

MPH Practicum

Table 1: Demographics

category	cancer	noncancer
Gender		
Male		
Female		
Median Age		
Smoking History		
Never		
Former		
Current		

Table 2: Exclusion criteria

	n
Initial	184
Excluded	
UTUC	5
No cysto	1
No follow up	4
Insufficient RNA for biomarker analysis	7
Duplicates	9
Total	158

Table 3: Smoking count

smoking	smoker	n
never	0	50
former	1	78
current	1	30

```
kable(smoking_count,
      booktabs = T,
      align = "l",
      caption = "Smoking count"
    )
```

```
kable(cysto_result_risk,
      booktabs = T,
      align = "l",
      caption = "Cysto results vs risk"
    )
```

```
kable(smoking_hx,
      booktabs = T,
      align = "l",
      caption = "Smoking history"
    )
```

Table 4: Cysto results vs risk

visual_suspicion	risk	bca	n
no	NoBC	0	81
yes	NoBC	0	11
yes	low	1	7
yes	intermediate	1	12
yes	high	1	47

Table 5: Smoking history

smoking	packyrs	yearquit	n
never	0	0	50
former	0	1	3
former	1	1	75
current	1	0	30

```
#logistic regression cancer = biomarkers
log_biomarkers <- glm(bca ~ ROB01 + CRH + IGF2, family = "binomial", data = detection)
summary(log_biomarkers)
```

```
##
## Call:
## glm(formula = bca ~ ROB01 + CRH + IGF2, family = "binomial",
##      data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4177  -0.4177  -0.2003   0.2754   2.3159
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  21.60163    3.30546   6.535 6.36e-11 ***
## ROB01        -0.10807    0.07557  -1.430 0.152722
## CRH          -0.28474    0.07745  -3.676 0.000236 ***
## IGF2         -0.27209    0.09409  -2.892 0.003831 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 214.737  on 157  degrees of freedom
## Residual deviance:  90.608  on 154  degrees of freedom
## AIC: 98.608
##
## Number of Fisher Scoring iterations: 6
```

```
#logistic regression cancer = model3
log_model3 <- glm(bca ~ MODEL3, family = "binomial", data = detection)
summary(log_model3)
```

```
##
## Call:
## glm(formula = bca ~ MODEL3, family = "binomial", data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1730  -0.4505  -0.2786   0.4570   2.4043
##
## Coefficients:
```

```
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.4050      0.5000  -6.809 9.80e-12 ***
## MODEL3       5.7850      0.7659   7.553 4.26e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 214.74  on 157  degrees of freedom
## Residual deviance: 108.60  on 156  degrees of freedom
## AIC: 112.6
##
## Number of Fisher Scoring iterations: 5
```

```
#logistic regression cancer = biomarkers + smoking(binary)
log_smoker <- glm(bca ~ ROB01 + CRH + IGF2 + smoker, family = "binomial", data = detection)
summary(log_smoker)
```

```
##
## Call:
## glm(formula = bca ~ ROB01 + CRH + IGF2 + smoker, family = "binomial",
##      data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73058  -0.34312  -0.08536   0.18277   2.25117
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  21.93236    3.82240   5.738 9.59e-09 ***
## ROB01        -0.07658    0.07960  -0.962 0.336025
## CRH          -0.34115    0.08798  -3.878 0.000105 ***
## IGF2         -0.31032    0.10409  -2.981 0.002871 **
## smoker       2.50849    0.83424   3.007 0.002639 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 214.737  on 157  degrees of freedom
## Residual deviance:  78.957  on 153  degrees of freedom
## AIC: 88.957
##
## Number of Fisher Scoring iterations: 7
```

```
#logistic regression cancer = biomarkers + smoking(factor)
log_smoking <- glm(bca ~ ROB01 + CRH + IGF2 + smoking, family = "binomial", data = detection)
summary(log_smoking)
```

```
##
## Call:
## glm(formula = bca ~ ROB01 + CRH + IGF2 + smoking, family = "binomial",
##      data = detection)
```

```

##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.95643  -0.30678  -0.07892   0.18497   2.34767
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   22.83002    4.08118   5.594 2.22e-08 ***
## ROB01         -0.10029    0.08321  -1.205  0.22810
## CRH           -0.35229    0.08988  -3.920 8.87e-05 ***
## IGF2          -0.30478    0.10458  -2.914  0.00356 **
## smokingformer  2.34762    0.85759   2.737  0.00619 **
## smokingcurrent 3.13124    1.05069   2.980  0.00288 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 214.737  on 157  degrees of freedom
## Residual deviance:  77.869  on 152  degrees of freedom
## AIC: 89.869
##
## Number of Fisher Scoring iterations: 7

```

```
#R-squared cancer = biomarkers  
pR2(log_biomarkers)['McFadden']
```

```
## fitting null model for pseudo-r2
```

```
## McFadden  
## 0.5780512
```

```
#R-squared cancer = model3  
pR2(log_model3)['McFadden']
```

```
## fitting null model for pseudo-r2
```

```
## McFadden  
## 0.4942864
```

```
#R-squared cancer = biomarkers + smoking(binary)  
pR2(log_smoker)['McFadden']
```

```
## fitting null model for pseudo-r2
```

```
## McFadden  
## 0.6323096
```

```
#R-squared cancer = biomarkers + smoking(factor)  
pR2(log_smoking)['McFadden']
```

```
## fitting null model for pseudo-r2
```

```
## McFadden  
## 0.6373741
```

```

# c statistics cancer = biomarkers
predictions_biomarkers <- predict(log_biomarkers, type = "response")
roc(detection$bca, predictions_biomarkers)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

##
## Call:
## roc.default(response = detection$bca, predictor = predictions_biomarkers)
##
## Data: predictions_biomarkers in 92 controls (detection$bca 0) < 66 cases (detection$bca 1).
## Area under the curve: 0.9463

# c statistics cancer = model3
predictions_model3 <- predict(log_model3, type = "response")
roc(detection$bca, predictions_model3)

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

##
## Call:
## roc.default(response = detection$bca, predictor = predictions_model3)
##
## Data: predictions_model3 in 92 controls (detection$bca 0) < 66 cases (detection$bca 1).
## Area under the curve: 0.9269

# c statistics cancer = biomarkers + smoking(binary)
predictions_smoker <- predict(log_smoker, type = "response")
roc(detection$bca, predictions_smoker)

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

##
## Call:
## roc.default(response = detection$bca, predictor = predictions_smoker)
##
## Data: predictions_smoker in 92 controls (detection$bca 0) < 66 cases (detection$bca 1).
## Area under the curve: 0.9601

# c statistics cancer = biomarkers + smoking(factor)
predictions_smoking <- predict(log_smoking, type = "response")
roc(detection$bca, predictions_smoking)

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

```

```
##  
## Call:  
## roc.default(response = detection$bca, predictor = predictions_smoking)  
##  
## Data: predictions_smoking in 92 controls (detection$bca 0) < 66 cases (detection$bca 1).  
## Area under the curve: 0.9615
```



```
#BIC cancer = biomarkers  
BIC(log_biomarkers)
```

```
## [1] 110.8582
```

```
#BIC cancer = model3  
BIC(log_model3)
```

```
## [1] 118.7204
```

```
#BIC cancer = biomarkers + smoking(binary)  
BIC(log_smoker)
```

```
## [1] 104.2695
```

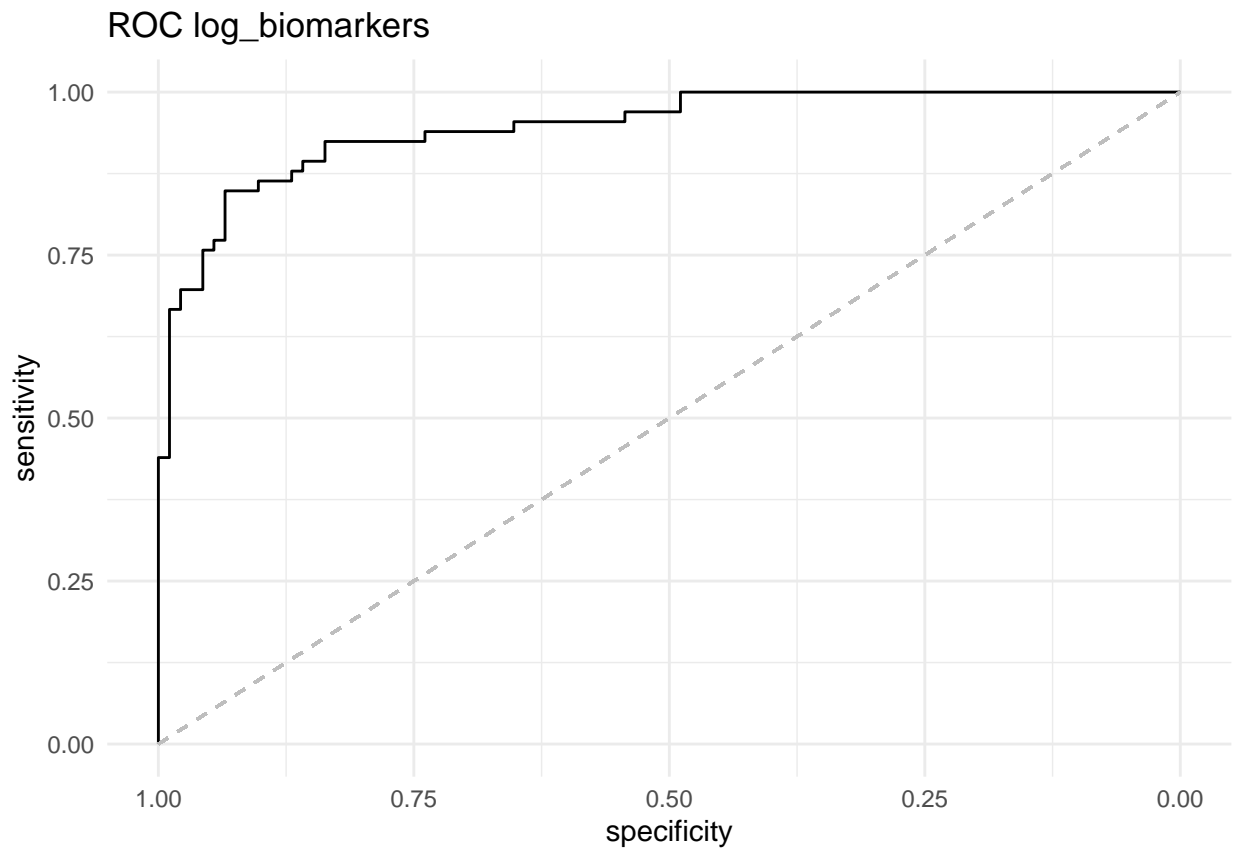
```
#BIC cancer = biomarkers + smoking(factor)  
BIC(log_smoking)
```

```
## [1] 108.2446
```

```
#ROC cancer = biomarkers
ggroc(roc(detection$bca, predictions_biomarkers)) +
  theme_minimal() +
  ggtitle("ROC log_biomarkers") +
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed")
```

```
## Setting levels: control = 0, case = 1
```

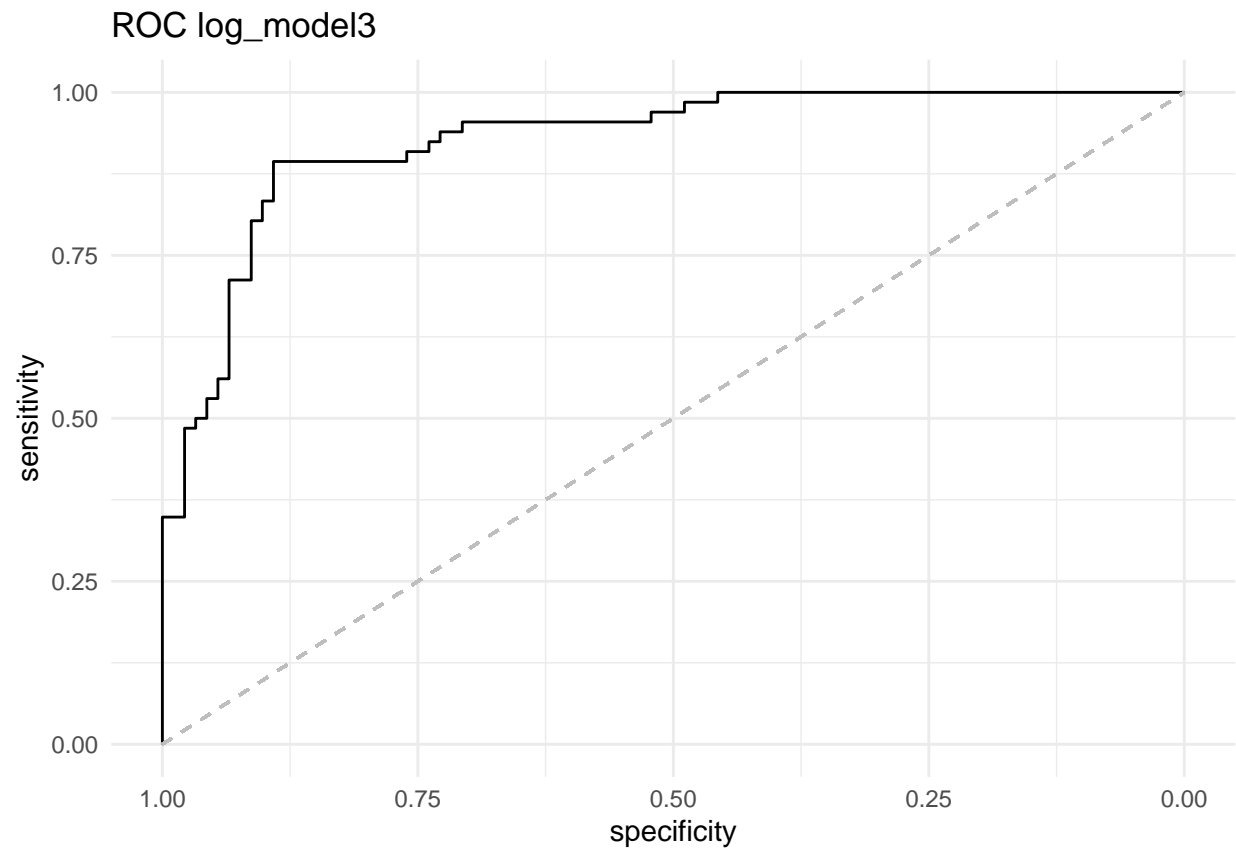
```
## Setting direction: controls < cases
```



```
#ROC cancer = model3
ggroc(roc(detection$bca, predictions_model3)) +
  theme_minimal() +
  ggtitle("ROC log_model3") +
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed")
```

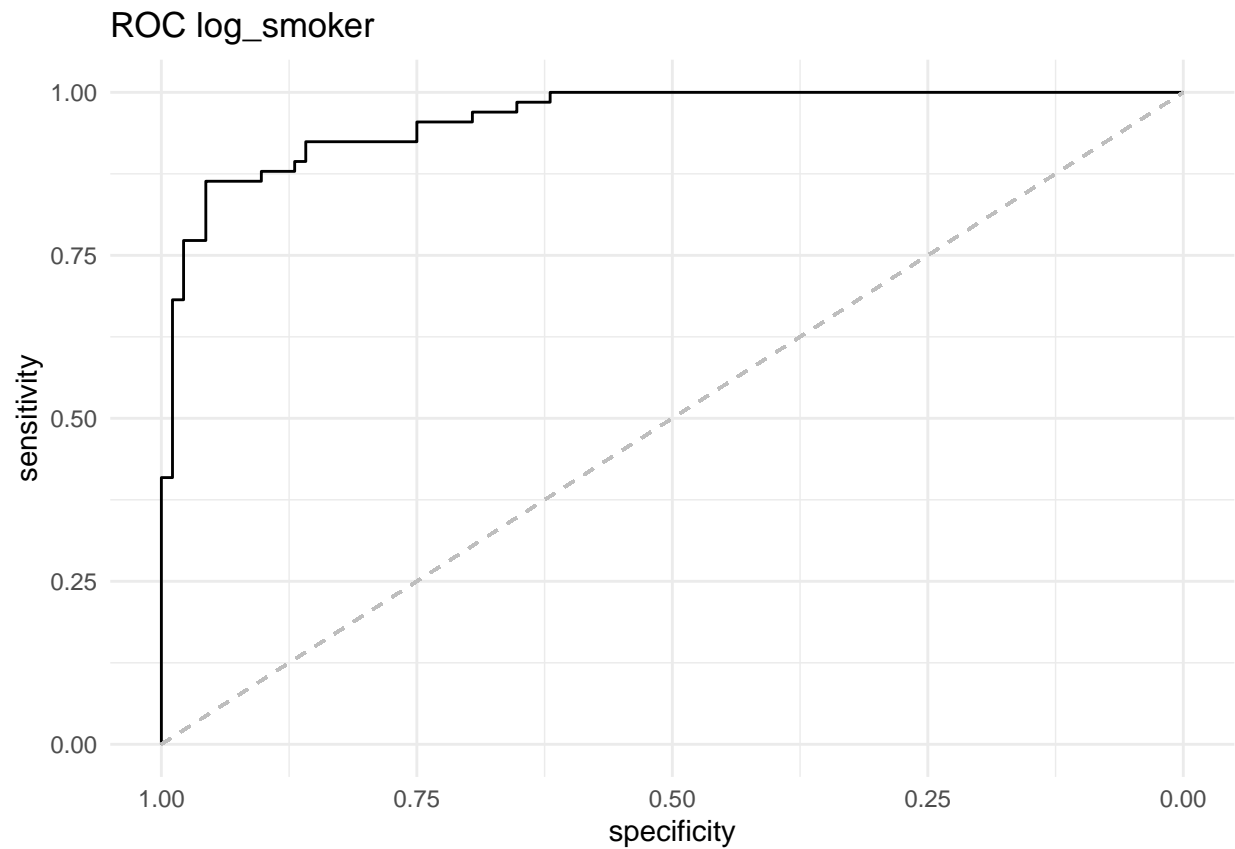
```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```



```
#ROC cancer = biomarkers + smoking(binary)  
ggroc(roc(detection$bca, predictions_smoker)) +  
  theme_minimal() +  
  ggtitle("ROC log_smoker") +  
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed")
```

```
## Setting levels: control = 0, case = 1  
## Setting direction: controls < cases
```



```
#ROC cancer = biomarkers + smoking(factor)  
ggroc(roc(detection$bca, predictions_smoking)) +  
  theme_minimal() +  
  ggtitle("ROC log_smoking") +  
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed")
```

```
## Setting levels: control = 0, case = 1  
## Setting direction: controls < cases
```

