Homework 7

Piotr Mankowski

March 2, 2018

1.

Suppose we are interested in evaluating associations with all-cause mortality using a logistic regression model. Perform a logistic regression analysis with indicator of death within 5 years of study enrollment as the response and with creatinine, age, indicator for ever smoked, and indicator of white race as predictors.

a.

Provide an interpretation of the exponentiated intercept term in the logistic regression model.

The exponentiated intercept in our logistic regression model is $2.13e-04$. This number represents the estimated odds of being dead at 5 years after the study for individuals with 0 values for the covariates of creatinine, age, smoking status, and an indicator for being white - in other words, a newborn, non-smoking, non-white individual with 0 mg/dl of creatinine. The interpretation of estimate is not scientifically useful.

b.

Provide an interpretation of the exponentiated age slope in the logistic regression model.

The exponentiated age slope for our model, with a value of 1.07, represents the odds ratio of the odds of 5-year mortality for individuals differing by 1 year of age, but with all other co-variates kept constant. 1-year older individuals are expected to have a 7% higher odds of mortality.

c.

From the logistic regression model, is there evidence of an association between death within 5 years of study enrollment and creatinine after adjusting for the other predictors? Give full inference.

Methods: To evaluate an association between 5 year all-cause mortality and creatinine, we fit a multivariate logistic regression model to estimate the odds ratio of death between groups differing by $1 mg/dl$ difference in serum creatinine, while adjusting for age, smoking status, and sex in our analysis. We report the exponentiated slope from this model as an estimate of this odds ratio. Also, we report a 95% confidence interval and the p-value for the hypothesis that the creatinine slope is equal to zero, and the exponentiated slope (creatinine odds ratio) equals 1.**

Results: We estimate the odds ratio of death for groups differing in serum creatinine level by 1 mg/dl but homogonous in age, smoking status, and white race status to be 5.22, with a 95% CI of $(2.68, 10.16)$. Individuals with creatinine levels 1 mg/dl higher but homogonous in the other covariates have an estimated 422% higher odds of 5-year mortality, and these results would not be suprising if the true value was between 168% and 920%. At the $\alpha = 0.05$ level, we reject the null hypothesis of no association between mortality with a p-value of 0.00005.

d.

From the logistic regression model, is there evidence of an association between death within 5 years of study enrollment and age after adjusting for the other predictors? Provide full inference.

Methods: We fit a multivariate logistic regression model with 5-year all-cause mortality as the response and age, sex, serum creatinine levels, smoking status, and race status as predictors. As an estimate of the odds for mortality in groups differing in age at time of study by 1 year, we report the exponentiated age slope from this model, a 95% confidence interval, and p-value for the hypothesis that the age slope is equal to zero, and the exponentiated slope (creatinine odds ratio) equals 1.**

Results: We estimate the odds ratio of death for groups differing in age by one year but homogonous in serum creatinine, smoking status, and white race status to be 1.07. Individuals 1 year apart in age but homogonous in creatinine, race status, and smoking status have an estimated 7% higher odds of 5-year mortality, and these results would not be suprising if the true value were between 3% and 11%. At the $\alpha = 0.05$ level, we reject the null hypothesis of no association between mortality with a p-value of 0.0002.

e.

From the logistic regression model, what is the best estimate of the odds of dying within 5 years of study enrollment for a non-white race individual who is 76 years old, has previously smoked, and has a creatinine level of 0.95.

The estimated odds of dying within 5 years for this type of individual is 0.239.

f.

From the logistic regression model, what is the best estimate of the probability of dying within 5 years of study enrollment for a white race individual who is 69 years old, has never smoked, and has a creatinine level of 1.2.

The estiamated odds for such an idividual are 0.113. These odds correspond to a probability of mortality of $p = o/1 + o = 0.113/1.113 = .102$

2.

Now perform a logistic regression analysis with indicator of death within 5 years of study enrollment as the response and creatinine, age, indicator of ever smoked, indicator of white race, ***and sex*** *as predictors.*

a.

Provide an interpretation of the exponentiated intercept term in the logistic regression model.

The intercept from this model of $2.28e-04$ represents odds of 5-year mortality of an individual who is a white non-smoking newborn with a serum creatinine level of $0mg/dl$ - a purely hypothetical type of individual. The interpretation is not useful scientifically.

b.

Provide an interpretation of the exponentiated sex slope in the logistic regression model. The exponentiated sex slope of 1.31 represents the estimated odds ratio between populations of males and females consistant in smoking status, creatinine serum levels, white race indicator, and age. The males are expected to have 31% higher odds of death than the females.

c.

Provide full inference for an association between all-cause mortality within 5 years and sex using the logistic regression model.

Methods: We examined the association between all-cause 5-year-mortality and sex while adjusting for serum creatinine level, age, smoking status, and white status by fitting a logistic regression with 5-year mortality as the response and the sex, creatinine, age, smoking status, and white status as predictors. We report the exponentiated slope for sex to estimate the odds ratio between populations of males and females where the other co-variates are held constant, and compute a 95% confidence interval for this slope estimate and the p-value from a t-test for the hypothesis that the slope is equal to zero (and the exponentiated slope is 1).**

Results: We estimate that the odds ratio between males and females homogenous for age, creatinine level, smoking status, and white race status is 1.31. This suggests the males have 31% higher odds of death within 5 years of the study than the females when the other covariates are adjusted for. This estimate would not be suprising if the true odds ratio was within the 95% CI of $(0.836, 2.05)$, or the odds for males were between 16.4% lower and 105% higher than for the group of females. At the 0.05 level, we cannot reject the null hypothesis that the adjusted odds ratio between males and females is 1, since the p-value of the t-test is 0.240.

d.

Provide full inference for an association between all-cause mortality within 5 year years and creatinine using the logistic regression.

Methods: We examined the association between all-cause 5-year-mortality and serum creatinine levels while adjusting for sex, age, smoking status, and white race status by fitting a logistic regression with serum creatinine level as the response and sex, age, smoking status, and white race status as predictors. We report the exponentiated slope for sex as an estimate of the odds ratio between populations of individuals differing by $1mg/dl$ in serum creatinine but homogenous in the other variables. We compute a 95% confidence interval for this estimate, and report the p-value from a t-test for the hypothesis that the slope is equal to zero - and the exponentiated slope is 1 - signifying a lack of association.

Results: We estimate that the odds ratio between individuals $1mg/dl$ apart in creatinine level but homogenous for age, sex, smoking status, and white race status is 4.46. This suggests higher-creatinine individuals are expected to have 346% higher odds of death within 5 years of the study when the other covariates are adjusted for. This estimate would not be suprising if the true odds ratio was within the 95% CI of $(2.21, 9.03)$, or the odds the higher-creatinine populations were between 121% and 803% higher. At the 0.05 level, we can reject the null hypothesis that the adjusted odds ratio is 1 and find evidence for a statistically significant association since the p-value of the t-test for this hypothesis is < 0.00005.

e.

Is sex a confounder, precision variable, both or neither for the associations between all-cause mortality within 5 years of study enrollment and each of the other four predictors of interest: creatinine, age, indicator of ever smoked, and indicator of white race? Explain and provide evidence to support your reasoning.

Mortality vs Creatinine, unadjusted and adjusted for sex

For sex to confound the association between 5-year all cause mortality and serum creatinine levels, we need evidence of an association between sex and 5-year mortality (the outcome), and an association with creatinine levels, the exposure variable. This would mean that the association between creatinine levels and mortality should be different before and after controlling for sex. Linear regression on creatinine given sex suggest and association, and the unadjusted logistic regression of creatinine on mortality also suggests an association. In addition, the odds ratio estimates for the adjusted and unadjusted analysis - 4.92 and 5.99 respectively, suggesting the association between mortality and creatinine levels is different when adjusting for sex. This suggests sex is a confounding variable. It is not a precision variable, since sex is associated with creatine, the predictor of interest.

Mortality vs age, unadjusted and adjusted for sex

For sex to confound the association between 5-year all-cause mortality and age, we need evidence of an association between sex and 5-year mortality (the outcome), and an association between sex and age, the exposure variable. This would mean that the association between age and mortality should be different before and after controlling for sex.

Linear regression on age given sex suggest no association; additionally, the age slope estimates are almost identical between the adjusted and non-adjusted analyses, suggesting no evidence of confounding.

For sex to be a potential precision variable for this association, it should be associated with the outcome (mortality), and not the predictor (age). In addition, adjusting for sex should decrease the variance of the outcome. After running linear regression of age on sex, logistic regression of mortality on sex, and the adjusted and un-adjusted regression analysis of mortality vs age, we find evidence for all of these criteria, except the increase of precision between unadjusted and adjusted analyses.

Mortality vs smoking status, unadjusted and adjusted for sex

For sex to confound the association between 5-year all-cause mortality and age, we need evidence of an association between sex and 5-year mortality (the outcome), and an association between sex and smoking status, the exposure variable. This would mean that the association between age and mortality should be different before and after controlling for sex.

Logistic regression of sex on smoking status shows evidence for an association, and we previously showed evidence of a significant association between sex mortality. In addition, the estimated exponentiated slope for smoking status is 1.36 in the unadjusted case, and 1.19 when adjusting for sex. This evidence suggests sex

confounds the mortality-smoking status association. Since sex is associated with the POI smoking status, it is not a precision variable.

Mortality vs indicator of white race, unadjusted and adjusted for sex

For sex to confound the association between 5-year all-cause mortality and white race, we need evidence of an association between sex and 5-year mortality (the outcome), and an association between sex and the POI, white race indicator. After performing logistic regression of sex on the white race indicator, we do not find evidence of a significant association, suggesting sex is not a confounder.

For sex to be a potential precision variable for this association, it should be associated with the outcome (mortality), and not the predictor (white race). We find evidence for both using logistic regression. In addition, adjusting for sex should decrease the variance of the outcome, but not change the estimated slope. After running logistic regression of sex on white race, logistic regression of mortality on sex, and the adjusted and un-adjusted regression analysis of mortality vs white race, we find evidence of these criteria. The unadjusted slope of 0.681 is close to the adjusted slope 0.671, the 95% CIs are a bit smaller in the adjusted analysis, and there is an association between sex and mortality, but not sex and white race, suggesting sex might be a precision variable for the mortality vs. white race indicator association.

3.

Now suppose we are interested in evaluating associations with all-cause mortality using a Poisson regression analysis. Perform a Poisson regression analysis with indicator of death within 5 years as the response and creatinine, age, indicator of ever smoked, indicator of white race, and sex as predictors.

a.

Provide an interpretation of the exponentiated intercept term of the Poisson regression model.

The intercept in the Poisson regression model is the log probability of death from any cause within 5 years for newborn non-white males who hever smoked and have a serum creatinine level of 0 mg/dl. Exponentiating this intercept, we estimate the risk of death to be 0.00142 for this hypothetical group, which is not meaningful scientifically.

b.

Provide an interpretation of the exponentiated creatinine slope in the Poisson regression model.

The slope in the Poisson regression model is the difference in log probabilities of 5-year all-cause mortality between subjects differing by $1mg/dl$ in serum creatinine levels, but homogonous in sex, age, smoking status, and white status. The exponentiated slope of 2.12 represents the relative risk between such groups, and suggests individuals with the higher creatinine but constant in the other covariates are 112% more likely to die witin 5 years.

c.

Provide full inference with the Poisson regression model for an association between all-cause mortality within 5 years of enrollment and creatinine after adjusting for the other predictors.

Methods: We fit a multivariate Poisson regression with 5-year all-cause mortality as the response and the variables of creatinine serum concentration, age, sex, smoking status, and race as predictors. We report the slope for creatinine to estimate the relative risk of death for groups differing by $1mg/dl$ in creatinine but homogenous in age, sex, race status, and smoking status. We compute a 95% CI and p-value from the hypothesis that the relative risk is equal to 1.

Results: We estimate that for individuals differing in creatinine levels by $1mg/dl$ but homogenous in age, white status, sex, and smoking status, those with higher creatinine have a 111% higher relative risk of death within 5 years; these results would not be suprising if the true relative risk was within the 95% confience interval of 51% to 196% This finding is statistically significant at the 0.05 level, with a p value of < 0.00005.

d.

Provide full inference with the Poisson regression model for an association between all-cause mortality within 5 years of enrollment and sex after adjusting for the other predictors.

Methods: We fit a multivariate Poisson regression with 5-year all-cause mortality as the response and the variables of creatinine serum concentration, age, sex, smoking status, and race as predictors. We report the slope for sex to estimate the relative risk of death between males and females otherwise homogenous in age, creatinine level, race status, and smoking status. We compute a 95% CI and p-value from the hypothesis that the relative risk is equal to 1, and no adjusted association between mortality and sex exists.

Results: We estimate the exponentiated sex slope to 1.35, suggesting males have a 35% higher relative risk of mortality than females of the same age, white status, smoking status, and creatinine level. These results would not be suprising if the true relative risk was within the 95% confience interval of -5.58% to 94%. This finding has a p-value of 0.0995 and is not statistically significant at the 0.05 level; we cannot reject the null hypothesis that the exponentiated sex slope is equal to 1.

Appendix

Setup code

```
knitr::opts_chunk$set(echo = FALSE, results = 'hide', warning = FALSE)

require(uwIntroStats)
library(ggplot2)
library(data.table)

mri <- as.data.table(read.table("../data/mri.txt", header = TRUE))
mri[, obstime_yr:=obstime/365]
mri[, dead_at_5yr:=ifelse((obstime_yr <= 5.0 & death == 1), 1, 0), by='ptid']
mri[, white:=as.integer(race == 1)]
mri[, smoked:=as.integer(packyrs > 0)]
```

Question 1 code

```
q1data <- mri[!is.na(crt)]
mod1 <- regress("odds", dead_at_5yr ~ crt + age + smoked + white, data = q1data)
mod1
```

```
## ( 1  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ crt + age + smoked +
##     white, data = q1data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4044  -0.6060  -0.4969  -0.3791   2.4595
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -8.456     1.433     1.428             35.06 1
## [2] crt            1.652     0.3439    0.3394            23.68 1
## [3] age           0.06787   0.01805   0.01832            13.73 1
## [4] smoked         0.2961    0.2204    0.2216             1.79 1
## [5] white         -0.3874    0.2357    0.2362             2.69 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0002
## [4] smoked         0.1819
## [5] white          0.1014
##
## Transformed Model:
##                   e(Est)     e(95%L)    e(95%H)          F stat    df
## [1] Intercept     2.125e-04  1.288e-05  3.508e-03           35.06 1
```

```
## [2] crt             5.217      2.679      10.16             23.68 1
## [3] age             1.070      1.032      1.109             13.73 1
## [4] smoked          1.345     0.8703      2.077              1.79 1
## [5] white          0.6788     0.4270      1.079              2.69 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0002
## [4] smoked         0.1819
## [5] white          0.1014
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 653.14  on 731  degrees of freedom
## Residual deviance: 603.35  on 727  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 613.35
##
## Number of Fisher Scoring iterations: 4
```

```
testC <- c(0, 1, 0, 0, 0)
lincom(mod1, testC)
```

```
##
## H0: 1*crt   =  0
## Ha: 1*crt  !=  0
##       e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,]  5.2167    0.3394  2.6792 10.1578 4.867    1.39e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
testC <- c(0, 0, 1, 0, 0)
lincom(mod1, testC)
```

```
##
## H0: 1*age   =  0
## Ha: 1*age  !=  0
##       e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,] 1.07023   0.01832 1.03243 1.10942 3.706    0.000227 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
testC <- c(1, 0.95, 76, 1, 0)
lincom(mod1, testC)
```

```
##
## H0: 0.95*(Intercept)+76*crt+1*age+0.95*smoked   =  0
## Ha: 0.95*(Intercept)+76*crt+1*age+0.95*smoked  !=  0
##      e(Est) Std. Err. e(95%L) e(95%H)      T Pr(T > |t|)
## [1,] 0.2387    0.2341  0.1508  0.3779 -6.121    1.52e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
testC <- c(1, 1.2, 69, 0, 1)
lincom(mod1, testC)
```

```
##
```

```
## H0: 1.2*(Intercept)+69*crt+1*age+1.2*white   =  0
## Ha: 1.2*(Intercept)+69*crt+1*age+1.2*white  !=  0
##       e(Est) Std. Err. e(95%L) e(95%H)      T Pr(T > |t|)
## [1,] 0.11324   0.24749 0.06966 0.18409 -8.801      <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 2 code

```
mod2 <- regress("odds", dead_at_5yr ~ crt + age + smoked + white + male, data = mri)
mod2
```

```
## ( 3  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ crt + age + smoked +
##     white + male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3734  -0.6178  -0.5013  -0.3799   2.4908
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -8.388     1.435     1.444            33.74 1
## [2] crt            1.497     0.3647    0.3583           17.45 1
## [3] age           0.06753   0.01806   0.01850           13.33 1
## [4] smoked         0.2606    0.2225    0.2256            1.33 1
## [5] white         -0.3903    0.2361    0.2375            2.70 1
## [6] male           0.2680    0.2320    0.2279            1.38 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0003
## [4] smoked         0.2485
## [5] white          0.1007
## [6] male           0.2401
##
## Transformed Model:
##                  e(Est)     e(95%L)    e(95%H)          F stat    df
## [1] Intercept     2.277e-04  1.337e-05  3.876e-03           33.74 1
## [2] crt             4.469      2.211      9.030            17.45 1
## [3] age             1.070      1.032      1.109            13.33 1
## [4] smoked          1.298     0.8333      2.021             1.33 1
## [5] white          0.6769     0.4247      1.079             2.70 1
## [6] male            1.307     0.8357      2.045             1.38 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0003
```

```
## [4] smoked         0.2485
## [5] white          0.1007
## [6] male           0.2401
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 653.14  on 731  degrees of freedom
## Residual deviance: 602.02  on 726  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 614.02
##
## Number of Fisher Scoring iterations: 5
```

```
testC <- c(0, 0, 0, 0, 0, 1)
lincom(mod2, testC)
```

```
##
## H0: 1*male   =  0
## Ha: 1*male  !=  0
##      e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,] 1.3073    0.2279  0.8357  2.0450 1.176        0.24
```

```
testC <- c(0, 1, 0, 0, 0, 0)
lincom(mod2, testC)
```

```
##
## H0: 1*crt   =  0
## Ha: 1*crt  !=  0
##      e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,] 4.4685    0.3583  2.2113  9.0299 4.178     3.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
regress("mean", crt ~ male, data=mri)
```

```
## ( 2  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "mean", formula = crt ~ male, data = mri)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4986 -0.1297 -0.0297  0.1014  2.8014
##
## Coefficients:
##                  Estimate  Naive SE  Robust SE    95%L      95%H
## [1] Intercept      0.9297   0.01418   0.01361       0.9030    0.9564
## [2] male           0.2689   0.02007   0.02007       0.2295    0.3083
##                     F stat    df Pr(>F)
## [1] Intercept         4665.22 1  < 0.00005
## [2] male               179.59 1  < 0.00005
##
## Residual standard error: 0.2716 on 731 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1973, Adjusted R-squared:  0.1962
```

```
## F-statistic: 179.6 on 1 and 731 DF,  p-value: < 2.2e-16
```

```
# [2] male           0.2689   0.02007   0.02007       0.2295    0.3083        179.59 1  < 0.00005

unadjusted <- regress("odds", dead_at_5yr ~ crt, data = mri)
# [2] crt            5.986     3.116     11.50             28.97 1  < 0.00005

adjusted <- regress("odds", dead_at_5yr ~ crt + male, data = mri)
# [2] crt            4.915     2.467     9.791             20.57 1  < 0.00005
# [3] male           1.364     0.8814    2.111              1.95 1    0.1633


unadjusted
```

```
## ( 2  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ crt, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0917  -0.5964  -0.5492  -0.4643   2.2837
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -3.605     0.4012    0.4008            80.89 1
## [2] crt            1.789     0.3398    0.3325            28.97 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
##
## Transformed Model:
##                   e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept     0.02719   0.01238   0.05973            80.89 1
## [2] crt           5.986     3.116     11.50              28.97 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 656.75  on 732  degrees of freedom
## Residual deviance: 624.70  on 731  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 628.7
##
## Number of Fisher Scoring iterations: 4
```

```
adjusted
```

```
## ( 2  cases deleted due to missing values)
##
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ crt + male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0178  -0.6340  -0.4736  -0.4069   2.3786
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -3.564     0.4001    0.4029            78.27 1
## [2] crt            1.592     0.3637    0.3511            20.57 1
## [3] male           0.3104    0.2271    0.2224             1.95 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] male           0.1633
##
## Transformed Model:
##                  e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept     0.02832   0.01284   0.06246            78.27 1
## [2] crt            4.915     2.467     9.791            20.57 1
## [3] male           1.364     0.8814    2.111             1.95 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] male           0.1633
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 656.75  on 732  degrees of freedom
## Residual deviance: 622.82  on 730  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 628.82
##
## Number of Fisher Scoring iterations: 4
```

```
regress("mean", age ~ male, data=mri)
```

```
##
## Call:
## regress(fnctl = "mean", formula = age ~ male, data = mri)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4065 -3.7268 -0.7268  3.2732 24.2732
##
## Coefficients:
##                  Estimate  Naive SE  Robust SE    95%L      95%H
## [1] Intercept      74.41     0.2839    0.2737       73.87     74.94
## [2] male           0.3203    0.4023    0.4024      -0.4697    1.110
##                     F stat    df Pr(>F)
## [1] Intercept        73899.43 1  < 0.00005
```

```
## [2] male                 0.63 1    0.4263
##
## Residual standard error: 5.453 on 733 degrees of freedom
## Multiple R-squared:  0.0008641,  Adjusted R-squared:  -0.000499
## F-statistic: 0.6335 on 1 and 733 DF,  p-value: 0.4263
# [2] male           0.3203    0.4023    0.4024      -0.4697    1.110           0.63 1    0.4263

regress("odds", dead_at_5yr ~ male, data=mri)

##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6924  -0.6924  -0.4978  -0.4978   2.0734
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -2.026     0.1622    0.1625           155.46 1
## [2] male           0.7194    0.2064    0.2067            12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## Transformed Model:
##                  e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1319   0.09588    0.1815           155.46 1
## [2] male           2.053     1.368     3.081             12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 644.86  on 733  degrees of freedom
## AIC: 648.86
##
## Number of Fisher Scoring iterations: 4
# [2] male           2.053     1.368     3.081             12.11 1     5e-04

unadjusted <- regress("odds", dead_at_5yr ~ age, data = mri)
# [2] age             1.073      1.037     1.111              16.15 1     1e-04

adjusted <- regress("odds", dead_at_5yr ~ age + male, data = mri)
# [2] age             1.072      1.035     1.111              14.77 1     1e-04
# [3] male            2.029      1.349     3.054              11.56 1     7e-04


unadjusted
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ age, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1997  -0.6341  -0.5391  -0.4722   2.1507
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -6.942     1.301     1.339            26.90 1
## [2] age           0.07065   0.01707   0.01758           16.15 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] age             1e-04
##
## Transformed Model:
##                  e(Est)     e(95%L)    e(95%H)         F stat    df
## [1] Intercept     9.664e-04  6.981e-05  0.01338            26.90 1
## [2] age             1.073      1.037     1.111             16.15 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] age             1e-04
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 640.75  on 733  degrees of freedom
## AIC: 644.75
##
## Number of Fisher Scoring iterations: 4
```

adjusted

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ age + male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3254  -0.6438  -0.5489  -0.3926   2.3091
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -7.258     1.320     1.361            28.43 1
## [2] age           0.06961   0.01725   0.01811           14.77 1
## [3] male           0.7077    0.2087    0.2082           11.56 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] age             1e-04
## [3] male            7e-04
```

```
##
## Transformed Model:
##                  e(Est)     e(95%L)    e(95%H)         F stat    df
## [1] Intercept     7.046e-04  4.869e-05  0.01020            28.43 1
## [2] age             1.072      1.035     1.111             14.77 1
## [3] male            2.029      1.349     3.054             11.56 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] age             1e-04
## [3] male            7e-04
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 628.84  on 732  degrees of freedom
## AIC: 634.84
##
## Number of Fisher Scoring iterations: 4
```

```
regress("odds", smoked ~ male, data=mri)
```

```
## ( 1  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "odds", formula = smoked ~ male, data = mri)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.480  -1.111   0.902   0.902   1.245
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -0.1575    0.1044    0.1046             2.27 1
## [2] male           0.8465    0.1524    0.1526            30.78 1
##                  Pr(>F)
## [1] Intercept      0.1325
## [2] male         < 0.00005
##
## Transformed Model:
##                  e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.8543    0.6957    1.049              2.27 1
## [2] male           2.332     1.728     3.146             30.78 1
##                  Pr(>F)
## [1] Intercept      0.1325
## [2] male         < 0.00005
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1005.98  on 733  degrees of freedom
## Residual deviance:  974.38  on 732  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 978.38
```

```
##
## Number of Fisher Scoring iterations: 4
# [2] male           2.332     1.728     3.146            30.78 1  < 0.00005

regress("odds", dead_at_5yr ~ male, data=mri)
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6924  -0.6924  -0.4978  -0.4978   2.0734
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -2.026     0.1622    0.1625           155.46 1
## [2] male           0.7194    0.2064    0.2067            12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## Transformed Model:
##                  e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1319   0.09588    0.1815           155.46 1
## [2] male           2.053     1.368     3.081             12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 644.86  on 733  degrees of freedom
## AIC: 648.86
##
## Number of Fisher Scoring iterations: 4
```

```
# [2] male           2.053     1.368     3.081            12.11 1     5e-04

unadjusted <- regress("odds", dead_at_5yr ~ smoked, data = mri)
# [2] smoked         1.361     0.9090    2.038             2.25 1    0.1342

adjusted <- regress("odds", dead_at_5yr ~ smoked + male, data = mri)
# [2] smoked         1.188     0.7841    1.801             0.66 1    0.4154
# [3] male           1.958     1.291     2.970            10.04 1    0.0016


unadjusted
```

```
## ( 1  cases deleted due to missing values)
##
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ smoked, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5496  -0.5496  -0.5496   1.8471   1.9823
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -1.814     0.1608    0.1610           126.93 1
## [2] smoked         0.3082    0.2053    0.2056             2.25 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] smoked         0.1342
##
## Transformed Model:
##                   e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1630    0.1189    0.2236           126.93 1
## [2] smoked         1.361     0.9090    2.038              2.25 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] smoked         0.1342
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 653.86  on 733  degrees of freedom
## Residual deviance: 651.57  on 732  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 655.57
##
## Number of Fisher Scoring iterations: 4
```

adjusted

```
## ( 1  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ smoked + male,
##     data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7056  -0.7056  -0.6533  -0.5193   2.1084
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -2.108     0.1928    0.1920           120.51 1
## [2] smoked         0.1726    0.2107    0.2118             0.66 1
## [3] male           0.6721    0.2109    0.2121            10.04 1
```

```
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] smoked         0.4154
## [3] male           0.0016
##
## Transformed Model:
##                   e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1215   0.08332    0.1771           120.51 1
## [2] smoked         1.188     0.7841    1.801              0.66 1
## [3] male           1.958     1.291     2.970             10.04 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] smoked         0.4154
## [3] male           0.0016
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 653.86  on 733  degrees of freedom
## Residual deviance: 641.09  on 731  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 647.09
##
## Number of Fisher Scoring iterations: 4
```

```
regress("mean", white ~ male, data=mri)
```

```
##
## Call:
## regress(fnctl = "mean", formula = white ~ male, data = mri)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7814  0.2186  0.2186  0.2249  0.2249
##
## Coefficients:
##                   Estimate   Naive SE  Robust SE    95%L      95%H
## [1] Intercept      0.7751    0.02166   0.02177       0.7323    0.8178
## [2] male          6.353e-03  0.03069   0.03069     -0.05389   0.06660
##                     F stat    df Pr(>F)
## [1] Intercept         1268.03 1  < 0.00005
## [2] male                 0.04 1     0.836
##
## Residual standard error: 0.416 on 733 degrees of freedom
## Multiple R-squared:  5.846e-05,  Adjusted R-squared:  -0.001306
## F-statistic: 0.04286 on 1 and 733 DF,  p-value: 0.836
```

```
# [2] male          6.353e-03  0.03069   0.03069     -0.05389   0.06660          0.04 1     0.836
```

```
regress("odds", dead_at_5yr ~ male, data=mri)
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ male, data = mri)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -0.6924  -0.6924  -0.4978  -0.4978   2.0734
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -2.026     0.1622    0.1625           155.46 1
## [2] male           0.7194    0.2064    0.2067            12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## Transformed Model:
##                  e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1319   0.09588    0.1815           155.46 1
## [2] male           2.053     1.368     3.081             12.11 1
##                  Pr(>F)
## [1] Intercept    < 0.00005
## [2] male            5e-04
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 644.86  on 733  degrees of freedom
## AIC: 648.86
##
## Number of Fisher Scoring iterations: 4
```

```
# [2] male           2.053     1.368     3.081             12.11 1     5e-04

unadjusted <- regress("odds", dead_at_5yr ~ white, data = mri)
# [2] white          0.6806    0.4371    1.060              2.91 1    0.0884

adjusted <- regress("odds", dead_at_5yr ~ white + male, data = mri)
# [2] white          0.6712    0.4276    1.054              3.01 1    0.0831
# [3] male           2.065     1.377     3.098             12.32 1    0.0005


unadjusted
```

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ white, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6840  -0.5744  -0.5744  -0.5744   1.9407
##
## Coefficients:
##
## Raw Model:
##                  Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -1.333     0.1928    0.1930           47.71 1
## [2] white         -0.3848    0.2252    0.2255            2.91 1
```

```
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] white          0.0884
##
## Transformed Model:
##                   e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.2636    0.1804    0.3850            47.71 1
## [2] white          0.6806    0.4371    1.060              2.91 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] white          0.0884
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 654.66  on 733  degrees of freedom
## AIC: 658.66
##
## Number of Fisher Scoring iterations: 4
```

adjusted

```
##
## Call:
## regress(fnctl = "odds", formula = dead_at_5yr ~ white + male,
##     data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7907  -0.6636  -0.5719  -0.4747   1.9449
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -1.728     0.2306    0.2420            50.97 1
## [2] white         -0.3987    0.2275    0.2297             3.01 1
## [3] male           0.7252    0.2069    0.2066            12.32 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] white          0.0831
## [3] male           0.0005
##
## Transformed Model:
##                   e(Est)    e(95%L)   e(95%H)         F stat    df
## [1] Intercept      0.1777    0.1105    0.2858            50.97 1
## [2] white          0.6712    0.4276    1.054              3.01 1
## [3] male           2.065     1.377     3.098             12.32 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] white          0.0831
## [3] male           0.0005
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 657.47  on 734  degrees of freedom
## Residual deviance: 641.90  on 732  degrees of freedom
## AIC: 647.9
##
## Number of Fisher Scoring iterations: 4
```

Question 3 code

```
mod3 <- regress('rate', dead_at_5yr ~ crt + age + smoked + white + male, data = mri)
mod3
```

```
## ( 3  cases deleted due to missing values)
##
##
## Call:
## regress(fnctl = "rate", formula = dead_at_5yr ~ crt + age + smoked +
##     white + male, data = mri)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0895  -0.5787  -0.4925  -0.3910   1.9752
##
## Coefficients:
##
## Raw Model:
##                   Estimate  Naive SE  Robust SE       F stat    df
## [1] Intercept     -6.560     1.183     1.026             40.86 1
## [2] crt            0.7488    0.1923    0.1704            19.30 1
## [3] age           0.05044   0.01524   0.01344            14.09 1
## [4] smoked         0.2698    0.1973    0.1813             2.21 1
## [5] white         -0.3143    0.2056    0.1797             3.06 1
## [6] male           0.3018    0.2031    0.1829             2.72 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0002
## [4] smoked         0.1372
## [5] white          0.0807
## [6] male           0.0995
##
## Transformed Model:
##                   e(Est)     e(95%L)    e(95%H)         F stat    df
## [1] Intercept     1.416e-03  1.889e-04  0.01062            40.86 1
## [2] crt             2.115      1.513     2.955             19.30 1
## [3] age             1.052      1.024     1.080             14.09 1
## [4] smoked          1.310     0.9175     1.869              2.21 1
## [5] white          0.7303     0.5132     1.039              3.06 1
## [6] male            1.352     0.9442     1.937              2.72 1
##                   Pr(>F)
## [1] Intercept    < 0.00005
## [2] crt          < 0.00005
## [3] age            0.0002
## [4] smoked         0.1372
```

```
## [5] white          0.0807
## [6] male           0.0995
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 433.99  on 731  degrees of freedom
## Residual deviance: 396.57  on 726  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 648.57
##
## Number of Fisher Scoring iterations: 6
```

```
lincom(mod3, c(0,1,0,0,0,0))
```

```
##
## H0: 1*crt   =  0
## Ha: 1*crt  !=  0
##      e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,] 2.1146    0.1704  1.5132  2.9548 4.394    1.28e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lincom(mod3, c(0,0,0,0,0,1))
```

```
##
## H0: 1*male   =  0
## Ha: 1*male  !=  0
##      e(Est) Std. Err. e(95%L) e(95%H)     T Pr(T > |t|)
## [1,] 1.3523    0.1829  0.9442  1.9366 1.649      0.0995 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```