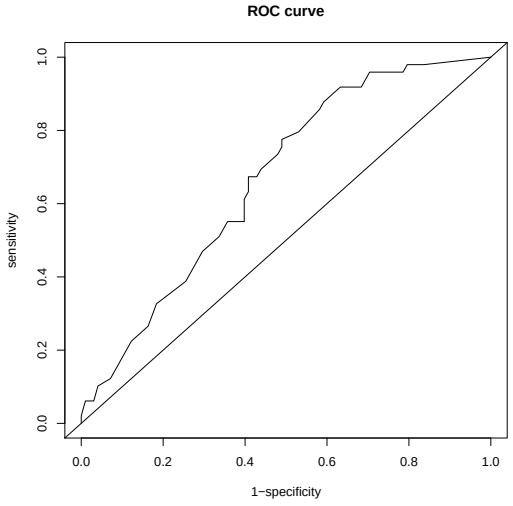
Stats for Clinical Trials, Math 150
Jo Hardin
Info on ROC curves

| | | Truth | | |
|---|---|---|---|---|
| | | Significant | Null | |
| Test | Significant | True Positive | False Positive | P* |
| | Null | False Negative | True Negative | N* |
| | | P | N | |

- type I error = FP
- type II error = FN
- sensitivity = power = true positive rate (TPR) = TP / P = TP / (TP+FN)
- false positive rate (FPR) = FP / N = FP / (FP + TN)
- specificity = 1 - FPR = TN / (FP + TN)
- accuracy (acc) = (TP+TN) / (P+N)
- positive predictive value (PPV) = precision = TP / (TP + FP)
- negative predictive value (NPV) = TN / (TN + FN)
- false discovery rate = 1 - PPV = FP / (FP + TP)



Below is the code for creating ROC curves in R:

```
   install.packages("ROCR")
library(ROCR)
smok.pred <- prediction(fitted(smok.log),lung.c)
smok.perf <- performance(smok.pred,measure="tpr",x.measure="fpr")
plot(smok.perf,xlab="1-specificity",ylab="sensitivity",main="ROC curve")
abline(a=0,b=1)
```

Note that R has many options within the `performance` function. I've listed only a few of them below. For more see:

```
?performance
```

- acc: Accuracy. P(Yhat = Y). Estimated as: (TP+TN)/(P+N).
- err: Error rate. P(Yhat $\neq$ Y). Estimated as: (FP+FN)/(P+N).
- fpr: False positive rate. P(Yhat = + | Y = -). Estimated as: FP/N.
- tpr: True positive rate. P(Yhat = + | Y = +). Estimated as: TP/P.
- sens: Sensitivity. Same as tpr.
- fnr: False negative rate. P(Yhat = - | Y = +). Estimated as: FN/P.
- tnr: True negative rate. P(Yhat = - | Y = -). Estimated as: TN/N.
- spec: Specificity. Same as tnr.
- ppv: Positive predictive value. P(Y = + | Yhat = +). Estimated as: TP/(TP+FP).
- prec: Precision. Same as ppv.
- npv: Negative predictive value. P(Y = - | Yhat = -). Estimated as: TN/(TN+FN).
- rpp: Rate of positive predictions. P(Yhat = +). Estimated as: (TP+FP)/(TP+FP+TN+FN).
- rnp: Rate of negative predictions. P(Yhat = -). Estimated as: (TN+FN)/(TP+FP+TN+FN).
- odds: Odds ratio. (TP*TN)/(FN*FP). Note that odds ratio produces Inf or NA values for all cutoffs corresponding to FN=0 or FP=0. This can substantially decrease the plotted cutoff region.
- auc: Area under the ROC curve. This is equal to the value of the Wilcoxon-Mann-Whitney test statistic and also the probability that the classifier will score are randomly drawn positive sample higher than a randomly drawn negative sample. Since the output of auc is cutoff-independent, this measure cannot be combined with other measures into a parametric curve. The partial area under the ROC curve up to a given false positive rate can be calculated by passing the optional parameter fpr.stop=0.5 (or any other value between 0 and 1) to performance.