Info on ROC curves

<table>
<thead>
<tr>
<th>Test</th>
<th>Significant</th>
<th>Null</th>
<th>P*</th>
<th>N*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>True Positive</td>
<td>False Negative</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>False Positive</td>
<td>True Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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)

ROC curve

![ROC curve](image-url)
Below is the code for creating ROC curves in R:

```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:

```r
?performance
```

- `acc`: Accuracy. $P(\hat{Y} = Y)$. Estimated as: $(TP+TN)/(P+N)$.
- `err`: Error rate. $P(\hat{Y} \neq Y)$. Estimated as: $(FP+FN)/(P+N)$.
- `fpr`: False positive rate. $P(\hat{Y} = + | Y = -)$. Estimated as: $FP/N$.
- `tpr`: True positive rate. $P(\hat{Y} = + | Y = +)$. Estimated as: $TP/P$.
- `sens`: Sensitivity. Same as `tpr`.
- `fnr`: False negative rate. $P(\hat{Y} = - | Y = +)$. Estimated as: $FN/P$.
- `spec`: Specificity. Same as `tnr`.
- `ppv`: Positive predictive value. $P(Y = + | \hat{Y} = +)$. Estimated as: $TP/(TP+FP)$.
- `prec`: Precision. Same as `ppv`.
- `rpp`: Rate of positive predictions. $P(\hat{Y} = +)$. Estimated as: $(TP+FP)/(TP+FP+TN+FN)$.
- `rnp`: Rate of negative predictions. $P(\hat{Y} = -)$. 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 a 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`. 

2