Week 10 Assignments

Zoucha, Michael

11/2/2021

Assignment 1

- The effect of DsponeaBefore is statistically significant and positive (beta = 1.37, 95% CI [0.38, 2.31], p = 0.005; Std. beta = 1.37, 95% CI [0.38, 2.31])
- The effect of Size is statistically significant and positive (beta = 1.65, 95% CI [0.43, 2.85], p = 0.007; Std. beta = 1.65, 95% CI [0.43, 2.85])
- The effect of Diabetes is statistically significant and positive (beta = 0.93, 95% CI [0.02, 1.78], p = 0.037; Std. beta = 0.93, 95% CI [0.02, 1.78])
- The effect of Smoking is statistically significant and positive (beta = 1.08, 95% CI [0.18, 2.16], p = 0.030; Std. beta = 1.08, 95% CI [0.18, 2.16])

```
##
## Call:
## glm(formula = Risk1Yr ~ Diagnosis + DsponeaBefore + Size + Diabetes +
##     Smoking + AGE + Performance, family = binomial(link = "logit"),
##     data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3848  -0.5292  -0.4656  -0.2708   2.5236
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.858e+01  2.400e+03  -0.008   0.9938
## DiagnosisDGN2      1.437e+01  2.400e+03   0.006   0.9952
## DiagnosisDGN3      1.394e+01  2.400e+03   0.006   0.9954
## DiagnosisDGN4      1.442e+01  2.400e+03   0.006   0.9952
## DiagnosisDGN5      1.630e+01  2.400e+03   0.007   0.9946
## DiagnosisDGN6     -1.665e-01  2.672e+03   0.000   1.0000
## DiagnosisDGN8      1.804e+01  2.400e+03   0.008   0.9940
## DsponeaBeforeTRUE  6.231e-01  6.253e-01   0.996   0.3190
## SizeOC12           1.892e-01  3.844e-01   0.492   0.6225
## SizeOC13           3.731e-01  7.816e-01   0.477   0.6331
## SizeOC14           1.134e+00  7.899e-01   1.436   0.1510
## DiabetesTRUE       2.178e-01  6.406e-01   0.340   0.7338
## SmokingTRUE        1.313e+00  6.749e-01   1.946   0.0517 .
## AGE                8.314e-03  2.110e-02   0.394   0.6935
## PerformancePRZ1    6.793e-01  4.808e-01   1.413   0.1577
## PerformancePRZ2    3.672e-01  9.307e-01   0.395   0.6932
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 271.01  on 358  degrees of freedom
## Residual deviance: 243.34  on 343  degrees of freedom
## AIC: 275.34
##
## Number of Fisher Scoring iterations: 15
```

```
##             Predicted_Value
## Actual_Value FALSE TRUE
##        FALSE   310    4
##        TRUE     40    5
```

```
## [1] 0.8774373
```

Assignment 2

```
##      label             x                y         
##  Min.   :0.000   Min.   : -5.20   Min.   : -4.019  
##  1st Qu.:0.000   1st Qu.: 19.77   1st Qu.: 21.207  
##  Median :0.000   Median : 41.76   Median : 44.632  
##  Mean   :0.488   Mean   : 45.07   Mean   : 45.011  
##  3rd Qu.:1.000   3rd Qu.: 66.39   3rd Qu.: 68.698  
##  Max.   :1.000   Max.   :104.58   Max.   :106.896
```

```
## 
## Call:
## glm(formula = label ~ x + y, family = "binomial", data = train_data)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -1.3658  -1.1672  -0.9614   1.1650   1.4004  
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)   
## (Intercept)  0.415200   0.143553   2.892 0.003824 **
## x           -0.002376   0.002237  -1.062 0.288036   
## y           -0.007953   0.002302  -3.456 0.000549 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 1382.9  on 997  degrees of freedom
## Residual deviance: 1367.6  on 995  degrees of freedom
## AIC: 1373.6
## 
## Number of Fisher Scoring iterations: 4
```

```
##             Predicted_Value
## Actual_Value FALSE TRUE
##            0   289  222
##            1   202  285
```

```
## [1] 0.5751503
```