stat431 assignment02

Yiming Shen 20891774

20/10/2023

Question 1

(a)

```
y <- rep(c(0,1,0,1,0,1,0,1), c(78,22,46,54,71,40,20,60))
x1 <- rep(c(0,0,1,1,0,0,1,1), c(78,22,46,54,71,40,20,60))
x2 <- rep(c(0,0,0,0,1,1,1,1), c(78,22,46,54,71,40,20,60))
model <- glm(y ~ x1 + x2 + x1*x2, family = binomial(link = logit))
summary(model)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x1 * x2, family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6651  -0.9454  -0.7049   1.1101   1.7402
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2657     0.2414  -5.243 1.58e-07 ***
## x1            1.4260     0.3139   4.543 5.55e-06 ***
## x2            0.6919     0.3120   2.217   0.0266 *
## x1:x2         0.2464     0.4520   0.545   0.5856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 538.14  on 390  degrees of freedom
## Residual deviance: 478.45  on 387  degrees of freedom
## AIC: 486.45
##
## Number of Fisher Scoring iterations: 4
```

```
c <- 1.96
# beta0 95%
beta0_L <- -1.2657 - c*(0.2414)
beta0_U <- -1.2657 + c*(0.2414)
beta0_L
```

```
## [1] -1.738844
```

```r
beta0_U
```

```
## [1] -0.792556
```

```r
# beta1 95%
beta1_L <- 1.4260 - c*(0.3139)
beta1_U <- 1.4260 + c*(0.3139)
beta1_L
```

```
## [1] 0.810756
```

```r
beta1_U
```

```
## [1] 2.041244
```

```r
# beta2 95%
beta2_L <- 0.6919 - c*(0.3120)
beta2_U <- 0.6919 + c*(0.3120)
beta2_L
```

```
## [1] 0.08038
```

```r
beta2_U
```

```
## [1] 1.30342
```

```r
# beta3 95%
beta3_L <- 0.2464 - c*(0.4520)
beta3_U <- 0.2464 + c*(0.4520)
beta3_L
```

```
## [1] -0.63952
```

```r
beta3_U
```

```
## [1] 1.13232
```

```r
# Fitted Values Estimated and Confidence Interval
beta_est <- c(-1.2657,1.4260,0.6919,0.2464)

## when X1=1 & X2=1 ##
x_vector11 <- c(1,1,1,1)
beta_fun <- t(x_vector11)%*%beta_est
pie11 <- exp(beta_fun) / (1 + exp(beta_fun))
pie11
```

```
##           [,1]
## [1,] 0.7499977
```

```r
# 95% CI for beta_function
beta_fun_L <- beta_fun - c * sqrt(t(x_vector11) %*% summary(model)$cov.unscaled %*% x_vector11)
beta_fun_U <- beta_fun + c * sqrt(t(x_vector11) %*% summary(model)$cov.unscaled %*% x_vector11)
# 95% CI for fitted value
pie11_L <- exp(beta_fun_L) / (1 + exp(beta_fun_L))
pie11_U <- exp(beta_fun_U) / (1 + exp(beta_fun_U))
pie11_L
```

```
##           [,1]
## [1,] 0.6439456
```

```r
pie11_U
```

```
##         [,1]
## [1,] 0.83267
```

```r
## when X1=1 & X2=0 ##
x_vector10 <- c(1,1,0,0)
beta_fun <- t(x_vector10)%*%beta_est
pie10 <- exp(beta_fun) / (1 + exp(beta_fun))
pie10
```

```
##           [,1]
## [1,] 0.5399894
```

```r
# 95% CI for beta_function
beta_fun_L <- beta_fun - c * sqrt(t(x_vector10) %*% summary(model)$cov.unscaled %*% x_vector10)
beta_fun_U <- beta_fun + c * sqrt(t(x_vector10) %*% summary(model)$cov.unscaled %*% x_vector10)
# 95% CI for fitted value
pie10_L <- exp(beta_fun_L) / (1 + exp(beta_fun_L))
pie10_U <- exp(beta_fun_U) / (1 + exp(beta_fun_U))
pie10_L
```

```
##           [,1]
## [1,] 0.4420219
```

```r
pie10_U
```

```
##           [,1]
## [1,] 0.6349612
```

```r
## when X1=0 & X2=1 ##
x_vector01 <- c(1,0,1,0)
beta_fun <- t(x_vector01)%*%beta_est
pie01 <- exp(beta_fun) / (1 + exp(beta_fun))
pie01
```

```
##           [,1]
## [1,] 0.3603605
```

```r
# 95% CI for beta_function
beta_fun_L <- beta_fun - c * sqrt(t(x_vector01) %*% summary(model)$cov.unscaled %*% x_vector01)
beta_fun_U <- beta_fun + c * sqrt(t(x_vector01) %*% summary(model)$cov.unscaled %*% x_vector01)
# 95% CI for fitted value
pie01_L <- exp(beta_fun_L) / (1 + exp(beta_fun_L))
pie01_U <- exp(beta_fun_U) / (1 + exp(beta_fun_U))
pie01_L
```

```
##           [,1]
## [1,] 0.2766204
```

```r
pie01_U
```

```
##           [,1]
## [1,] 0.4535563
```

```r
## when X1=0 & X2=0 ##
x_vector00 <- c(1,0,0,0)
beta_fun <- t(x_vector00)%*%beta_est
pie00 <- exp(beta_fun) / (1 + exp(beta_fun))
```

```r
pie00
```

```
##           [,1]
## [1,] 0.2199942
```

```r
# 95% CI for beta_function
beta_fun_L <- beta_fun - c * sqrt(t(x_vector00) %*% summary(model)$cov.unscaled %*% x_vector00)
beta_fun_U <- beta_fun + c * sqrt(t(x_vector00) %*% summary(model)$cov.unscaled %*% x_vector00)
# 95% CI for fitted value
pie00_L <- exp(beta_fun_L) / (1 + exp(beta_fun_L))
pie00_U <- exp(beta_fun_U) / (1 + exp(beta_fun_U))
pie00_L
```

```
##           [,1]
## [1,] 0.1494593
```

```r
pie00_U
```

```
##          [,1]
## [1,] 0.311621
```

estimate of beta0 = -1.2657: estimated log odds of response Y when X1=X2=0 ; 95%C.I.: [-1.738844,-0.792556]
estimate of beta1 = 1.4260: estimated log odds ratio of response Y when X1=1 vs X1=0 while keeping X2=0 ; 95%C.I.:[0.810756,2.041244]
estimate of beta2 = 0.6919: estimated log odds ratio of response Y when X2=1 vs X2=0 while keeping X1=0 ; 95%C.I.:[0.08038,1.30342]
estimate of beta3 = 0.2464: estimated difference between log odds ratio of response Y when X1=1 vs X1=0 while keeping X2=1 and log odds ratio of response Y when X1=1 vs X1=0 while keeping X2=0 ; 95%C.I:[-0.63952,1.13232]
estimate of fitted value (when X1=1 X2=1) = 0.7499977 ; 95%C.I.:[0.6439456,0.83267]
estimate of fitted value (when X1=1 X2=0) = 0.5399894; 95%C.I.:[0.4420219,0.6349612]
estimate of fitted value (when X1=0 X2=1) = 0.3603605; 95%C.I.:[0.2766204,0.4535563]
estimate of fitted value (when X1=0 X2=0) = 0.2199942; 95%C.I.:[0.1494593,0.311621]

(b)

```r
model <- lm(y ~ x1*x2)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 * x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7500 -0.3604 -0.2200  0.4600  0.7800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.22000    0.04620   4.762 2.71e-06 ***
## x1           0.32000    0.06533   4.898 1.42e-06 ***
## x2           0.14036    0.06369   2.204   0.0281 *
## x1:x2        0.06964    0.09412   0.740   0.4598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.462 on 387 degrees of freedom
## Multiple R-squared:  0.1466, Adjusted R-squared:   0.14
## F-statistic: 22.17 on 3 and 387 DF,  p-value: 2.882e-13
```

```r
c <- qt(0.975, 400-5)

# beta0 95%
beta0_L <- 0.22000 - c*(0.04620)
beta0_U <- 0.22000 + c*(0.04620)
beta0_L
```

```
## [1] 0.1291714
```

```r
beta0_U
```

```
## [1] 0.3108286
```

```r
# beta1 95%
beta1_L <- 0.32000 - c*(0.06533)
beta1_U <- 0.32000 + c*(0.06533)
beta1_L
```

```
## [1] 0.191562
```

```r
beta1_U
```

```
## [1] 0.448438
```

```r
# beta2 95%
beta2_L <- 0.14036 - c*(0.06369)
beta2_U <- 0.14036 + c*(0.06369)
beta2_L
```

```
## [1] 0.01514623
```

```r
beta2_U
```

```
## [1] 0.2655738
```

```r
# beta3 95%
beta3_L <- 0.06964 - c*(0.09412)
beta3_U <- 0.06964 + c*(0.09412)
beta3_L
```

```
## [1] -0.1153988
```

```r
beta3_U
```

```
## [1] 0.2546788
```

```r
# fitted values and confidence intervals
# when X1=1 X2=1
predict(model, interval = "confidence")[312,]
```

```
##       fit       lwr       upr
## 0.7500000 0.6484546 0.8515454
```

```r
# when X1=1 X2=0
predict(model, interval = "confidence")[101,]
```

```
##       fit       lwr       upr
## 0.5400000 0.4491751 0.6308249
```

```r
# when X1=0 X2=1
predict(model, interval = "confidence")[201,]
```

```
##       fit       lwr       upr
## 0.3603604 0.2741532 0.4465676
```

```r
# when X1=0 X2=0
predict(model, interval = "confidence")[1,]
```

```
##       fit       lwr       upr
## 0.2200000 0.1291751 0.3108249
```

estimate of beta0 =0.22 : estimated mean of response Y when X1=X2=0 ; 95%C.I.: [0.1291714,0.3108286]
estimate of beta1 =0.32 : estimated mean of response Y when X1=1 vs X1=0 while keeping X2=0 ; 95%C.I.:[0.191562,0.448438]
estimate of beta2 = 0.14036: estimated mean of response Y when X2=1 vs X2=0 while keeping X1=0 ; 95%C.I.:[0.01514623,0.2655738]
estimate of beta3 = 0.06964: estimated difference between mean of response Y when X1=1 vs X1=0 while keeping X2=1 and mean of response Y when X1=1 vs X1=0 while keeping X2=0 ; 95%C.I:[-0.1153988,0.2546788]
estimate of fitted value (when X1=1 X2=1) = 0.75 ; 95%C.I.:[0.6484546,0.8515454]
estimate of fitted value (when X1=1 X2=0) = 0.54; 95%C.I.:[0.4491751,0.6308249]
estimate of fitted value (when X1=0 X2=1) = 0.3603604; 95%C.I.:[0.2741532,0.4465676]
estimate of fitted value (when X1=0 X2=0) = 0.22; 95%C.I.:[0.1291751,0.3108249]

(d)

For analysis in (a):
pros: Since all assumption of logistic regression model are satisfied, so those maximum likelihood estimators in analysis (a) are valid.
For analysis in (b):
pros: Compared with analysis (a), those coefficients have easier interpretation. (mean difference vs. odds ratio difference)
cons: The assumption of normality and constant variance are violated, so those estimators might be not valid.
Propose a change: weighted least square regression
For analysis in (c):
pros: Compared with analysis (a), those coefficients have easier interpretation. (mean difference vs. odds ratio difference). Besides, all assumptionn of model are satisfied, so those estimators in analysis (c) are valid.

Qurstion (2)

```r
### (b)
midpoint <-c(1.34, 1.60, 1.75, 1.85, 1.95, 2.00, 2.14, 2.25, 2.34)
survived <- c(13,19,67,45,71,50,35,7,1)
died <- c(0,0,2,5,8,20,31,49,12)
resp <- cbind(survived, died)
# fit logit link
logit_model <- glm(resp ~ midpoint, family = binomial(link = "logit"))
summary(logit_model)
```

```
##
## Call:
## glm(formula = resp ~ midpoint, family = binomial(link = "logit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4339  -1.0324  -0.1424   0.4234   1.5489
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   21.989      2.113   10.41   <2e-16 ***
## midpoint     -10.397      1.021  -10.19   <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: 198.7115  on 8  degrees of freedom
## Residual deviance:   8.8634  on 7  degrees of freedom
## AIC: 37.402
##
## Number of Fisher Scoring iterations: 5
```

```r
# fit probit link
probit_model <- glm(resp ~ midpoint, family = binomial(link = "probit"))
summary(probit_model)
```

```
##
## Call:
## glm(formula = resp ~ midpoint, family = binomial(link = "probit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6284  -0.8056  -0.1565   0.2099   1.6548
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  12.5444     1.1127   11.27   <2e-16 ***
## midpoint     -5.9364     0.5408  -10.98   <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: 198.712  on 8  degrees of freedom
## Residual deviance:  10.133  on 7  degrees of freedom
## AIC: 38.672
##
## Number of Fisher Scoring iterations: 5
```

```r
# fit log-log link
cloglog_model <- glm(resp ~ midpoint, family = binomial(link = "cloglog"))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

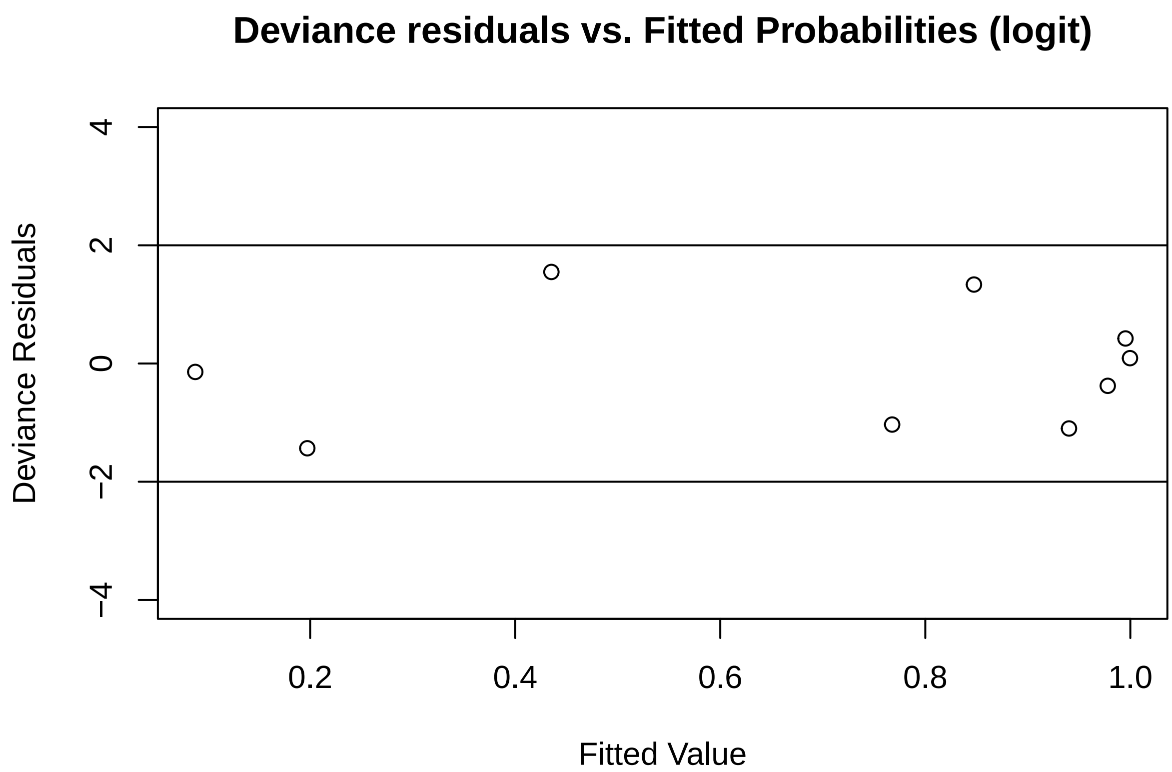
```r
summary(cloglog_model)
```

```
##
## Call:
## glm(formula = resp ~ midpoint, family = binomial(link = "cloglog"))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.69253  -1.05163   0.00000   0.08249   2.32167
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  11.0497     1.1005  10.041   <2e-16 ***
## midpoint     -5.4184     0.5508  -9.838   <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: 198.712  on 8  degrees of freedom
## Residual deviance:  19.207  on 7  degrees of freedom
## AIC: 47.746
##
## Number of Fisher Scoring iterations: 7
```

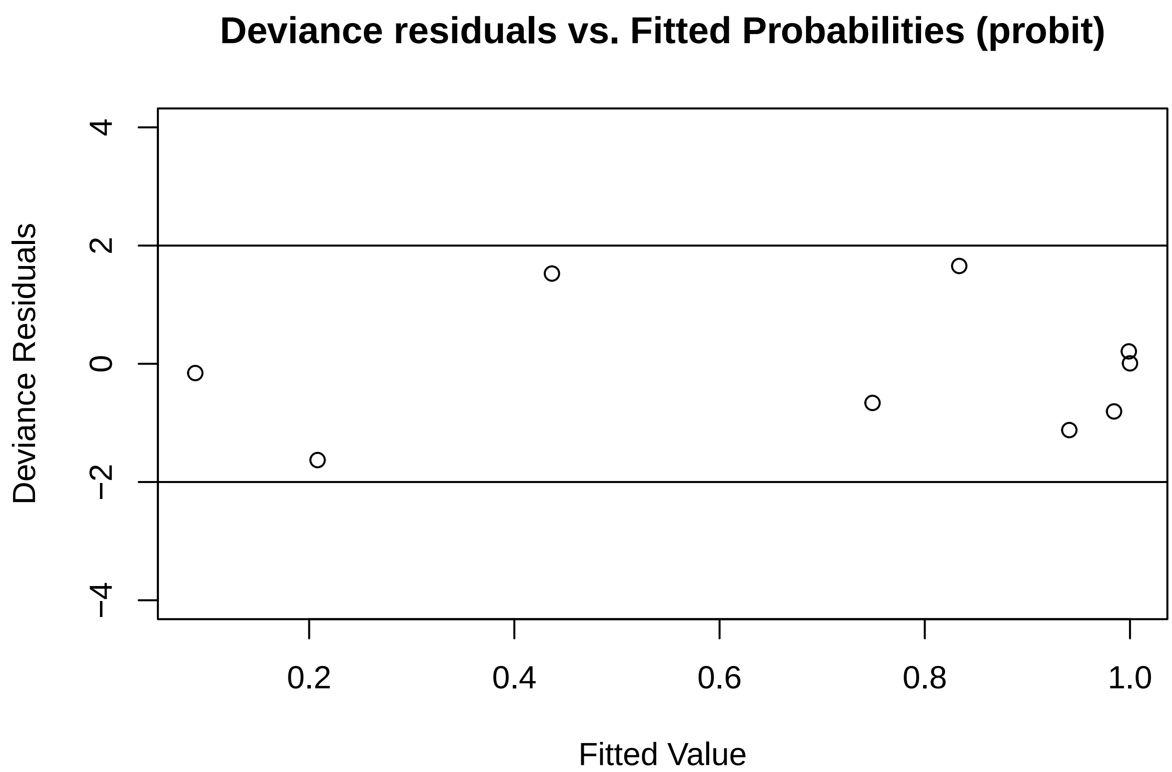
Interpretation:
For logit_model:
beta0_hat=21.989: the estimated log odds when the midpoints is zero (actually third-degree burn area is zero)
beta1_hat=-10.397: the estimated log odds ratio when the midpoint increases one unit
For probit_model:
beta0_hat=12.5444: the inverse CDF of N(0,1) for the probability of surviving when the midpoint is zero (burn area = 0)
beta1_hat=-5.9364: the estimated change of inverse CDF of N(0,1) for the probability of surviving when the midpoint increases one unit
For cloglog_model:
beta0_hat=11.0497: the estimated complimentary log log of the probability of surviving when the midpoint is zero (burn area = 0)
beta1_hat=-5.4184: the estimated change of complimentary log log of the probability of surviving when the midpoint is zero (burn area = 0)

(c)

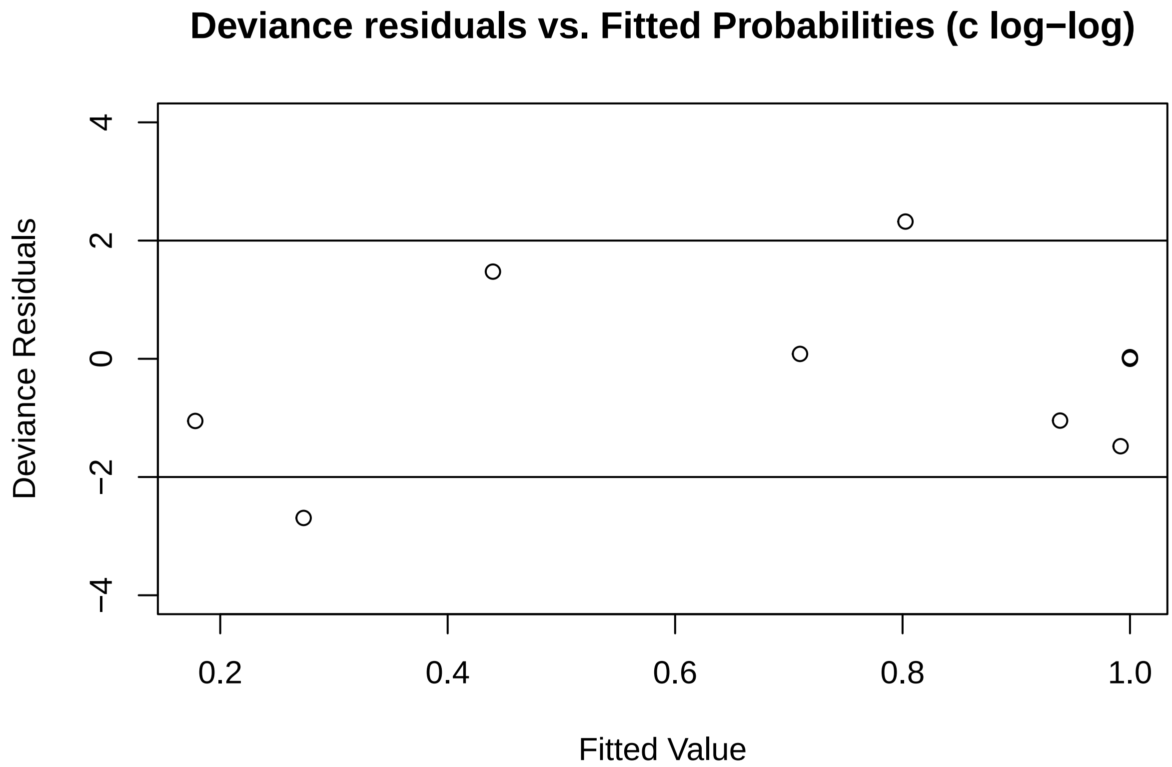
```
# logit model
rd1 <- residuals.glm(logit_model,"deviance")
fv1 <- logit_model$fitted.values
plot(fv1,rd1,xlab="Fitted Value", ylim=c(-4,4),
     ylab="Deviance Residuals",main="Deviance residuals vs. Fitted Probabilities (logit)")
 abline(h=-2) ; abline(h=2) ;
```




```
# probit model
rd2 <- residuals.glm(probit_model,"deviance")
fv2 <- probit_model$fitted.values
plot(fv2,rd2,xlab="Fitted Value", ylim=c(-4,4),
    ylab="Deviance Residuals",main="Deviance residuals vs. Fitted Probabilities (probit)")
abline(h=-2) ; abline(h=2)
```




```r
# c log-log model
rd3 <- residuals.glm(cloglog_model,"deviance")
fv3 <- cloglog_model$fitted.values
plot(fv3,rd3,xlab="Fitted Value", ylim=c(-4,4),
    ylab="Deviance Residuals",main="Deviance residuals vs. Fitted Probabilities (c log-log)")
abline(h=-2) ; abline(h=2)
```



Conclusion: Based on 3 plots, we find that logit model is the best. Since for c log-log model, there are 2 points outside [-1.96, 1.96] range. Besides, all plots in logit model are closer to line 0 compared to probit model.

(d)

```r
# we select logit model
pie <- 0.8
beta0_hat <- 21.989
beta1_hat <- -10.397

x <- (log(pie/(1-pie))-beta0_hat) / beta1_hat
x
```

```
## [1] 1.981601
```

```r
area <- exp(2) - 1
area
```

```
## [1] 6.389056
```

Therefore, the area is estimated to be 6.389056

Question 3

```r
# Save the original .csv file in your R Working Directory
# and then run this code block to input the data and
# prepare it for our analysis.
COVIDdata = read.csv("journal.pone.0245327.s010.csv")
# Limit the data to students from NCSU and a restricted set
# of explanatory variables
COVIDdata_NCSU = COVIDdata[(!is.na(COVIDdata$Source) & (COVIDdata$Source ==
  "NCState")), names(COVIDdata) %in% c("Health_General", "Hrs_Screen",
  "Hrs_Outdoor", "Hrs_Exercise", "Class_Self", "Infected_Any",
  "BMI", "Educ_College_Grad", "Age", "Classification_High",
  "Ethnoracial_Group_White1_Asian2", "Age_18to25")]

# Remove observations with missing Ethnoracial data (all
# other variable are complete)
COVIDdata_NCSU = COVIDdata_NCSU[!is.na(COVIDdata_NCSU$Ethnoracial_Group_White1_Asian2),]

# clean up non-integer class values
COVIDdata_NCSU$Class_Self <- round(COVIDdata_NCSU$Class_Self)
# Create factor variables where necessary
COVIDdata_NCSU$Infected_Any = factor(COVIDdata_NCSU$Infected_Any)
COVIDdata_NCSU$Educ_College_Grad = factor(COVIDdata_NCSU$Educ_College_Grad)
COVIDdata_NCSU$Ethnoracial_Group_White1_Asian2 = factor(COVIDdata_NCSU$Ethnoracial_Group_White1_Asian2)
COVIDdata_NCSU$Age_18to25 = factor(COVIDdata_NCSU$Age_18to25)


# str(COVIDdata_NCSU) # Display data set structure,
# commented out to save space
```

(a)

```r
# Fit a main effects logistic regression model
modelA = glm(Classification_High ~ +Age + Ethnoracial_Group_White1_Asian2 +
  Class_Self + Health_General + BMI + Hrs_Screen + Hrs_Outdoor +
  Hrs_Exercise + Educ_College_Grad + Infected_Any, family = binomial(link = "logit"),
  data = COVIDdata_NCSU)

summary(modelA)
```

```
##
## Call:
## glm(formula = Classification_High ~ +Age + Ethnoracial_Group_White1_Asian2 +
##     Class_Self + Health_General + BMI + Hrs_Screen + Hrs_Outdoor +
##     Hrs_Exercise + Educ_College_Grad + Infected_Any, family = binomial(link = "logit"),
##     data = COVIDdata_NCSU)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6842  -1.0830  -0.8592   1.1886   1.7050
##
## Coefficients:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         0.592785   0.508024   1.167  0.24327
```

```
## Age25 to 32                         0.226878   0.161655   1.403  0.16048
## Age33 to 44                         0.605891   0.332882   1.820  0.06874 .
## Age45 to 54                         0.657726   0.556496   1.182  0.23724
## Age55 to 64                       -12.687259 324.743737  -0.039  0.96884
## Ethnoracial_Group_White1_Asian21    0.287848   0.198502   1.450  0.14703
## Ethnoracial_Group_White1_Asian22    0.439006   0.233072   1.884  0.05962 .
## Class_Self                         -0.163984   0.061946  -2.647  0.00812 **
## Health_General                     -0.239811   0.058939  -4.069 4.73e-05 ***
## BMI                                -0.004303   0.013186  -0.326  0.74419
## Hrs_Screen                          0.027334   0.021813   1.253  0.21018
## Hrs_Outdoor                        -0.054803   0.048116  -1.139  0.25471
## Hrs_Exercise                        0.017909   0.072284   0.248  0.80432
## Educ_College_Grad1                  0.036405   0.145139   0.251  0.80195
## Infected_Any1                       0.444331   0.138834   3.200  0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1809.2  on 1311  degrees of freedom
## Residual deviance: 1753.7  on 1297  degrees of freedom
## AIC: 1783.7
##
## Number of Fisher Scoring iterations: 11
```

```r
c <- 1.96
# estimated class_self = -0.163984
exp_class_self <- exp(-0.163984)
exp_class_self_L <- exp(-0.163984 - c*0.061946)
exp_class_self_U <- exp(-0.163984 + c*0.061946)
exp_class_self
```

```
## [1] 0.8487556
```

```r
exp_class_self_L
```

```
## [1] 0.7517149
```

```r
exp_class_self_U
```

```
## [1] 0.9583235
```

```r
# estimated infected_any = 0.444331
exp_infected_any <- exp(0.444331)
exp_infected_any_L <- exp(0.444331 - c*0.138834)
exp_infected_any_U <- exp(0.444331 + c*0.138834)
exp_infected_any
```

```
## [1] 1.559447
```

```r
exp_infected_any_L
```

```
## [1] 1.187935
```

```r
exp_infected_any_U
```

```
## [1] 2.047144
```

```r
# estimated Ethnoracial_Group_White1_Asian21 = 0.287848
exp_asian21 <- exp(0.287848)
exp_asian21_L <- exp(0.287848 - c*0.198502)
exp_asian21_U <- exp(0.287848 + c*0.198502)
exp_asian21
```

```
## [1] 1.333555
```

```r
exp_asian21_L
```

```
## [1] 0.9037379
```

```r
exp_asian21_U
```

```
## [1] 1.967792
```

```r
# estimated Ethnoracial_Group_White1_Asian22 = 0.439006
exp_asian22 <- exp(0.439006)
exp_asian22_L <- exp(0.439006 - c*0.233072)
exp_asian22_U <- exp(0.439006 + c*0.233072)
exp_asian22
```

```
## [1] 1.551165
```

```r
exp_asian22_L
```

```
## [1] 0.9823426
```

```r
exp_asian22_U
```

```
## [1] 2.449361
```

Interpretation:

estimated exp class self=0.8487556: when remaining other factors unchanged, as the self-reported class increases one level, the odds of high psychological impact is expected to be 0.8487556 times as original odds. The 95% C.I is [0.7517149,0.9583235]

estimated exp infected any=1.559447: when remaining other factors unchanged, the odds of high psychological impact if knowing some infected is 1.559447 times as the odds of high psychological impact without knowing some infected. The 95% C.I is [1.187935,2.047144]

estimated exp asian21=1.333555: when remainning other factors unchanged, the odds of high psychological impact on while people is 1.333555 times as the odds of high psychological impact on black or hispanic. The 95% C.I is [0.9037379,1.967792]

estimated exp asian22=1.551165: when remaining other factors unchanged, the odds of high psychological impact on asian is 1.551165 times of the odds as high psychological impact on black or hispanic. The 95% C.I is [0.9823426,2.449361]

(c)

```
# WE found that beta1 - beta2 menas the 22-32 vs. 33-44
estimated_diff <- 0.226878 - 0.605891
# find variance for 22-32, 33-44 and their covariance
summary(modelA)$cov.unscaled
```

```
##                                    (Intercept)  Age25 to 32  Age33 to 44
## (Intercept)                       0.2580885311  0.0005859404  0.0078417931
## Age25 to 32                       0.0005859404  0.0261324461  0.0054664518
## Age33 to 44                       0.0078417931  0.0054664518  0.1108105523
## Age45 to 54                       0.0038094470  0.0061630844  0.0066144150
## Age55 to 64                      -0.0038369375  0.0060650203  0.0060329528
## Ethnoracial_Group_White1_Asian21 -0.0337678178  0.0008455126  0.0019015027
## Ethnoracial_Group_White1_Asian22 -0.0381689196 -0.0018917480  0.0010647506
## Class_Self                       -0.0088620953  0.0010673847  0.0004085332
## Health_General                   -0.0129402437 -0.0007111176 -0.0014935468
## BMI                              -0.0048115239 -0.0001968862 -0.0004482746
## Hrs_Screen                       -0.0045231060 -0.0001355533  0.0000320836
## Hrs_Outdoor                      -0.0018452208 -0.0001581507 -0.0007170866
## Hrs_Exercise                     -0.0014946800  0.0006374048  0.0011495826
## Educ_College_Grad1               -0.0004388881 -0.0045136715 -0.0005570443
## Infected_Any1                    -0.0036637502  0.0013445121  0.0028733662
##                                    Age45 to 54   Age55 to 64
## (Intercept)                       0.0038094470 -3.836938e-03
## Age25 to 32                       0.0061630844  6.065020e-03
## Age33 to 44                       0.0066144150  6.032953e-03
## Age45 to 54                       0.3096881979  6.059934e-03
## Age55 to 64                       0.0060599342  1.054585e+05
## Ethnoracial_Group_White1_Asian21  0.0039257898 -1.066692e-02
## Ethnoracial_Group_White1_Asian22  0.0055338625 -3.286452e-03
## Class_Self                        0.0004282546  7.415884e-03
## Health_General                   -0.0007625476 -1.326620e-03
## BMI                              -0.0006030043 -4.792722e-04
## Hrs_Screen                        0.0004913689 -3.425009e-04
## Hrs_Outdoor                       0.0006407953  1.936567e-03
## Hrs_Exercise                     -0.0011100033  7.592434e-04
## Educ_College_Grad1               -0.0051785339  3.055961e-03
## Infected_Any1                     0.0037996132  4.381107e-03
##                                  Ethnoracial_Group_White1_Asian21
## (Intercept)                                         -0.0337678178
## Age25 to 32                                          0.0008455126
## Age33 to 44                                          0.0019015027
## Age45 to 54                                          0.0039257898
## Age55 to 64                                         -0.0106669231
## Ethnoracial_Group_White1_Asian21                     0.0394028735
## Ethnoracial_Group_White1_Asian22                     0.0334410465
## Class_Self                                          -0.0023710155
## Health_General                                      -0.0001028943
## BMI                                                  0.0001782523
## Hrs_Screen                                           0.0003349741
## Hrs_Outdoor                                         -0.0008681931
## Hrs_Exercise                                         0.0002241730
## Educ_College_Grad1                                   0.0005815714
```

```
## Infected_Any1                                         0.0009374802
##                                    Ethnoracial_Group_White1_Asian22    Class_Self
## (Intercept)                                           -3.816892e-02 -8.862095e-03
## Age25 to 32                                           -1.891748e-03  1.067385e-03
## Age33 to 44                                            1.064751e-03  4.085332e-04
## Age45 to 54                                            5.533863e-03  4.282546e-04
## Age55 to 64                                           -3.286452e-03  7.415884e-03
## Ethnoracial_Group_White1_Asian21                       3.344105e-02 -2.371015e-03
## Ethnoracial_Group_White1_Asian22                       5.432245e-02 -1.605764e-03
## Class_Self                                            -1.605764e-03  3.837340e-03
## Health_General                                        -7.392678e-05 -5.394259e-04
## BMI                                                    3.052118e-04  6.823006e-05
## Hrs_Screen                                             1.554307e-04 -2.761036e-05
## Hrs_Outdoor                                            7.795445e-04  9.968718e-05
## Hrs_Exercise                                          -2.671169e-04 -7.946834e-05
## Educ_College_Grad1                                    -1.980868e-03 -4.307830e-05
## Infected_Any1                                          2.983463e-03 -3.178784e-04
##                                    Health_General           BMI    Hrs_Screen
## (Intercept)                         -1.294024e-02 -4.811524e-03 -4.523106e-03
## Age25 to 32                         -7.111176e-04 -1.968862e-04 -1.355533e-04
## Age33 to 44                         -1.493547e-03 -4.482746e-04  3.208360e-05
## Age45 to 54                         -7.625476e-04 -6.030043e-04  4.913689e-04
## Age55 to 64                         -1.326620e-03 -4.792722e-04 -3.425009e-04
## Ethnoracial_Group_White1_Asian21    -1.028943e-04  1.782523e-04  3.349741e-04
## Ethnoracial_Group_White1_Asian22    -7.392678e-05  3.052118e-04  1.554307e-04
## Class_Self                          -5.394259e-04  6.823006e-05 -2.761036e-05
## Health_General                       3.473838e-03  1.417538e-04  9.284716e-05
## BMI                                  1.417538e-04  1.738821e-04 -7.656849e-06
## Hrs_Screen                           9.284716e-05 -7.656849e-06  4.758188e-04
## Hrs_Outdoor                         -1.954902e-04 -3.763373e-05  2.241826e-04
## Hrs_Exercise                        -4.777996e-04 -1.326362e-05  1.402887e-04
## Educ_College_Grad1                  -3.878136e-04 -7.754580e-05  1.798010e-05
## Infected_Any1                       -1.437905e-04 -6.506259e-05  9.005371e-05
##                                     Hrs_Outdoor  Hrs_Exercise Educ_College_Grad1
## (Intercept)                        -1.845221e-03 -1.494680e-03      -4.388881e-04
## Age25 to 32                        -1.581507e-04  6.374048e-04      -4.513672e-03
## Age33 to 44                        -7.170866e-04  1.149583e-03      -5.570443e-04
## Age45 to 54                         6.407953e-04 -1.110003e-03      -5.178534e-03
## Age55 to 64                         1.936567e-03  7.592434e-04       3.055961e-03
## Ethnoracial_Group_White1_Asian21 -8.681931e-04  2.241730e-04       5.815714e-04
## Ethnoracial_Group_White1_Asian22  7.795445e-04 -2.671169e-04      -1.980868e-03
## Class_Self                          9.968718e-05 -7.946834e-05      -4.307830e-05
## Health_General                     -1.954902e-04 -4.777996e-04      -3.878136e-04
## BMI                                -3.763373e-05 -1.326362e-05      -7.754580e-05
## Hrs_Screen                          2.241826e-04  1.402887e-04       1.798010e-05
## Hrs_Outdoor                         2.315143e-03 -1.583764e-03      -1.997339e-04
## Hrs_Exercise                       -1.583764e-03  5.224935e-03       1.837769e-06
## Educ_College_Grad1                 -1.997339e-04  1.837769e-06       2.106546e-02
## Infected_Any1                      -2.844259e-04  3.721965e-04       1.000053e-03
##                                    Infected_Any1
## (Intercept)                        -3.663750e-03
## Age25 to 32                         1.344512e-03
## Age33 to 44                         2.873366e-03
## Age45 to 54                         3.799613e-03
```

```
## Age55 to 64                          4.381107e-03
## Ethnoracial_Group_White1_Asian21  9.374802e-04
## Ethnoracial_Group_White1_Asian22  2.983463e-03
## Class_Self                        -3.178784e-04
## Health_General                    -1.437905e-04
## BMI                               -6.506259e-05
## Hrs_Screen                         9.005371e-05
## Hrs_Outdoor                       -2.844259e-04
## Hrs_Exercise                       3.721965e-04
## Educ_College_Grad1                 1.000053e-03
## Infected_Any1                      1.927479e-02
```

```
var_diff <- 0.0261324461 + 0.1108105523 - 2*0.0054664518
# 95% C.I for difference
L <- estimated_diff - 1.96*sqrt(var_diff)
U <- estimated_diff + 1.96*sqrt(var_diff)
# 95% C.I for exp(difference)
exp_L <-exp(L)
exp_U <-exp(U)
exp_L
```

```
## [1] 0.3413756
```

```
exp_U
```

```
## [1] 1.372654
```

Therefore, the 95% C.I is [0.3413756,1.372654]

(d)

```r
modelD <- glm(Classification_High ~ +Age + Ethnoracial_Group_White1_Asian2 +
  factor(Class_Self) + Health_General + BMI + Hrs_Screen + Hrs_Outdoor +
  Hrs_Exercise + Educ_College_Grad + Infected_Any, family = binomial(link = "logit"),
  data = COVIDdata_NCSU)
summary(modelD)
```

```
##
## Call:
## glm(formula = Classification_High ~ +Age + Ethnoracial_Group_White1_Asian2 +
##     factor(Class_Self) + Health_General + BMI + Hrs_Screen +
##     Hrs_Outdoor + Hrs_Exercise + Educ_College_Grad + Infected_Any,
##     family = binomial(link = "logit"), data = COVIDdata_NCSU)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7402  -1.0808  -0.8551   1.1883   1.7142
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       0.317136   0.510183   0.622  0.53420
## Age25 to 32                       0.216686   0.161996   1.338  0.18103
## Age33 to 44                       0.605888   0.332536   1.822  0.06845 .
## Age45 to 54                       0.685885   0.555882   1.234  0.21725
## Age55 to 64                     -12.608327 324.743754  -0.039  0.96903
## Ethnoracial_Group_White1_Asian21   0.295319   0.198855   1.485  0.13752
## Ethnoracial_Group_White1_Asian22   0.448961   0.233757   1.921  0.05478 .
## factor(Class_Self)2               0.056837   0.215463   0.264  0.79194
## factor(Class_Self)3              -0.277177   0.190497  -1.455  0.14567
## factor(Class_Self)4              -0.429777   0.207625  -2.070  0.03846 *
## factor(Class_Self)5              -0.169289   0.554759  -0.305  0.76025
## Health_General                   -0.236694   0.058990  -4.012 6.01e-05 ***
## BMI                              -0.004032   0.013199  -0.305  0.76001
## Hrs_Screen                        0.028191   0.021858   1.290  0.19715
## Hrs_Outdoor                      -0.056294   0.048141  -1.169  0.24226
## Hrs_Exercise                      0.021721   0.072318   0.300  0.76391
## Educ_College_Grad1                0.032739   0.145383   0.225  0.82183
## Infected_Any1                     0.441932   0.138956   3.180  0.00147 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1809.2  on 1311  degrees of freedom
## Residual deviance: 1751.6  on 1294  degrees of freedom
## AIC: 1787.6
##
## Number of Fisher Scoring iterations: 11
```

(e)

```
beta <- modelA$coefficients
x_vector <- c(1,0,1,0,0,1,0,1,5,28,
              median(COVIDdata_NCSU$Hrs_Screen),
              median(COVIDdata_NCSU$Hrs_Outdoor),
              median(COVIDdata_NCSU$Hrs_Exercise),1,0)
beta_fun <- t(x_vector)%*%beta
pie_est <- exp(beta_fun) / (exp(beta_fun)+1)
pie_est
```

```
##           [,1]
## [1,] 0.5537449
```

Therefore, the estimated probability is 0.5537449

Question 4

```r
N <- 10000
x <- rbinom(N,1,0.25)
z <- rnorm(N)
beta0 <- -2
beta1 <- 1
beta2 <- 0.5
n <- 1000
num_iterations <- 2000

# Create a list to store results
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0 + beta1*x + beta2*z) / (1 + exp(beta0 + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)

  sample <- data[sample(N, n),]

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat <- mean(beta0_vector)
beta0_hat
```

```
## [1] -2.00844
```

```r
beta1_hat <- mean(beta1_vector)
beta1_hat
```

```
## [1] 1.003283
```

```r
beta2_hat <- mean(beta2_vector)
beta2_hat
```

```
## [1] 0.5032083
```

```r
beta0_sd <- sd(beta0_vector)
beta0_sd
```

```
## [1] 0.1194726
```

```r
beta1_sd <- sd(beta1_vector)
beta1_sd
```

```
## [1] 0.1847332
```

```r
beta2_sd <- sd(beta2_vector)
beta2_sd
```

```
## [1] 0.09125081
```

(a)

```r
# let beta0_1 = -1
beta0_1 <- -1
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_1 + beta1*x + beta2*z) / (1 + exp(beta0_1 + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)

  sample <- data[sample(N, n),]

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_1 <- mean(beta0_vector)
beta0_hat_1
```

```
## [1] -1.002082
```

```r
beta1_hat_1 <- mean(beta1_vector)
beta1_hat_1
```

```
## [1] 0.9987938
```

```r
beta2_hat_1 <- mean(beta2_vector)
beta2_hat_1
```

```
## [1] 0.5032699
```

```r
beta0_sd_1 <- sd(beta0_vector)
beta0_sd_1
```

```
## [1] 0.08387976
```

```r
beta1_sd_1 <- sd(beta1_vector)
beta1_sd_1
```

```
## [1] 0.1549929
```

```r
beta2_sd_1 <- sd(beta2_vector)
beta2_sd_1
```

```
## [1] 0.07398962
```

```r
# let beta0_2 = 0
beta0_2 <- 0
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_2 + beta1*x + beta2*z) / (1 + exp(beta0_2 + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)

  sample <- data[sample(N, n),]

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_2 <- mean(beta0_vector)
beta0_hat_2
```

```
## [1] -0.001999287
```

```r
beta1_hat_2 <- mean(beta1_vector)
beta1_hat_2
```

```
## [1] 1.00897
```

```r
beta2_hat_2 <- mean(beta2_vector)
beta2_hat_2
```

```
## [1] 0.5028059
```

```r
beta0_sd_2 <- sd(beta0_vector)
beta0_sd_2
```

```
## [1] 0.07527366
```

```r
beta1_sd_2 <- sd(beta1_vector)
beta1_sd_2
```

```
## [1] 0.1619011
```

```r
beta2_sd_2 <- sd(beta2_vector)
beta2_sd_2
```

```
## [1] 0.07172014
```

```r
# let beta0_3 = -3
beta0_3 <- -3
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))
```

```r
# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_3 + beta1*x + beta2*z) / (1 + exp(beta0_3 + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)

  sample <- data[sample(N, n),]

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_3 <- mean(beta0_vector)
beta0_hat_3
```

```
## [1] -3.023142
```

```r
beta1_hat_3 <- mean(beta1_vector)
beta1_hat_3
```

```
## [1] 1.013354
```

```r
beta2_hat_3 <- mean(beta2_vector)
beta2_hat_3
```

```
## [1] 0.5067475
```

```r
beta0_sd_3 <- sd(beta0_vector)
beta0_sd_3
```

```
## [1] 0.1822404
```

```r
beta1_sd_3 <- sd(beta1_vector)
beta1_sd_3
```

```
## [1] 0.2610926
```

```r
beta2_sd_3 <- sd(beta2_vector)
beta2_sd_3
```

```
## [1] 0.1271901
```

Conclusion: After repeat simulation with different beta0, I find that the estimates of beta1 and beta2 are almost unchanged and pretty close to real values no matter the value of beta0, which shows that they are unbiased Meanwhile, as the beta0 gets smaller, the standard deviation of estimates of beta1 and beta2 will increase, which are unbiased and the uncertainty is affected by values of beta0.

(c)

```r
# when beta0 = -1
beta0_1cc <- -1
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_1cc + beta1*x + beta2*z) / (1 + exp(beta0_1cc + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_1cc <- mean(beta0_vector)
beta0_hat_1cc
```

```
## [1] -1.00338
```

```r
beta1_hat_1cc <- mean(beta1_vector)
beta1_hat_1cc
```

```
## [1] 1.001733
```

```r
beta2_hat_1cc <- mean(beta2_vector)
beta2_hat_1cc
```

```
## [1] 0.5103914
```

```r
beta0_sd_1cc <- sd(beta0_vector)
beta0_sd_1cc
```

```
## [1] 0.1029451
```

```r
beta1_sd_1cc <- sd(beta1_vector)
beta1_sd_1cc
```

```
## [1] 0.2269133
```

```r
beta2_sd_1cc <- sd(beta2_vector)
beta2_sd_1cc
```

```
## [1] 0.109139
```

```r
# when beta0 = 0
beta0_2cc <- 0
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_2cc + beta1*x + beta2*z) / (1 + exp(beta0_2cc + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_2cc <- mean(beta0_vector)
beta0_hat_2cc
```

```
## [1] -0.001832343
```

```r
beta1_hat_2cc <- mean(beta1_vector)
beta1_hat_2cc
```

```
## [1] 1.009092
```

```r
beta2_hat_2cc <- mean(beta2_vector)
beta2_hat_2cc
```

```
## [1] 0.5025666
```

```r
beta0_sd_2cc <- sd(beta0_vector)
beta0_sd_2cc
```

```
## [1] 0.08580688
```

```r
beta1_sd_2cc <- sd(beta1_vector)
beta1_sd_2cc
```

```
## [1] 0.236969
```

```r
beta2_sd_2cc <- sd(beta2_vector)
beta2_sd_2cc
```

```
## [1] 0.100667
```

```r
# when beta0 = -3
beta0_3cc <- -3
```

```r
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, exp(beta0_3cc + beta1*x + beta2*z) / (1 + exp(beta0_3cc + beta1*x + beta2*z)))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "logit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_3cc <- mean(beta0_vector)
beta0_hat_3cc
```

```
## [1] -3.047579
```

```r
beta1_hat_3cc <- mean(beta1_vector)
beta1_hat_3cc
```

```
## [1] 1.019752
```

```r
beta2_hat_3cc <- mean(beta2_vector)
beta2_hat_3cc
```

```
## [1] 0.511233
```

```r
beta0_sd_3cc <- sd(beta0_vector)
beta0_sd_3cc
```

```
## [1] 0.2575465
```

```r
beta1_sd_3cc <- sd(beta1_vector)
beta1_sd_3cc
```

```
## [1] 0.3726368
```

```r
beta2_sd_3cc <- sd(beta2_vector)
beta2_sd_3cc
```

```
## [1] 0.1870315
```

Conclusion, based on different beta0 values, we find that the estimates of beta1 and beta2 are relatively fixed and close to the real value no matter the values of beta0, which means that they are unbiased. Meanwhile, the estimate of beta0 will be changed with the change of beta0 value, which is biased. The standard deviation of estimates are relatively fixed no matter the values of beta0, which is unbiased.

(d)

In the situation that number of values generated by random sampling simulation study is not balanced. For example, sometimes we have a lot of results with Y=1 but only a few results with Y=0. In such cases, we can use case control simulation study to reduce the uncertainty.

(e)

```r
N <- 10000
num_iterations <- 2000
beta1 <- 1
beta2 <- 0.5
x <- rbinom(N,1,0.25)
z <- rnorm(N)

# use probit link
# when beta0 = -1
beta0_1pl <- -1
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, pnorm(beta0_1pl + beta1*x + beta2*z))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "probit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_1pl <- mean(beta0_vector)
beta0_hat_1pl
```

```
## [1] -1.004929
```

```r
beta1_hat_1pl <- mean(beta1_vector)
beta1_hat_1pl
```

```
## [1] 1.005478
```

```r
beta2_hat_1pl <- mean(beta2_vector)
beta2_hat_1pl
```

```
## [1] 0.5046245
```

```r
beta0_sd_1pl <- sd(beta0_vector)
beta0_sd_1pl
```

```
## [1] 0.07522873
```

```r
beta1_sd_1pl <- sd(beta1_vector)
beta1_sd_1pl
```

```
## [1] 0.1464388
```

```r
beta2_sd_1pl <- sd(beta2_vector)
beta2_sd_1pl
```

```
## [1] 0.07431727
```

```r
# when beta0 = 0
beta0_2pl <- 0
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, pnorm(beta0_2pl + beta1*x + beta2*z))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "probit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}

# investigate bias and uncertainty
beta0_hat_2pl <- mean(beta0_vector)
beta0_hat_2pl
```

```
## [1] 0.001275068
```

```r
beta1_hat_2pl <- mean(beta1_vector)
beta1_hat_2pl
```

```
## [1] 1.008934
```

```r
beta2_hat_2pl <- mean(beta2_vector)
beta2_hat_2pl
```

```
## [1] 0.5044835
```

```r
beta0_sd_2pl <- sd(beta0_vector)
beta0_sd_2pl
```

```
## [1] 0.05376338
```

```r
beta1_sd_2pl <- sd(beta1_vector)
beta1_sd_2pl
```

```
## [1] 0.1557974
```

```r
beta2_sd_2pl <- sd(beta2_vector)
beta2_sd_2pl
```

```
## [1] 0.06601984
```

```r
# when beta0 = -3
beta0_3pl <- -3
beta0_vector <- c(rep(NULL,num_iterations))
beta1_vector <- c(rep(NULL,num_iterations))
beta2_vector <- c(rep(NULL,num_iterations))

# Run the loop
for (i in 1:num_iterations) {

  y <- rbinom(N, 1, pnorm(beta0_3pl + beta1*x + beta2*z))
  data <- data.frame(x=x, z=z, y=y)
  data1 <- data[data$y==1,]
  data0 <- data[data$y==0,]

  sample1 <- data1[sample(N, 500),]
  sample0 <- data0[sample(N, 500),]
  sample <- rbind(sample1,sample0)

  model <- glm(y ~ x + z, family = binomial(link = "probit"), data = sample)

  beta0_vector[i] <- summary(model)$coefficients[1,1]
  beta1_vector[i] <- summary(model)$coefficients[2,1]
  beta2_vector[i] <- summary(model)$coefficients[3,1]
}
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
# investigate bias and uncertainty
beta0_hat_3pl <- mean(beta0_vector)
beta0_hat_3pl
```

```
## [1] -4.247273
```

```
beta1_hat_3pl <- mean(beta1_vector)
beta1_hat_3pl
```

```
## [1] 2.001583
```

```
beta2_hat_3pl <- mean(beta2_vector)
beta2_hat_3pl
```

```
## [1] 0.5876529
```

```
beta0_sd_3pl <- sd(beta0_vector)
beta0_sd_3pl
```

```
## [1] 3.927422
```

```
beta1_sd_3pl <- sd(beta1_vector)
beta1_sd_3pl
```

```
## [1] 2.301005
```

```
beta2_sd_3pl <- sd(beta2_vector)
beta2_sd_3pl
```

```
## [1] 1.163747
```

Conclusion: The relationship we found by probit link is different with by logit link. As the value of beta0 decreases, the bias of beta1 and beta2's estimates get larger, which are unbiased. And the standard deviation get smaller actually, which are unbiased as well.