Notes

Arthur Lui

19 March 2014

1 Reading Missing Data in R

```
read.table(na.string="") # => NA if ""
```

2 Cross Validation

- Fit the data using training set

- Evaluate the resultant model using test data

2.1 Steps

Goal:

- Predict whether an individual will be high / low risk

- To validate our model

1. obtain a model using training data

2. obtain cox model coefficients from that model

3. obtain risk scores $(e^{X\beta})$ for the subjects in our test data

4. Dichotomize risk scores (high/low risk groups, by median) = predictions

5. Compare predictions with reality (by plotting the KM curves for the two groups)

Further the 2 curves => better the model. Better distinguishing of two groups.
Risk scores are hazard ratio.

2.2 R Code