Nithin Das 10422784 Date: 11/21/2019

**ASSIGNMENT 5**

**Question 12.23**

**PerformabinarylogisticregressionanalysisusingtheParentalHIVdatato model the probability of having been absent from school without a reason (variable HOOKEY). Find the variables that best predict whether an adolescent had been absent withouta reason or not. Assess goodness-of-ﬁtfor the ﬁnalmodel(overallandinﬂuenceofpatterns).**

The variables that best predict the results are:

AGE -age

HOWREL -religious or spiritual

BSI53 – others not giving proper credit for achievements

BSI37 – Feeling weak in parts of your body

BSI40 – Having urges to beat, injure or harm someone

BSI14 – Feeling lonely even when with people

BSI10 – Feeling that most people cannot be trusted

AGEMAR- age first had marijuana

Call:

glm(formula = HOOKEY ~ AGE + HOWREL + BSI53 + BSI37 + BSI40 +

BSI14 + BSI10 + AGEMAR, family = binomial(link = "logit"),

data = clean\_data)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.2457 -0.4446 0.0940 0.5377 2.8257

Coefficients:

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

(Intercept) -13.78637 2.28028 -6.046 1.49e-09 \*\*\*

AGE 0.85455 0.14515 5.887 3.92e-09 \*\*\*

HOWREL1 -2.39312 0.82704 -2.894 0.003809 \*\*

HOWREL2 -0.44201 0.61610 -0.717 0.473106

HOWREL3 1.34804 0.77406 1.742 0.081594 .

BSI531 -0.59024 0.61786 -0.955 0.339425

BSI532 -0.19773 1.33149 -0.148 0.881949

BSI533 -2.19155 1.33096 -1.647 0.099640 .

BSI534 -9.24212 2.44098 -3.786 0.000153 \*\*\*

BSI371 1.61431 0.55604 2.903 0.003693 \*\*

BSI372 -0.24851 0.96578 -0.257 0.796935

BSI373 0.94570 1.04289 0.907 0.364509

BSI374 2.76632 1.35897 2.036 0.041790 \*

BSI401 2.10338 0.65602 3.206 0.001345 \*\*

BSI402 1.58631 1.27914 1.240 0.214924

BSI403 4.29877 1.62679 2.642 0.008230 \*\*

BSI404 1.46638 1.79125 0.819 0.412995

BSI141 1.40181 0.56092 2.499 0.012450 \*

BSI142 0.89101 1.16901 0.762 0.445947

BSI143 -1.75713 0.84543 -2.078 0.037674 \*

BSI144 2.09455 1.11581 1.877 0.060495 .

BSI101 0.33334 0.48974 0.681 0.496091

BSI102 -0.04353 0.84142 -0.052 0.958738

BSI103 2.42578 0.82807 2.929 0.003396 \*\*

BSI104 0.60849 1.00888 0.603 0.546420

AGEMAR 0.14861 0.03708 4.008 6.12e-05 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 340.90 on 251 degrees of freedom

Residual deviance: 172.81 on 226 degrees of freedom

AIC: 224.81

Number of Fisher Scoring iterations: 6

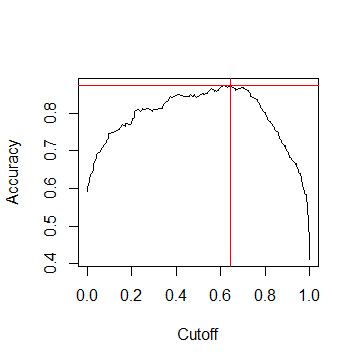
All the variables are significant with p values<0.05

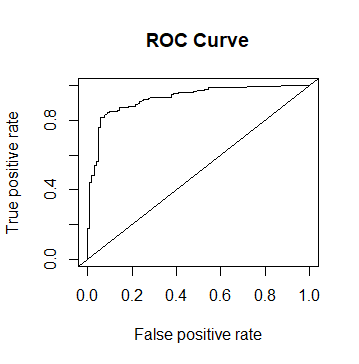
**Pseudo R^2 value**: 0.4931

If the value is close to 0, then the model has weak predictive power. In this case, the model has strong predictive power.

**Question 12.24**

**For the model in 12.23 ﬁnd an appropriate cutoff point to discriminate between adolescents who were absent without a reason and those who were not. Assess how well the model predicts the outcome using sensitivity, speciﬁcity, and the ROCcurve.**





**Confusion Matrix**

predicted\_values 0 1

0 94 24

1 9 125

The appropriate cut off point is at 0.645

Sensitivity at this point: 0.9126

Specificity at this point: 0.8389

Area Under Curve(AUC): 0.9262