6210HW5

Qinyuan Xing

2021/11/8

Contents

Q1

```
covid = data.frame(read.csv("covid_data_pdb_2.csv"))
covid$StateAbbr = as.factor(covid$StateAbbr)
covid$State = as.factor(covid$State)
covid$State_name = as.factor(covid$State_name)
covid$division = as.factor(covid$division)
covid$County = as.character(covid$County)
covid$TransHigh = as.factor(100000*(covid$covid_count_sep17-covid$covid_count_sep10)/covid$Tot_Populati
covid$PovHigh = as.factor((covid$pct_Prs_Blw_Pov_Lev_ACS_14_18>20))
```

```
cat("\n(a)\n")
```

```
##
## (a)
```

```
t1 = xtabs(~PovHigh+TransHigh,data=covid)
rownames(t1) = c("No","Yes")
colnames(t1) = c("No","Yes")
t1
```

```
##        TransHigh
## PovHigh  No  Yes
##     No   176 2297
##     Yes   19  640
```

```
cat("\nThe joint and marginal probabilities:\n")
```

```
##
## The joint and marginal probabilities:
```

```r
addmargins(prop.table(t1))
```

```
##         TransHigh
## PovHigh       No     Yes     Sum
##     No   0.05619 0.73340 0.78959
##     Yes  0.00607 0.20434 0.21041
##     Sum  0.06226 0.93774 1.00000
```

```r
cat("\n(b)\n")
```

```
##
## (b)
```

```r
mt1=prop.table(t1,margin = 1)
mt1
```

```
##         TransHigh
## PovHigh      No    Yes
##     No   0.0712 0.9288
##     Yes  0.0288 0.9712
```

```r
cat("\nThe conditional probabilities of High transmission given High or NOT Poverty=0.9712 or 0.9288,\n
```

```
##
## The conditional probabilities of High transmission given High or NOT Poverty=0.9712 or 0.9288,
```

```r
cat("\n(c)\n")
```

```
##
## (c)
```

```r
diff = mt1[2,2]-mt1[1,2]
diff
```

```
## [1] 0.0423
```

```r
cat("\nThe difference is because for the county with high poverty rate, the living and medical condition
```

```
##
## The difference is because for the county with high poverty rate, the living and medical condition ar
```

```r
se_diff = sqrt(mt1[2,2]*(1-mt1[2,2])/sum(t1[2,]) + mt1[1,2]*(1-mt1[1,2])/sum(t1[1,]))
l = diff - qnorm(1-(.05/2))*se_diff
u = diff + qnorm(1-(.05/2))*se_diff
cat("\nThe 95% Wald CI is:\n")
```

```
##
## The 95% Wald CI is:
```

```r
c(l,u)
```

```
## [1] 0.0260 0.0586
```

```r
cat("\n(d)\n")
```

```
##
## (d)
```

```r
r = mt1[2,2]/mt1[1,2]
r
```

```
## [1] 1.05
```

```r
cat("\nThe relative risk is",r,"which means a county with High poverty rate has 1.05 times probability
```

```
##
## The relative risk is 1.05 which means a county with High poverty rate has 1.05 times probability to
```

```r
logr = log(r)
logr
```

```
## [1] 0.0446
```

```r
se_logr = sqrt((1-mt1[2,2])/t1[2,2] + (1-mt1[1,2])/t1[1,2])
se_logr
```

```
## [1] 0.00872
```

```r
lr = logr - qnorm(1-(.05/2))*se_logr
ur = logr + qnorm(1-(.05/2))*se_logr
c(lr,ur)
```

```
## [1] 0.0275 0.0617
```

```r
explr = exp(lr)
expur = exp(ur)
cat("\nThe 95% Wald CI of RR is:\n")
```

```
##
## The 95% Wald CI of RR is:
```

```r
c(explr,expur)
```

```
## [1] 1.03 1.06
```

```
cat("\n(e)\n")
```

```
##
## (e)
```

```
OR = (t1[1,1]*t1[2,2]) / (t1[1,2]*t1[2,1])
OR
```

```
## [1] 2.58
```

```
cat("\nThe Odd Ratio is",OR,"which means the county with High poverty rate has 2.58 times odds compared
```

```
##
## The Odd Ratio is 2.58 which means the county with High poverty rate has 2.58 times odds compared wit
```

```
logOR = log(OR)
logOR
```

```
## [1] 0.948
```

```
se_logOR = sqrt(sum(1/t1))
se_logOR
```

```
## [1] 0.246
```

```
lo = logOR - qnorm(1-(.05/2))*se_logOR
uo = logOR + qnorm(1-(.05/2))*se_logOR
c(lo,uo)
```

```
## [1] 0.467 1.429
```

```
explo = exp(lo)
expuo = exp(uo)
cat("\nThe 95% Wald CI of ORis:\n")
```

```
##
## The 95% Wald CI of ORis:
```

```
c(explo,expuo)
```

```
## [1] 1.59 4.18
```

```
cat("\n(f)\n")
```

```
##
## (f)
```

```
cat("\nThere is an association between High poverty and High Covid because the diffrence of propotions
```

```
##
## There is an association between High poverty and High Covid because the diffrence of propotions is n
```

```
cat("\n(g)\n")
```

```
##
## (g)
```

```
x = matrix(c(124,4,11,3,40,12,2297,640), nrow=2)
rownames(x) = c("Not High Poverty","High Poverty")
colnames(x) = c("Low","Moderate","Substantial","High")
x
```

```
##                  Low Moderate Substantial High
## Not High Poverty 124       11          40 2297
## High Poverty       4        3          12  640
```

```
pi_hat = x/sum(x)
cat("\nThe joint and marginal probabilities:\n")
```

```
##
## The joint and marginal probabilities:
```

```
addmargins(prop.table(x))
```

```
##                      Low Moderate Substantial  High  Sum
## Not High Poverty 0.03960 0.003513     0.01278 0.734 0.79
## High Poverty     0.00128 0.000958     0.00383 0.204 0.21
## Sum              0.04088 0.004471     0.01661 0.938 1.00
```

```
cat("\n(h)\n")
```

```
##
## (h)
```

```
x_marg = apply(x,1,sum)
y_marg = apply(x,2,sum)
mu_hat = (x_marg %*% t(y_marg)) / sum(x)
mu_hat
```

```
##        Low Moderate Substantial High
## [1,] 101.1    11.05        41.1 2319
## [2,]  26.9     2.95        10.9  618
```

```r
cat("\n(i)\n")
```

```
##
## (i)
```

```r
X2 = sum((x-mu_hat)^2/mu_hat)
X2
```

```
## [1] 25.8
```

```r
qchisq(.95,(dim(x)[1]-1)*(dim(x)[2]-1))
```

```
## [1] 7.81
```

```r
1-pchisq(X2,(dim(x)[1]-1)*(dim(x)[2]-1))
```

```
## [1] 1.03e-05
```

```r
cat("\n25.8>7.81 so we reject the H0:High poverty and Covid Transmission are independent. p-value is 1.0
```

```
##
## 25.8>7.81 so we reject the H0:High poverty and Covid Transmission are independent. p-value is 1.03e-0
```

```r
chisq.test(x)
```

```
##
##  Pearson's Chi-squared test
##
## data:  x
## X-squared = 26, df = 3, p-value = 1e-05
```

```r
cat("\nThe result above proved our calculations.\n")
```

```
##
## The result above proved our calculations.
```

```r
cat("\n(j)\n")
```

```
##
## (j)
```

```r
G2 = 2*(sum(x*log(x/mu_hat)))
G2
```

```
## [1] 36.6
```

```r
1-pchisq(G2,(dim(x)[1]-1)*(dim(x)[2]-1))
```

```
## [1] 5.67e-08
```

```r
cat("\n36.6>7.81 so we reject the H0:High poverty and Covid Transmission are independent. p-value is 5.
```

```
##
## 36.6>7.81 so we reject the H0:High poverty and Covid Transmission are independent. p-value is 5.67e-
```

```r
library(DescTools)
GTest(x)
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  x
## G = 37, X-squared df = 3, p-value = 6e-08
```

```r
cat("\nThe result above proved our calculations.\n")
```

```
##
## The result above proved our calculations.
```

```r
cat("\n(k)\n")
```

```
##
## (k)
```

```r
cat("The standardized Pearson residuals are:\n")
```

```
## The standardized Pearson residuals are:
```

```r
chisq.test(x)$stdres
```

```
##                    Low Moderate Substantial  High
## Not High Poverty  5.08   -0.035      -0.362 -3.97
## High Poverty     -5.08    0.035       0.362  3.97
```

```r
cat("\nThe results show that Low and High Covid transmission have association with Poverty rate.\n")
```

```
##
## The results show that Low and High Covid transmission have association with Poverty rate.
```

```r
cat("\n(l)\n")
```

```
##
## (l)
```

```
x
```

```
##                   Low Moderate Substantial High
## Not High Poverty 124       11           40 2297
## High Poverty       4        3           12  640
```

```
qchisq(.95,1)
```

```
## [1] 3.84
```

```
qchisq(.95,2)
```

```
## [1] 5.99
```

```
GTest(matrix(c(11,3,40,12,2297,640),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  matrix(c(11, 3, 40, 12, 2297, 640), nrow = 2)
## G = 0.05, X-squared df = 2, p-value = 1
```

```
cat("\nWe cannot reject the independence between High Poverty and modest or substantial or High transmis
```

```
##
## We cannot reject the independence between High Poverty and modest or substantial or High transmission
```

```
GTest(matrix(c(124,4,51,15),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  matrix(c(124, 4, 51, 15), nrow = 2)
## G = 18, X-squared df = 1, p-value = 2e-05
```

```
cat("\nWe can find the difference between Low and modest/substantial transmission in terms of the povert
```

```
##
## We can find the difference between Low and modest/substantial transmission in terms of the poverty ra
```

```
GTest(matrix(c(51,15,2297,640),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  matrix(c(51, 15, 2297, 640), nrow = 2)
## G = 0.03, X-squared df = 1, p-value = 0.9
```

```r
cat("\nWe cannot find the difference between High and modest/substantial transmission in terms of the po
```

```
##
## We cannot find the difference between High and modest/substantial transmission in terms of the pover
```

```r
GTest(matrix(c(124,4,2348,655),nrow=2))
```

```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  matrix(c(124, 4, 2348, 655), nrow = 2)
## G = 37, X-squared df = 1, p-value = 2e-09
```

```r
cat("\nWe can find the difference between Low and modest/substantial/High transmission in terms of the p
```

```
##
## We can find the difference between Low and modest/substantial/High transmission in terms of the pove
```

```r
cat("\nFinally we find that Low tansmission rate has association with Poverty rate.\n")
```

```
##
## Finally we find that Low tansmission rate has association with Poverty rate.
```

Q2

```r
cat("\n(a)\n")
```

```
##
## (a)
```

```r
covid2 = data.frame(read.csv("three_way_COVID_table.csv"))
t2 = xtabs(count~region+covid_rate_high+vacc_rate_high,data=covid2)
x_midwest = xtabs(count~vacc_rate_high+covid_rate_high,data=covid2,subset = region=="Midwest")
cat("\nMidwest:\n")
```

```
##
## Midwest:
```

```r
x_midwest
```

```
##                covid_rate_high
## vacc_rate_high FALSE TRUE
##          FALSE   113  771
##          TRUE      5  166
```

```
OddsRatio(x_midwest)
```

```
## [1] 4.87
```

```
x_northeast = xtabs(count~vacc_rate_high+covid_rate_high,data=covid2,subset = region=="Northeast")
cat("\nNortheast:\n")
```

```
##
## Northeast:
```

```
x_northeast
```

```
##               covid_rate_high
## vacc_rate_high FALSE TRUE
##         FALSE     1   66
##         TRUE      3  148
```

```
OddsRatio(x_northeast)
```

```
## [1] 0.747
```

```
x_south = xtabs(count~vacc_rate_high+covid_rate_high,data=covid2,subset = region=="South")
cat("\nSouth:\n")
```

```
##
## South:
```

```
x_south
```

```
##               covid_rate_high
## vacc_rate_high FALSE TRUE
##         FALSE    14 1054
##         TRUE      3   90
```

```
OddsRatio(x_south)
```

```
## [1] 0.398
```

```
x_west = xtabs(count~vacc_rate_high+covid_rate_high,data=covid2,subset = region=="West")
cat("\nWest:\n")
```

```
##
## West:
```

```
x_west
```

```
##               covid_rate_high
## vacc_rate_high FALSE TRUE
##         FALSE    17  271
##         TRUE     18  125
```

```
OddsRatio(x_west)
```

[1] 0.436

```
cat("\nThe results show that there is an association between
vaccination and COVID transmission level for each of the regions because the OR is not 1.\n")
```

```
##
## The results show that there is an association between
## vaccination and COVID transmission level for each of the regions because the OR is not 1.
```

```
cat("\nThe OR in Midwest shows that the odds of low covid rate for low vaccine rate is 4.87 times the od
```

```
##
## The OR in Midwest shows that the odds of low covid rate for low vaccine rate is 4.87 times the odds :
```

```
cat("\nThe OR in West shows that the odds of low covid rate for low vaccine rate is 0.436 times the odds
```

```
##
## The OR in West shows that the odds of low covid rate for low vaccine rate is 0.436 times the odds for
```

```
cat("\n(b)\n")
```

```
##
## (b)
```

```
model = glm(count~covid_rate_high*vacc_rate_high*region,data=covid2, family=poisson())
summary(model)
```

```
##
## Call:
## glm(formula = count ~ covid_rate_high * vacc_rate_high * region,
##     family = poisson(), data = covid2)
##
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##                                              Estimate Std. Error
## (Intercept)                                    4.7274     0.0941
## covid_rate_highTRUE                            1.9203     0.1007
## vacc_rate_highTRUE                            -3.1179     0.4570
## regionNortheast                               -4.7274     1.0044
## regionSouth                                   -2.0883     0.2833
## regionWest                                    -1.8942     0.2601
## covid_rate_highTRUE:vacc_rate_highTRUE         1.5822     0.4649
## covid_rate_highTRUE:regionNortheast            2.2694     1.0126
## covid_rate_highTRUE:regionSouth                2.4010     0.2873
## covid_rate_highTRUE:regionWest                 0.8486     0.2696
```

```
## vacc_rate_highTRUE:regionNortheast                              4.2166     1.2418
## vacc_rate_highTRUE:regionSouth                                  1.5775     0.7833
## vacc_rate_highTRUE:regionWest                                   3.1751     0.5685
## covid_rate_highTRUE:vacc_rate_highTRUE:regionNortheast         -1.8733     1.2536
## covid_rate_highTRUE:vacc_rate_highTRUE:regionSouth             -2.5023     0.7956
## covid_rate_highTRUE:vacc_rate_highTRUE:regionWest              -2.4132     0.5850
##                                                               z value Pr(>|z|)
## (Intercept)                                                     50.25  < 2e-16 ***
## covid_rate_highTRUE                                             19.06  < 2e-16 ***
## vacc_rate_highTRUE                                             -6.82  8.9e-12 ***
## regionNortheast                                                -4.71  2.5e-06 ***
## regionSouth                                                    -7.37  1.7e-13 ***
## regionWest                                                     -7.28  3.3e-13 ***
## covid_rate_highTRUE:vacc_rate_highTRUE                          3.40  0.00067 ***
## covid_rate_highTRUE:regionNortheast                             2.24  0.02501 *
## covid_rate_highTRUE:regionSouth                                 8.36  < 2e-16 ***
## covid_rate_highTRUE:regionWest                                  3.15  0.00164 **
## vacc_rate_highTRUE:regionNortheast                              3.40  0.00069 ***
## vacc_rate_highTRUE:regionSouth                                  2.01  0.04403 *
## vacc_rate_highTRUE:regionWest                                   5.58  2.3e-08 ***
## covid_rate_highTRUE:vacc_rate_highTRUE:regionNortheast         -1.49  0.13507
## covid_rate_highTRUE:vacc_rate_highTRUE:regionSouth             -3.15  0.00166 **
## covid_rate_highTRUE:vacc_rate_highTRUE:regionWest              -4.13  3.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance:  5.3522e+03  on 15  degrees of freedom
## Residual deviance: -1.7542e-14  on  0  degrees of freedom
## AIC: 121.5
##
## Number of Fisher Scoring iterations: 3
```

```r
resids = residuals(model,type="pearson")
resids
```

```
##        1         2         3         4         5         6         7         8
## -5.88e-14 -1.64e-14 -2.03e-14 -2.87e-14  3.15e-14  1.03e-15  4.67e-15 -5.70e-15
##        9        10        11        12        13        14        15        16
## -3.50e-14 -1.97e-14  8.99e-15  5.17e-15 -3.45e-15  4.19e-15 -2.42e-14 -4.68e-11
```

```r
cat("\nThe residual deviance and pearson residuals are zero because the model fits well.\n")
```

```
##
## The residual deviance and pearson residuals are zero because the model fits well.
```

```r
cat("\n(c)\n")
```

```
##
## (c)
```

```r
model1 = glm(count~covid_rate_high*vacc_rate_high+region,data=covid2, family=poisson())
summary(model1)
```

```
##
## Call:
## glm(formula = count ~ covid_rate_high * vacc_rate_high + region,
##     family = poisson(), data = covid2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
##   -9.62    -3.30    -1.51     4.74    13.04
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                          3.9777     0.0866   45.95  < 2e-16 ***
## covid_rate_highTRUE                  2.7021     0.0858   31.50  < 2e-16 ***
## vacc_rate_highTRUE                  -1.6094     0.2034   -7.91  2.5e-15 ***
## regionNortheast                     -1.5768     0.0744  -21.19  < 2e-16 ***
## regionSouth                          0.0957     0.0425    2.25    0.024 *
## regionWest                          -0.8952     0.0572  -15.66  < 2e-16 ***
## covid_rate_highTRUE:vacc_rate_highTRUE  0.2016   0.2091    0.96    0.335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5352.15  on 15  degrees of freedom
## Residual deviance:  562.78  on  9  degrees of freedom
## AIC: 666.3
##
## Number of Fisher Scoring iterations: 5
```

```r
qchisq(.95,model1$df.residual)
```

```
## [1] 16.9
```

```r
1-pchisq(model1$deviance,model$df.residual)
```

```
## [1] 0
```

```r
cat("\n562.78>16.9 so we reject H0: The new model is better,p-value =0\n")
```

```
##
## 562.78>16.9 so we reject H0: The new model is better,p-value =0
```

```r
cat("\n(d)\n")
```

```
##
## (d)
```

```r
cat("\nCovid rate Not High, vaccine rate Not High, Midwest are my reference group.\n")
```

```
##
## Covid rate Not High, vaccine rate Not High, Midwest are my reference group.
```

```r
cat("\nThe coefficient for the interaction covid_rate_highTRUE:vacc_rate_highTRUE is 0.2016. Thus the e
reference group) is exp(0.2016) = 1.22.\n")
```

```
##
## The coefficient for the interaction covid_rate_highTRUE:vacc_rate_highTRUE is 0.2016. Thus the estim
## reference group) is exp(0.2016) = 1.22.
```

```r
cat("\n(e)\n")
```

```
##
## (e)
```

```r
model2=glm(count~covid_rate_high+vacc_rate_high+region+covid_rate_high:vacc_rate_high:region,data=covid
summary(model2)
```

```
##
## Call:
## glm(formula = count ~ covid_rate_high + vacc_rate_high + region +
##     covid_rate_high:vacc_rate_high:region, family = poisson(),
##     data = covid2)
##
## Deviance Residuals:
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##
## Coefficients: (6 not defined because of singularities)
##                                                     Estimate Std. Error
## (Intercept)                                            3.948      0.312
## covid_rate_highTRUE                                    1.938      0.252
## vacc_rate_highTRUE                                    -0.774      0.108
## regionNortheast                                       -0.115      0.113
## regionSouth                                           -0.612      0.131
## regionWest                                            -0.284      0.118
## covid_rate_highFALSE:vacc_rate_highFALSE:regionMidwest  0.780      0.326
## covid_rate_highTRUE:vacc_rate_highFALSE:regionMidwest   0.762      0.138
## covid_rate_highFALSE:vacc_rate_highTRUE:regionMidwest  -1.565      0.519
## covid_rate_highTRUE:vacc_rate_highTRUE:regionMidwest     NA         NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionNortheast -3.833    1.048
## covid_rate_highTRUE:vacc_rate_highFALSE:regionNortheast -1.581     0.183
## covid_rate_highFALSE:vacc_rate_highTRUE:regionNortheast -1.961     0.635
## covid_rate_highTRUE:vacc_rate_highTRUE:regionNortheast   NA         NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionSouth   -0.697      0.417
## covid_rate_highTRUE:vacc_rate_highFALSE:regionSouth     1.687      0.154
## covid_rate_highFALSE:vacc_rate_highTRUE:regionSouth    -1.463      0.639
## covid_rate_highTRUE:vacc_rate_highTRUE:regionSouth       NA         NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionWest    -0.831      0.355
## covid_rate_highTRUE:vacc_rate_highFALSE:regionWest       NA         NA
```

```
## covid_rate_highFALSE:vacc_rate_highTRUE:regionWest            NA         NA
## covid_rate_highTRUE:vacc_rate_highTRUE:regionWest             NA         NA
##                                                       z value Pr(>|z|)
## (Intercept)                                             12.66  < 2e-16 ***
## covid_rate_highTRUE                                      7.69  1.5e-14 ***
## vacc_rate_highTRUE                                      -7.16  8.3e-13 ***
## regionNortheast                                        -1.02  0.30999
## regionSouth                                            -4.68  2.9e-06 ***
## regionWest                                             -2.40  0.01660 *
## covid_rate_highFALSE:vacc_rate_highFALSE:regionMidwest   2.39  0.01671 *
## covid_rate_highTRUE:vacc_rate_highFALSE:regionMidwest    5.53  3.3e-08 ***
## covid_rate_highFALSE:vacc_rate_highTRUE:regionMidwest   -3.01  0.00258 **
## covid_rate_highTRUE:vacc_rate_highTRUE:regionMidwest      NA       NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionNortheast -3.66  0.00025 ***
## covid_rate_highTRUE:vacc_rate_highFALSE:regionNortheast  -8.63  < 2e-16 ***
## covid_rate_highFALSE:vacc_rate_highTRUE:regionNortheast  -3.09  0.00203 **
## covid_rate_highTRUE:vacc_rate_highTRUE:regionNortheast     NA       NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionSouth    -1.67  0.09471 .
## covid_rate_highTRUE:vacc_rate_highFALSE:regionSouth     10.94  < 2e-16 ***
## covid_rate_highFALSE:vacc_rate_highTRUE:regionSouth     -2.29  0.02197 *
## covid_rate_highTRUE:vacc_rate_highTRUE:regionSouth        NA       NA
## covid_rate_highFALSE:vacc_rate_highFALSE:regionWest     -2.34  0.01927 *
## covid_rate_highTRUE:vacc_rate_highFALSE:regionWest        NA       NA
## covid_rate_highFALSE:vacc_rate_highTRUE:regionWest        NA       NA
## covid_rate_highTRUE:vacc_rate_highTRUE:regionWest         NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5.3522e+03  on 15  degrees of freedom
## Residual deviance: 1.1480e-13  on  0  degrees of freedom
## AIC: 121.5
##
## Number of Fisher Scoring iterations: 3
```

```
model3 = glm(count~covid_rate_high+vacc_rate_high+region,data=covid2, family=poisson())
summary(model3)
```

```
##
## Call:
## glm(formula = count ~ covid_rate_high + vacc_rate_high + region,
##     family = poisson(), data = covid2)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
##  -9.50   -3.59   -1.48    4.52   13.11
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)           3.9434     0.0802   49.18   <2e-16 ***
## covid_rate_highTRUE   2.7386     0.0782   35.01   <2e-16 ***
## vacc_rate_highTRUE   -1.4193     0.0472  -30.09   <2e-16 ***
## regionNortheast      -1.5768     0.0744  -21.19   <2e-16 ***
```

```
## regionSouth              0.0957      0.0425     2.25    0.024 *
## regionWest               -0.8952     0.0572    -15.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 5352.15  on 15  degrees of freedom
## Residual deviance:  563.75  on 10  degrees of freedom
## AIC: 665.2
##
## Number of Fisher Scoring iterations: 5
```

```r
qchisq(.95,model3$df.residual)
```

```
## [1] 18.3
```

```r
1-pchisq(model3$deviance,model2$df.residual)
```

```
## [1] 0
```

```r
anova(model3,model2,test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: count ~ covid_rate_high + vacc_rate_high + region
## Model 2: count ~ covid_rate_high + vacc_rate_high + region + covid_rate_high:vacc_rate_high:region
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        10        564
## 2         0          0 10      564   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
cat("\n563.73>18.3 so we reject the H0: model without three-way interaction is better. So the associati
region\n")
```

```
##
## 563.73>18.3 so we reject the H0: model without three-way interaction is better. So the association be
## region
```

```r
cat("\n(f)\n")
```

```
##
## (f)
```

```r
cat("\n(g)\n")
```

```
##
## (g)
```

```r
cat("The model with only 1 three-way interaction is better. conditional independence")
```

```
## The model with only 1 three-way interaction is better. conditional independence
```