

DESCRIPTIVE AND INFERENTIAL STATISTICS

BY

AMUA, UCHECHUKWU ROSELINE

SUPERVISORS: PROF. FEMIN YALCIN AND MR. NAIL. SENBAS

**DEPARTMENT OF COMPUTER SCIENCE
CONTEMPORARY TECHNOLOGY UNIVERSITY**

NOVEMBER, 2022

CONTENTS

- Introduction
- Data Set
- Descriptive Statistics
- Visualizations
- Interval Estimation
- Simple Linear Regression
- Interpretations
- References
- Acknowledgements

INTRODUCTION

- Descriptive Statistics tells us about, displays, and summarizes the basic parameters or characteristics of a data set conducted in a given study.
- Inferential Statistics helps us to come to conclusions and make predictions based on our observations or data..
- Inferential Statistics are useful for making estimations or testing hypothesis.

DATA SET

This data set was gotten from records of Covid-19 vaccinations in Obio Akpor Local Government Area of Rivers State, Nigeria between a period in August 2021 to October 2021 for first-dose ****moderna**** vaccinations only.

Date	HCWs	M	HCWs	FOther	FLWs	Mother	FLWs	Elderly	7	50years	Melderly	7	50years	M	18-49years	M	18-49years	FAEFI	Serious	AEFI	Unserious	accine	Dials	Rcd	accine	Dials	Used
23-Aug-21	1	5		2		6		7		10		71		46	0		0		12		12		12		12		12
24-Aug-21	5	3		6		1		58		26		117		88	0		0		25		25		25		25		25
25-Aug-21	6	6		1		1		52		39		106		118	0		0		26		26		26		26		26
26-Aug-21	9	2		34		28		68		39		159		105	0		0		34		34		34		34		34
27-Aug-21	11	10		36		20		52		46		146		141	0		0		36		36		36		36		36
28-Aug-21	1	1		4		6		27		17		121		94	0		0		22		22		22		22		22
29-Aug-21	0	0		0		0		14		6		24		25	0		0		6		6		6		6		6
30-Aug-21	8	8		25		35		76		53		201		164	0		0		42		42		42		42		42
31-Aug-21	4	6		1		5		51		31		172		121	0		0		30		30		30		30		30
1-Sep-21	2	8		8		8		72		36		158		80	0		0		28		28		28		28		28
2-Sep-21	5	3		11		5		25		15		64		53	0		0		13		13		13		13		13
3-Sep-21	6	7		2		0		39		32		105		92	0		0		21		21		21		21		21
4-Sep-21	3	8		3		0		28		34		107		109	0		0		21		21		21		21		21
5-Sep-21	0	0		0		0		14		11		22		16	0		0		5		5		5		5		5
6-Sep-21	9	19		7		7		73		52		158		160	0		0		36		36		36		36		36
7-Sep-21	9	3		22		26		66		45		215		165	0		0		43		43		43		43		43
8-Sep-21	8	9		3		1		65		42		124		116	0		0		31		31		31		31		31
9-Sep-21	5	20		1		0		32		6		48		40	0		0		12		12		12		12		12
10-Sep-21	0	0		0		0		13		11		12		22	0		0		5		5		5		5		5
11-Sep-21	2	0		23		28		36		33		139		90	0		0		30		30		30		30		30
12-Sep-21	0	0		0		0		0		1		3		6	0		0		1		1		1		1		1
13-Sep-21	0	0		0		0		1		1		4		7	0		0		1		1		1		1		1
14-Sep-21	0	8		51		34		58		30		133		125	0		0		33		33		33		33		33
15-Sep-21	0	0		0		0		0		2		5		2	0		0		1		1		1		1		1
16-Sep-21	5	4		40		31		85		70		252		261	0		0		58		58		58		58		58
17-Sep-21	2	8		39		28		89		77		215		248	0		0		53		53		53		53		53
18-Sep-21	2	1		26		20		26		17		109		106	0		0		27		27		27		27		27
19-Sep-21	0	0		0		0		12		13		34		37	0		0		11		11		11		11		11
20-Sep-21	2	3		27		38		87		72		171		145	0		0		36		36		36		36		36
21-Sep-21	0	0		27		19		30		32		111		116	0		0		26		26		26		26		26
22-Sep-21	0	0		7		10		4		5		4		1	0		0		3		3		3		3		3
23-Sep-21	0	0		22		15		16		19		28		27	0		0		10		10		10		10		10
24-Sep-21	1	1		38		24		37		45		48		31	0		0		18		18		18		18		18
25-Sep-21	0	0		8		13		11		8		30		22	0		0		9		9		9		9		9
26-Sep-21	0	0		0		0		7		6		0		0	0		0		1		1		1		1		1
27-Sep-21	0	0		43		38		44		53		62		47	0		0		20		20		20		20		20
28-Sep-21	0	36		29		31		30		75		83		83	0		0		22		22		22		22		22
29-Sep-21	0	0		33		27		87		67		54		69	0		0		27		27		27		27		27
30-Sep-21	2	0		52		26		48		49		131		108	0		0		34		34		34		34		34
1-Oct-21	2	0		14		21		13		17		18		14	0		0		8		8		8		8		8
2-Oct-21	0	0		0		0		95		55		15		16	0		0		14		14		14		14		14
3-Oct-21	2	0		0		0		12		8		40		41	0		0		8		8		8		8		8
4-Oct-21	0	0		86		71		44		56		127		131	0		0		43		43		43		43		43
5-Oct-21	1	1		36		19		122		108		93		108	0		0		38		38		38		38		38
6-Oct-21	2	0		20		21		98		85		88		98	0		0		33		33		33		33		33
7-Oct-21	0	0		20		21		110		93		72		80	0		0		31		31		31		31		31
8-Oct-21	0	1		15		13		75		72		71		70	0		0		25		25		25		25		25
9-Oct-21	2	5		8		5		24		19		33		32	0		0		10		10		10		10		10
10-Oct-21	0	0		0		0		1		2		4		6	0		0		1		1		1		1		1
11-Oct-21	0	0		11		18		34		38		74		76	0		0		21		21		21		21		21
12-Oct-21	2	2		17		17		32		31		71		85	0		0		20		20		20		20		20
13-Oct-21	0	1		17		15		41		32		25		43	0		0		18		18		18		18		18
14-Oct-21	0	10		0		16		16		15		38		46	0		0		10		10		10		10		10
15-Oct-21	2	3		11		23		24		18		75		66	0		0		19		19		19		19		19
16-Oct-21	0	1		11		23		21		16		64		55	0		0		18		18		18		18		18
17-Oct-21	0	0		0		0		29		7		36		40	0		0		16		16		16		16		16
18-Oct-21	1	3		0		0		19		12		35		44	0		0		9		9		9		9		9
19-Oct-21	0	0		0		0		24		16		0		0	0		0		4		4		4		4		4
20-Oct-21	4	3		19		11		27		30		90		99	0		0		24		24		24		24		24
21-Oct-21	0	0		2		7		28		27		23		24	0		0		9		9		9		9		9
22-Oct-21	1	24		2		20		36		25		107		84	0		0		27		27		27		27		27
23-Oct-21	0	0		1		2		8		9		12		0	0		0		3		3		3		3		3
24-Oct-21	0	0		1		1		12		11		40		29	0		0		8		8		8		8		8
25-Oct-21	0	1		16		4		29		22		52		61	0		0		18		18		18		18		18
26-Oct-21	1	21		2		21		39		30		111		110	0		0		27		27		27		27		27
27-Oct-21	0	0		6		2		5		2		7		5	0		0		3		3		3		3		3
28-Oct-21	2	9		49		27		48		49		162		140	0		0		36		36		36		36		36
29-Oct-21	0	0		0		1		13		15		19		17	0		0		5		5		5		5		5
30-Oct-21	0	0		12		10		7		2		33		28	0		0		6		6		6		6		6
31-Oct-21	0	0		0		0		5		3		9		15	0		0		3		3		3		3		3

DESCRIPTIVE STATISTICS

Summary Statistics

	HCWs M	HCWs F	Other FLWs M	Other FLWs F	Elderly 50+ M	Elderly 50+ F	Youths 18-49 M	Youths 18-49 F
count	70	70	70	70	70	70	70	70
mean	1.86	2.57	15.21	13.24	37.58	30.18	76.88	70.41
std	2.75	4.12	17.15	13.56	29.45	24.16	61.32	56.18
min	0	0	0	0	0	1	0	0
25%	0	0	1	1	13.25	11.25	25.75	25.5
50%	0.5	0.5	10.5	10	29.5	26.5	67.5	63
75%	2.	3	23.75	21	52	44.25	115.5	107.5
max	11	20	86	71	122	108	252	261

DESCRIPTIVE STATISTICS

Covariance Matrix

	HCWs M	HCWs F	Other FLWs M	Other FLWs F	Elderly 50+ M	Elderly 50+ F	Youths 18-49 M	Youths 18-49 F
HCWs M	7.57	7.42	4.62	3.02	30.78	17.36	107.89	89.38
HCWs F	7.42	16.9	1.93	-1.71	36.16	17.05	126.70	112.42
Other FLWs M	4.62	1.93	294.28	211.05	230.81	242.95	563.94	553.20
Other FLWs F	3.02	-1.706	211.05	183.95	178.07	183.74	455.57	424.01
Elderly 50+ M	30.78	36.16	230.81	178.07	867.32	663.42	1165.17	1116.68
Elderly 50+ F	17.36	17.05	242.95	183.74	663.42	583.66	886.67	904.46
Youths 18-49 M	107.89	126.70	563.94	455.57	1165.17	886.67	3759.93	3285.45
Youths 18-49 F	89.38	112.42	553.20	424.01	1116.68	904.46	3285.45	3157.06

DESCRIPTIVE STATISTICS

Correlation Matrix

	HCWs M	HCWs F	Other FLWs M	Other FLWs F	Elderly 50+ M	Elderly 50+ F	Youths 18-49 M	Youths 18-49 F
HCWs M	1	0.65	0.09	0.08	0.38	0.26	0.64	0.58
HCWs F	0.65	1	0.03	-0.03	0.30	0.17	0.50	0.48
Other FLWs M	0.09	0.03	1	0.91	0.45	0.58	0.53	0.57
Other FLWs F	0.08	-0.03	0.91	1	0.44	0.56	0.55	0.56
Elderly 50+ M	0.38	0.30	0.45	0.44	1	0.93	0.64	0.67
Elderly 50+ F	0.26	0.17	0.58	0.56	0.93	1	0.59	0.66
Youths 18-49 M	0.64	0.50	0.53	0.55	0.64	0.59	1	0.95
Youths 18-49 F	0.58	0.48	0.57	0.56	0.67	0.66	0.95	1

VISUALIZATIONS

STACKED HISTOGRAMS

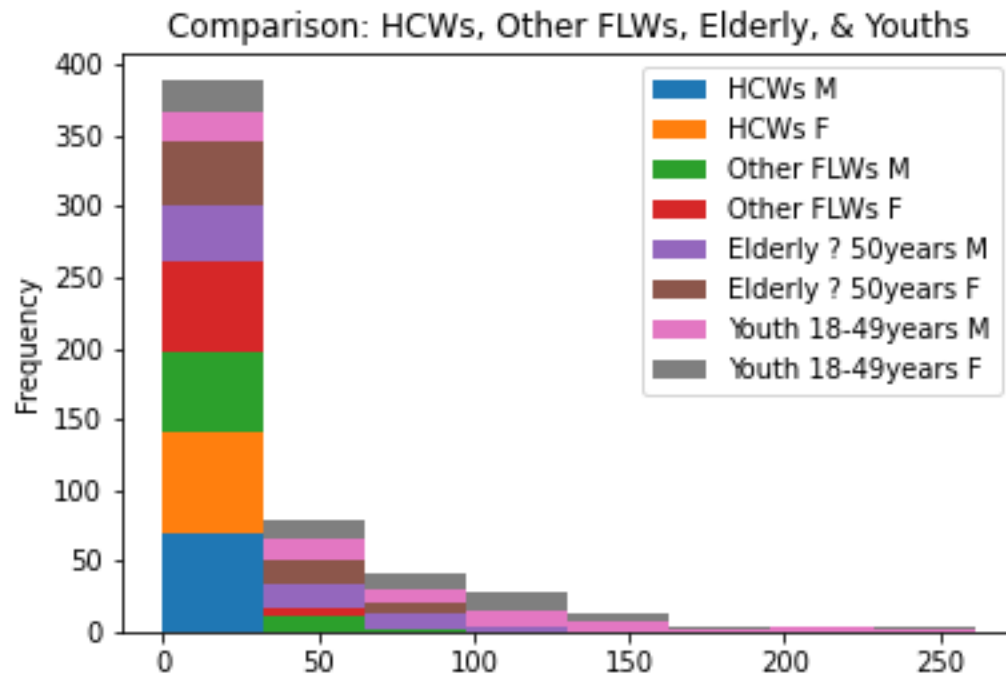
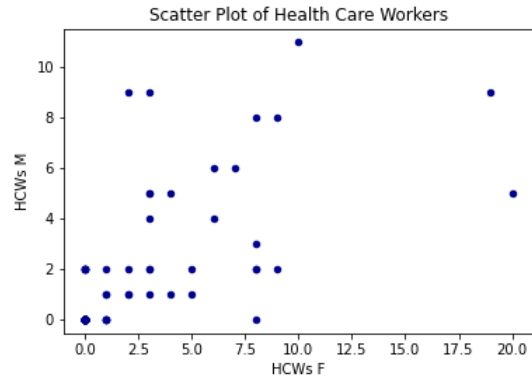


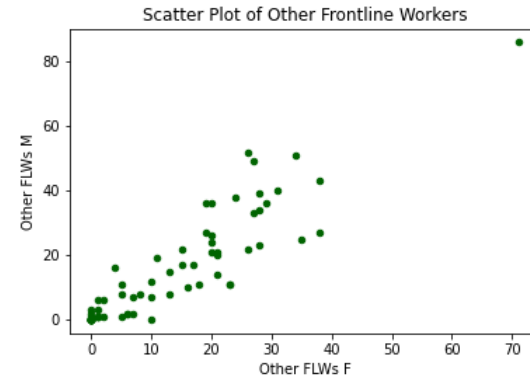
Figure One: Frequency Distribution of Features

VISUALIZATIONS

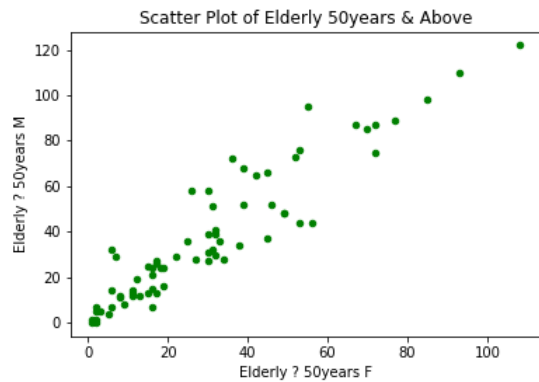
SCATTER PLOTS



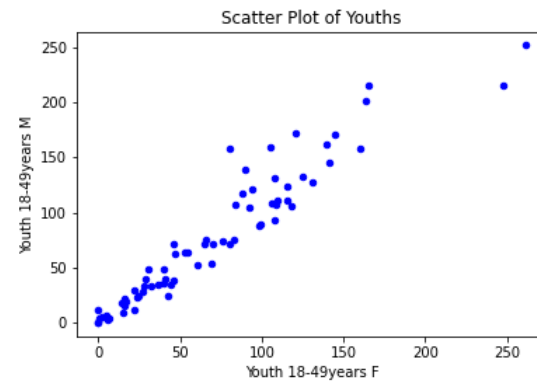
Scatterplot of Healthcare Workers



Scatterplot of Other Frontline Workers



Scatterplot of Elderly 50years +



Scatterplot of Youths 18-49years

VISUALIZATIONS

BOX COMPARISON OF FEATURES

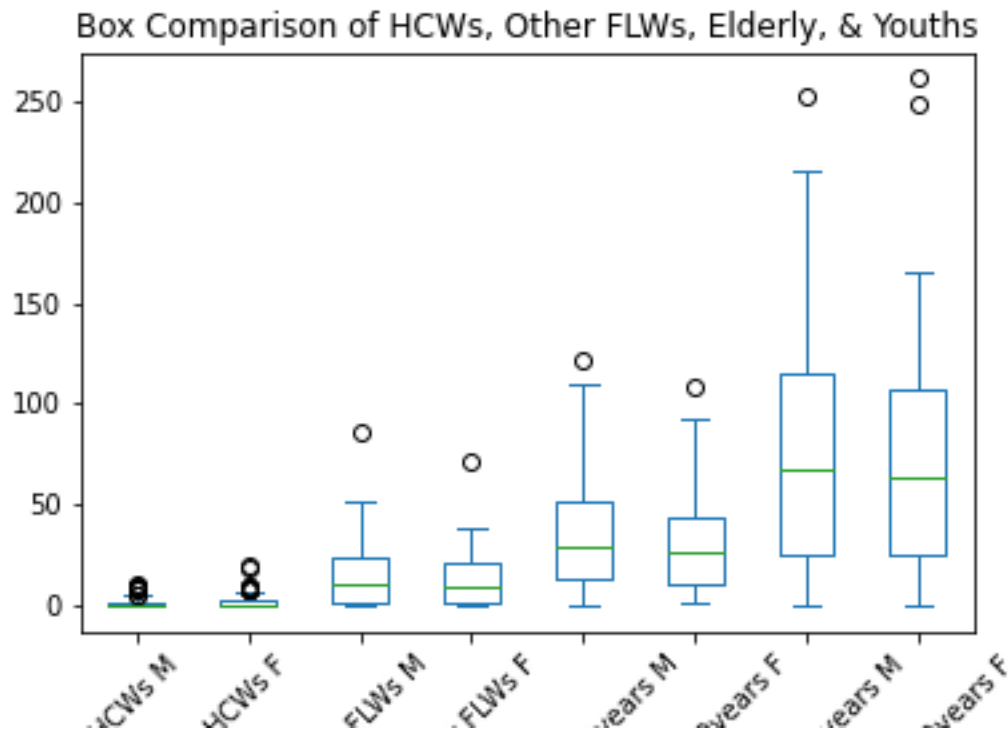


Figure 3: Box plots

VISUALIZATIONS

Heat Map Display of Covariance Matrix of Features

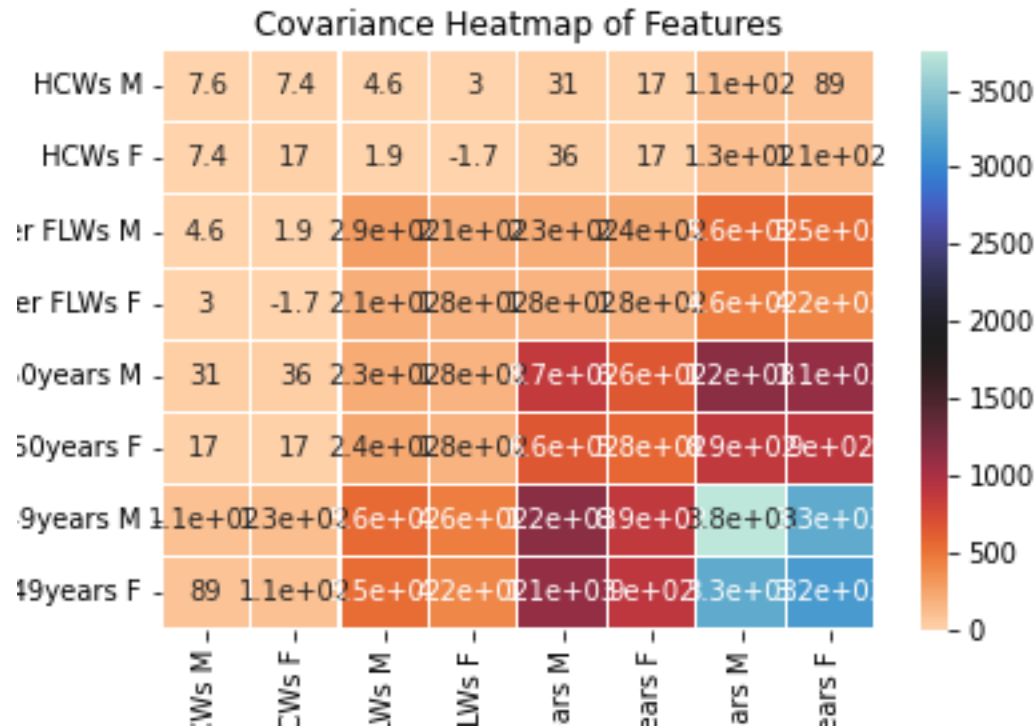


Figure 4: Covariance Heatmap

VISUALIZATIONS

Heat Map Display of Correlation Matrix of Features

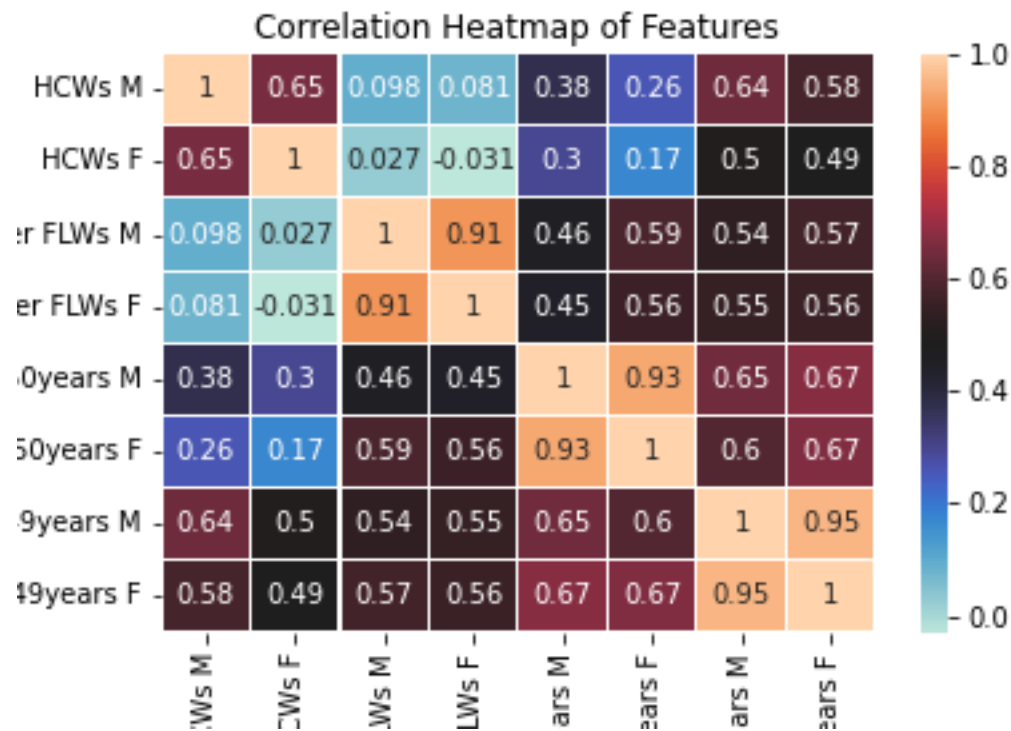


Figure 4: Correlation Heatmap

INTERVAL ESTIMATION

Feature One: Healthcare Workers

A 95% Confidence Interval for Mu When Sigma is Known

Sample size = $n = 40$

\bar{X} = sample mean

σ = population standard deviation

$R_f = z_{\text{critical}}$

$Se = \sigma / \text{square root of } n$

$Me = R_f * Se$

$CI = (\bar{X} - Me, \bar{X} + Me)$

A 95% CI for μ : (0.738, 2.912)

Feature Two: Youths

A 99% Confidence Interval for Mu When Sigma is unknown

Sample size = $n_1 = 30$

Sample_mean = sample mean

Sample_std = sample standard deviation($ddof=1$)

$R_f = t_{\text{critical}}$

$Se = \text{sample_std} / \text{square root of } n_1$

$Me = R_f * Se$

$LCL = \text{Sample_mean} - Me$,

$UCL = \text{Sample_mean} + Me$

$CI = (LCL, UCL)$

A 99% CI for μ : (56.52 , 115.67)

SIMPLE LINEAR REGRESSION

A Comparison Between the Output Variable & the Predicted Output

Date	Y	Y-Predicted
23-Aug-2021	12	14.15
24-Aug-2021	25	23.82
25-Aug-2021	26	30.71
26-Aug-2021	34	27.77
27-Aug-2021	36	36.00
28-Oct-2021	36	35.77
29-Oct-2021	5	7.4
30-Oct-2021	6	10

INTERPRETATIONS

- ✓The descriptive statistics section gave a summary statistics of the data set.
- ✓The distribution of data in the stacked histogram showing comparison among features is right-skewed. Also, more females than their male counterparts in all features were vaccinated.
- ✓Some outliers can be seen at the tail end for some features.
- ✓The scattered plot for health workers showed a weak correlation among variables.
- ✓The scattered plot for other frontline workers showed a not strong relationship between variables.
- ✓The scattered plot for elderly 50+ showed a not so strong relationship among variables.
- ✓The scattered plot for youths showed a very strong relationship between variables.
- ✓The comparison box plots among features showed that all of the features have a right-skewed distribution. We can also see the outliers, some of which are very close to the range of values while others are farther from the range. The length of the box represents the *interquartile range IQR*. The longer the length of box the longer the distance between the third quartile and the first quartile. Hence, the larger the *IQR*. The edge of the lower whisker represents the minimum value, the lower edge of the box represents the first quartile, the middle line represents the second quartile or median, the upper edge of the box represents the third quartile, while the edge of the upper whisker represents the maximum value.
- ✓The covariance heat map makes it easy for us to visualize and understand the relationship between variables. The change in colour indicates an increase in covariance.

INTERPRETATIONS

- ✓The larger the covariance, the closer the relationship as can be seen in the relationship between "Youth 18-49years M" and "Youth 18-49years F".
- ✓The correlation heat map showed the relationship between features.
- ✓From the color scale it is easy to see this relationship. The correlation coefficient usually between -1 and 1 tells us the degree to which two features move in sequence.
- ✓From the correlation heat map, the features ****Youth 18-49years M**** and ****Youth 18-49years F**** had the strongest relationship between them with a correlation coefficient of 0.95
- ✓Whereas, the features ****HCWs F****(female healthcare workers) and ****Other FLWs F****(Other female frontline workers) had the weakest relationship between them with a correlation coefficient of -0.03
- ✓The first feature estimated the mean of the subset **Healthcare Workers** when sigma is *known*. We obtained a 95% confidence interval that our mu of feature one **healthcare workers** when sigma is known falls in the range(0.863, 3.037).
- ✓The second feature estimated the mean of the subset *Youths* when sigma is *unknown*.
- ✓We obtained a 99% confidence interval that our mu of feature two **youths** when sigma is unknown falls in the range(55.218, 123.582).
- ✓From our simple linear regression model, we obtained an Rsquared of 0.906. This is about 90.6% proportion of the variance in our dependent variable that is predictable from our independent variable.

REFERENCES

Adi Bronshtein (2017). Article on Simple and Multiple Linear Regression.

Retrieved from <http://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-C92842516819>

Contemporary Technology University (2022). *Probability & Statistics Essentials*. Contech.

Walpole, R.E, Myers, R.H, Myers, S.L, and Ye, K. (2007). Probability & Statistics for Engineers & Scientists, 9th edition, Pearson Education, inc.

ACKNOWLEDGEMENTS

- ✓ My appreciation goes to my instructor Professor Femin Yalcin for doing a great job of delivering the lectures in a friendly and easy to understand way through out the entire duration of the course.
- ✓ My appreciation also goes to my Teaching Assistant, Mr. Nail Senbas, for always being supportive and ready to provide assistance when needed.
- ✓ Lastly to my classmates, family, and friends, who one way or the other supported me through the length of the course.
- ✓ I thank you all.