Week 9

Live Discussion Session

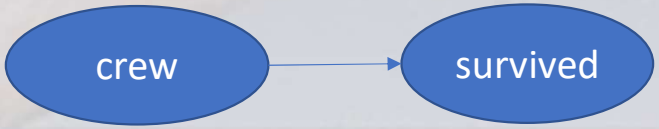
Starts at 2.05pm

The basic Titanic 'death' statistics

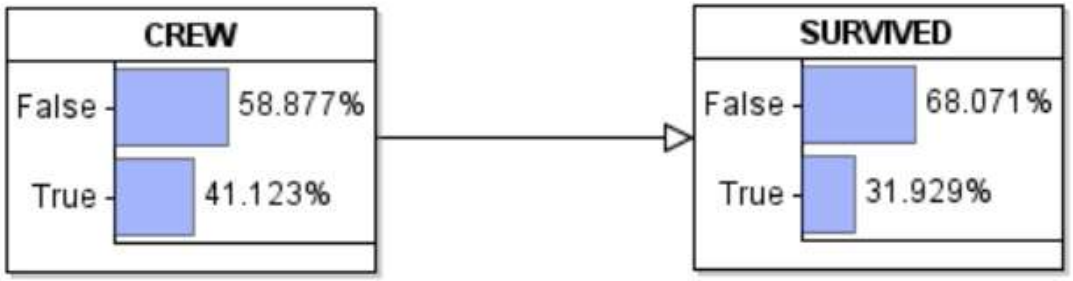
	Total	Died	Survived
Adult Female passengers	408	104	304
Adult Female staff	22	5	17
Child Female passengers	56	28	28
Adult Male passengers	780	647	133
Adult Male staff	886	691	195
Child male passengers	56	28	28
Total Females	486	137	349
Total males	1722	1366	356
Total passengers	1300	807	493
Total crew	908	696	212
Total	**2208**	**1503**	**705**

QUESTION: What is the probability a survivor was a crew member? (note: This is NOT the same as the probability a crew member survived)

Hint: create this model in AgenaRisk (and note that every NPT entry can be taken straight from the tabledata as AgenaRisk will do the necessary 'normalisation' to ensure probabilities sum to 1):



QUESTION: What is the probability a survivor was a man?



False	1300.0
True	908.0

CREW	False	True
False	807.0	696.0
True	493.0	212.0

2021 Exam Question

The following algorithm is 'learnt' from a subset of the dataset of passengers on the Titanic cruise liner which sank after hitting an iceberg on 15 April 1912:

If Sex = "Male" then Probability (survive) = 0.2
If Sex = "Female" and Class = 1 or 2 then Probability (survive) = 0.8
If Sex = "Female" and Class = 3 then Probability (survive) = 0.6

The relevant information in the different test dataset is summarized as:

	Male	Female Class 1 or 2	Female Class 3
Survived	75	75	60
Did not survive	225	15	50

Based on this test set data, the accuracy of the algorithm **for cut-off value 0.1** can be represented in the following format, where "YES" means survive and "NO" means not survive.

	Number predicted YES	Number predicted NO	Total
Number YES's	210	0	210
Number NO's	290	0	290

This enables us to compute:

Sensitivity:**100%;** Specificity: **0%**; False positive rate: **100%**; Accuracy:**42%**

a) For ***each*** of the different cut-off values 0.5, 0.7, 0.9 complete the following table and fill in all the missing ?? values

	Number predicted YES	Number predicted NO	Total
Number YES's	??	??	210
Number NO's	??	??	290

Sensitivity: **??%**; Specificity: **??%**; False positive rate: **??%**; Accuracy: **??%**

You will need to complete three tables and in each case the sensitivity specificity, false positive and accuracy percentages (8 marks each). [24 marks]

b) Sketch the ROC curve for this algorithm. [6 marks]

	Male	Female Class 1 or 2	Female Class 3
Survived	75	75	60
Did not survive	225	15	50

If Sex = “Male” then Probability (survive) = 0.2
If Sex = “Female” and Class = 1 or 2 then Probability (survive) = 0.8
If Sex = “Female” and Class = 3 then Probability (survive) = 0.6

	Number predicted YES	Number predicted NO	Total
Number YES’s	*A*	*B*	*A+B*
Number NO’s	*C*	*D*	*C+D*

Cut-off 0.1

	Number predicted YES	Number predicted NO	Total
Number YES’s	210	0	210
Number NO’s	290	0	290

Everyone have prob(survive) >0.1 so everyone predicted to survive

Sensitivity:**100%;** Specificity: **0%**; False positive rate: **100%**;
Accuracy:**42%**

Sensitivity = *A*/(*A+B*)

Specificity = *D*/(*C+D*)

Accuracy = (*A+D*)/(*A+B+C+D*)

Cut-off 0.5

	Number predicted YES	Number predicted NO	Total
Number YES’s	135	75	210
Number NO’s	65	225	290

No men have prob(survive) >0.5 so no men predicted to survive.
But all women have prob(survive)>0.5 so all women pred to survive

Sensitivity:**64%;** Specificity: **78%**; False positive rate: **22%**;
Accuracy:**72%**

Cut-off 0.7

	Number predicted YES	Number predicted NO	Total
Number YES’s	75	135	210
Number NO’s	15	275	290

No men have prob(survive) >0.7 so no men predicted to survive.
No women in class 3 have prob(survive)>0.7 but all women in class 1 or 2 pred to survive

Sensitivity:**36%;** Specificity: **95%**; False positive rate: **5%**; Accuracy:**70%**

Cut-off 0.9

	Number predicted YES	Number predicted NO	Total
Number YES’s	0	210	210
Number NO’s	0	290	290

Nobody has prob(survive) >0.9 so nobody predicted to survive.

Sensitivity:**0%;** Specificity: **100%**; False positive rate: **0%**; Accuracy:**58%**

CUT OFF 0.1

Sensitivity:**100%;** Specificity: **0%**;
False positive rate: **100%**;

CUT OFF 0.5

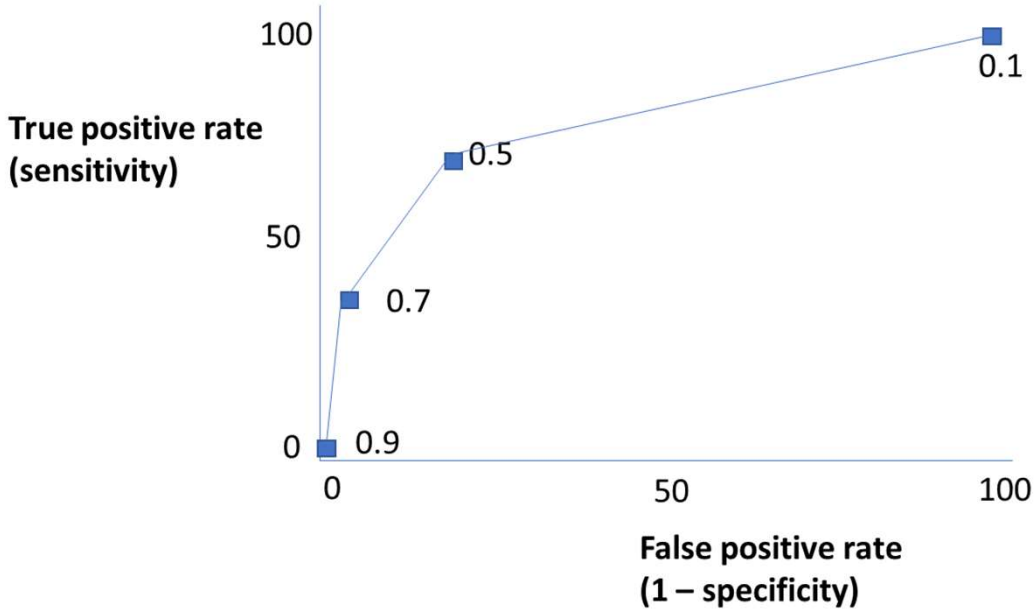
Sensitivity:**64%;** Specificity: **78%**;
False positive rate: **22%**;

CUT OFF 0.7

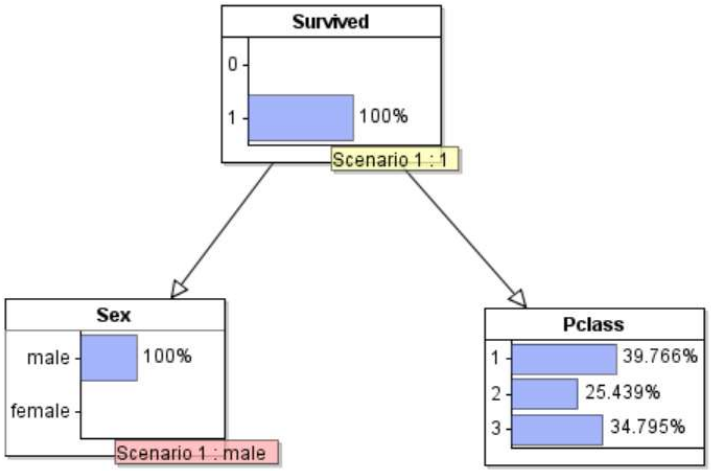
Sensitivity:**36%;** Specificity: **95%**;
False positive rate: **5%**;

CUT OFF 0.9

Sensitivity:**0%;** Specificity: **100%**;
False positive rate: **0%**;



Naïve Bayes – learnt from data



Causal Bayes – learnt from data