5/28/2021

Assignment 3

STAT3175

Rajiv Mehta
45433062

1 PART A

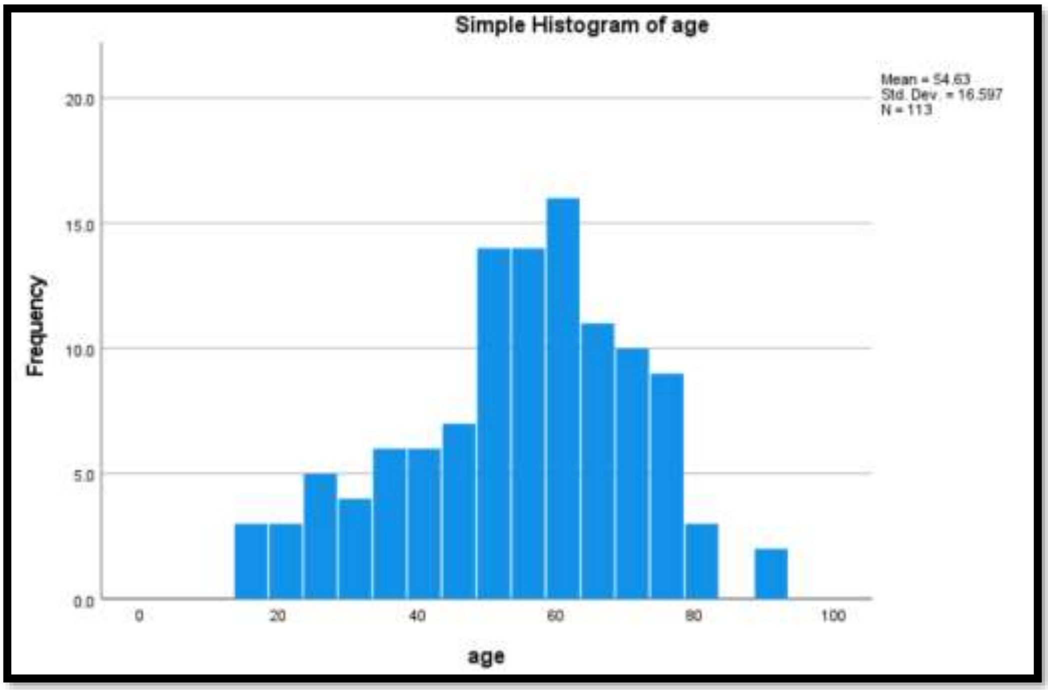


Figure 1. Simple Histogram of Age

The simple histogram of age as seen in Figure 1, indicates that the data is normally distributed and therefore, requires no transformation.

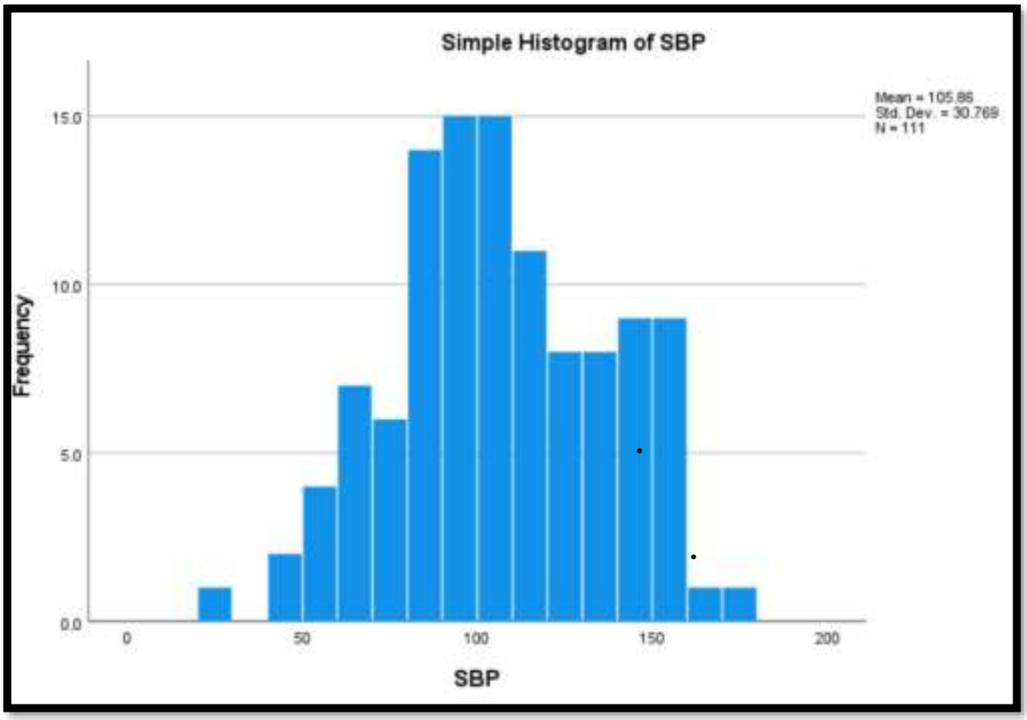


Figure 2. Simple Histogram of SBP

The simple histogram of SBP as seen in Figure 2, indicates that SBP is approximately normal and does not require any transformations.

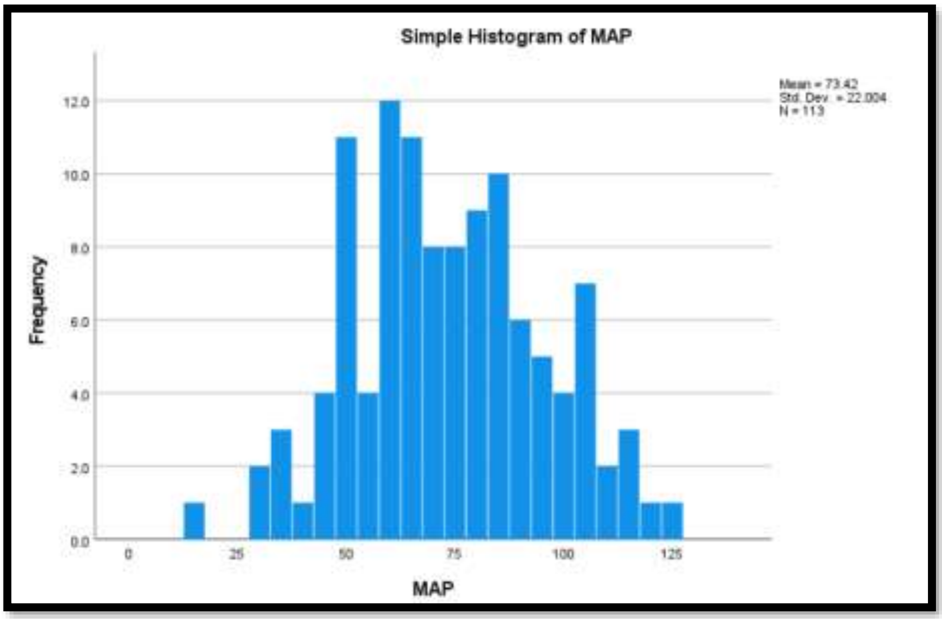


Figure 3. Simple Histogram of MAP

The simple histogram of MAP as seen in Figure 3, indicates that the data follows a normal distribution and therefore, will require no transformations.

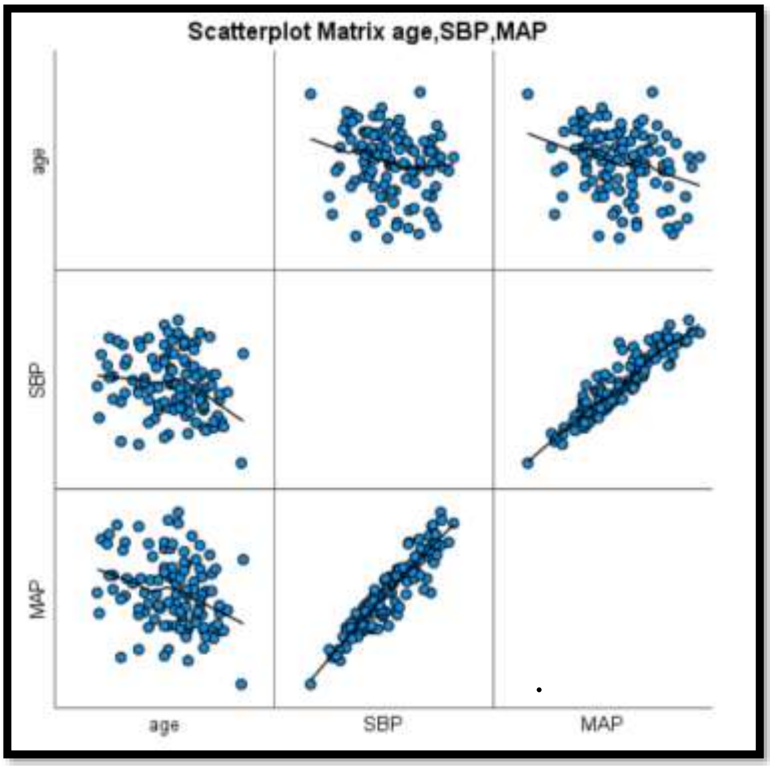


Figure 4. Scatter Plot Matrix

The scatterplot matrix shown in figure 4, indicates the following:

- Age and SBP has a weak negative linear relationship.
- Age and MAP has a weak negative linear relationship.
- SBP and MAP has a very strong positive linear relationship.

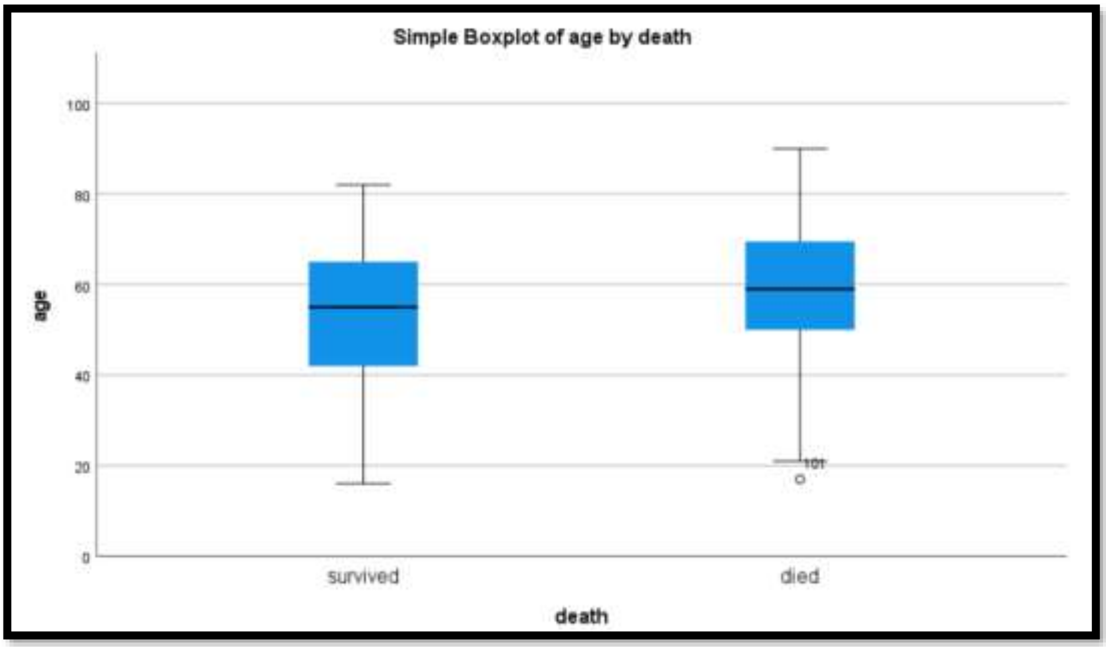


Figure 5. Boxplot of Age by Death

The boxplot of Age with Death as seen in Figure 5, indicates that the higher the age of the patient the more likely they are to have died. The ranges of both plots are very close, indicating that age might not have a significant effect on death. Note, that there is an outlier but it shall not be removed from the dataset as it is not extreme.

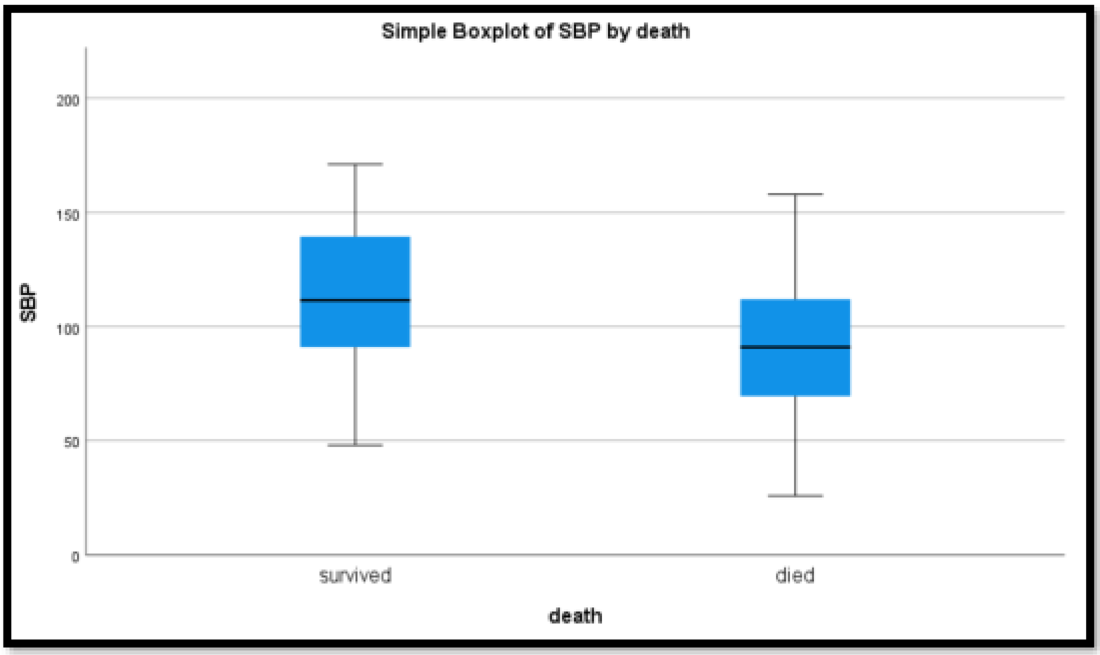


Figure 6. Boxplot of SBP by death

The boxplot of SBP by age as seen in Figure 6, indicates that the higher the SBP level the more likely you are to survive. The survived data range is approximately the same as the Died data range.

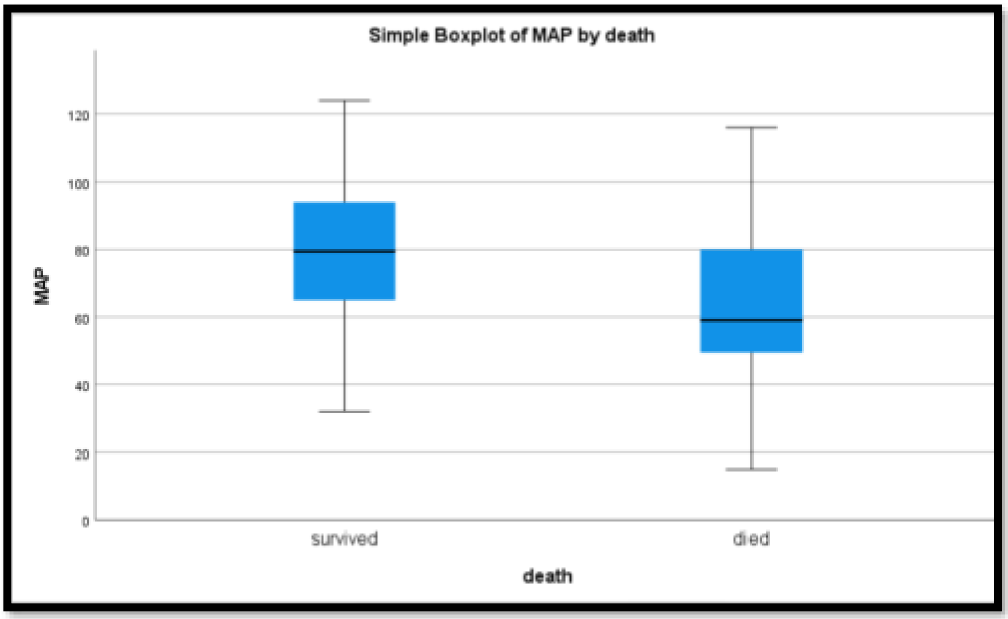


Figure 7. Boxplot of MAP by death

The boxplot of MAP and death as seen in Figure 7, indicates that the higher the MAP the chance of survival is higher. The range of Death is significantly larger than that of Survived.

Between-Subjects Factors

		Value Label	N
shock	2	non-shock	34
	3	hypovolemic	17
	4	cardiogenic	20
	5	bacterial	16
	6	neurogenic	16
	7	other	10
sex	1	male	59
	2	female	54

Figure 8. Frequency table

The frequency table as seen in Figure 8, indicates that the non-shock has the largest 'N' value and therefore, should become the reference category for our regression. As the differences between number of males and females are similar the reference category can stay as female.

Between-Subjects Factors

		N
Shock2	3.00	17
	4.00	20
	5.00	16
	6.00	16
	7.00	10
	10.00	34

Figure 9. Frequency table for shock after recode

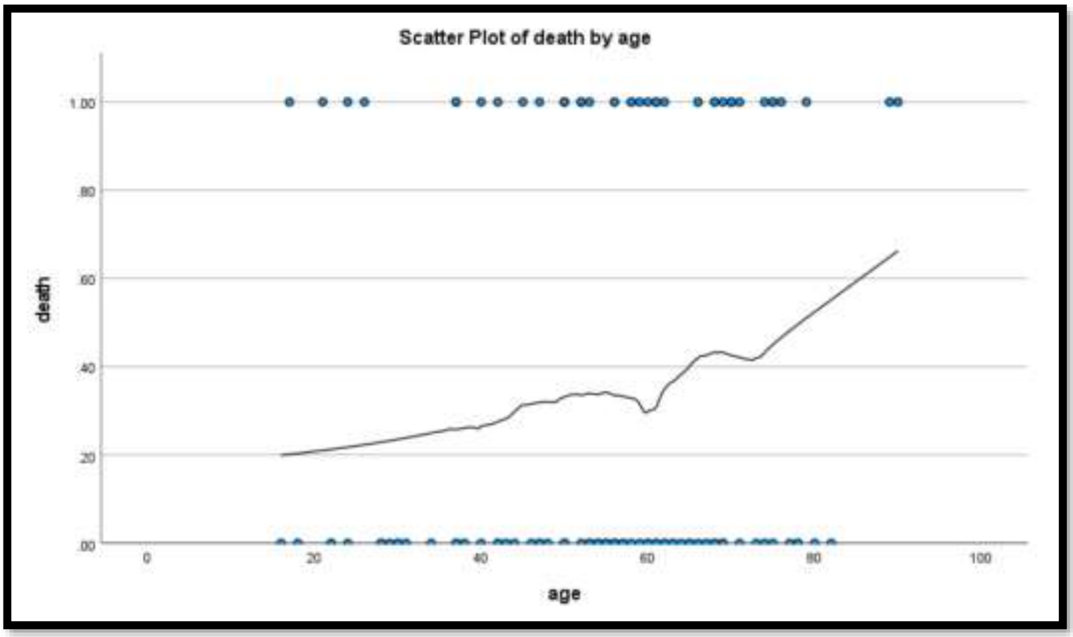


Figure 10. Scatterplot of Age vs Death

Figure 9, indicates that as age increases the chances of death also increases. Though, it can be seen that the chance of dying are still relatively low until the patient is >= 70.

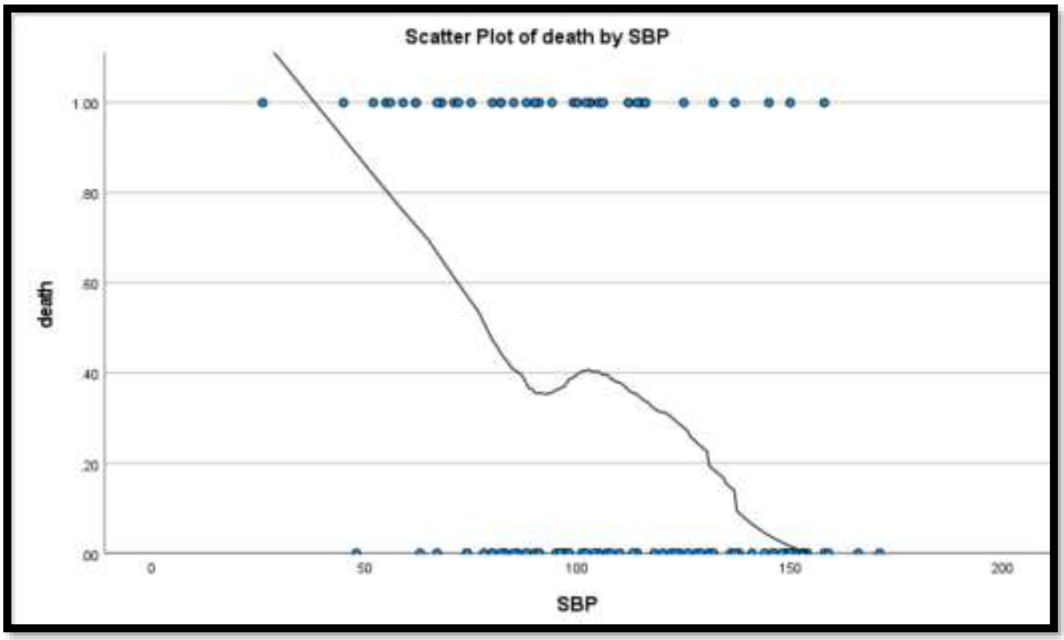


Figure 11. Scatterplot of SBP vs Death

Figure 10, indicates that as the SBP level increases the chances of Survival increases. It can be seen that when SBP is >=100 the chances of death are low and when it reaches 150 the chances of death are negligible.

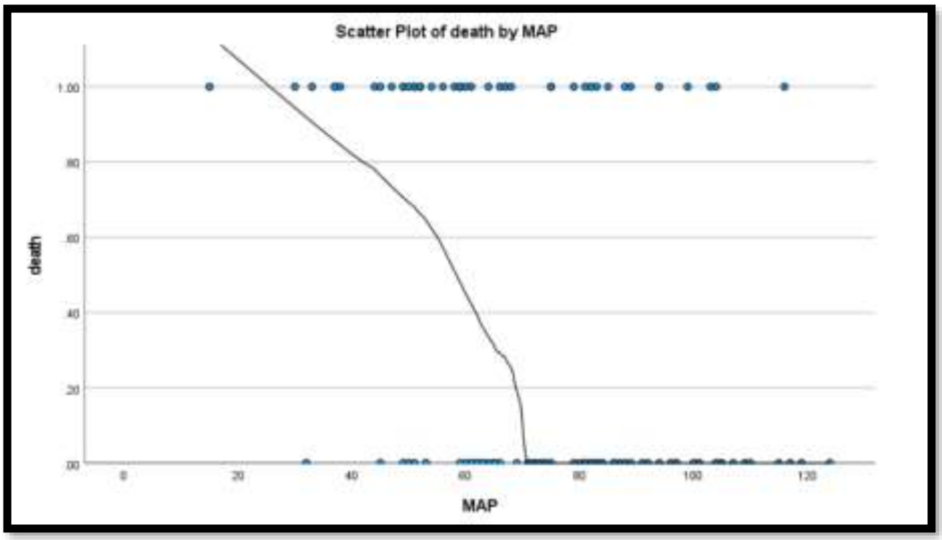


Figure 12. Scatterplot of MAP

Figure 11, indicates that as MAP increases the rate of survival also increases. It can be seen that when MAP is 60 the chances of dying are low and then when the value >=70. The chances of death are negligible.

death * sex Crosstabulation

			sex		
			male	female	Total
death	survived	Count	42	28	70
		% within sex	71.2%	51.9%	61.9%
	died	Count	17	26	43
		% within sex	28.8%	48.1%	38.1%
Total		Count	59	54	113
		% within sex	100.0%	100.0%	100.0%

Figure 13. Crosstabulation for death*sex

The crosstabulation of the death and sex as seen in Figure 13, indicates that the death rate for patients was much higher if they were female almost having a 50-50 split with death and survived.

death * shock Crosstabulation

			shock						
			non-shock	hypovolemic	cardiogenic	bacterial	neurogenic	other	Total
death	survived	Count	31	7	10	10	9	3	70
		% within shock	91.2%	41.2%	50.0%	62.5%	56.3%	30.0%	61.9%
	died	Count	3	10	10	6	7	7	43
		% within shock	8.8%	58.8%	50.0%	37.5%	43.8%	70.0%	38.1%
Total		Count	34	17	20	16	16	10	113
		% within shock	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 14. Crosstabulation for death*shock

The crosstabulation of death and shock as seen in Figure 14, indicates that the different types of shock will influence the chances of death differently to each other. If the patient falls under non-shock then the chances of survival are very high at 91.2%. If the patient suffered from hypovolemic shock, then the chances of survival were significantly lower at 41.2%. If the patient suffered cardiogenic shock then there was a 50% chance of survival which indicates that it might not influence the probability of death for the patient. If the patient suffered a bacterial shock, then the chance of survival was also higher at 62.5%. If the patient suffered neurogenic shock then there was almost an even chance of death and survival, with the patient having a higher chance at survival 56.3%. Lastly, if the patient suffered and other kind of shock, then the patient had a low chance of survival, with only 30% surviving.

2 PART B

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	MAP	-.040	.011	13.341	1	<.001	.961
	Constant	2.354	.786	8.975	1	.003	0.000260

a. Variable(s) entered on step 1: MAP.

Variables in the Equation

Variable		B	S.E.	Wald	df	Sig.	Exp(B)
		Statistics					
Step 1[a]	SBP	-.026	.008	11.984	1	0.000537	.974
	Constant	2.238	.789	8.042	1	.005	9.374

a. Variable(s) entered on step 1: SBP.

As MAP is smaller, we will use assume MAP gives the smaller AIC and BIC values.

Model	**N**	**P**	**-2L(p)**	**AIC = -2L(p) + 2p**	**BIC = -2L(p) + p*log(n)**
Shock2	113	6	126.372	138.372	154.736
Sex	113	2	145.642	149.642	155.097
Age	113	2	148.321	152.642	157.776
SBP	113	2	134.018	138.018	143.473
MAP	113	2	133.784	137.018	143.238
MAP + Shock	113	7	118.593	132.59	151.685
MAP + Sex	113	3	130.455	136.455	144.637
MAP + Age	113	3	133.670	139.67	147.852
MAP + Shock + Sex	113	8	116.592	132.59	154.411
MAP + Shock + Age	113	8	118.478	134.478	156.297
MAP + Sex + Age	113	4	130.009	138.009	149.479
MAP + Shock + Sex + Age	113	9	116.293	134.293	158.839

Table 1. Model Selection results

For this regression analysis we will go with best model under AIC, highlighted in yellow as seen in Table 1.

Model Information

Dependent Variable	death[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models survived as the response, treating died as the reference category.

Case Processing Summary

	N	Percent
Included	113	100.0%
Excluded	0	0.0%
Total	113	100.0%

Categorical Variable Information

			N	Percent
Dependent Variable	death	survived	70	61.9%
		died	43	38.1%
		Total	113	100.0%
Factor	Shock2	3.00	17	15.0%
		4.00	20	17.7%
		5.00	16	14.2%
		6.00	16	14.2%
		7.00	10	8.8%
		10.00	34	30.1%
		Total	113	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Covariate	MAP	113	15	124	73.42	22.004

Goodness of Fit[a]

	Value	df	Value/df
Deviance	102.637	90	1.140
Scaled Deviance	102.637	90	
Pearson Chi-Square	90.092	90	1.001
Scaled Pearson Chi-Square	90.092	90	
Log Likelihood[b]	-55.020		
Akaike's Information Criterion (AIC)	124.040		
Finite Sample Corrected AIC (AICC)	125.107		
Bayesian Information Criterion (BIC)	143.132		
Consistent AIC (CAIC)	150.132		

Dependent Variable: death
Model: (Intercept), MAP, Shock2

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
31.544	6	<.001

Dependent Variable: death
Model: (Intercept), MAP, Shock2

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	4.703	1	.030
MAP	6.905	1	.009
Shock2	12.087	5	.034

Dependent Variable: death
Model: (Intercept), MAP, Shock2

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.167	1.0964	-2.316	1.982	.023	1	.879
MAP	.031	.0117	.008	.053	6.905	1	.009
[Shock2=3.00]	-2.168	.8190	-3.773	-.562	7.005	1	.008
[Shock2=4.00]	-2.043	.7753	-3.563	-.524	6.948	1	.008
[Shock2=5.00]	-1.293	.8235	-2.907	.321	2.467	1	.116
[Shock2=6.00]	-1.753	.8125	-3.346	-.161	4.657	1	.031
[Shock2=7.00]	-2.847	.9496	-4.708	-.986	8.992	1	.003
[Shock2=10.00]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: death
Model: (Intercept), MAP, Shock2

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

$$Y_i \sim Bernoulli\ (\pi_i) independently, i = 1, 2\ and\ j = 1,2, \ldots ,5$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = -0.167 + 0.031x_1 - 2.168x_{2,1} - 2.043x_{2,2} - 1.293x_{2,3} - 1.753x_{2,4} - 2.847x_{2,5}$$

Where:

- $x_1 = MAP$
- $x_{2,j}\ for\ j = 1,2, \ldots ,5 = \{0 = No\ shock\}\ \{when\ 1 = Shock\}$ where j:
  - 1=hypovolemic shock,
  - 2=cardiogenic shock,
  - 3=bacterial shock,
  - 4=neurogenic shock,
  - 5=other types of shock

3 PART C

Parameter	Beta	Exp(Beta)
MAP	0.031	1.031
Hypovolemic Shock	-2.168	0.1144
Cardiogenic Shock	-2.043	0.1296
Bacterial Shock	-1.293	0.2744
Neurogenic Shock	-1.753	0.1733
Other Types of Shock	-2.847	0.05801

Table 2. Odds of parameter compared to No Shock

In the above model, the reference category for mortality is default to death. Therefore, the final model’s predictors as seen in Part B, indicate the following:

- Mean Arterial Pressure (MAP): The coefficient of MAP in the final model is a positive value. This indicates that for every 1mmHg increase the chance of survival for the patient also increases by 3.1%. This matches the results obtained in the Scatterplot seen in Figure 12, in Part A as both the model and the graph showcases that an increase in MAP will increase the chances of survival.
- Shock: The coefficient for each category of shock in the final model are negative values, indicating a reduction of the patients odds of survival by their respective percentages if the patient suffered that type of shock. The results obtained for the best model heavily differ from the crosstabulation of death and shock as seen in Figure 14, in Part A.
    - Hypovolemic shock – In the final model, it can be seen that if the patient suffered from Hypovolemic shock, then the chances of death would increase by 0.1144 times the odds of death compared to non-shock. This differs from the indication given from the crosstabulation table, as Hypovolemic shock has a 58.8% chance of death. This indicates that though, both results prove that the Hypovolemic shock cause a higher chance of death in the patient, the odds highly differ.
    - Cardiogenic shock – Like the Hypovolemic shock as discussed previously, cardiogenic shock according to the final model increases the odds of death by 0.1296 times compared to non-shock. This differs highly with the crosstabulation table seen in Figure 14, as in that table it indicates that cardiogenic shock has a 50% chance of death, meaning that it should have no effect on the model. Therefore, the model and crosstabulation table indicate very different effects on Mortality.
    - Bacterial shock – In the final model, if the patient suffered from a bacterial shock then the odds of death increase by 0.2744 compared to non-shock. This matches with the results in Part A as the chance odds of survival is indicated as 62.5%.
    - Neurogenic shock – In the final model, if the patient suffered from neurogenic shock then their odds of death increased by 0.1733 compared to non-shock. This matches somewhat with the results in Part A, as the chance of death is 43.8% so the possibility of death exists.
    - Other types of shock – in the final model, if the patient suffered from other types of shock, then their odds of death increases by 0.05801 compared to non-shock. This matches perfectly with the results in part A, as other types of shock have a high chance of death, 70% according to Figure 14.

4 PART D

4.1 FITTED ODDS

$$\frac{\pi_i}{1-\pi_i} = e^{-0.617\,+\,0.031*110-2.168*0-\,2.043*1-1.293*0-1.753*0-2.847*0}$$

$$\frac{\pi_i}{1-\pi_i} = 2.1170$$

4.2 PREDICTED PROBABILITY

$$\frac{\pi_i}{1-\pi_i} = 2.1170$$

$$\pi_i = 2.1170 - 2.1170\pi_i$$

$$\pi_i\,(1 + 2.1170) = 2.1170$$

$$\pi_i = \frac{2.1170}{1 + 2.1170}$$

$$\pi_i = 0.67918$$

Therefore, this patient has a 67.92% chance of surviving or a 32.08% chance of death.

5 PART E

Classification Table[a]

Observed			Predicted: death - survived	Predicted: death - died	Percentage Correct
Step 1	death	survived	57	13	81.4
		died	19	24	55.8
	Overall Percentage				71.7

a. The cut value is .500

Figure 15. Classification table with cut value 0.5

$$Prior\ Probability = \frac{19 + 24}{19 + 57 + 24 + 13}$$

$$= \frac{43}{113} = 0.38053$$

Classification Table[a]

Observed			Predicted: death - survived	Predicted: death - died	Percentage Correct
Step 1	death	survived	43	27	61.4
		died	11	32	74.4
	Overall Percentage				66.4

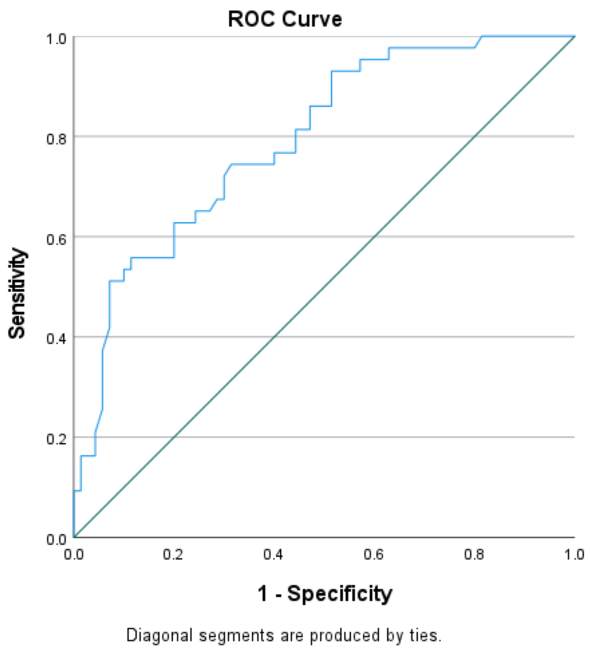
a. The cut value is .381

Figure 16. Classification table with cut value 0.381

$$Sensitivity = \frac{32}{32 + 11} = 0.74419$$

$$Specificity = \frac{43}{43 + 27} = 0.6143$$

6 PART F



Area Under the Curve

Test Result Variable(s): Predicted probability

Area
.792

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Figure 17. ROC curve

As the area under the curve is greater the 0.7, the model has a good predictive ability on the effect that MAP and the different categories of Shock have on the mortality of patients.