

MPH Practicum

Table 1: 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 2: Smoking count

smoking	smoker	n
never	0	49
former	1	77
current	1	32

Table 3: 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 4: Smoking history

smoking	packyrs	yearquit	n
never	0	0	48
never	1	1	1
former	0	0	9
former	0	1	17
former	1	0	14
former	1	1	37
current	0	0	10
current	0	1	1
current	1	0	19
current	1	1	2

```
#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 ~ ROBO1 + CRH + IGF2 + smoker, family = "binomial", data = detection)
summary(log_smoker)
```

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2 + smoker, family = "binomial",
##      data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71568  -0.36504  -0.09492   0.19290   2.25202
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 22.26933    3.82099   5.828 5.6e-09 ***
## ROBO1       -0.08918    0.07885  -1.131 0.25802
## CRH         -0.34050    0.08746  -3.893 9.9e-05 ***
## IGF2        -0.30266    0.10282  -2.944 0.00324 **
## smoker       2.25740    0.79740   2.831 0.00464 **
## ---
## 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:  80.683  on 153  degrees of freedom
## AIC: 90.683
##
## Number of Fisher Scoring iterations: 7
```

```
#logistic regression cancer = biomarkers + smoking(factor)
log_smoking <- glm(bca ~ ROBO1 + CRH + IGF2 + smoking, family = "binomial", data = detection)
summary(log_smoking)
```

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2 + smoking, family = "binomial",
##      data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89774  -0.32341  -0.08834   0.17613   2.33698
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 23.03348    4.06802   5.662 1.50e-08 ***
## ROBO1       -0.10832    0.08226  -1.317 0.18790
## CRH         -0.34982    0.08907  -3.928 8.58e-05 ***
```

```
## IGF2          -0.29933    0.10306   -2.904   0.00368 **
## smokingformer  2.12364    0.82073    2.588   0.00967 **
## smokingcurrent 2.75137    1.00268    2.744   0.00607 **
## ---
## 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:  79.947  on 152  degrees of freedom
## AIC: 91.947
##
## 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
## 0.6242701
```

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

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

```
## McFadden
## 0.6276985
```

```
#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: 0.9578

# 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: 0.9587

```

```
#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] 105.9959
```

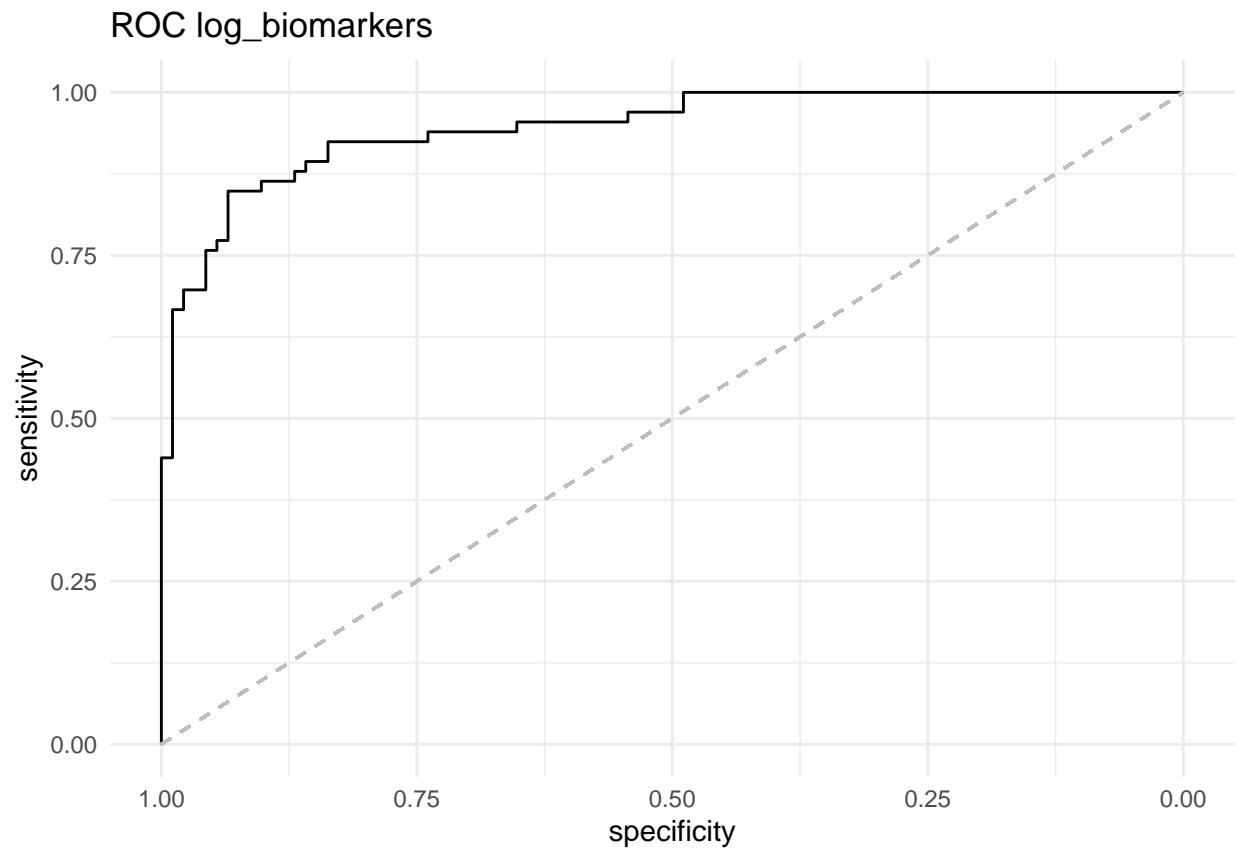
```
#BIC cancer = biomarkers + smoking(factor)  
BIC(log_smoking)
```

```
## [1] 110.3223
```

```
#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")
```

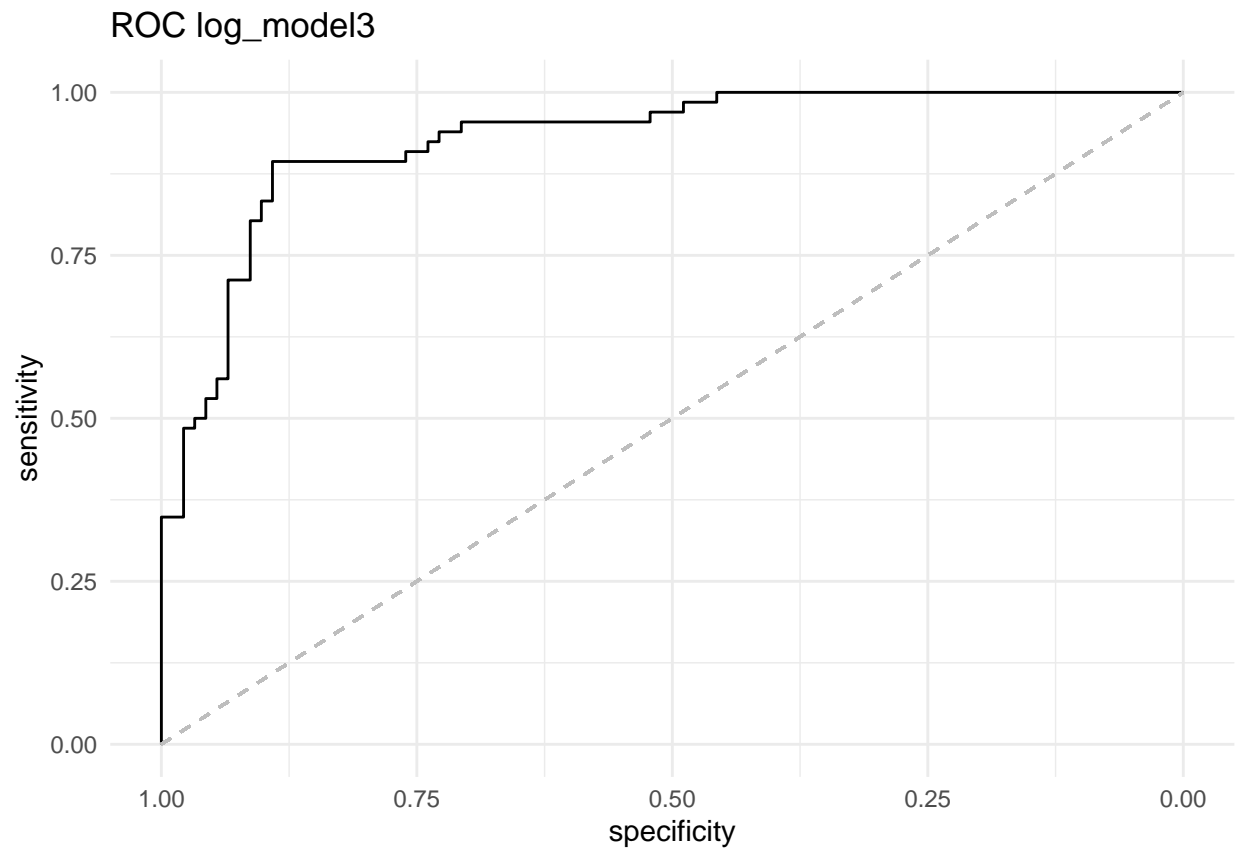
```
## Setting levels: control = 0, case = 1
```

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



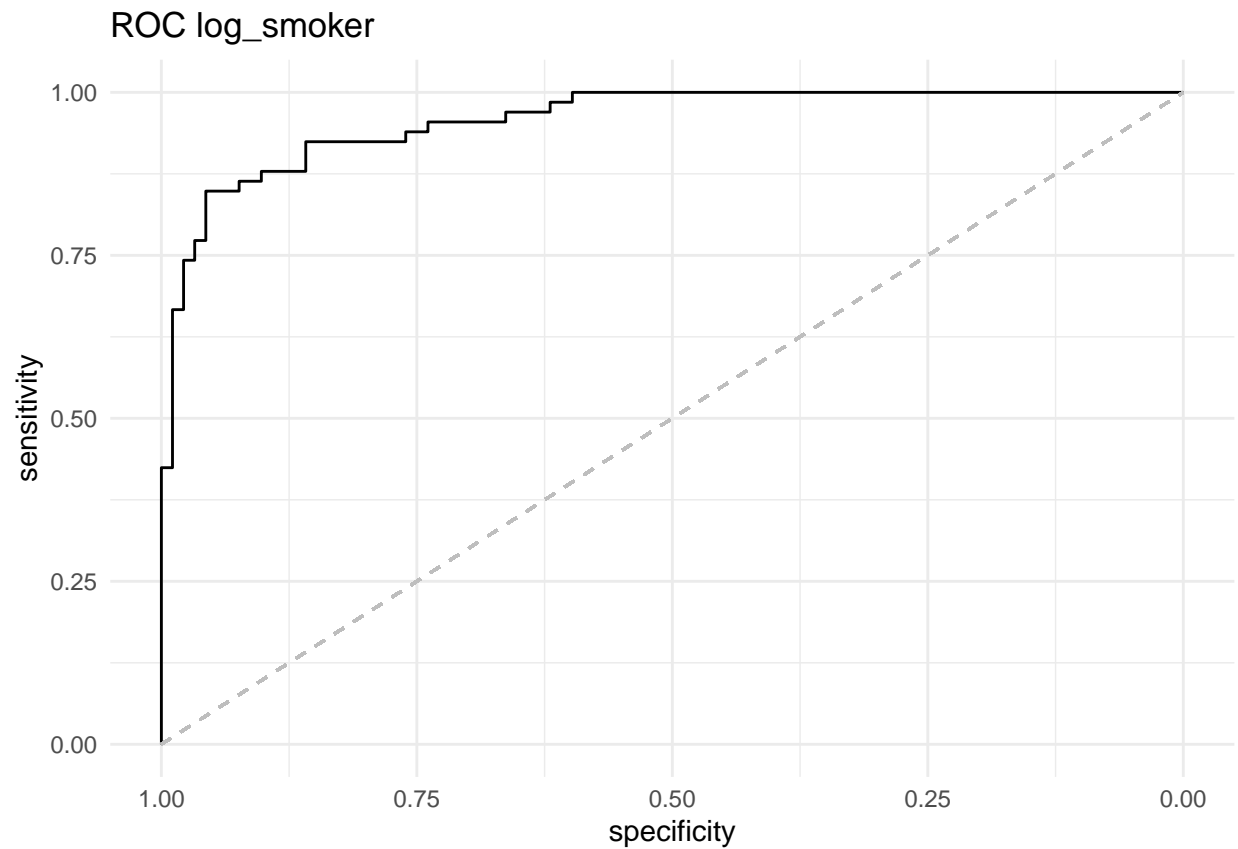
```
#ROC cancer = model3
ggroc(roc(detection$bca, predictions_model3)) +
  theme_minimal() +
  ggtitle("ROC log_model3") +
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed")
```

```
## 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
```

