MPH Practicum

```
## 'summarise()' has grouped output by 'bca'. You can override using the '.groups'
## argument.
```

Table 1: Demographics

	Cancer	No Cancer
Total, n(%)		
	66(42.3)	90(57.7)
Gender, n(%)		
Male	65(41.7)	85(54.5)
Female	1(0.6)	5(3.2)
Age, median(range)		
	74(56-90+)	72(36-90+)
BMI, mean(range)		
	29(14-59)	29(18-44)
Smoking History, n(%)		
Never	9(5.8)	41(26.3)
Former	40(25.6)	38(24.4)
Current	17(10.9)	11(7.1)
Indication, n(%)		
Gross hematuria	53(34.0)	37(23.7)
Microhematuria	3(1.9)	29(18.6)
LUTS	1(0.6)	9(5.8)
None	9(5.8)	15(9.6)

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
Chewing Tobacco	2
Total	**156**

Table 3: Smoking count

smoking	smoker	n
never	0	50
former	1	78
current	1	28

Table 4: Cysto results vs risk

visual_suspicion	risk	bca	n
no	NoBC	0	79
yes	NoBC	0	11
yes	low	1	7
yes	intermediate	1	12
yes	high	1	47

```
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 5: Smoking history

smoking	packyrs	yearquit	n
never	1	0	50
former	0	1	3
former	1	1	75
current	1	0	28

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

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2, family = "binomial",
##     data = detection)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4161  -0.4192  -0.2026   0.2805   2.3084
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 21.50422    3.30282   6.511 7.47e-11 ***
## ROBO1       -0.10765    0.07546  -1.427 0.153705
## CRH         -0.28281    0.07740  -3.654 0.000258 ***
## IGF2        -0.27137    0.09419  -2.881 0.003965 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.555  on 155  degrees of freedom
## Residual deviance:  90.387  on 152  degrees of freedom
## AIC: 98.387
##
## 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
## -2.1708  -0.4510  -0.2793   0.4588   2.3941
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.3757     0.5010  -6.738 1.61e-11 ***
## MODEL3        5.7498     0.7663   7.503 6.22e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.55  on 155  degrees of freedom
## Residual deviance: 108.28  on 154  degrees of freedom
```

```
## AIC: 112.28
##
## Number of Fisher Scoring iterations: 5
```

```
#logistic regression cancer = smoking(factor)
log_smoking_alone <- glm(bca ~ smoking, family = "binomial", data = detection)
summary(log_smoking_alone)
```

```
##
## Call:
## glm(formula = bca ~ smoking, family = "binomial", data = detection)
##
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -1.367  -1.199  -0.630   1.156   1.852
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.5163     0.3681  -4.119  3.8e-05 ***
## smokingformer    1.5676     0.4322   3.627 0.000287 ***
## smokingcurrent   1.9517     0.5341   3.654 0.000258 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.55  on 155  degrees of freedom
## Residual deviance: 192.74  on 153  degrees of freedom
## AIC: 198.74
##
## Number of Fisher Scoring iterations: 4
```

```
#logistic regression cancer = biomarkers + smoker(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
## -2.72750  -0.32807  -0.08422   0.19743   2.23921
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 21.77255    3.81629   5.705 1.16e-08 ***
## ROBO1       -0.07591    0.07948  -0.955  0.33953
## CRH         -0.33797    0.08788  -3.846  0.00012 ***
## IGF2        -0.30909    0.10418  -2.967  0.00301 **
## smoker       2.50507    0.83066   3.016  0.00256 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.555  on 155  degrees of freedom
## Residual deviance:  78.663  on 151  degrees of freedom
## AIC: 88.663
##
## Number of Fisher Scoring iterations: 7
```

```r
#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
## -2.97079  -0.28780  -0.07754   0.19542   2.33832
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    22.66773    4.06783   5.572 2.51e-08 ***
## ROBO1          -0.10143    0.08312  -1.220 0.222340
## CRH            -0.34860    0.08965  -3.888 0.000101 ***
## IGF2           -0.30219    0.10490  -2.881 0.003968 **
## smokingformer   2.32782    0.85402   2.726 0.006416 **
## smokingcurrent  3.18029    1.05073   3.027 0.002472 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.555  on 155  degrees of freedom
## Residual deviance:  77.401  on 150  degrees of freedom
## AIC: 89.401
##
## Number of Fisher Scoring iterations: 7
```

```r
#logistic regression cancer = biomarkers + pack years
log_packyears <- glm(bca ~ ROBO1 + CRH + IGF2 + smoke_packs, family = "binomial", data = detection)
summary(log_packyears)
```

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2 + smoke_packs, family = "binomial",
##     data = detection)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q       Max
```

```
## -2.4097  -0.2785  -0.1296   0.2072   2.4301
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 21.99295    3.71056   5.927 3.08e-09 ***
## ROBO1       -0.09807    0.08094  -1.212  0.22569
## CRH         -0.38028    0.09751  -3.900 9.62e-05 ***
## IGF2        -0.20909    0.09824  -2.128  0.03331 *
## smoke_packs  0.03264    0.01169   2.793  0.00523 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 209.212  on 152  degrees of freedom
## Residual deviance:  79.553  on 148  degrees of freedom
##    (3 observations deleted due to missingness)
## AIC: 89.553
##
## Number of Fisher Scoring iterations: 6
```

```
#logistic regression cancer = biomarkers + smoker(binary) + pack year
log_smoker_packyears <- glm(bca ~ ROBO1 + CRH + IGF2 + smoker + smoke_packs, family = "binomial", data =
summary(log_smoker_packyears)
```

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2 + smoker + smoke_packs,
##     family = "binomial", data = detection)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.67576  -0.29447  -0.07743   0.19686   2.12313
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 22.07585    3.96614   5.566 2.61e-08 ***
## ROBO1       -0.07247    0.08207  -0.883 0.377261
## CRH         -0.38165    0.09856  -3.872 0.000108 ***
## IGF2        -0.27164    0.10842  -2.505 0.012231 *
## smoker       2.00322    0.95211   2.104 0.035380 *
## smoke_packs  0.01700    0.01356   1.253 0.210127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 209.212  on 152  degrees of freedom
## Residual deviance:  74.641  on 147  degrees of freedom
##    (3 observations deleted due to missingness)
## AIC: 86.641
##
## Number of Fisher Scoring iterations: 7
```

```r
#logistic regression cancer = biomarkers + smoking(factor) + pack year
log_smoking_packyears <- glm(bca ~ ROBO1 + CRH + IGF2 + smoking + smoke_packs, family = "binomial", data
summary(log_smoking_packyears)
```

```
##
## Call:
## glm(formula = bca ~ ROBO1 + CRH + IGF2 + smoking + smoke_packs,
##     family = "binomial", data = detection)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79640  -0.29059  -0.07515   0.19372   2.17186
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    22.36720    4.05569   5.515 3.49e-08 ***
## ROBO1          -0.08476    0.08568  -0.989 0.322580
## CRH            -0.38054    0.09845  -3.865 0.000111 ***
## IGF2           -0.27072    0.10893  -2.485 0.012944 *
## smokingformer   1.99115    0.95767   2.079 0.037602 *
## smokingcurrent  2.40023    1.25578   1.911 0.055959 .
## smoke_packs     0.01444    0.01451   0.996 0.319416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 209.212  on 152  degrees of freedom
## Residual deviance:  74.396  on 146  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 88.396
##
## Number of Fisher Scoring iterations: 7
```

term	estimate	std.error	statistic	p.value
(Intercept)	22.6677	4.0678	5.5724	0.0000
ROBO1	-0.1014	0.0831	-1.2203	0.2223
CRH	-0.3486	0.0897	-3.8884	0.0001
IGF2	-0.3022	0.1049	-2.8807	0.0040
smokingformer	2.3278	0.8540	2.7257	0.0064
smokingcurrent	3.1803	1.0507	3.0268	0.0025

term	estimate	std.error	statistic	p.value
(Intercept)	22.3672	4.0557	5.5150	0.0000
ROBO1	-0.0848	0.0857	-0.9892	0.3226
CRH	-0.3805	0.0985	-3.8652	0.0001
IGF2	-0.2707	0.1089	-2.4853	0.0129
smokingformer	1.9911	0.9577	2.0792	0.0376
smokingcurrent	2.4002	1.2558	1.9114	0.0560
smoke_packs	0.0144	0.0145	0.9957	0.3194

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

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

##  McFadden
## 0.5747574
```

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

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

##  McFadden
## 0.4905778
```

```r
#R-squared cancer = smoking(factor)
pR2(log_smoking_alone)['McFadden']
```

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

##   McFadden
## 0.09322478
```

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

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

##  McFadden
## 0.6299162
```

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

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

##  McFadden
## 0.6358526
```

```r
#R-squared cancer = biomarkers + packyears
pR2(log_packyears)['McFadden']
```

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

##  McFadden
## 0.6197486
```

```
#R-squared cancer = biomarkers + smoker + packyears
pR2(log_smoker_packyears)['McFadden']
```

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

```
##   McFadden
## 0.6432253
```

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

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

```
## McFadden
## 0.644396
```

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

```
## [1] 110.5868
```

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

```
## [1] 118.3799
```

```r
#BIC cancer = smoking
BIC(log_smoking_alone)
```

```
## [1] 207.8891
```

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

```
## [1] 103.9124
```

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

```
## [1] 107.7004
```

```r
#BIC cancer = biomarkers + pack years
BIC(log_packyears)
```

```
## [1] 104.7052
```

```r
#BIC cancer = biomarkers + smoker + pack years
BIC(log_smoker_packyears)
```

```
## [1] 104.824
```

```r
#BIC cancer = biomarkers + smoking + pack years
BIC(log_smoking_packyears)
```

```
## [1] 109.6095
```

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

## Setting direction: controls < cases

##
## Call:
## roc.default(response = detection$bca, predictor = log_biomarkers$fitted.values,     plot = T, col =
##
## Data: log_biomarkers$fitted.values in 90 controls (detection$bca 0) < 66 cases (detection$bca 1).
## Area under the curve: 0.9456

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

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

## Setting direction: controls < cases
```