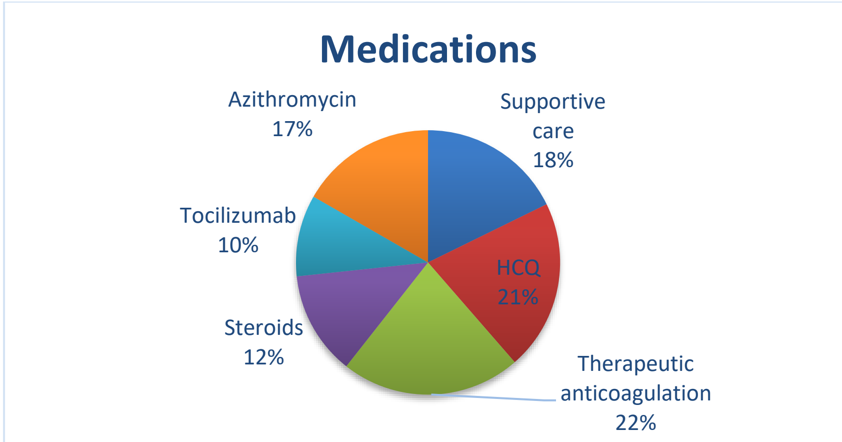
** Highly significant as p-value < 0.01, * Significant as p-value < 0.05, NS ** Nonsignificant as p-value > 0.05

Title:









(2) Project-III on Oncology Data

Table 1: Patients, clinical, tumor and anti-cancer treatment characteristics

Characteristics of the Patients	**All Patients (N=1694)**
Age	**N (%)**
> 65	381 (22.5)
≤ 65	1313 (77.5)
Gender	
Male	461 (27.2)
Female	1233 (72.8)
BMI	
Underweight	111 (6.6)
Normal	494 (29.4)
Overweight	492 (29.2)
Obese	591 (34.9)
Co-morbidities	
Yes	664 (39.2)
No	1030 (60.8)
ECOG	

Table 2: Outcome characteristics

Patients Outcome characteristics	N (%)
30-day mortality	**patients died (N=59, 3.5%)**
Disease progression	35 (71.4)
Sepsis	10 (20.4)
Pneumonia	2 (4.1)
Bleeding	1 (2.1)
Other	1 (2.1)

Figure: Regression, forest plot for 30-day mortality:

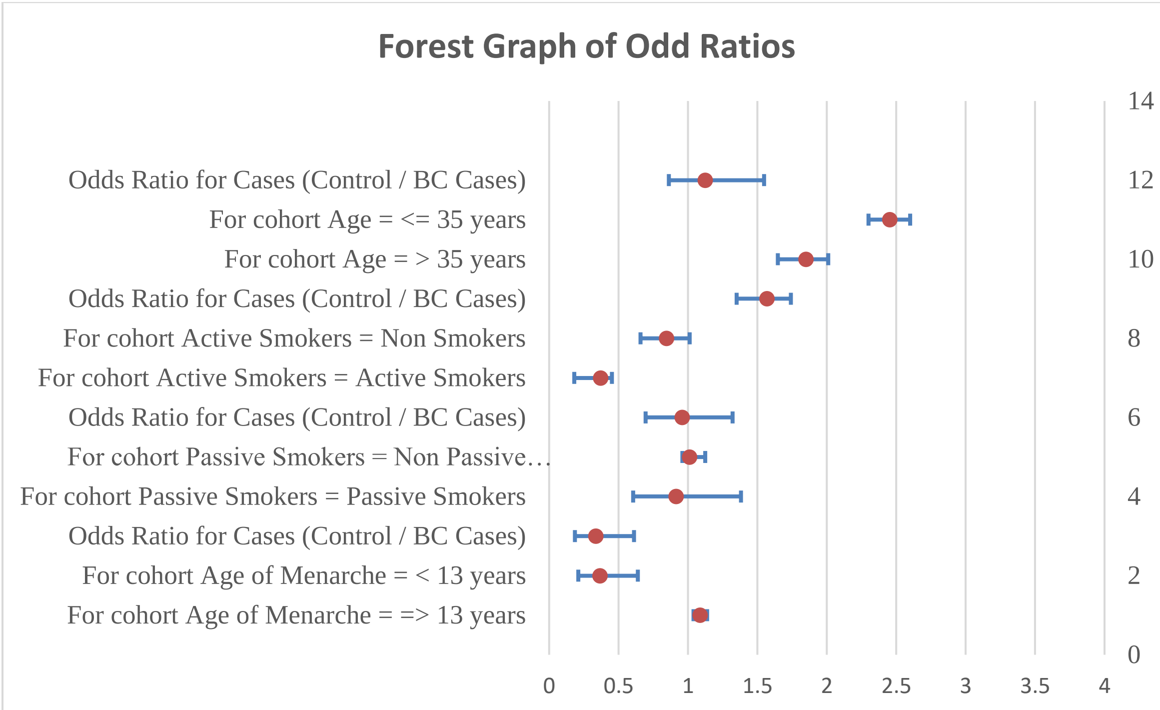
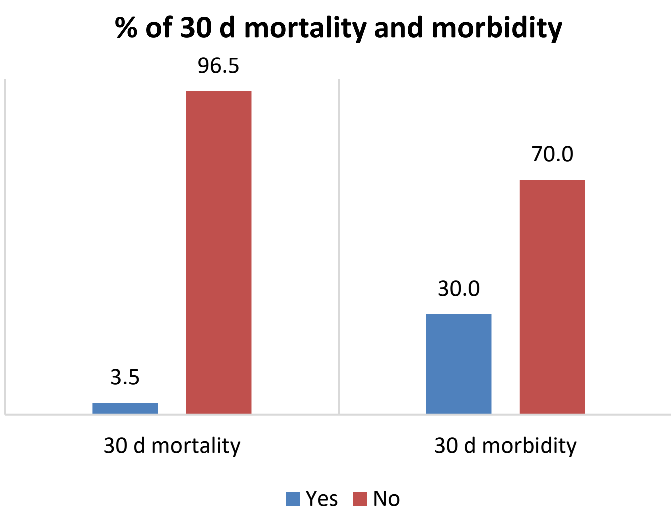
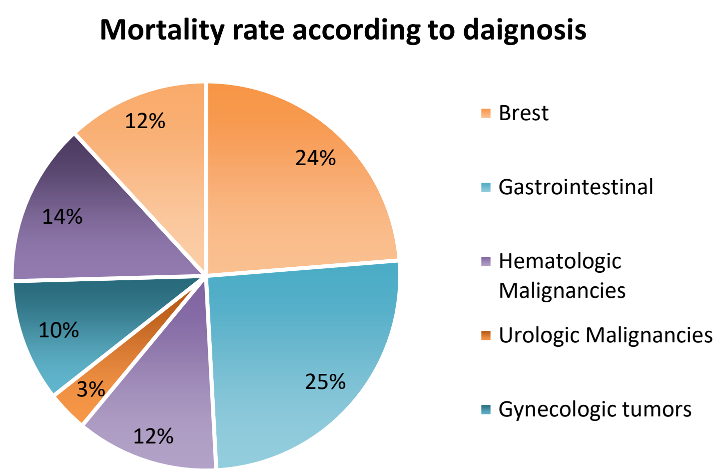


Figure 1: Percentage of 30-day mortality & morbidity and mortality rate according to diagnosis.





Statistical analysis:

logistic regression analysis was conducting to assess any associations between the explanatory variables and 30-day mortality. the results of these logistic regression analyses as adjusted odds ratios (OR) that reflect the effect of each variable in our multivariable regression model, alongside the unadjusted OR and proportion of patients with 30-day mortality. We used Z-tests to examine significance and a p value of 0·01 for statistical significance associated with the OR. We used the IBM SPSS version 26 to diagnose co-linearity between variables in the model. For each model, the mean variance inflation factor was lower than 1·04,

(4) Project-IV on Oncology Data

Cox hazard regression Model for Breast Cancer

The relationship between the hazard rate and a set of covariates for breast cancer (BC) is expressed as by Cox hazard regression model as

$$h(T)_{BC} = h_o(\mathrm{T})\,\mathrm{e}^{-0.491X_1+0.139X_2+2.314X_3+0.78X_4+0.531X_5+0.339X_6+0.323X_7+0.661X_8+1.254X_9+0.135X_{10}}$$

Where $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9 and X_{10}$ represent given birth, Age at first birth, Current menopausal Status, First degree family history of Breast Cancer, First degree family history relation of Breast cancer, First degree family history of ovarian Cancer, Hysterectomy, History of endometriosis, History of uterine fibroids and BMI respectively. T is the age of the patient and $h_o(\mathrm{T})$ is the baseline hazard when all covariates are equal to zero.

Cox hazard regression Model for Endometrial Cancer

The relationship between the hazard rate and a set of covariates for endometrial cancer (EC) is expressed as by Cox hazard regression model as

$$h(T)_{EC} = h_o(\mathrm{T})\,\mathrm{e}^{0.127X_1 + 1.003X_2 + 0.55X_3 + 1.884X_4 + 1.047X_5 + 0.601X_6}$$

Where $X_1, X_2, X_3, X_4, X_5 and X_6$ represent Age of menarche, Age at menopause, Breast biopsy, First degree family history of ovarian Cancer, First degree family history relation of Ovarian cancer and History of endometriosis respectively. T is the age of the patient and $h_o(\mathrm{T})$ is the baseline hazard rate when all covariates are equal to zero.

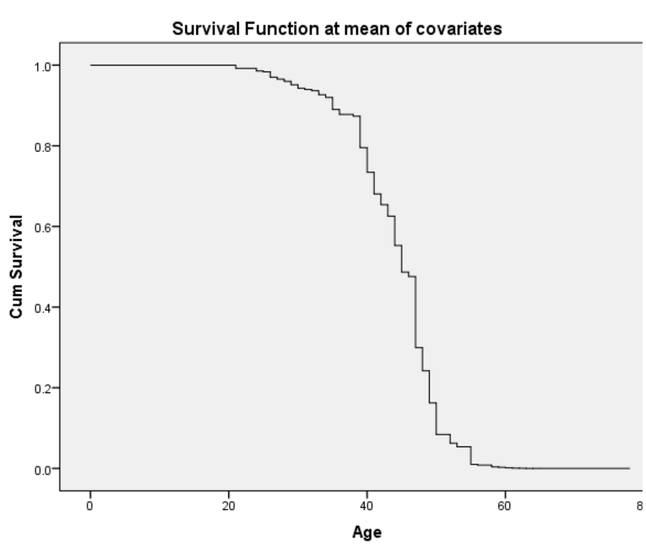
Test of Cox Proportional Hazard Regression Model Fit

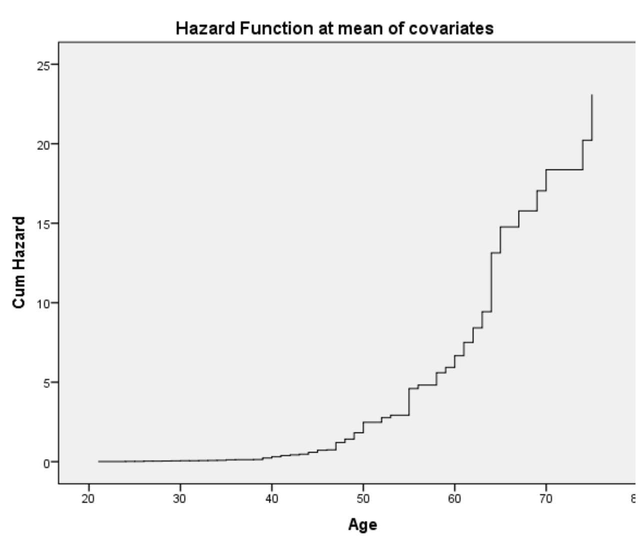
Change From Previous Block		
Chi-square	df	Sig.
189.328	6	.000**

** Highly significant at 5% level of significant as P-value <0.01

Cox Proportional Hazard Regression Model for Breast Cancer

Variables in Model	Hazar Ratio	Standard Error	Test Statistic	P-value	Expected Hazar Ratio	95% CI for Exp(B)	
	B				Ex(B)	Lower Bound	Upper Bound
Age of menarch	0.127	.065	3.785	.052	.881	.775	1.001
Age at menopause	1.003	.087	131.487	.000	.367	.309	.436
Breast biopsy	0.550	.189	8.452	.004	1.732	1.196	2.509
First degree family history of ovarian Cancer	1.884	.862	4.781	.029	6.582	1.216	35.641
First degree family history relation of Ovarian cancer	1.047	.670	2.438	.118	.351	.094	1.306
History of endometriosis	0.601	.270	4.951	.026	.548	.323	.931





Cox hazard regression Model for Ovarian Cancer

The relationship between the hazard rate and a set of covariates for ovarian cancer (OC) is expressed as by Cox hazard regression model as

$$h(T)_{OC} = h_o(\mathrm{T})\,\mathrm{e}^{1.207X_1 + 0.341X_2 + 1.579X_3 + 1.332X_4 + 0.954X_5 + 0.535X_6 + 0.253X_7}$$

Where $X_1, X_2, X_3, X_4, X_5, X_6 and X_7$ represent Current Menopausal Status, Age at menopause, First degree family history of Breast Cancer, First degree family history relation of Breast cancer, Hysterectomy, History of endometriosis and Age of menarche respectively. T is the age of the patient and $h_o(\mathrm{T})$ is the baseline hazard rate when all covariates are equal to zero.

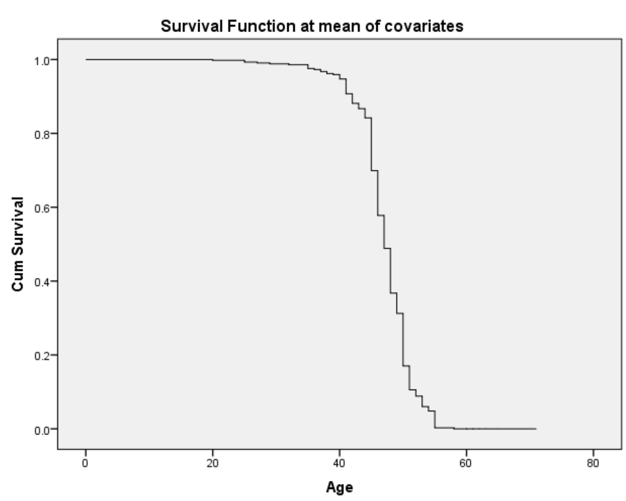
Test of Cox Proportional Hazard Regression Model Fit

Change From Previous Block		
Chi-square	df	Sig.
189.328	6	.000**

** Highly significant at 5% level of significant as P-value <0.01

Cox Proportional Hazard Regression Model for Ovarian Cancer

Variables in Model	Hazar Ratio B	Standard Error	t-test	P-value	Expected Hazar Ratio Ex(B)	95% CI for Exp(B) Lower Bound	Upper Bound
Current menopausal Status	1.207	.300	16.247	.000	.299	.166	.538
Age at menopause	.341	.128	7.068	.008	.711	.553	.914
First degree family history of BC	1.579	.634	6.207	.013	4.849	1.400	16.788
First degree family history relation of BC	1.332	.555	5.752	.016	.264	.089	.784
Hystrectomy	.954	.233	16.748	.000	2.596	1.644	4.099
History of endometriosis	.535	.265	4.076	.043	1.707	1.016	2.869
Age of menarche	.253	.106	5.716	.017	.777	.631	.955



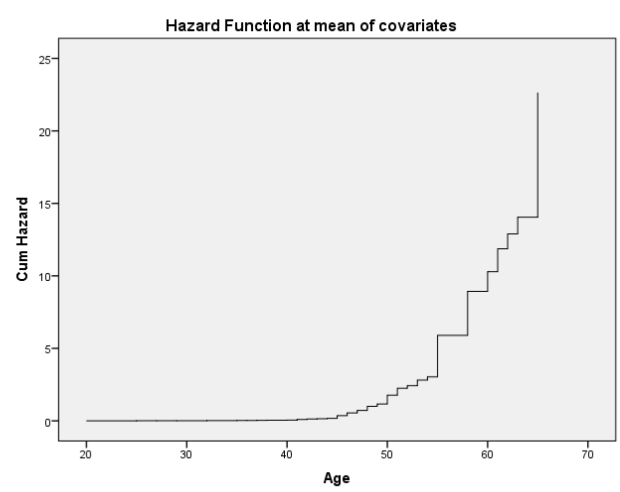


Table Title: Descriptive Statistics along with measure of Association of Cancer types with factors and test of columns proportions between factors

Characteristics	Subcategory	Breast Cancer Cases, N	Breast Cancer Percent	Endometrial Cancer Cases, N	Endometrial Cancer Percent	Ovarian Cancer Cases, N	Ovarian Cancer Percent	Chi-Square P-value
Age	<= 25	2_a	0.6%	10_b	5.2%	$4_{a, b}$	2.5%	0.000**
	26-45	151_a	43.6%	102_a	52.6%	68_a	43.0%	
	46-65	188_a	54.3%	71_b	36.6%	85_a	53.8%	
	65 =>	5_a	1.4%	11_b	5.7%	1_a	0.6%	
BMI	< 18.5	5_a	1.4%	10_b	5.2%	9_b	5.7%	0.000**
	18.5-22.9	21_a	6.1%	43_b	22.2%	32_b	20.3%	
	23-24.9	103_a	29.8%	27_b	13.9%	23_b	14.6%	
	=> 25	217_a	62.7%	114_a	58.8%	94_a	59.5%	
Age of menarche	< 12 years old	5_a	1.4%	0_a		16_b	10.1%	0.000**
	12 years old	39_a	11.3%	17_a	8.8%	65_b	41.1%	
	13 years old	159_a	46.0%	62_b	32.0%	43_b	27.2%	
	14 years old	88_a	25.4%	47_a	24.2%	32_a	20.3%	
	15 years old	32_a	9.2%	25_a	12.9%	2_b	1.3%	

** There is highly significant association between the cancer types and above tabulated factors as p-value <0.001. Also above table shows results of pairwise comparisons of column proportions and indicates which pairs of columns (for a given row) are significantly different. Significant differences are indicated in the crosstabulation table with APA-style formatting using subscript letters and are calculated at the 0.05 significance level. As in the interval <= 25 years the proportion of breast cancer is significantly different from endometrial cancer (having different superscripts) at 5% level of significant while ovarian cancer have both superscripts (a,b) which indicates that its proportion in this interval is same as BC and EC.

Similarly discuss the other results……………………..

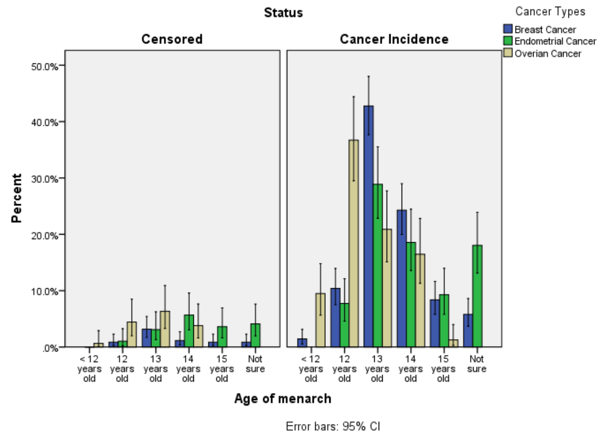
Table Title: Descriptive Statistics along with measure of Association of Cancer types with factors and test of columns proportions between factors

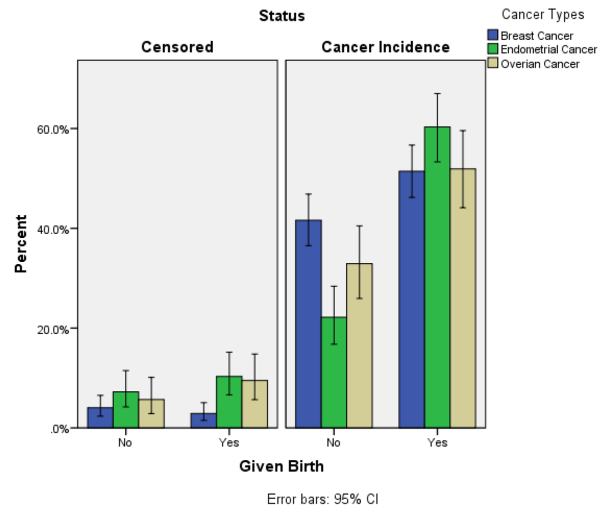
Characteristics	Subcategory	Breast Cancer		Endometrial Cancer		Ovarian Cancer		Chi-Square P-value
		Cases, N	Percent	Cases, N	Percent	Cases, N	Percent	
Age at first birth	No children	159_a	46.0%	58_b	29.9%	$61_{a,b}$	38.6%	0.000**
	< 16 years	36_a	10.4%	5_b	2.6%	$11_{a,b}$	7.0%	
	16-19 years	81_a	23.4%	19_b	9.8%	46_a	29.1%	
	20-24 years	52_a	15.0%	75_b	38.7%	27_a	17.1%	
	25-29 years	15_a	4.3%	26_b	13.4%	$11_{a,b}$	7.0%	
	30-34 years	2_a	0.6%	11_b	5.7%	$2_{a,b}$	1.3%	
	40 or more years	1_a	0.3%	0_a		0_a		
Oral Contraceptive use	Never use	285_a	82.4%	167_a	86.1%	128_a	81.0%	0.256NS
	Less than 1 year	45_a	13.0%	17_a	8.8%	17_a	10.8%	
	1-4 years	14_a	4.0%	9_a	4.6%	9_a	5.7%	
	5-9 years	2_a	0.6%	1_a	0.5%	4_a	2.5%	
Current menopausal Status	Premenopausal	153_a	44.2%	75_a	38.7%	73_a	46.2%	0.000**
	Menopause	34_a	9.8%	68_b	35.1%	42_b	26.6%	
	Post menopause	159_a	46.0%	51_b	26.3%	43_b	27.2%	
Age at menopause	Still menstruating	153_a	44.2%	73_a	37.6%	74_a	46.8%	0.000**
	<40 years	11_a	3.2%	34_b	17.5%	6_a	3.8%	
	40-44 years	20_a	5.8%	23_b	11.9%	$12_{a,b}$	7.6%	
	45-49 years	64_a	18.5%	30_a	15.5%	30_a	19.0%	
	50-54 years	48_a	13.9%	28_a	14.4%	22_a	13.9%	
	55 or more years	50_a	14.5%	6_b	3.1%	$14_{a,b}$	8.9%	
History of benign breast disease	No	39_a	11.3%	161_b	83.0%	109_c	69.0%	0.000**
	Yes	307_a	88.7%	33_b	17.0%	49_c	31.0%	

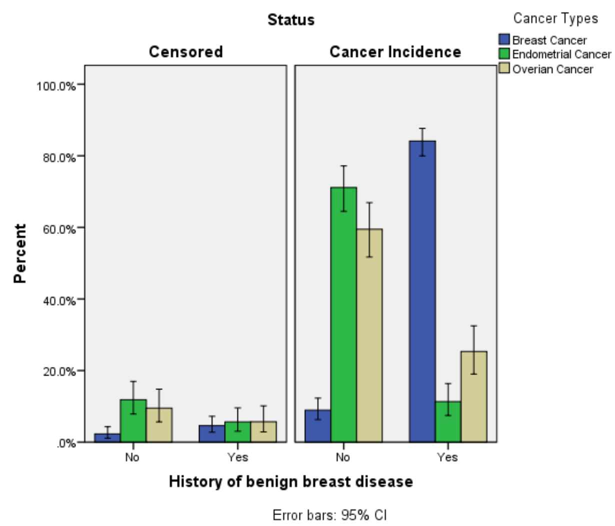
** There is highly significant association between the cancer types and above tabulated factors as p-value <0.001.

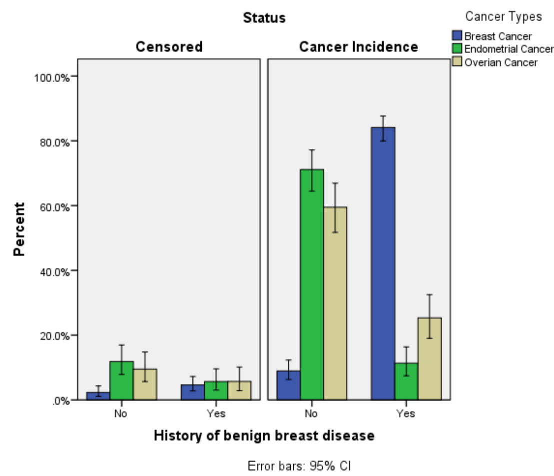
NS There is non significant association between the cancer types and above tabulated factors as p-value > 0.05.

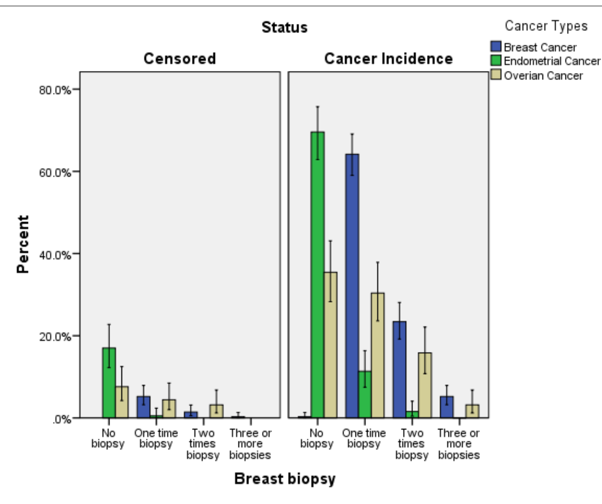
Data Visualizations

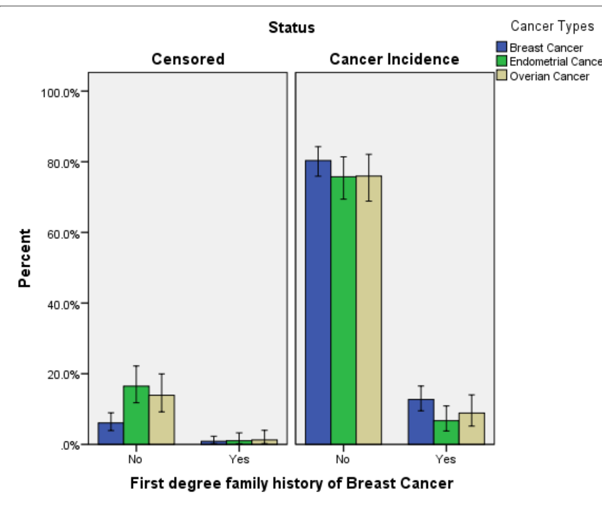


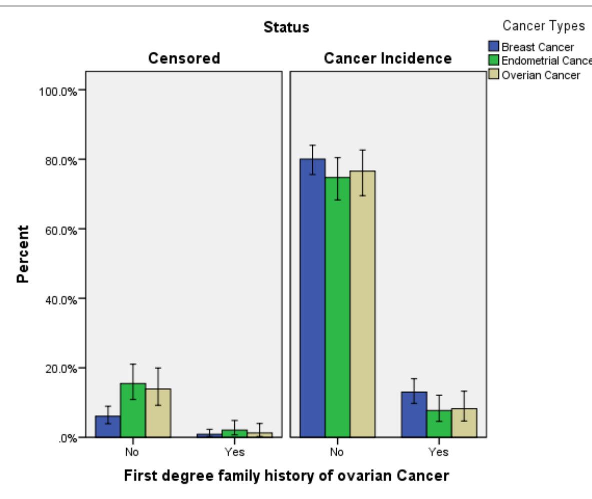


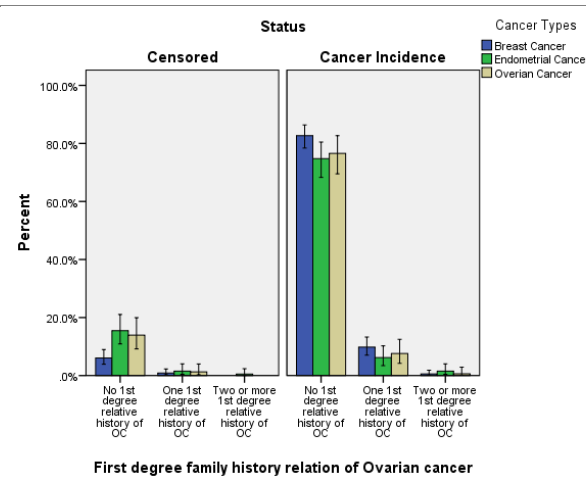




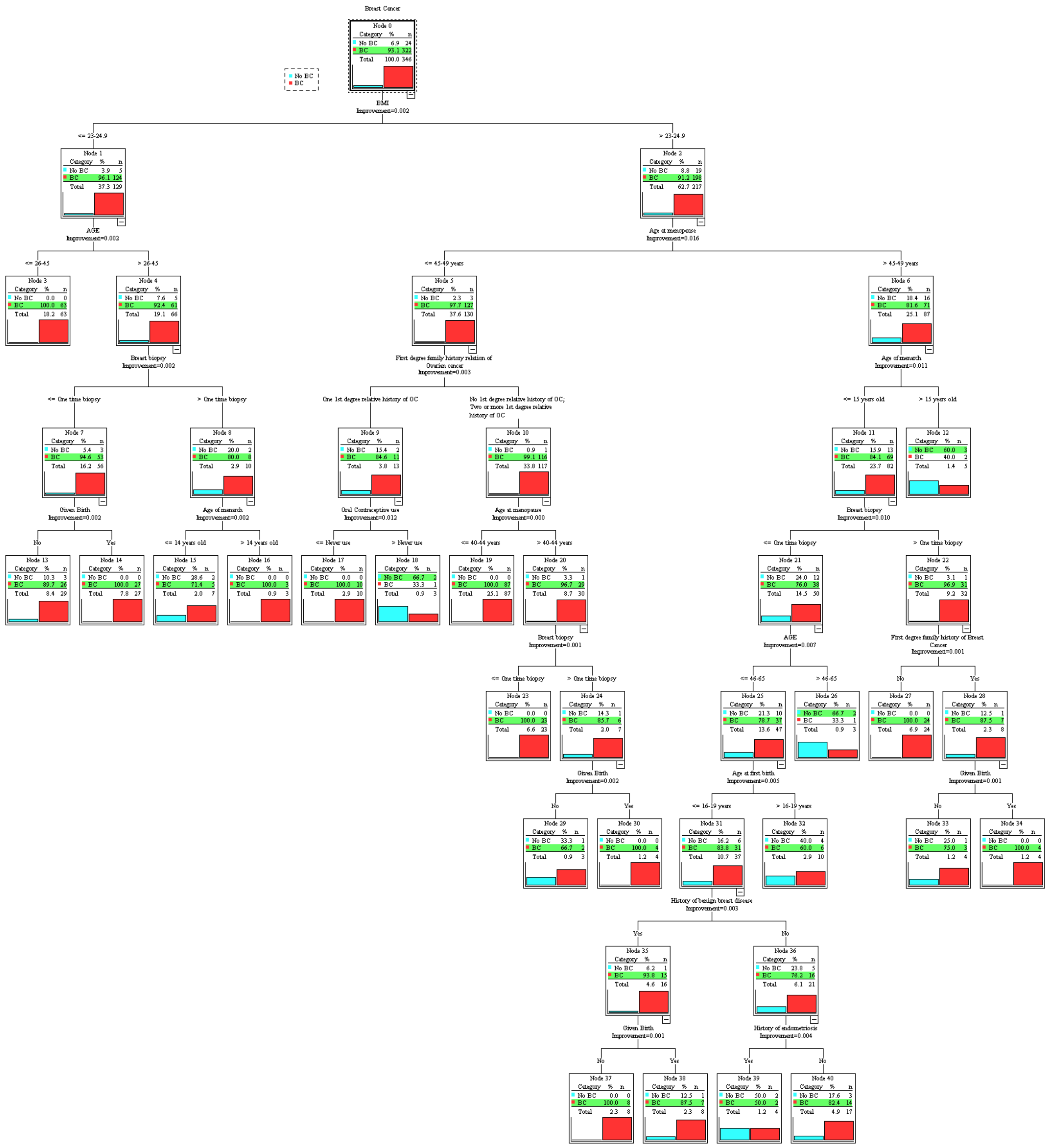


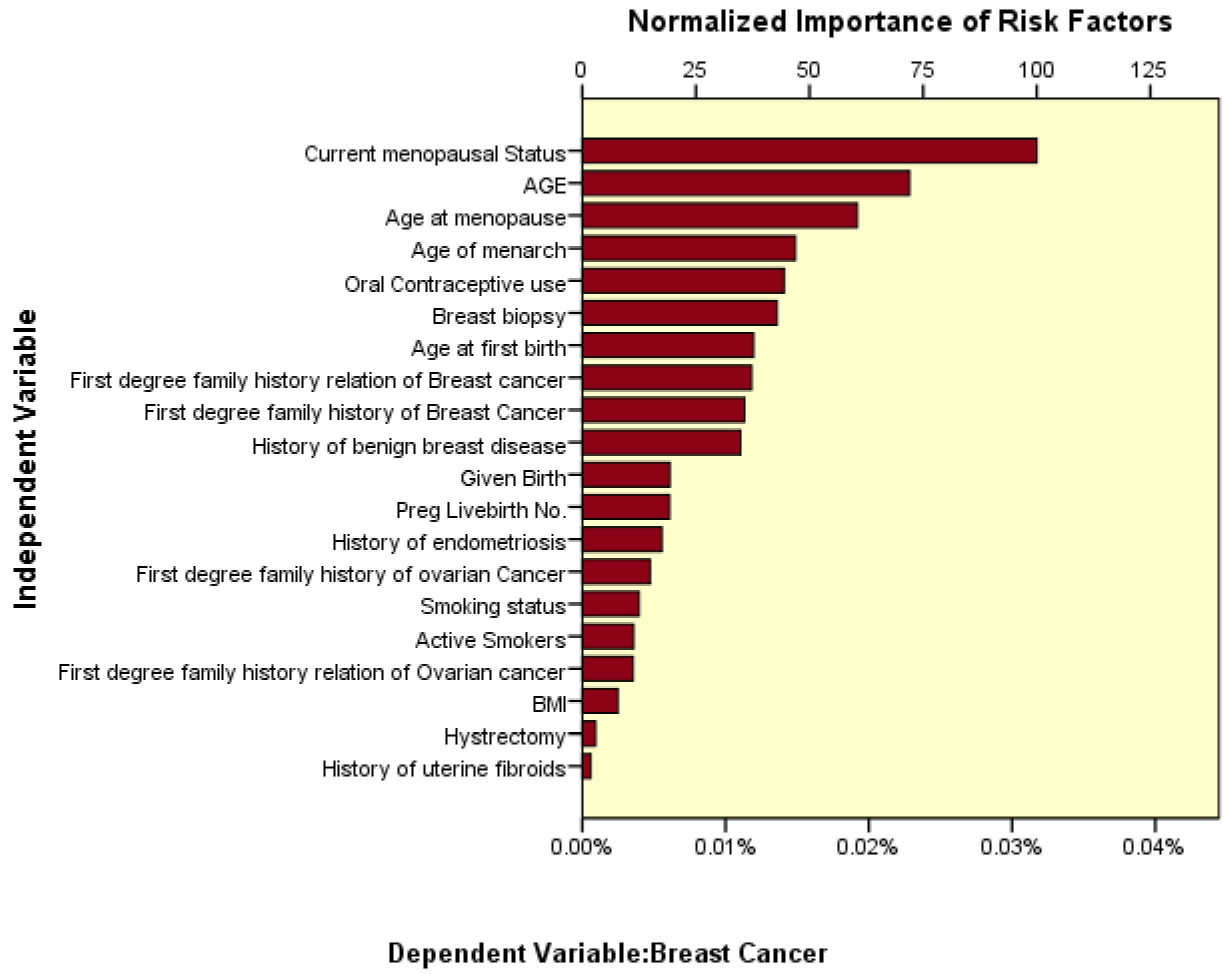






Classification and Regression Tree of Breast Cancer




Normalized Importance of Risk Factors
0
25
50
75
100
125
Independent Variable
Current menopausal Status
AGE
Age at menopause
Age of menarch
Oral Contraceptive use
Breast biopsy
Age at first birth
First degree family history relation of Breast cancer
First degree family history of Breast Cancer
History of benign breast disease
Given Birth
Preg Livebirth No.
History of endometriosis
First degree family history of ovarian Cancer
Smoking status
Active Smokers
First degree family history relation of Ovarian cancer
BMI
Hystrectomy
History of uterine fibroids
0.00%
0.01%
0.02%
0.03%
0.04%
Dependent Variable:Breast Cancer

Classification and Regression Tree of Ovarian Cancer

Some Recently Completed Projects (Business Management)

Project-I: Panel Data Regression Analysis

EMPERICAL RESULTS

Table 1. Correlation Analysis

Variables	ROCE	CR	QR	RTP	PTP	ITP
ROCE	1					
CR	0.0714	1				
QR	0.0302	0.9049	1			
RTP	-0.5101	-0.0374	-0.0586	1		
PTP	0.0648	-0.2210	-0.1908	0.2838	1	
ITP	-0.2070	0.1691	-0.0938	0.2927	0.2948	1

Table 1 provided the degree of relationship between all variables under studies. The positive sign of the correlation coefficient represents direct relationship between indicators while the negative sign is for indirect relationship. There is direct relationship between profitability and CR as the value of correlation coefficient is 0.0714. There is direct relationship between profitability and QR as the value of correlation coefficient is 0.0302. There is indirect relationship between profitability and RTP as the value of correlation coefficient is -0.5101. There is direct relationship between profitability and PTP as the value of correlation coefficient is 0.0648. There is indirect relationship between profitability and ITP as the value of correlation coefficient is -0.2070.

Panel Data Regression Model

When we need to analyze the data sets with multiple observations of cross-sectional units like profitability and firms over the period of time, we can use panel data that is a branch of time series analysis.

Panel data models of two types:

1. Homogeneous panel data models that assume that model parameters are same for all the firms.
2. Heterogeneous panel data models that assume that model parameters vary across firms.

The model that I have decided to use for analysis of panel data is

$$(\text{ROCE})_{it} = \beta_0 + \beta_1(\text{CR})_{it} + \beta_2(\text{QR})_{it} + \beta_3(\text{RTP})_{it} + \beta_4(\text{PTP})_{it} + \beta_5(\text{ITP})_{it} + \varepsilon_{it};$$

$$i = 1,2,3,\ldots,\text{N}; \quad \text{t} = 1,2,3,\ldots,\text{T}; \qquad (1)$$

The subscript i in the model is a cross-sectional unit such as a company and t represents the time dimension.

Where (ROCE) is return on capital employed our dependent variable, following are independent variables (CR) Current Ratio, (QR) Quick Ratio, (RTP) Receivable Turnover Period, (PTP) Payable Turnover Period, (ITP) Inventory Turnover Period Inventory and ε_{it} is the error term.

Empirical Panel Data Modeling

Empirical model is developed to analyze the impact of working capital management on profitability of the selected companies. For this purpose, panel data of 20 companies recorded from 2015 to 2019 is used to develop this model empirically. After implementation of full model with fixed effects, to capture the heterogeneity and with random effects to capture time component, we have these empirical models:

Empirical Model-I

$$(\text{ROCE})_{it} = 27.0906 + 6.9188(\text{CR})_{it} - 7.3996(\text{QR})_{it} - 0.2440(\text{RTP})_{it} + 0.1268(\text{PTP})_{it} - 0.1851(\text{ITP})_{it} \qquad (2)$$

$$i = 1,2,3,\ldots,100; \qquad \text{t} = 1,2,3,4,5$$

Table 2. Panel Data Regression Full Model with Fixed Effects

ROCE	Coef.	Std. Error	t-test	P-value	95% Conf. Interval
CR	6.9188	11.3368	0.61	0.544	(-15.6654 , 29.5029)
QR	-7.3997	15.2368	-0.49	0.629	(-37.7528 , 22.9535)
RTP	-0.2441	0.1239	-1.97	0.053	(-0.4908 , 0.0028)
PTP	0.1268	0.0417	3.04	0.003	(0.0437 , 0.2099)
ITP	-0.1851	0.0583	-3.18	0.002	(-0.3011 , -0.0689)
Constant	27.0906	9.1643	2.96	0.004	(8.8343 , 45.3469)

F-test	P-value	R-square
4.2100	0.0020	0.2293

The above table showed that the proposed model in equation (1) is highly significant as the p-value is 0.0020 <0.01, 1% level of significance. It explained the overall 22.93% variation as the R-square value is presented there. The empirically estimated parameters of the proposed model are presented as coefficients in the second column of the table 2 which showed that if one unit of CR is increased keeping the effect of other as constant then there will be on average 6.9188 unit increase in ROCE. Similarly, if one unit of QR is increased keeping the effect of other as constant then there will be on average 7.3997 unit decrease in ROCE, if one unit of RTP is increased keeping the effect of other as constant then there will be on average 0.2441 unit decrease in ROCE, if one unit of PTP is increased keeping the effect of other as constant then there will be on average 0.1268 unit increase in ROCE, if one unit of ITP is increased keeping the effect of other as constant then there will be on average 0.1851 unit decrease in ROCE. The interpretation of the constant term is sometime existing, and it is interpreted as there will be 27.0906 units of ROCE if no increment is made in any variable.

Table 4. Hausman Test Results

ROCE	(b) Fixed Effects	(B) Random Effects	(b-B) Difference	Std. Error
CR	6.9188	13.2944	-6.3756	5.8449
QR	-7.3997	-13.8871	6.4874	8.6803
RTP	-0.2441	-0.3471	0.1031	0.0504
PTP	0.1268	0.1279	-0.011	0.0216
ITP	-0.1851	-0.1933	0.0082	0.0225

Chi-square	P-value
6.99	0.216

The results of Hausman test presented in the table 4 suggested the empirical Model-II should be used as the p-value is 0.216 > 0.05.

Project-II: Panel Data Regression Analysis

Results and Discussions

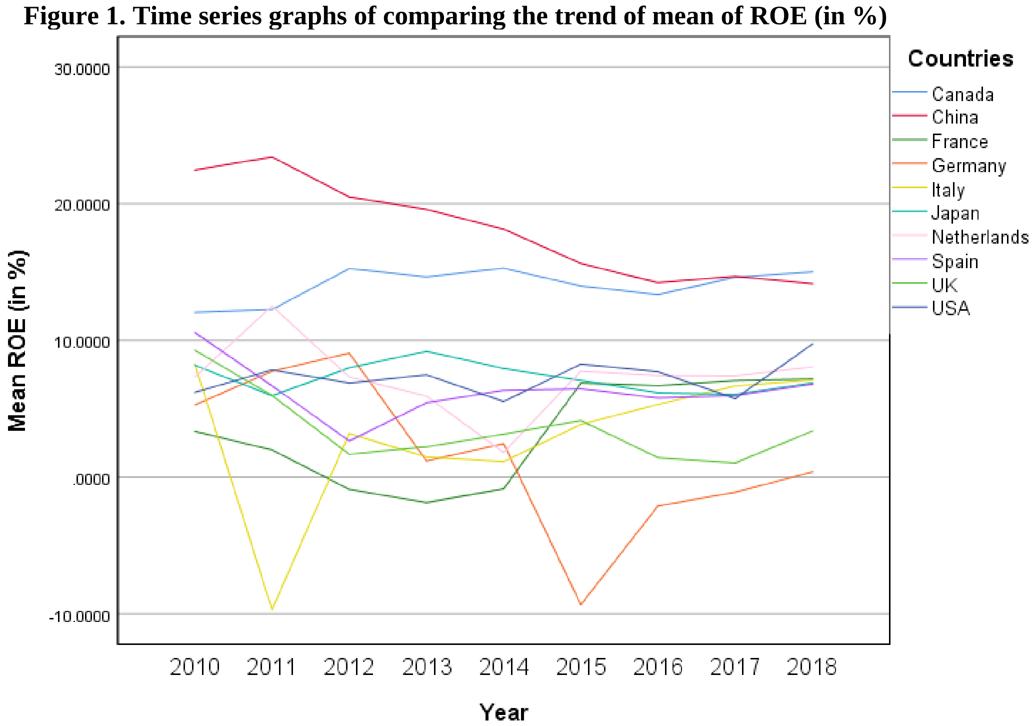
Table 1. Descriptive Statistics of Countries

Countries	Descriptive Statistics						
	Variables	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Canada	ROA (in %)	0.660	0.960	0.841	0.087	-0.381	-0.718
	ROE (in %)	11.150	17.020	14.052	1.845	0.093	-1.087
	Unemployment	5.831	8.056	6.991	0.622	-0.258	0.158
	Interest rate	-0.257	3.684	1.042	1.222	1.009	0.579
	Exchange rate	81.526	101.564	91.257	8.618	0.033	-2.003
	GDP	1528.243	1847.209	1703.562	117.403	-0.240	-1.570
	Inflation	-0.877	3.236	1.709	1.187	-1.001	0.879
China	ROA (in %)	0.790	2.230	1.196	0.225	1.692	7.146
	ROE (in %)	10.690	37.200	18.091	4.715	1.283	3.866
	Unemployment	3.600	4.672	4.053	0.227	0.231	0.894
	Interest rate	-1.402	4.521	1.961	2.227	-0.324	-1.569
	Exchange rate	6.143	6.770	6.459	0.235	0.021	-1.639
	GDP	6087.165	13894.817	10079.671	2305.914	-0.139	-0.756
	Inflation	-0.003	8.076	3.291	2.571	0.694	-0.713
France	ROA (in %)	-0.350	0.750	0.171	0.273	-0.370	0.231
	ROE (in %)	-14.100	18.710	3.284	8.239	-0.898	0.530
	Unemployment	8.811	10.354	9.568	0.579	0.043	-1.596
	Interest rate						
	Exchange rate						
	GDP	2438.208	2861.408	2682.622	152.646	-0.382	-1.273
	Inflation	0.522	1.162	0.852	0.250	-0.221	-1.611
Germany	ROA (in %)	-0.390	0.240	0.022	0.187	-1.325	2.482
	ROE (in %)	-9.330	9.070	1.511	5.589	-0.563	0.648
	Unemployment	3.384	6.966	4.917	1.106	0.444	0.121
	Interest rate						
	Exchange rate	92.521	100.000	96.475	2.476	-0.197	-1.060
	GDP	3360.550	3949.549	3636.385	210.807	0.121	-1.361
	Inflation	0.646	1.969	1.391	0.436	-0.297	-0.801
Italy	ROA (in %)	-1.200	0.770	0.188	0.562	-2.175	5.729
	ROE (in %)	-9.690	8.280	3.039	5.375	-1.865	4.255
	Unemployment	8.359	12.683	10.846	1.560	-0.868	-0.394
	Interest rate	1.766	3.951	3.060	0.789	-0.461	-1.230
	Exchange rate						
	GDP	1835.899	2291.991	2063.644	146.121	-0.303	-0.514
	Inflation	0.436	1.607	1.038	0.373	0.201	-0.209
Japan	ROA (in %)	0.170	0.630	0.376	0.089	0.377	1.900
	ROE (in %)	4.100	10.610	7.276	1.462	0.259	0.164
	Unemployment	2.400	5.100	3.691	0.835	0.113	-0.958
	Interest rate	-0.982	3.561	1.371	1.461	-0.100	-0.861
	Exchange rate	69.417	101.139	83.660	12.346	0.578	-1.529
	GDP	4389.476	6203.213	5244.510	606.183	0.525	-1.033

	Inflation	-1.895	2.145	-0.092	1.306	0.409	-0.709
Netherlands	ROA (in %)	0.100	0.460	0.349	0.119	-1.227	1.208
	ROE (in %)	1.810	12.520	7.289	2.741	-0.181	3.276
	Unemployment	3.830	7.416	5.777	1.226	-0.029	-1.106
	Interest rate	0.176	1.803	0.498	0.526	2.361	5.727
	Exchange rate	95.589	100.236	98.373	1.729	-0.479	-1.590
	GDP	765.265	914.105	850.476	51.756	-0.497	-0.811
	Inflation	0.194	2.208	0.979	0.648	0.584	0.101
Spain	ROA (in %)	0.170	0.660	0.439	0.128	-0.659	2.998
	ROE (in %)	2.650	10.580	6.302	2.036	0.551	3.428
	Unemployment	15.255	26.094	21.194	3.597	-0.274	-0.800
	Interest rate						
	Exchange rate	93.697	100.400	97.577	2.523	-0.518	-1.388
	GDP	1195.119	1478.773	1345.328	91.228	-0.350	-0.490
	Inflation	-0.223	1.381	0.393	0.543	0.865	-0.148
UK	ROA (in %)	-0.140	1.050	0.223	0.305	0.968	0.334
	ROE (in %)	-2.900	17.100	3.588	5.083	0.901	0.220
	Unemployment	1.172	2.382	1.461	0.405	1.500	0.890
	Interest rate	-1.509	-1.018	-1.284	0.144	0.377	-0.456
	Exchange rate	0.608	0.777	0.675	0.060	0.674	-1.288
	GDP	2475.244	3063.803	2759.893	166.030	0.233	-0.277
	Inflation	0.581	2.140	1.745	0.462	-1.805	2.640
USA	ROA (in %)	-0.430	1.420	0.670	0.414	-0.365	-0.063
	ROE (in %)	-3.610	12.530	7.270	3.719	-1.007	0.893
	Unemployment	3.896	9.633	6.510	1.985	0.231	-1.415
	Interest rate	1.137	2.486	1.834	0.494	-0.126	-1.695
	Exchange rate	1.000	1.000	1.000	0.000		
	GDP	14992.053	20580.223	17558.927	1774.312	0.179	-1.077
	Inflation	1.069	2.360	1.694	0.448	-0.293	-1.317

Table 1 represented the country wise descriptive statistics comprising of minimum value, maximum value, mean, standard deviation, skewness, and kurtosis of the variables under study. Skewness is a measure of the asymmetry of the probability distribution of a random variable about its mean. kurtosis identifies whether the tails of a given distribution contain extreme values. Some says for skewness (−1,1) and (−2,2) for kurtosis is an acceptable range for being normally distributed. If skewness is less than −1 or greater than +1, the distribution is highly skewed. These two measures are used to see the normality of the data. From the table above it can be seen that our almost all data is normally distribute.

Figure 1. Time series graphs of comparing the trend of mean of ROE (in %)



From figure 1 it is cleared that the China has highest ROA (in %) profitability form 2010 to 2017, after 2017 The Canada took this place but both the China and the Canada have greater profitablity than rest of the countries. The China attained the highest value in 2011 and have downward trend form 2011 upto 2016 and again had increased the profitablity in 2017. The Italy got the minimum profit in 2011 and increased the profitabilty in 2012, again lose in from 2012 to 2014 then improved its profitabiltiy onward. The Germany got its maximum profitability in 2012 and minimum in 2015. It is cleared that the Germany is the country with minimum profitablity.

Figure 2. Time series graphs of comparing the trend of mean of ROA (in %)

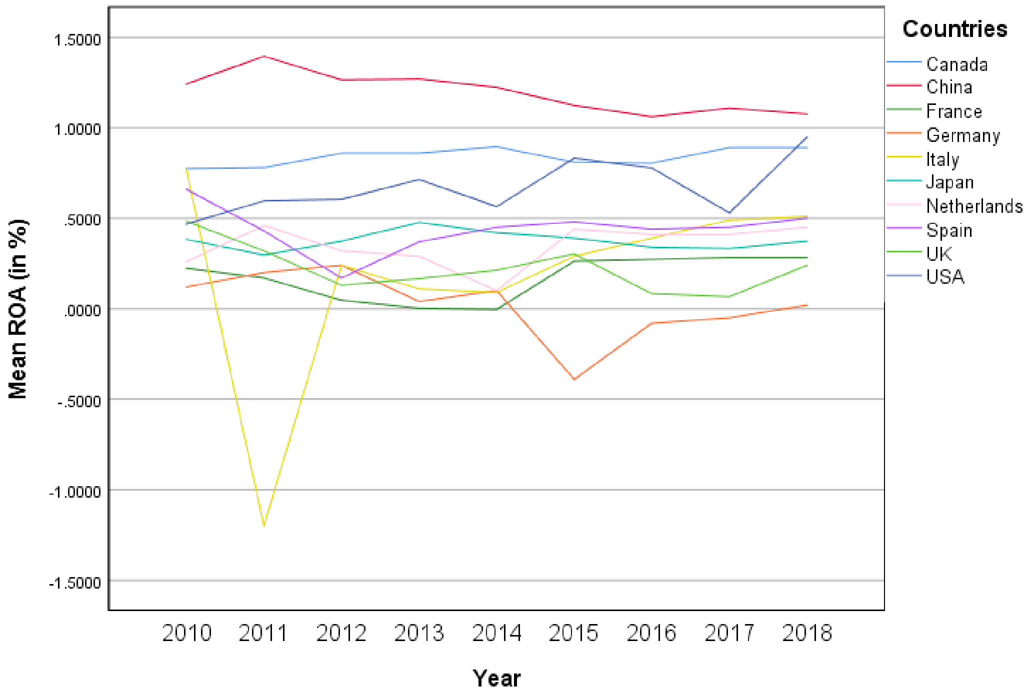


Figure 3. Time series graphs of comparing the trend of mean of unemployment

Research Hypotheses

Hypothesis that I have developed is based on these five variables are:

Null hypothesis-I:

H0 = There is no relationship between working ROA (in %) and independent variables: unemployment, interest rate, .exchange rate, GDP, inflation

Alternative hypothesis

H1 = There is indirect relationship between unemployment and ROA (in %).

H2 = There is direct relationship between interest rate and ROA (in %).

H3 = There is indirect relationship between exchange rate and ROA (in %).

H4 = There is direct relationship between GDP and ROA (in %)

H5 = There is indirect relationship between inflation and ROA (in %).

Correlation Analysis

Table 3. Correlation Analysis of ROA (in %) with other Independent variables

Correlations					
Variables	**ROA (in %)**	**Unemployment**	**Interest rate**	**Exchange rate**	**GDP**
ROA (in %)	1				
Unemployment	-0.171**	1			
	0.009				
Interest rate	0.258**	0.383**	1		
	0.000	0.000			
Exchange rate	-0.332**	0.366**	0.011	1	
	0.000	0.000	0.888		
GDP	0.397**	-0.185**	0.370**	-.638**	1
	0.000	0.005	0.000	0.000	
Inflation	-0.438**	-0.230**	-.499**	-0.408**	0.169**

	0.000	0.000	0.000	0.000	0.010
**. Correlation is significant at the 0.01 level (2-tailed).					

Table 3 provided the degree of relationship between all variables under studies. The significant positive sign of the correlation coefficient represents direct relationship between indicators while the significant negative sign is for indirect relationship. The correlation coefficient of ROA (in %) and unemployment is -0.171, highly statistically significant as the p-value is < 0.01, its negative sign ensured that there is indirect relationship between ROA (in %) and unemployment. The correlation coefficient of ROA (in %) and interest rate is 0.258, highly statistically significant as the p-value is < 0.01, its positive sign ensured that there is direct relationship between ROA (in %) and interest rate. The correlation coefficient of ROA (in %) and exchange rate is -0.332, highly statistically significant as the p-value is < 0.01, its negative sign ensured that there is indirect relationship between ROA (in %) and exchange rate. The correlation coefficient of ROA (in %) and GDP is 0.397, highly statistically significant as the p-value is < 0.01, its positive sign ensured that there is direct relationship between ROA (in %) and GDP. The correlation coefficient of ROA (in %) and inflation is -0.438, highly statistically significant as the p-value is < 0.01, its negative sign ensured that there is indirect relationship between ROA (in %) and inflation.

Regression Analysis of ROA (in %) with other independent variables

Table 5. Variance Inflation factor for Multicollinearity

	VIF	1/VIF
GDP	3.411	.293
Interest Rate	2.265	.441
Inflation	1.808	.553
Unemployment	1.596	.627
Exchange Rate	1.185	.844
Mean VIF	2.053	.

There is no multicollinearity between the variables.

Table 6. Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Source	chi2	df	p
Heteroskedasticity	26.530	20	0.149
Skewness	1.470	5	0.917
Kurtosis	2.920	1	0.088
Total	30.920	26	0.231

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of ROA (in%)

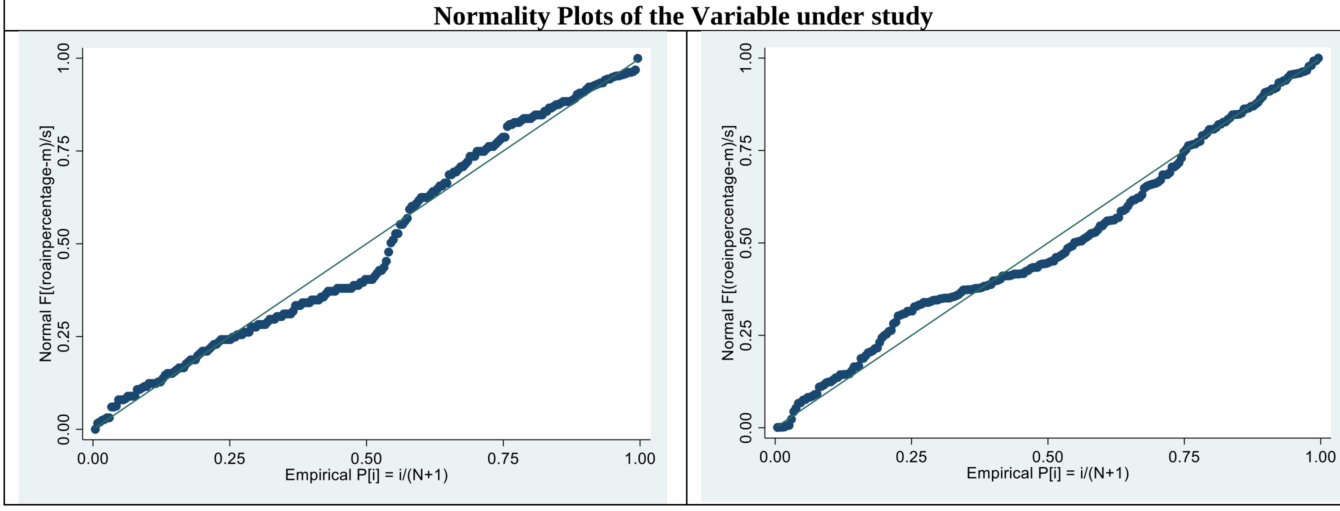
chi2(1) = 3.04

Prob > chi2 = 0.0810

The Breusch-Pagan / Cook-Weisberg test for heteroskedasticity suggested that there is no heteroskedasticity as p-value = 0.0810 > 0.05.

Table 7. Normal Probability Plots of the variables under study

Normality Plots of the Variable under study



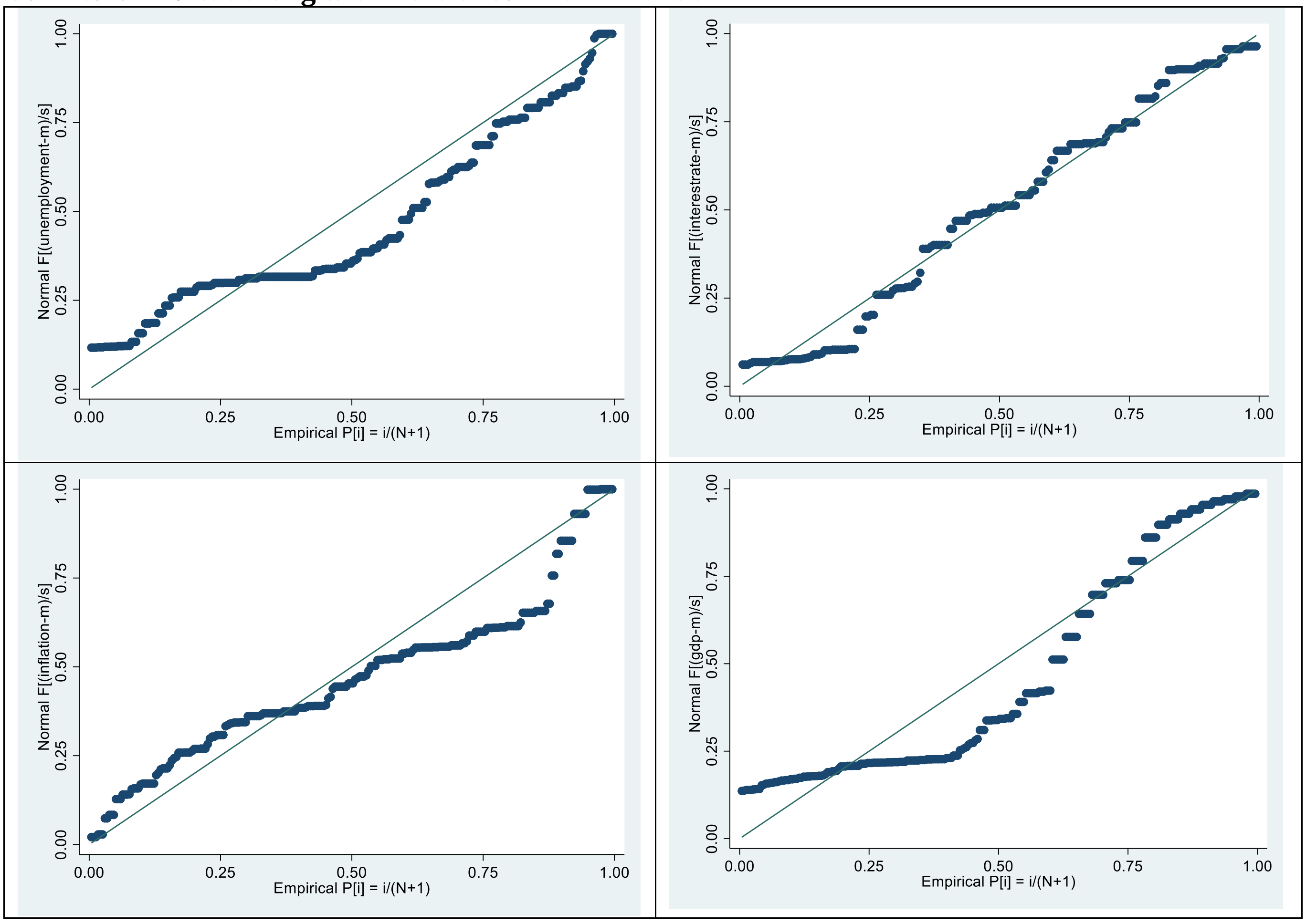


Table 7 showed that the data on all variables is approximately normally distributed.

Regression Analysis

The model that I have decided to use for analysis of panel data is

$$(\text{ROA(in\%)})_{it} = \beta_0 + \beta_1(\text{Unemployment})_{it} + \beta_2(\text{Interest rate})_{it} + \beta_3(\text{Exchagerate})_{it} + \beta_4(\text{GDP})_{it} + \beta_5(\text{Inflation})_{it} + \varepsilon_{it}$$

$i = 1, 2, 3, \ldots, \text{N}; \quad \text{t} = 1, 2, 3, \ldots, \text{T};$

The subscript i in the model is a cross-sectional unit such as a company and t represents the time dimension.

Where (ROA(in%)) is our dependent variable, following are independent variables (Unemployment), (Interest rate), (Exchange rate), (GDP), (Inflation), and ε_{it} is the error term.

Empirical Regression Modeling

Empirical model is developed to analyze the impact of working capital management on profitability of the selected companies. For this purpose, panel data of 10 countries and 26 banks recorded from 2010 to 2018 are used to develop this model empirically. After implementation of full regression model, we obtained the following empirical models.

Empirical Regression Model of ROA (in %)

$$(\text{ROA(in\%)})_{it} = 0.366 - 0.012(\text{Unemployment})_{it} + 0.1345(\text{Interest rate})_{it} - 0.0006(\text{Exchagerate})_{it} + 0.000172(\text{GDP})_{it} - 0.134(\text{Inflation})_{it} + \varepsilon_{it}$$

$i = 1, 2, 3, \ldots, \text{N}; \quad \text{t} = 1, 2, 3, \ldots, \text{T};$

Table 8. Normal Probability Plots of the variables under study

ROA (in%)	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Unemployment	-0.0124626	0.026	-0.47	0.635	-.064	0.039	
Interest Rate	0.1344976	0.02	6.63	0.000	.095	0.174	***
Exchange Rate	-.0005724	0.001	-0.40	0.687	-.003	0.002	
GDP	1.72e-6	7.71e-6	0.22	0.823	-0.0000134	0.0000168	
Inflation	-0.1336132	0.022	6.12	0.000	-.091	0.176	***
Constant	0.3656131	0.131	2.80	0.005	0.11	0.621	***

Mean dependent var	0.718	SD dependent var	0.453
Overall r-squared	0.579	Number of obs	180.000
Chi-square	74.465	Prob > chi2	0.000
R-squared within	0.115	R-squared between	0.686

**** p<.01, ** p<.05, * p<.1*

The above table showed that the proposed model is highly significant as the p-value of F-test is 0.000 <0.01, 1% level of significance. It explained the overall 68.6% variation as the R-square value is presented there. The empirically estimated parameters of the proposed model are presented as coefficients in the second column of the table 8 which showed that if one unit of unemployment is increased keeping the effect of other as constant then there will be on average 0.0125 unit decrease in ROA (in %). If one unit of interest rate is increased keeping the effect of other as constant, then there will be on average 0.1345 unit increase in ROA (in %), the coefficient of the interest rate is highly significant as p-value is 0.000 < 0.01. Other results can be interpreted in the similar way.

Project-III: Time Series Analysis

Time Series Forecasting using different Approaches with R

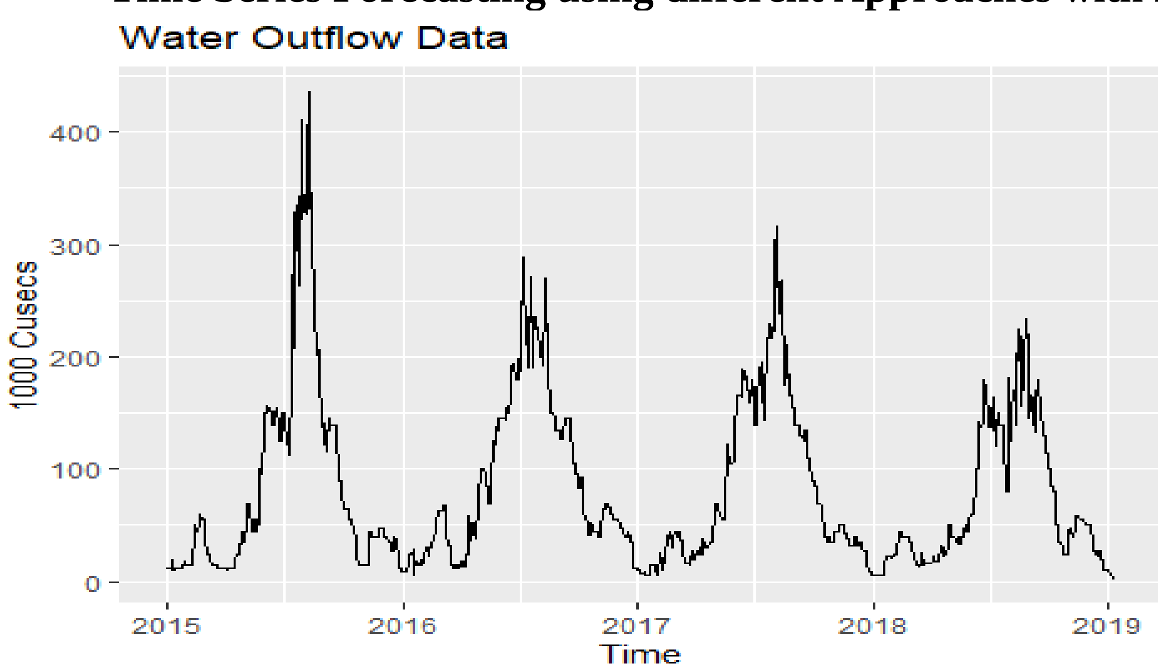


Table 4.1 Candidate SARIMA Models

Model	AIC	Model	AIC
ARIMA(0,1,0)(0,1,0)	8828.187	**ARIMA(1,1,4)(0,1,0)**	**8756.504**
ARIMA(0,1,1)(0,1,0)	8795.579	ARIMA(2,1,0)(0,1,0)	8790.907
ARIMA(0,1,2)(0,1,0)	8792.474	ARIMA(2,1,1)(0,1,0)	8770.703
ARIMA(0,1,3)(0,1,0)	8791.784	ARIMA(3,1,0)(0,1,0)	8787.185
ARIMA(0,1,4)(0,1,0)	8756.738	ARIMA(3,1,1)(0,1,0)	8772.593
ARIMA(0,1,5)(0,1,0)	8757.509	ARIMA(4,1,0)(0,1,0)	8774.229

ARIMA(1,1,0)(0,1,0)	8800.982	ARIMA(4,1,1)(0,1,0)	8765.531
ARIMA(1,1,1)(0,1,0)	8789.731	ARIMA(5,1,0)(0,1,0)	8773.098

Table 4.10 Mean Square Error of Artificial Neural Network

Method	**MSE**
ANN fit with (10,5) hidden nodes	3.4394

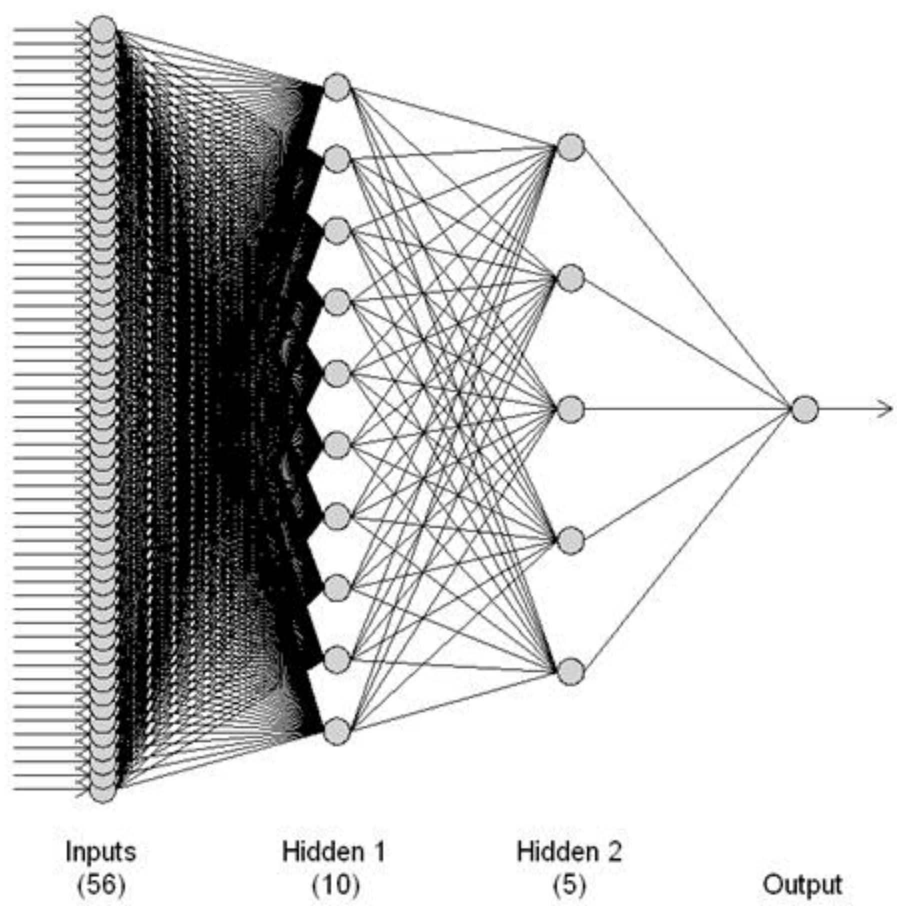


Figure 4.10:- Graphical presentation of Artificial Neural Network

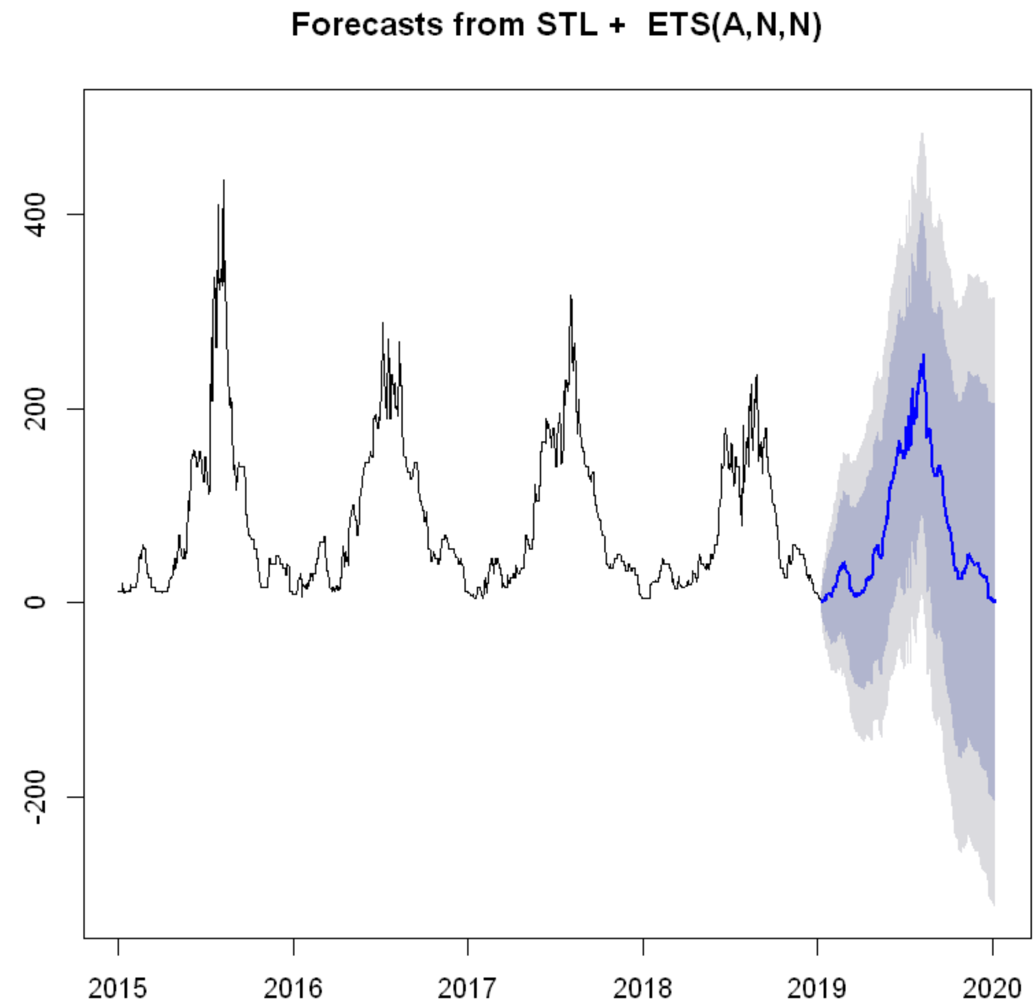
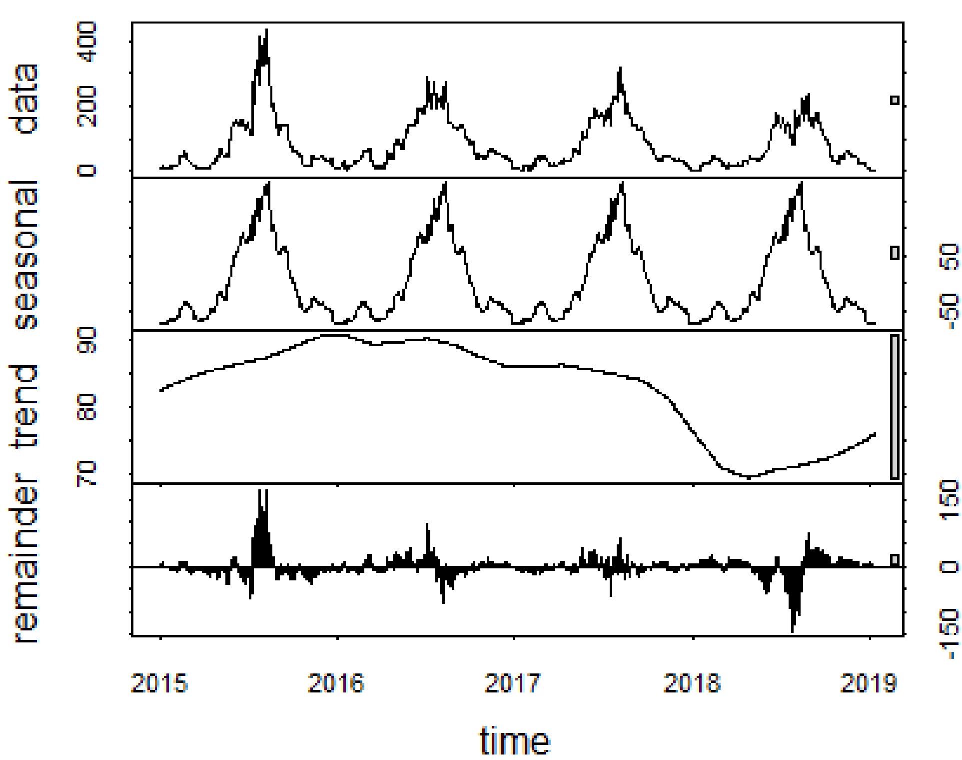
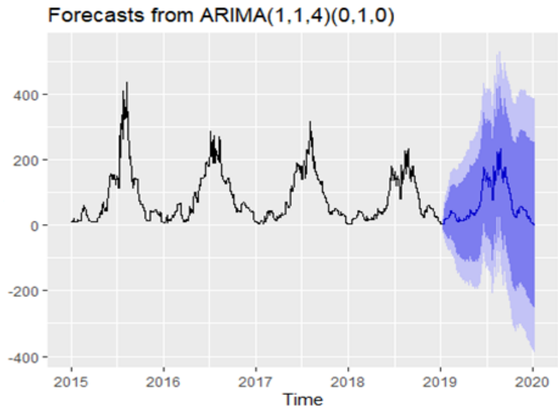


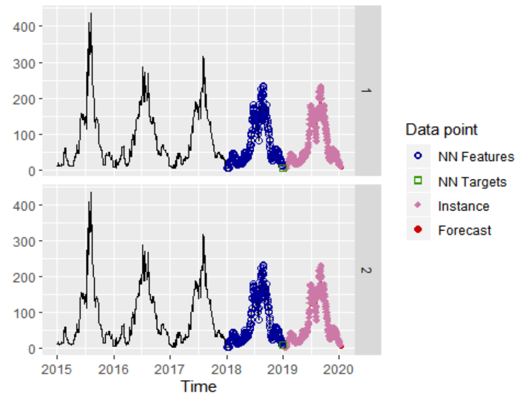
Figure 4.16:- Graph of Forecast using Artificial Neural Network

Forecasting using Non-parametric Technique 4.2

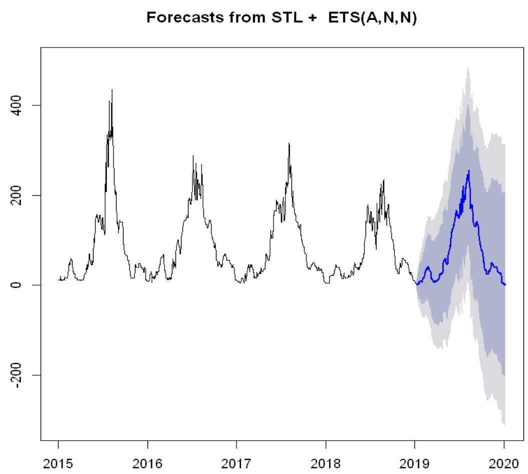


(a)	(b)





(c) (d)



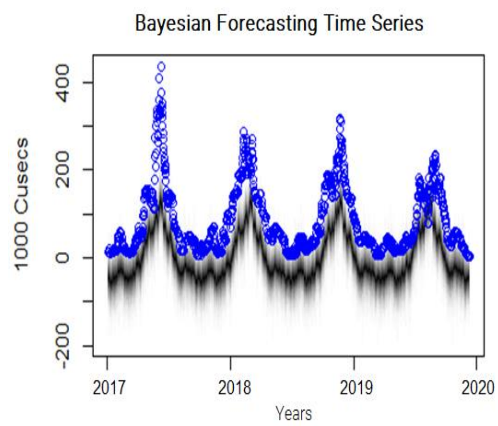
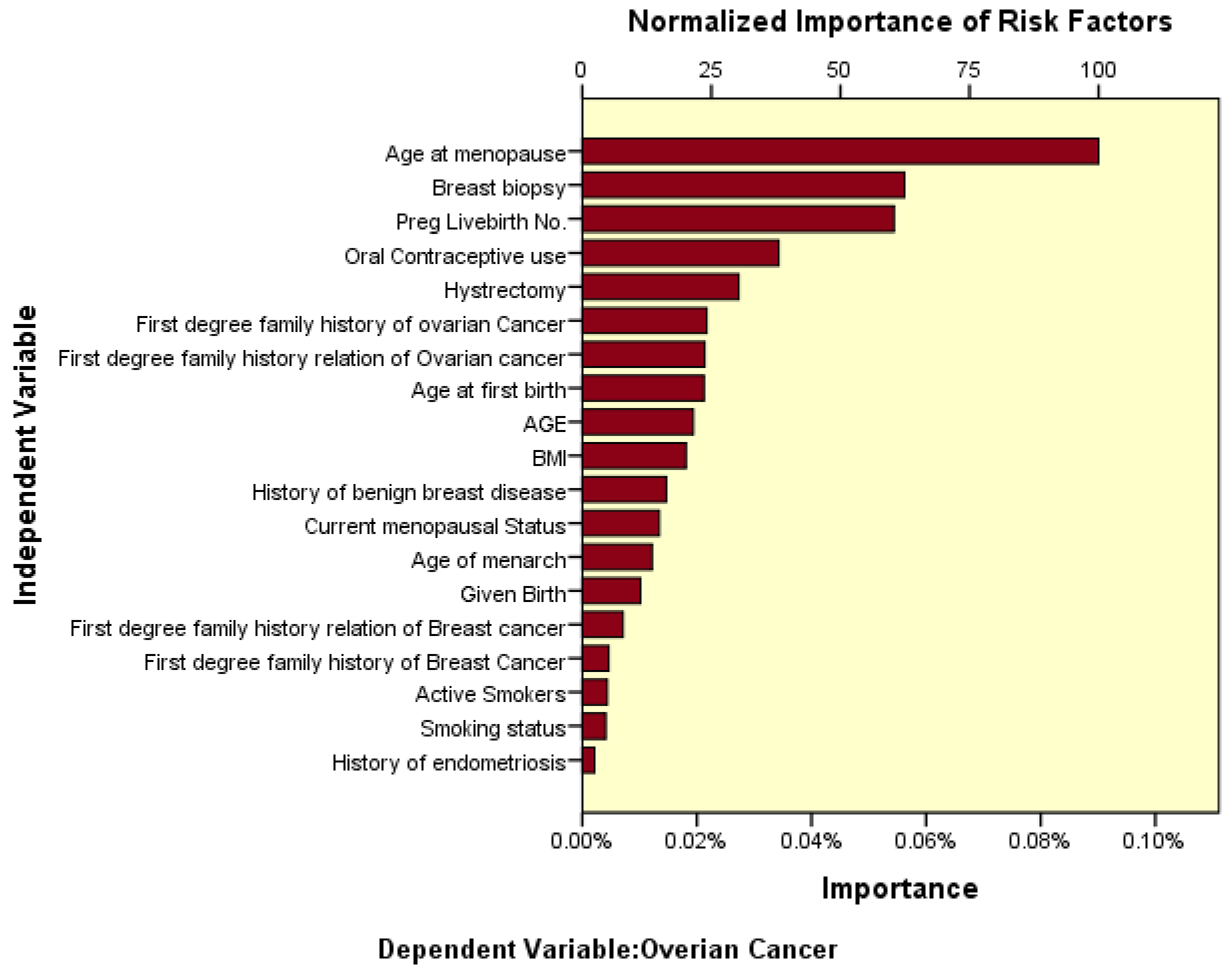
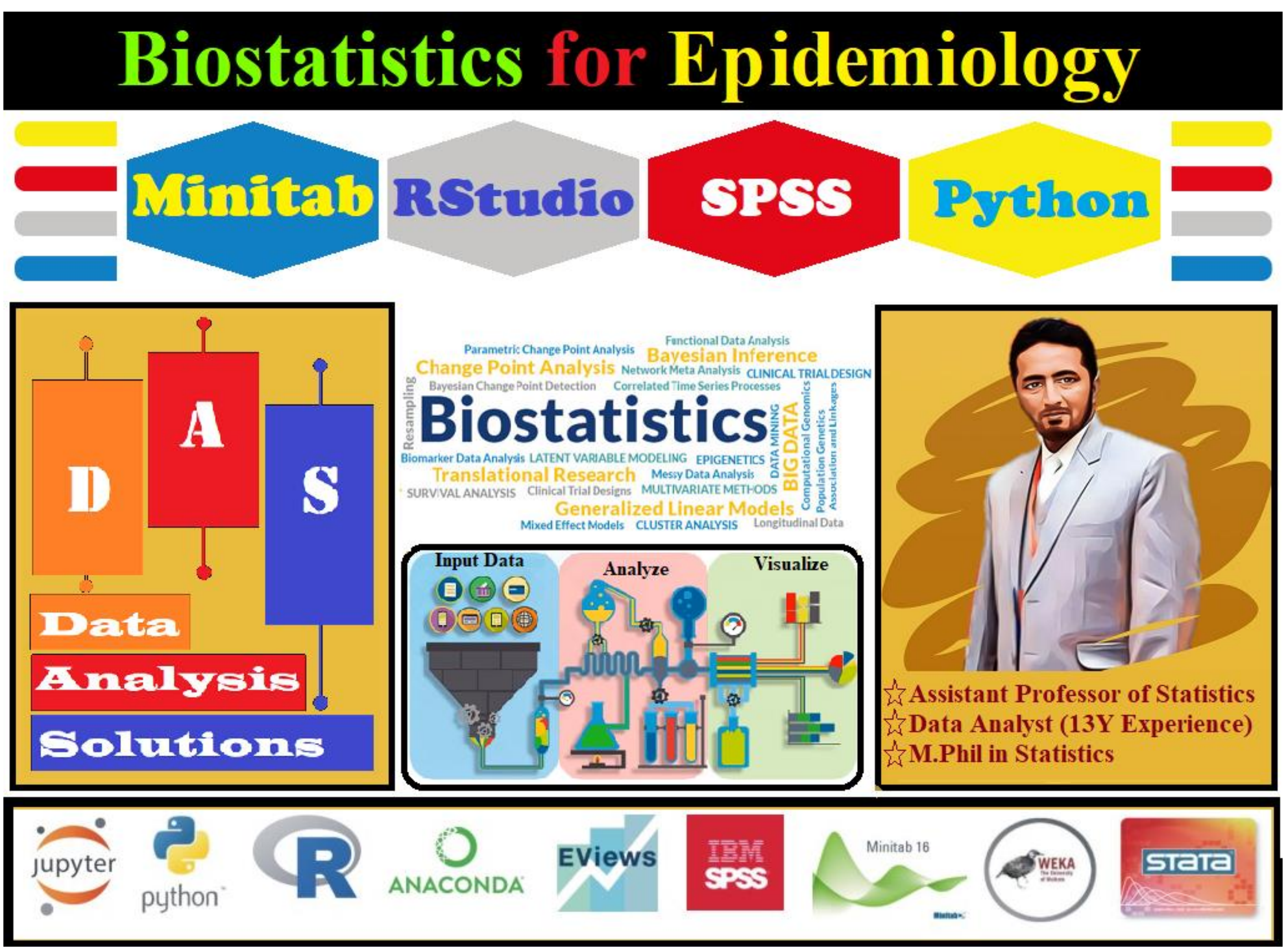


Table 5.1 Conclusions and Recommendations

Forecasting Methods	**RMSE**
SARIMA Model	10.8925
Bayesian Approach	8087.4049
Non-parametric Method KNN	180.3049
ANN with 5 Hidden nodes	11.0876
ANN fit with (10,5) hidden nodes	**3.4394**


Normalized Importance of Risk Factors
0
25
50
75
100
Independent Variable
Age at menopause
Breast biopsy
Preg Livebirth No.
Oral Contraceptive use
Hystrectomy
First degree family history of ovarian Cancer
First degree family history relation of Ovarian cancer
Age at first birth
AGE
BMI
History of benign breast disease
Current menopausal Status
Age of menarch
Given Birth
First degree family history relation of Breast cancer
First degree family history of Breast Cancer
Active Smokers
Smoking status
History of endometriosis
0.00%
0.02%
0.04%
0.06%
0.08%
0.10%
Importance
Dependent Variable:Overian Cancer


Biostatistics for Epidemiology
Minitab
RStudio
SPSS
Python
D
A
S
Data
Analysis
Solutions
Biostatistics
Bayesian Inference
Change Point Analysis
Translational Research
Generalized Linear Models
BIG DATA
Input Data
Analyze
Visualize
☆Assistant Professor of Statistics
☆Data Analyst (13Y Experience)
☆M.Phil in Statistics
jupyter
python
R
ANACONDA
EViews
IBM SPSS
Minitab 16
WEKA
STATA