Homework 06
Joseph Blubaugh
jblubau1@tamu.edu
STAT 659-700

4.1

a) $exp[log(\frac{\pi}{1-\pi})] = exp[-3.771 + .1449(8)] = .0734 \rightarrow \frac{.0734}{1+.0734} = .068$

b) $exp[log(\frac{\pi}{1-\pi})] = exp[-3.771 + .1449(26)] = .9964 \rightarrow \frac{.9964}{1+.9964} = .5$

c) LI=8: $.1499 * .068(1 - .068) = .009$
LI=26: $.1499 * .5(1 - .5) = .036$

d) $exp[-3.771 + .1449(14)] = .175 \rightarrow \frac{.175}{1+.175} = .15$
$exp[-3.771 + .1449(28)] = 1.331 \rightarrow \frac{1.331}{1+1.331} = .57$

e) $exp(.1449) = 1.16$

4.2

a) $Ho : li = 0, Ha : li > 0,\ .1449/.0593 = 2.44 > 1.96$
pvalue $= (1 - pnorm(2.44)) * 2 = .014$

b) $exp[.1449 \pm 1.96(.0593)]$
A one unit increase in li increases the odds of remission by1.029087, 1.2983938

c) $Ho : li = 1, Ha : li > 1,\ 8.3 > 3.84 = X^2_{1,.05}$
pvalue $= 1 - pchisq(8.3, 1) = .004$

d) $exp(.0425, .2846) = (1.04, 1.32)$
A one unit increase in li increases the odds of remission by .03, 1.32

4.5

a)

```
dta = data.frame(
  Temperature = c(66, 70, 69, 68, 67, 72, 73, 70, 57, 63, 70, 78,
                  67, 53, 67, 75, 70, 81, 76, 79, 75, 76, 58),
  TD = c(0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0,
         0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1)
)

summary(dta)
```

```
  Temperature          TD
 Min.   :53.00   Min.   :0.0000
 1st Qu.:67.00   1st Qu.:0.0000
 Median :70.00   Median :0.0000
 Mean   :69.57   Mean   :0.3043
 3rd Qu.:75.00   3rd Qu.:1.0000
 Max.   :81.00   Max.   :1.0000
```

```
(mdl = glm(TD ~ Temperature, family = binomial(), data = dta))
```

```
Call:  glm(formula = TD ~ Temperature, family = binomial(), data = dta)

Coefficients:
(Intercept)  Temperature
    15.0429      -0.2322

Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
Null Deviance:      28.27
Residual Deviance: 20.32    AIC: 24.32
```

```
anova(mdl, test = "Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: TD

Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                           22     28.267
Temperature  1    7.952        21     20.315 0.004804 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As temperature increases by one unit, the odds of a failure are multiplied by 0.7928171

b) $log(\frac{\pi(31)}{1-\pi(31)}) = 15.043 - .2322(31) = 7.84$
   $exp(7.84)/(1 + exp(7.84)) = .999$

c) $log(.5/(1 - .5)) = 0 = 15.04 - .2322x \rightarrow x = .64.77$
   $-.2322 * .5 * (1 - .5) = -.05$

d) For a one unit increase in temperature, the odds of stress are multiplied by exp(-.2322) = 0.7927875

e) i. Wald Test Statistic: -.2322 / .1082 = -2.145, 2*pnorm(-2.145) = .032

ii. Liklihood Ratio Test: 1 - pchisq(7.965, 1), 0.004769

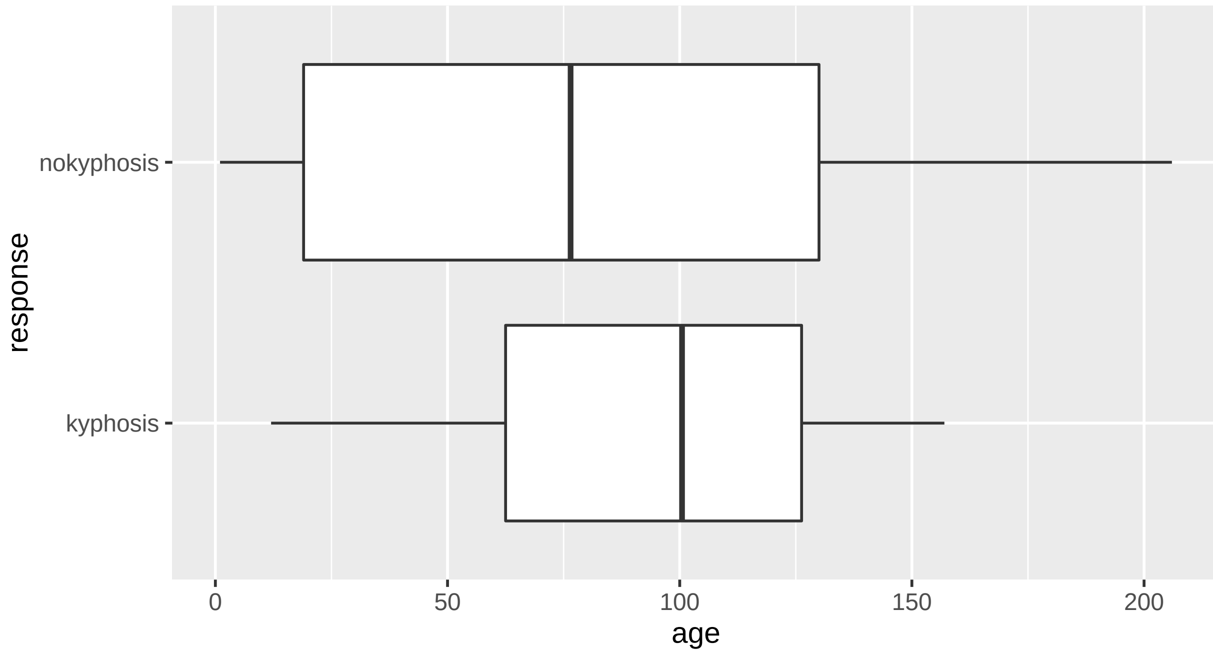
4.7

```
       response        age        
 kyphosis  :18   Min.   :  1.00  
 nokyphosis:22   1st Qu.: 35.50  
                 Median : 93.50  
                 Mean   : 85.95  
                 3rd Qu.:128.50  
                 Max.   :206.00  
```

```r
mdl = glm(response ~ age, family = binomial(), data = dta)
summary(mdl); anova(mdl, test = "Chisq")
```

```
Call:
glm(formula = response ~ age, family = binomial(), data = dta)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-1.4052  -1.2170   0.9482   1.0907   1.3126  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.572693   0.602395   0.951    0.342
age         -0.004296   0.005849  -0.734    0.463

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 55.051  on 39  degrees of freedom
Residual deviance: 54.504  on 38  degrees of freedom
AIC: 58.504

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: response

Terms added sequentially (first to last)

     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                    39     55.051         
age   1  0.54689        38     54.504   0.4596
```

a) Based on both the Wald and the Likelihood Ratio statistic, age does not significantly affect the odds of kyphosis occurring b)



c) The odds of kyphosis is not a linear function of age, extreme highs and lows of age have higher odds of kyphosis. The ages between 50 and 150 appear to minimize the odds of developing kyphosis

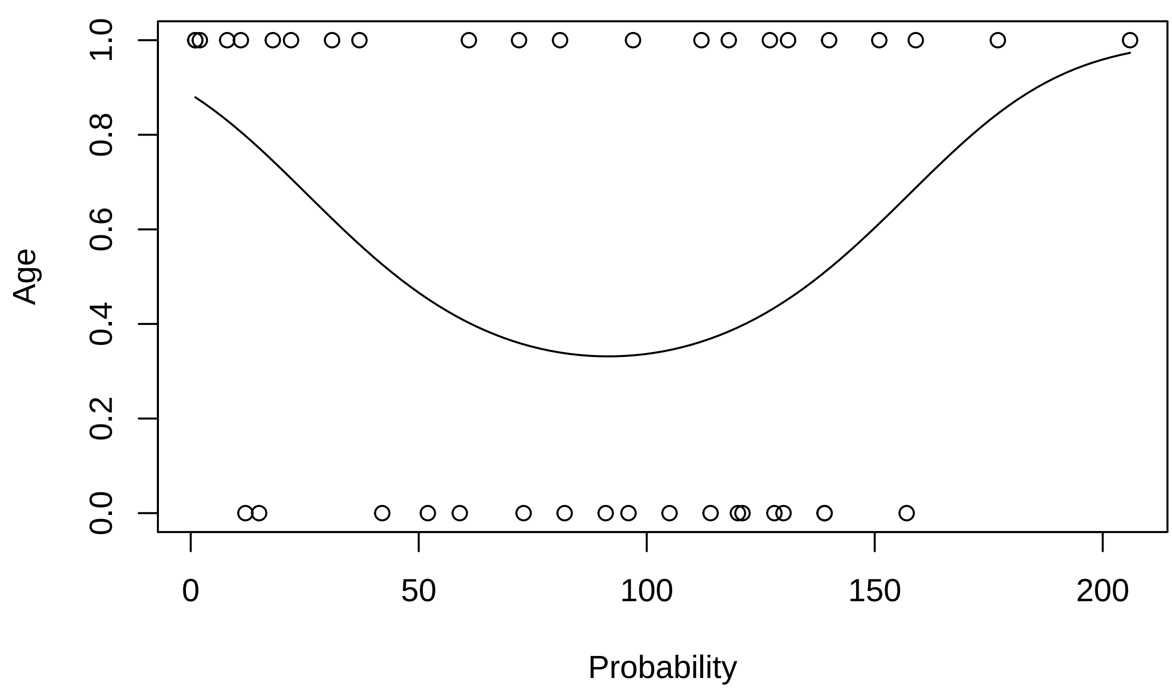
```r
mdl = glm(response ~ poly(age, 2), binomial(), data = dta)
summary(mdl); anova(mdl, test = "Chisq")
```

```
Call:
glm(formula = response ~ poly(age, 2), family = binomial(), data = dta)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.788  -1.012   0.507   1.009   1.482

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.3039     0.3599   0.844    0.398
poly(age, 2)1  -1.3212     2.4927  -0.530    0.596
poly(age, 2)2   6.2187     2.9659   2.097    0.036 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 55.051  on 39  degrees of freedom
```

```
Residual deviance: 48.228  on 37  degrees of freedom
AIC: 54.228

Number of Fisher Scoring iterations: 4
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: response

Terms added sequentially (first to last)

             Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                            39     55.051
poly(age, 2)  2   6.8231        37     48.228  0.03299 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(x = dta$age, y = as.numeric(dta$response) - 1, xlab = "Probability", ylab = "Age")
curve(predict(mdl, data.frame(age = x), type = "response"), add = TRUE)
```



4.8

a) Prediction Equation: $log(\frac{\pi(x)}{1-\pi(x)}) = -3.695 + 1.815x$

```
Call:
glm(formula = y ~ weight, family = binomial(), data = crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1108  -1.0749   0.5426   0.9122   1.6285

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6947     0.8802  -4.198 2.70e-05 ***
weight        1.8151     0.3767   4.819 1.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 195.74  on 171  degrees of freedom
AIC: 199.74

Number of Fisher Scoring iterations: 4
```

b)

```
predict(mdl, data.frame(weight = c(1.2, 2.44, 5.2)), type = "response")
```

```
        1         2         3
0.1799697 0.6757320 0.9968084
```

c)

$$\begin{aligned} log(\frac{\pi(.5)}{1-\pi(.5)}) &= -3.695 + 1.815x \\ log(1) &= -3.695 + 1.815x \\ 3.695 &= 1.815x \\ x &= 3.695/1.815 \\ &= 2.035 \end{aligned}$$

d) i. $1.815(.5)(1-.5) = .45$, ii. $.45/10 = .045$, iii. $.45/.58 = .77$

e) $exp[1.815 \pm 1.96(.376)] = (2.93, 12.83)$ A one unit increase in weight increases the odds of a having a satellite by 2.9 to 12.8 times.

f) $Ho : \theta = 1, Ha : \theta > 1, 4.81 > 1.96 = 1 - pnorm(4.81) = .0001$

4.9

a) $\log(\frac{\pi(x)}{1-\pi(x)}) = 1.0986 - .1226(color_2) - .7309(color_3) - 1.86(color_4)$ Color 1 is the default level in R so the coefficient is the intercept. The odds of having a satellite when the female crab is $color_1$ is exp(1.098) = 2.99

```
(mdl = glm(y ~ factor(color), family = binomial(), data = crabs))
```

```
Call:  glm(formula = y ~ factor(color), family = binomial(), data = crabs)

Coefficients:
   (Intercept)  factor(color)2  factor(color)3  factor(color)4
        1.0986         -0.1226         -0.7309         -1.8608

Degrees of Freedom: 172 Total (i.e. Null);  169 Residual
Null Deviance:      225.8
Residual Deviance: 212.1    AIC: 220.1
```

b) $13.7 > 7.8 = X^2_{3,.05} \rightarrow 1 - pchisq(225.76 - 212.06, 3) = .003$
Reject null and conclude that color is significant.

```
anova(mdl, test = "Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: y

Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                            172     225.76
factor(color)  3   13.698       169     212.06 0.003347 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) $log(\frac{\pi(x)}{1-\pi(x)}) = 2.363 - .714x$
As color increases by one unit, the odds of a female crab having a satellite increase by a factor of $exp(-7.14) = .48$

```
(mdl = glm(y ~ color, family = binomial(), data = crabs))
```

```
Call:  glm(formula = y ~ color, family = binomial(), data = crabs)
```

```
Coefficients:
(Intercept)        color
     2.3635      -0.7147

Degrees of Freedom: 172 Total (i.e. Null);  171 Residual
Null Deviance:      225.8
Residual Deviance: 213.3    AIC: 217.3
```

d) $12.461 > 3.84 = X^2_{1,.05} \rightarrow 1 - pchisq(12.461, 1) = .0004$ Color appears to have a significant factor on the probability of a female having a satellite.

```r
anova(mdl, test = "Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    172     225.76
color  1   12.461       171     213.30 0.0004156 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e) The advantage of having more power is because there are more degrees of freedom when you treat a variable as quantitative vs qualitative however a disadvantage of treated color as a quantitative variable implies that $color_2$ is greater than $color_1$ which doesn't make sense for something that should be qualitative and so you wont get an accurate representation between the effects of color.

4.15

```
## d)
library(DescTools)
library(reshape2)

dta = data.frame(
  District = c("NC", "NC", "NE", "NE", "NW", "NW", "SE", "SE", "SW", "SW"),
  Race = rep(c("Black", "White"), 5),
  Yes = c(24, 47, 10, 45, 5, 57, 16, 54, 7, 59),
  No = c(9, 12, 3, 8, 4, 9, 7, 10, 4, 12)
)

dta$Proportion = with(dta, Yes / (Yes + No))

mdl = glm(cbind(dta$Yes, dta$No) ~ Race, family = binomial(), data = dta); anova(mdl, test = "(
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(dta$Yes, dta$No)

Terms added sequentially (first to last)

     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                     9    10.6649
Race  1   8.0772         8     2.5876 0.004483 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dta2 = melt(dta[, 1:4], value.name = "Freq", variable.name = "Merit")

BreslowDayTest(xtabs(Freq ~ Race + Merit + District, data = dta2))
```

```
    Breslow-Day test on Homogeneity of Odds Ratios

data:  xtabs(Freq ~ Race + Merit + District, data = dta2)
X-squared = 2.1507, df = 4, p-value = 0.7081
```

Reject that null that the odds ratios are the same based on the Liklihood Ratio Statistic. Fail to reject the null of the Breslow Day test that the odds ratios differ. These two tests conflict

a) Conditional on District we reject the null that the proportions of race and merit are the same.

```
library(lawstat)

cmh.test(xtabs(Freq ~ Race + Merit + District, data = dta2))
```

```

    Cochran-Mantel-Haenszel Chi-square Test

data:  xtabs(Freq ~ Race + Merit + District, data = dta2)
CMH statistic = 7.8149000, df = 1.0000000, p-value = 0.0051817, MH
Estimate = 0.4617300, Pooled Odd Ratio = 0.4469900, Odd Ratio of
level 1 = 0.6808500, Odd Ratio of level 2 = 0.5925900, Odd Ratio
of level 3 = 0.1973700, Odd Ratio of level 4 = 0.4232800, Odd
Ratio of level 5 = 0.3559300
```

b) Using the Wald Test, $2.91 > 1.96 \rightarrow (1 - pnorm(2.91)) * 2 = .003$

c) You can create confidence intervals for the odds ratios.

4.16

a) Model: $log(\frac{\pi(x)}{1-\pi(x)}) = -2.11 - .55(EI_I) - .429(SN_S) + .687(TF_T) + .2(JP_P)$
The indicator variables are set up as factor with the default level intercept equal to EI_E, SN_N, TF_F, JP_J

```
dta = data.frame(
  EI = c("E", "E", "E", "E", "E", "E", "E", "E", "I", "I", "I", "I", "I", "I", "I", "I"),
  SN = c("S", "S", "S", "S", "N", "N", "N", "N", "S", "S", "S", "S", "N", "N", "N", "N"),
  TF = c("T", "T", "F", "F", "T", "T", "F", "F", "T", "T", "F", "F", "T", "T", "F", "F"),
  JP = c("J", "P", "J", "P", "J", "P", "J", "P", "J", "P", "J", "P", "J", "P", "J", "P"),
  Y = c(10, 8 , 5 , 7 , 3 , 2 , 4 , 15, 17, 3 , 6 , 4 , 1 , 5 , 1 , 6),
  N = c(67, 34, 101, 72, 20, 16, 27, 65, 123, 49, 132, 102, 12, 30, 30, 73)
)

(mdl = glm(cbind(dta$Y, dta$N) ~ EI + SN + TF + JP, family = binomial(), data = dta))
```

```

Call:  glm(formula = cbind(dta$Y, dta$N) ~ EI + SN + TF + JP, family = binomial(),
    data = dta)

Coefficients:
(Intercept)          EII          SNS          TFT          JPP
    -2.1140      -0.5550      -0.4292       0.6873       0.2022

Degrees of Freedom: 15 Total (i.e. Null);  11 Residual
Null Deviance:      30.49
Residual Deviance: 11.15    AIC: 73.99
```

b)

```
predict(mdl, data.frame(EI = "E", SN = "S", TF = "T", JP = "J"), type = "response")
```

```
       1
0.135186
```

c) This is the combination of variables which all coefficients are positive so it will have the highest probability

```
summary(mdl)
```

```

Call:
glm(formula = cbind(dta$Y, dta$N) ~ EI + SN + TF + JP, family = binomial(),
    data = dta)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2712  -0.8062  -0.1063   0.1124   1.5807

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1140     0.2715  -7.788 6.82e-15 ***
EII          -0.5550     0.2170  -2.558  0.01053 *
SNS          -0.4292     0.2340  -1.834  0.06664 .
TFT           0.6873     0.2206   3.116  0.00184 **
JPP           0.2022     0.2266   0.893  0.37209
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 30.488  on 15  degrees of freedom
Residual deviance: 11.149  on 11  degrees of freedom
AIC: 73.99

Number of Fisher Scoring iterations: 4
```

4.17

a)

$$
\begin{aligned}
log(\frac{\pi(x)}{1-\pi(x)}) =& -2.829 + .5805(e) + .597(t) \\
(I,F) =& -2.829 + 0 + .597 = -.232 \\
=& exp(-.232)/(1+exp(-.232)) \\
=& .44
\end{aligned}
$$

b) $exp(.5805) = 1.78$
Someone with an extrovert personality type is 1.78 times more likely that introverts to use alcohol frequently

c) $exp(.1589, 1.008) = (1.17, 2.74)$
At a 95% confidence level an extrovert is between 1.17 to 2.74 times more likely to use alcohol frequently

d) $exp(-1,008, -.1589) = (.364, 853). An introvert is between .364 and .854 times more likely to use alcohol frequently

e) I used the Liklihood Ratio test which tests the intercept only model first, then the intercept with EI, then with TF added. The results of the test show that both EI and TF are significant variables. The EI model is an improvement over the intercept only model and the TF model is an improvement over the EI model.

```
mdl = glm(cbind(dta$Y, dta$N) ~ EI + TF, family = binomial(), data = dta)
anova(mdl, test = "Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(dta$Y, dta$N)

Terms added sequentially (first to last)

     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                    15     30.488
EI    1   6.4521        14     24.036 0.011082 *
TF    1   7.6379        13     16.398 0.005715 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.19

a) The odds of supporting abortion are increased by $exp(.16) = 1.17$ when the gender is female

b)

$$
\begin{aligned}
\text{i) } log(\frac{\pi(x)}{1-\pi(x)}) =& \alpha + \beta_2^G + \beta_2^R + \beta_2^P \\
=& -.11 + 0 - .66 - 1.67 = -2.44 \\
\pi(x) =& \frac{exp(-2.44)}{1+exp(-2.44)} \\
=& .08
\end{aligned}
$$

$$
\begin{aligned}
\text{ii) } log(\frac{\pi(x)}{1-\pi(x)}) =& \alpha + \beta_1^G + \beta_3^R + \beta_1^P \\
=& -.11 + .16 + 0 + .84 \\
\pi(x) =& \frac{exp(.84)}{1+exp(.84)} \\
=& .698
\end{aligned}
$$

c) $B_2^G = -.16$

d) $B_1^G = .08, B_2^G = -.08$

4.22

a) For each unit increase in weight, the odds of a female having a satellite increase by a factor of $exp(.43) = 1.53$. When a female is color2 the odds of her attracting a satellite are increased by a factor of $exp(-.008) = .99$ than if the crab is color 1. When a female is color3 the odds of her attracting a satellite are increased by a factor of $exp(-.137) = .871$ than if the crab is color1. When a female is color4 the odds of her attracting a satellite are increased by a factor of $exp(-.66) = .51$ then if she is color1.

```
Call:  glm(formula = cbind(crabs$y, crabs$n) ~ weight + factor(color),
    family = binomial(), data = crabs)

Coefficients:
   (Intercept)          weight  factor(color)2  factor(color)3
      -1.43628         0.43364        -0.00849        -0.13745
factor(color)4
      -0.66410

Degrees of Freedom: 172 Total (i.e. Null);  168 Residual
Null Deviance:      72.31
Residual Deviance: 64.59    AIC: 228.5
```

b) The likelihood Ratio test for color shows that controlling for Weight, color is insignificant in determining the probability of a female crab having a satellite.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(crabs$y, crabs$n)

Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                            172     72.305
weight         1   5.5825       171     66.723  0.01814 *
factor(color)  3   2.1342       168     64.589  0.54503
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) For each unit increase in weight, the odds of a female having a satellite increase by a factor of $exp(.44) = 1.55$. For a one unit increase in color, the odds of a female having a satellite increases by a factor of $exp(-.2) = .818$. The liklihood Ratio Test for color controlling for weight also shows that weight is insignificant in determining the probability that a female will have a satellite.

```
Call:  glm(formula = cbind(crabs$y, crabs$n) ~ weight + color, family = binomial(),
    data = crabs)

Coefficients:
(Intercept)       weight        color
    -1.0685       0.4418      -0.2070

Degrees of Freedom: 172 Total (i.e. Null);  170 Residual
Null Deviance:      72.31
Residual Deviance: 65.19    AIC: 225.1

Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(crabs$y, crabs$n)

Terms added sequentially (first to last)

       Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                      172     72.305
weight  1   5.5825       171     66.723  0.01814 *
color   1   1.5343       170     65.189  0.21547
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.24

a) For a one minute increase in duration, the odds of a patient having a sore throat increases by a factor of $exp(.068) = 1.07$. Having a tracheal tube increases the odds of having a sore throat by a factor of $exp(-1.65) = .19$ vs having a laryngeal mask.

```
Call:  glm(formula = Y ~ D + T, family = binomial(), data = dta)

Coefficients:
(Intercept)            D            T
   -1.41734      0.06868     -1.65895

Degrees of Freedom: 34 Total (i.e. Null);  32 Residual
Null Deviance:      46.18
Residual Deviance: 30.14    AIC: 36.14
```

b) A 95% confidence interval for the odds increase in duration is between 1.0169989, 1.1278848

c)

i. When T = 1, $log(\frac{\pi(x)}{1-\pi(x)}) = .049 + (.0284 + .0746)x + -4.47$. The odds of a patient having a sore throat increases by a factor of $exp(.103) = 1.1$.

ii. When T = 0, $log(\frac{\pi(x)}{1-\pi(x)}) = .049 + .028x$. The odds of a patient having a sore throat increases by a factor of $exp(.028) = 1.028$

```
Call:  glm(formula = Y ~ D * T, family = binomial(), data = dta)

Coefficients:
(Intercept)            D            T          D:T
    0.04979      0.02848     -4.47224      0.07460

Degrees of Freedom: 34 Total (i.e. Null);  31 Residual
Null Deviance:      46.18
Residual Deviance: 28.32    AIC: 36.32
```

d) Based on the likelihood ratio test, the interaction term does not significantly affect the odds of a patient having a sore throat.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

     Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    34     46.180
D     1  12.5285        33     33.651 0.0004008 ***
T     1   3.5134        32     30.138 0.0608744 .
D:T   1   1.8169        31     28.321 0.1776844
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.30

If the athlete is white, the odds that the person will graduate increases by a factor of $exp(1.01) = 2.74$ vs someone who is black. If the athlete is male the odds the person will graduate increases by a factor of $exp(-.352) = .7$ vs that of a male. The 95% confidence level for the odds ratio between race and gender are $exp[log(.483) \pm 1.96 * sqrt(.018 + .0011 + .002 + .005)] = (.351, .662)$. We would conclude that the odds ratios between race and gender are significantly different and the odds of a white person graduating are higher than a black person.

```
  Grad No.Grad  Race Gender
1  498     298 White Female
2  878     747 White   Male
3   54      89 Black Female
4  197     463 Black   Male

Probability Prediction

        1         2         3         4
0.6257107 0.5402673 0.3771629 0.2985844



Call:
glm(formula = cbind(dta$Grad, dta$No.Grad) ~ Race + Gender, family = binomial(),
    data = dta)

Deviance Residuals:
        1          2          3          4
-0.004812   0.003270   0.011335  -0.005588

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) -0.50161    0.10004  -5.014 5.33e-07 ***
RaceWhite    1.01547    0.08723  11.641  < 2e-16 ***
GenderMale  -0.35244    0.08044  -4.381 1.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.8006e+02  on 3  degrees of freedom
Residual deviance: 1.9355e-04  on 1  degrees of freedom
AIC: 33.029

Number of Fisher Scoring iterations: 2
```

Additional Problem

```
Call:
glm(formula = cbind(dta2$Yes, dta2$No) ~ Gender + factor(Department),
    family = binomial(), data = dta2)

Deviance Residuals:
      1        2        3        4        5        6        7        8
 3.7189  -1.2487   0.2706  -0.0560  -0.9243   1.2533  -0.0858   0.0826
      9       10       11       12
-0.8509   1.2205   0.2052  -0.2076

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.68192    0.09911   6.880 5.97e-12 ***
GenderMale          -0.09987    0.08085  -1.235    0.217
factor(Department)2 -0.04340    0.10984  -0.395    0.693
factor(Department)3 -1.26260    0.10663 -11.841  < 2e-16 ***
factor(Department)4 -1.29461    0.10582 -12.234  < 2e-16 ***
factor(Department)5 -1.73931    0.12611 -13.792  < 2e-16 ***
factor(Department)6 -3.30648    0.16998 -19.452  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 877.056  on 11  degrees of freedom
Residual deviance:  20.204  on  5  degrees of freedom
AIC: 103.14

Number of Fisher Scoring iterations: 4
```

a)

```r
## Conditional Odds Ratios
exp(coef(mdl))
```

```
        (Intercept)          GenderMale factor(Department)2
         1.97767415          0.90495497          0.95753028
factor(Department)3 factor(Department)4 factor(Department)5
         0.28291804          0.27400567          0.17564230
factor(Department)6
         0.03664494
```

```r
## 95% Confidence Interval
exp(confint(mdl))
```

```
                         2.5 %     97.5 %
(Intercept)         1.62982591 2.40389206
GenderMale          0.77196157 1.05989382
factor(Department)2 0.77234827 1.18813231
factor(Department)3 0.22930959 0.34833301
factor(Department)4 0.22242464 0.33680469
factor(Department)5 0.13685898 0.22441389
factor(Department)6 0.02596498 0.05061965
```

b)

Breslow Day: $Ho : \theta_1 = 1, Ha : \theta_1 \neq 1$. Reject the null and conclude that the odds ratio between male and female are significantly different

```r
anova(mdl, test = "Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(dta2$Yes, dta2$No)

Terms added sequentially (first to last)

                   Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                  11     877.06
Gender              1    93.45        10     783.61 < 2.2e-16 ***
factor(Department)  5   763.40         5      20.20 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CMH: $Ho: \beta_1 = 0, Ha: \beta_1 \neq 0$ Conditional on department there is insufficient evidence to show that the odds ratio between men and women are different.

c) There is insufficient evidence to suggest that the common odds ratio is sufficient.

```r
dta2 = dta2[3:12,]

mdl = glm(cbind(dta2$Yes, dta2$No) ~ Gender + factor(Department),
          family = binomial(), data = dta2)
summary(mdl)
```

```

Call:
glm(formula = cbind(dta2$Yes, dta2$No) ~ Gender + factor(Department),
    family = binomial(), data = dta2)

Deviance Residuals:
      3        4        5        6        7        8        9       10
 0.5680  -0.1191  -0.3914   0.5239   0.5440  -0.5164  -0.4892   0.6868
     11       12
 0.5158  -0.5024

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.51349    0.11936   4.302 1.69e-05 ***
GenderMale           0.03069    0.08676   0.354    0.724
factor(Department)3 -1.14008    0.12188  -9.354  < 2e-16 ***
factor(Department)4 -1.19456    0.11984  -9.968  < 2e-16 ***
factor(Department)5 -1.61308    0.13928 -11.581  < 2e-16 ***
factor(Department)6 -3.20527    0.17880 -17.927  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 539.4581  on 9  degrees of freedom
Residual deviance:   2.5564  on 4  degrees of freedom
AIC: 71.791

Number of Fisher Scoring iterations: 3
```

```r
exp(confint(mdl))
```

```
Waiting for profiling to be done...

                        2.5 %     97.5 %
```

```
(Intercept)          1.3235598 2.11348814
GenderMale           0.8696401 1.22201575
factor(Department)3 0.2515380 0.40564777
factor(Department)4 0.2391342 0.38257688
factor(Department)5 0.1513330 0.26129574
factor(Department)6 0.0282632 0.05703262
```

4.34

When x = 0, $log(\frac{\pi(x)}{1-\pi(x)}) = \alpha \rightarrow \pi(x) = \frac{exp(\alpha)}{1+exp(\alpha)}$
When x = 1, $log(\frac{\pi(x)}{1-\pi(x)}) = \alpha + \beta(1) \rightarrow \pi(x) = \frac{exp(\alpha+\beta(1))}{1+exp(\alpha+\beta(1))}$
When x = 2, $log(\frac{\pi(x)}{1-\pi(x)}) = \alpha + \beta(2) \rightarrow \pi(x) = \frac{exp(\alpha+\beta(2))}{1+exp(\alpha+\beta(2))}$
When x = 3, $log(\frac{\pi(x)}{1-\pi(x)}) = \alpha + \beta(3) \rightarrow \pi(x) = \frac{exp(\alpha+\beta(3))}{1+exp(\alpha+\beta(3))}$

4.35

a) $exp[-(\alpha + \beta(-\alpha/\beta))^2/2] = 0 \rightarrow exp[0] = 1 \rightarrow \pi(x) = .5 = .4\beta$
b) $7.502/.302 = 24.8$
c) $.4 * .5 * (1 - .4) = .12$