Notes

Plot of dnr day and drn

#Given that we have reason to believe drn (do not rescussitate order) is not independant to DNRday I plot on a chart

#in order to ascertain indepandance since we see that they are clearly not independant I decide to remove DNRday in order to simplify the analysis

ax=sns.stripplot(x=X2['dnr'], y=X2['dnrday'],hue=X2['surv2m'])

sns.move\_legend(ax, "upper left", bbox\_to\_anchor=(1, 1))

A table of numbers and symbols

Description automatically generated

Ca

txt="Figure 1: Confusion Matrix and Metrics Model 1"

plt.figtext(0.5, 0.01, txt, wrap=True, horizontalalignment='center', fontsize=12)

plt.tight\_layout()

plt.show()

\*Figure 2: Correlation Matrix Independant Variables\*

dnr

Model 1:

Accuracy of the model: 0.85

Micro Precision: 0.8511806699615596

Micro Recall: 0.8511806699615596

r2: 0.2281363885195904

mse: 0.14881933003844042

bic: 4039.0100693185977

Model 2:

Model 3:

Model 4:

A screenshot of a graph

Description automatically generated

The relationship between age and mortality appears weakly positive with the mean age of dead patients higher than the survivors. We also can see a significant number of outliers in deaths around 20 in both survivors and dead patients.

Relationship between scoma and hospital deaths is clearly positive although we can see a number of outliers in the survivors having multiple positive values.

Weakly positive relationship between totcst and deaths but we see a large number of outliers in the case of survivors having in many cases extremely high costs. It is likely that this may need to be removed and, or, transformed. Wil consider proportion plot in case that bears out clearer results.

There is a clear positive relationship between deaths and avtisst with median avtisst much higher in case of hospital deaths.

Clear positive relationship between hospital deaths and adlsc although we see that the maximum adlsc values for both survivors and dead patients are similar the median is significantly higher in the case of the hospital deaths.

It is unclear whether there is any relationship between race\_black and hospital deaths or not. It is possible that is in fact spurious. I note that the correlation was negligible (Figure 3 earlier in section 4.1). This is worth looking at as a proportion ratio plot before making any further decisions.

Dnr after admission (drn\_drn\_after\_sadm) has a clear apparent positive relationship with deaths but I note we have outliers in the case of hospital deaths with many not having this value and many in the survivor group having this value.

Dnr before admission (drn\_drn\_before\_sadm) does not have a clear relationship with deaths from the chart alone and I note we have outliers in the case of hospital deaths with many not having this value and many in the survivor group having this value. As with race\_black it is worth looking at as a proportion ratio plot before making any further decisions.

None of the dzgroup related plots showed any clear relationship but will as with race\_black, totcst and drn\_drn\_before\_sadm consider proportion plots before making further decisions on whether to include or remove these features from the training set.