ACB spreadsheet verification: receiver-operator characteristics analysis

Ed Wilkes

This document describes the verification of the ROC analysis spreadsheet, written by Prof Anders Kallner (May 2018 version). Calculations performed by these spreadsheets were verified in an independent statistical software (the R statistical computing environment v3.4.1) by the author of this document. The R packages required to run this code are shown below. This code can be copied and pasted into an instance of R and, given the test data as input, reproduce the analysis in this document.

Required packages:

```
require(caret)
## Warning: package 'caret' was built under R version 3.4.2
require(dplyr)
require(ggplot2)
require(knitr)
require(pROC)
require(ROCR)
```

Reading data into R:

```
# Read in csv file: "2018-07 ACB ROC analyses, method evaluation - test
data 1.csv"
df <- read.csv(file.choose(), header = TRUE)
kable(df)
```

UID	result	outcome
50	104	D
53	93	D
90	104	D
91	85	D
95	81	D
113	54	H
120	94	D
121	113	D
132	99	D
133	84	D
134	80	H
142	93	D
149	104	D
157	94	D
163	92	H
174	80	H

UID	result	outcome
175	70	D
184	110	D
185	94	D
187	78	H
217	64	H
240	87	H
241	87	H
257	70	D
262	93	D
268	98	D
273	91	D
294	69	H
319	106	D
341	75	H
347	88	H
357	71	D
368	80	D
376	74	D
377	93	D
381	86	D
399	87	D
403	72	H
406	80	D
409	83	D
415	96	D
438	82	H
451	83	D
484	82	D
490	79	D
545	96	D
549	77	H
555	86	D
563	76	D
568	86	H
573	93	D
578	76	H
583	78	D
588	86	D
593	69	D
598	92	D

UID	result	outcome
603	94	D
608	87	D
613	96	D
618	85	H
623	82	H
628	98	D
633	93	D
638	81	D
643	108	D
648	107	D
653	79	H
658	95	D
663	83	H
668	82	D
673	79	D
678	83	H
683	112	H
688	87	D
693	67	H
698	97	D
703	84	D
708	75	H
713	90	H
718	90	H
723	74	H
728	89	D
733	71	H
738	103	D
743	102	D
748	104	D
753	78	D
758	79	D
763	73	H
768	95	D
773	103	D
778	75	H
783	88	D
788	81	H
793	60	H
798	81	D

UID	result	outcome
803	103	D
808	83	H
813	70	H
818	98	D

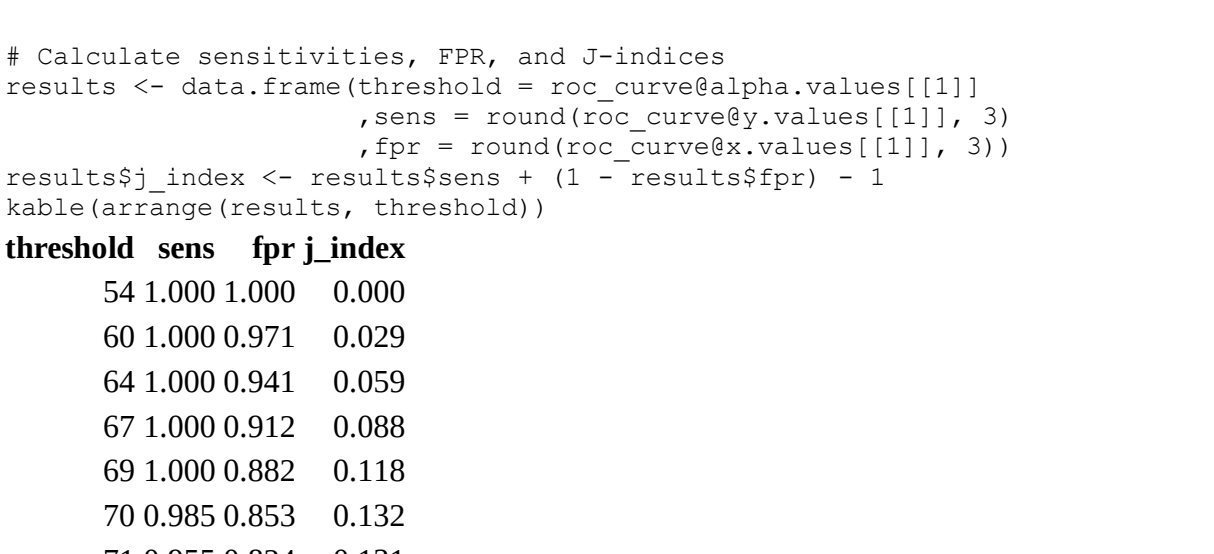
Check calculation of false positive rate and J-index at different thresholds:

These are the data presented in columns I:J from row 2 to 52 in sheets "*ROC (1)*" and "*ROC (2)*".

```
# Calculate ROC curve for test data set 1
roc_stats <- ROCR::prediction(predictions = df$result
                              ,labels = df$outcome
                              ,label.ordering = c("H", "D"))

roc_curve <- ROCR::performance(roc_stats
                               ,measure = "sens"
                               ,x.measure = "fpr")

plot(roc_curve)
abline(a = 0, b = 1, lty = 2)
```



```
# Calculate sensitivities, FPR, and J-indices
results <- data.frame(threshold = roc_curve@alpha.values[[1]]
                      ,sens = round(roc_curve@y.values[[1]], 3)
                      ,fpr = round(roc_curve@x.values[[1]], 3))
results$j_index <- results$sens + (1 - results$fpr) - 1
kable(arrange(results, threshold))
```

threshold	sens	fpr	j_index
54	1.000	1.000	0.000
60	1.000	0.971	0.029
64	1.000	0.941	0.059
67	1.000	0.912	0.088
69	1.000	0.882	0.118
70	0.985	0.853	0.132
71	0.955	0.824	0.131
72	0.939	0.794	0.145
73	0.939	0.765	0.174
74	0.939	0.735	0.204
75	0.924	0.706	0.218
76	0.924	0.618	0.306
77	0.909	0.588	0.321
78	0.909	0.559	0.350
79	0.879	0.529	0.350

threshold	sens	fpr	j_index
80	0.833	0.500	0.333
81	0.803	0.441	0.362
82	0.758	0.412	0.346
83	0.727	0.353	0.374
84	0.697	0.265	0.432
85	0.667	0.265	0.402
86	0.652	0.235	0.417
87	0.606	0.206	0.400
88	0.561	0.147	0.414
89	0.545	0.118	0.427
90	0.530	0.118	0.412
91	0.530	0.059	0.471
92	0.515	0.059	0.456
93	0.500	0.029	0.471
94	0.409	0.029	0.380
95	0.348	0.029	0.319
96	0.318	0.029	0.289
97	0.273	0.029	0.244
98	0.258	0.029	0.229
99	0.212	0.029	0.183
102	0.197	0.029	0.168
103	0.182	0.029	0.153
104	0.136	0.029	0.107
106	0.076	0.029	0.047
107	0.061	0.029	0.032
108	0.045	0.029	0.016
110	0.030	0.029	0.001
112	0.015	0.029	-0.014
113	0.015	0.000	0.015
Inf	0.000	0.000	0.000

The false positive rates (FPR, AKA 1 - specificity), sensitivities, and Youden indices (J-index) match those calculated in the spreadsheet.

Calculate confusion (2x2) matrix and associated statistics:

These are the data presented in cells K2:N14 and O3:R7 and are calculated based on the threshold with the highest J-index.

```
# Provide predictions given the optimal threshold
```

```
df$prediction <- ifelse(df$result >= 91
                        ,yes = "D"
                        ,no = "H")

df$prediction <- factor(df$prediction, levels = c("H", "D"))
df$outcome <- factor(df$outcome, levels = c("H", "D"))

# Calculate confusion (2x2) matrix statistics and AUROC
conf_mat <- caret::confusionMatrix(data = df$prediction
                                   ,reference = df$outcome
                                   ,positive = "D")
conf_mat
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  H  D
##          H 32 31
##          D  2 35
##
##                Accuracy : 0.67
##                  95% CI : (0.5688, 0.7608)
##     No Information Rate : 0.66
##     P-Value [Acc > NIR] : 0.4625
##
##                   Kappa : 0.3907
##  Mcnemar's Test P-Value : 1.093e-06
##
##             Sensitivity : 0.5303
##             Specificity : 0.9412
##          Pos Pred Value : 0.9459
##          Neg Pred Value : 0.5079
##              Prevalence : 0.6600
##          Detection Rate : 0.3500
##    Detection Prevalence : 0.3700
##       Balanced Accuracy : 0.7357
##
##        'Positive' Class : D
##
auc <- pROC::roc(response = df$outcome, predictor = df$result) %>%
  auc %>%
  round(3)
auc_ci <- pROC::roc(response = df$outcome, predictor = df$result) %>%
  ci.auc
auc_se <- ((auc_ci[[3]] - auc) / 1.96) %>%
  round(3)
lr_pos <- conf_mat$byClass[[1]] / (1 - conf_mat$byClass[[2]])
lr_neg <- (1 - conf_mat$byClass[[1]]) / conf_mat$byClass[[2]]
kable(data.frame(auc, auc_se, lr_pos, lr_neg))
```

auc	auc_se	lr_pos	lr_neg
0.785	0.047	9.015151	0.499053

The values for efficiency (i.e., accuracy), sensitivity, specificity, FPR (1 - specificity), PPV (post. prob. PV+), NPV (post. prob. PV-), prevalence, Cohen's κ , LR+, and LR- match those in the spreadsheet. The value for the AUROC matches that calculated in the spreadsheet.

Comparing the two ROC curves:

These are the data presented in cells B15:F22 in the "*Comparison*" sheet.

```
# Read in csv file: "2018-07 ACB ROC analyses, method evaluation - test
data 2.csv"
df_2 <- read.csv(file.choose(), header = TRUE)
roc_curve_1 <- pROC::roc(response = df$outcome
                         ,predictor = df$result
                         ,levels = c("H", "D"))
roc_curve_2 <- pROC::roc(response = df_2$outcome
                         ,predictor = df_2$result
                         ,levels = c("H", "D"))

# Perform null hypothesis significance testing of the two ROC curves
comparison <- data.frame(auc_1 = roc_curve_1$auc
                         ,auc_2 = roc_curve_2$auc) %>%
  mutate(diff_auc = roc_curve_1$auc - roc_curve_2$auc
         ,auc_se_1 = (ci.auc(roc_curve_1)[[3]] - auc_1) / 1.96
         ,auc_se_2 = (ci.auc(roc_curve_2)[[3]] - auc_2) / 1.96
         ,auc_var_1 = auc_se_1 ^ 2
         ,auc_var_2 = auc_se_2 ^ 2

         ,diff_auc_se = auc_se_1 - auc_se_2)
kable(comparison)
```

auc_1	auc_2	diff_auc	auc_se_1	auc_se_2	auc_var_1	auc_var_2	diff_auc_se
0.785205	0.5514706	0.2337344	0.0473155	0.0629425	0.0022388	0.0039618	-0.015627

```
pROC::roc.test(roc1 = roc_curve_1
               ,roc2 = roc_curve_2
               ,method = "delong"
               ,alternative = "two.sided")
##
##  DeLong's test for two ROC curves
##
## data:  roc_curve_1 and roc_curve_2
## D = 2.9683, df = 183.81, p-value = 0.003394
## alternative hypothesis: true difference in AUC is not equal to 0
## sample estimates:
## AUC of roc1 AUC of roc2
##   0.7852050   0.5514706
```

The values for ZZ (i.e., DD in the output shown above) were slightly different, most likely due to rounding errors. The null hypothesis (H0H0) was rejected in each case.

Conclusions:

1. Calculation of the FPR and J-index at various thresholds matched the values shown in the spreadsheet (tested in both "*ROC (1)*" and "*ROC (2)*")
2. Calculation of the confusion (2x2) matrix for the optimal threshold (91) matched the values shown in the spreadsheet (tested in both "*ROC (1)*" and "*ROC (2)*")"
3. Comparison of the ROC curves provided slightly different results – presumably due to rounding errors – but produced identical conclusions (i.e., rejection of H0H0)