ROC Curve for multiple biomarkers

**Multiple biomarkers – Link function – Sepsis data**

A preamble: Let us look at the ‘Sepsis’ data. There are several predictors having impact on the outcome of surgery. Some predictors are categorical. Each predictor is potentially a biomarker for the outcome. Is it possible to combine all predictors to form one biomarker? Yes, we can. Logistic regression with its link function paves the way. Let us start. Download the data.

> Sepsis <- read.delim("clipboard")

> dim(Sepsis)

[1] 106 6

> head(Sepsis)

Shock Malnutrition Alcoholism Age Infarction Death

1 0 0 0 56 0 0

2 0 0 0 80 0 0

3 0 0 0 61 0 0

4 0 0 0 26 0 0

5 0 0 0 53 0 0

6 0 1 0 87 0 1

Fit the logistic regression model.

> Sepsis1 <- glm(Death ~ ., data = Sepsis, family = binomial)

> summary(Sepsis1)

Call:

glm(formula = Death ~ ., family = binomial, data = Sepsis)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.3277 -0.4204 -0.0781 -0.0274 3.2946

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -9.75391 2.54170 -3.838 0.000124 \*\*\*

Shock 3.67387 1.16481 3.154 0.001610 \*\*

Malnutrition 1.21658 0.72822 1.671 0.094798 .

Alcoholism 3.35488 0.98210 3.416 0.000635 \*\*\*

Age 0.09215 0.03032 3.039 0.002374 \*\*

Infarction 2.79759 1.16397 2.403 0.016240 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 105.528 on 105 degrees of freedom

Residual deviance: 53.122 on 100 degrees of freedom

AIC: 65.122

Number of Fisher Scoring iterations: 7

The model is an adequate summary of the data.

Write the prediction model …..

Biomarker: X = ln(Pr(Death)/Pr(Survival)) = linear combination of the predictors. This is called link function of the model.

Spend time on this…

R gives the value of X (link) for every one in the data.

> Sepsis3 <- predict(Sepsis1, newdata = Sepsis, type = "link")

> head(Sepsis3)

1 2 3 4 5 6

-4.5933554 -2.3816910 -4.1325919 -7.3579358 -4.8698134 -0.5200411

Is this biomarker a good candidate for the development of a diagnostic test? Do some feasible studies.

Segregate the data according to the outcome.Identify those who died.

> Death <- subset(Sepsis3, Sepsis$Death == 1)

> length(Death)

[1] 21

Identify those who survived.

> Survival <- subset(Sepsis3, Sepsis$Death == 0)

> length(Survival)

[1] 85

Summary statistics of the biomarker X for those who died:

> summary(Death)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-5.4230 -0.5200 0.9918 0.7640 2.5230 3.9050

Summary statistics of X for those who survived:

> summary(Survival)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-8.1870 -6.7130 -4.1920 -4.2400 -1.6990 0.3467

Comment: The X values for those who died seems to be higher than those of survivors.

Do density curves:

> plot(density(Death), xlab = "Link Function", ylab = "Density", xlim = c(-9, 6),

+ main = "Nonparametric Density Curves", col = "red", lwd = 2)

> rug(Death, side = 1, col = "red")

> lines(density(Survival), col = "blue", lwd = 2)

> rug(Survival, side = 3, col = "blue")



Comment: The curves are reasonably well separated. Try to build a diagnostic test. Activate ‘pROC’ package.

Create a data frame of the data.

> DiagData <- data.frame(Link = c(Death, Survival), Outcome = rep(c("Death", "Survival"), c(21, 85)))

> dim(DiagData)

[1] 106 2

> head(DiagData)

Link Outcome

6 -0.52004113 Death

8 -0.04048587 Death

11 2.48064245 Death

19 3.46328865 Death

24 2.52316564 Death

29 1.58722911 Death

Crank out the package.

> MyROC <- roc(DiagData$Outcome, DiagData$Link, levels = c("Death", "Survival"),

+ direction = ">")

> names(MyROC)

[1] "percent" "sensitivities" "specificities"

[4] "thresholds" "direction" "cases"

[7] "controls" "fun.sesp" "auc"

[10] "call" "original.predictor" "original.response"

[13] "predictor" "response" "levels"

> MyROC1 <- data.frame(Sensitivity = MyROC$sensitivities, Specificity =

+ MyROC$specificities, Cutoffs = MyROC$thresholds)

> head(MyROC1)

Sensitivity Specificity Cutoffs

1 1 0.00000000 Inf

2 1 0.04761905 3.801340

3 1 0.09523810 3.580256

4 1 0.14285714 3.156271

5 1 0.19047619 2.826630

6 1 0.23809524 2.663586

> plot(MyROC, col = "red")

> title(main = "ROC Curve for Sepsis data")

> ci.auc(MyROC)

95% CI: 0.863-0.9958 (DeLong)

> AUC <- (0.863 + 0.9958)/2

> AUC

[1] 0.9294

> MyROC1$Dist <- (MyROC1$Sensitivity - 1)^2 + (MyROC1$Specificity - 1)^2

> min(MyROC1$Dist)

[1] 0.03715557

> which(MyROC1$Dist == 0.03715557)

integer(0)

> MyROC1

Sensitivity Specificity Cutoffs Dist

1 1.00000000 0.00000000 Inf 1.00000000

2 1.00000000 0.04761905 3.80133970 0.90702948

3 1.00000000 0.09523810 3.58025608 0.81859410

4 1.00000000 0.14285714 3.15627092 0.73469388

5 1.00000000 0.19047619 2.82662951 0.65532880

6 1.00000000 0.23809524 2.66358574 0.58049887

7 1.00000000 0.28571429 2.50190404 0.51020408

8 1.00000000 0.33333333 2.31039383 0.44444444

9 1.00000000 0.38095238 1.86368716 0.38321995

10 1.00000000 0.42857143 1.33558581 0.32653061

11 1.00000000 0.47619048 1.03786617 0.27437642

12 1.00000000 0.52380952 0.69051703 0.22675737

13 1.00000000 0.57142857 0.36798264 0.18367347

14 0.98823529 0.57142857 0.24527027 0.18381188

15 0.98823529 0.61904762 0.05166682 0.14526312

16 0.98823529 0.71428571 -0.08875328 0.08177106

17 0.97647059 0.71428571 -0.17160787 0.08218629

18 0.96470588 0.71428571 -0.22488906 0.08287833

19 0.94117647 0.71428571 -0.38181211 0.08509286

20 0.94117647 0.76190476 -0.53545996 0.06014955

21 0.92941176 0.76190476 -0.60891862 0.06167204

22 0.91764706 0.76190476 -0.68565248 0.06347135

23 0.91764706 0.80952381 -0.75654357 0.04306319

24 0.90588235 0.80952381 -0.83000224 0.04513931

25 0.89411765 0.80952381 -0.86056192 0.04749225

26 0.88235294 0.80952381 -0.98070671 0.05012201

27 0.87058824 0.80952381 -1.11182872 0.05302858

28 0.87058824 0.85714286 -1.13921104 0.03715557

29 0.85882353 0.85714286 -1.15571399 0.04033896

30 0.84705882 0.85714286 -1.20179033 0.04379917

31 0.83529412 0.85714286 -1.27524899 0.04753619

32 0.82352941 0.85714286 -1.37679760 0.05155003

33 0.81176471 0.85714286 -1.45698679 0.05584069

34 0.80000000 0.85714286 -1.49585828 0.06040816

35 0.77647059 0.85714286 -1.52651579 0.07037356

36 0.76470588 0.85714286 -1.57979698 0.07577149

37 0.75294118 0.85714286 -1.66255374 0.08144623

38 0.74117647 0.85714286 -1.71792818 0.08739778

39 0.72941176 0.85714286 -1.75470642 0.09362616

40 0.71764706 0.85714286 -1.82816509 0.10013135

41 0.71764706 0.90476190 -2.02395963 0.08879348

42 0.70588235 0.90476190 -2.20826499 0.09557549

43 0.69411765 0.90476190 -2.26154618 0.10263431

44 0.69411765 0.95238095 -2.32631654 0.09583159

45 0.65882353 0.95238095 -2.40196628 0.11866896

46 0.64705882 0.95238095 -2.43874453 0.12683505

47 0.63529412 0.95238095 -2.55669825 0.13527795

48 0.62352941 0.95238095 -2.70422536 0.14399768

49 0.61176471 0.95238095 -2.79637805 0.15299422

50 0.60000000 0.95238095 -2.91591305 0.16226757

51 0.58823529 0.95238095 -3.10021842 0.17181775

52 0.57647059 0.95238095 -3.25714146 0.18164474

53 0.55294118 0.95238095 -3.38607239 0.20212917

54 0.54117647 0.95238095 -3.51500332 0.21278660

55 0.52941176 0.95238095 -3.66253044 0.22372086

56 0.51764706 0.95238095 -3.94828658 0.23493193

57 0.50588235 0.95238095 -4.16216535 0.24641982

58 0.49411765 0.95238095 -4.28389143 0.25818453

59 0.48235294 0.95238095 -4.43862340 0.27022605

60 0.47058824 0.95238095 -4.54727902 0.28254439

61 0.45882353 0.95238095 -4.68550805 0.29513955

62 0.44705882 0.95238095 -4.82373707 0.30801152

63 0.43529412 0.95238095 -4.96196610 0.32116031

64 0.42352941 0.95238095 -5.19234780 0.33458591

65 0.41176471 0.95238095 -5.37665317 0.34828833

66 0.41176471 1.00000000 -5.45950776 0.34602076

67 0.40000000 1.00000000 -5.50558410 0.36000000

68 0.38823529 1.00000000 -5.65311122 0.37425606

69 0.36470588 1.00000000 -5.92027117 0.40359862

70 0.35294118 1.00000000 -6.05850020 0.41868512

71 0.34117647 1.00000000 -6.15995097 0.43404844

72 0.32941176 1.00000000 -6.34425634 0.44968858

73 0.29411765 1.00000000 -6.47318727 0.49826990

74 0.27058824 1.00000000 -6.61141629 0.53204152

75 0.24705882 1.00000000 -6.80501975 0.56692042

76 0.23529412 1.00000000 -6.98932512 0.58477509

77 0.21176471 1.00000000 -7.12755414 0.62131488

78 0.20000000 1.00000000 -7.21970683 0.64000000

79 0.17647059 1.00000000 -7.31185951 0.67820069

80 0.15294118 1.00000000 -7.40401219 0.71750865

81 0.12941176 1.00000000 -7.54224122 0.75792388

82 0.11764706 1.00000000 -7.68047024 0.77854671

83 0.08235294 1.00000000 -7.77262292 0.84207612

84 0.07058824 1.00000000 -7.86477561 0.86380623

85 0.03529412 1.00000000 -7.95692829 0.93065744

86 0.02352941 1.00000000 -8.04908097 0.95349481

87 0.01176471 1.00000000 -8.14123365 0.97660900

88 0.00000000 1.00000000 -Inf 1.00000000

> MyROC1[28, ]

Sensitivity Specificity Cutoffs Dist

28 0.8705882 0.8571429 -1.139211 0.03715557

> arrows(1, 0.8571429, 1, 0.8705882, col = "blue", lwd = 2)

> arrows(1, 1, 0.8571429, 0.8705882, col = "blue", lwd = 2)



Discussion: