Assignment 4

Bharath Dasari

21/09/2021

Data Cleansing

```
library(car)
```

```
## Loading required package: carData
```

```
#data$Region <- as.factor(data$Region)
data$Region_NE <- recode(data$Region,"1=1;c(2,3,4)=0")
data$Region_NC <- recode(data$Region,"2=1;c(1,3,4)=0")
data$Region_S <- recode(data$Region,"3=1;c(1,2,4)=0")
data$Region_W <- recode(data$Region,"4=1;c(1,3,2)=0")
data$PopulationDensity <-  data$`Total Population`/data$`Land area`
str(data)
```

```
## 'data.frame':    440 obs. of  22 variables:
##  $ ID                    : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ County                : chr  "Los_Angeles" "Cook" "Harris" "San_Diego" ...
##  $ State                 : chr  "CA" "IL" "TX" "CA" ...
##  $ Land area             : num  4060 946 1729 4205 790 ...
##  $ Total Population      : num  8863164 5105067 2818199 2498016 2410556 ...
##  $ POP aged 18-34        : num  32.1 29.2 31.3 33.5 32.6 28.3 29.2 27.4 27.1 32.6 ...
##  $ POP 65 or older       : num  9.7 12.4 7.1 10.9 9.2 12.4 12.5 12.5 13.9 8.2 ...
##  $ Physicians            : num  23677 15153 7553 5905 6062 ...
##  $ Hospital Beds         : num  27700 21550 12449 6179 6369 ...
##  $ Crimes                : num  688936 436936 253526 173821 144524 ...
##  $ High school graduates: num  70 73.4 74.9 81.9 81.2 63.7 81.5 70 65 77.1 ...
##  $ Bachelor degrees      : num  22.3 22.8 25.4 25.3 27.8 16.6 22.1 13.7 18.8 26.3 ...
##  $ Below poverty level   : num  11.6 11.1 12.5 8.1 5.2 19.5 8.8 16.9 14.2 10.4 ...
##  $ Unemployement         : num  8 7.2 5.7 6.1 4.8 9.5 4.9 10 8.7 6.1 ...
##  $ Percapita income      : num  20786 21729 19517 19588 24400 ...
##  $ Personal income       : num  184230 110928 55003 48931 58818 ...
##  $ Region                : num  4 2 3 4 4 1 4 2 3 3 ...
##  $ Region_NE             : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ Region_NC             : num  0 1 0 0 0 0 0 1 0 0 ...
##  $ Region_S              : num  0 0 1 0 0 0 0 0 1 1 ...
##  $ Region_W              : num  1 0 0 1 1 0 1 0 0 0 ...
##  $ PopulationDensity     : num  2183 5396 1630 594 3051 ...
```

Baseline Category Logit model with 1=NE as the baseline category

```
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
##
## Attaching package: 'VGAM'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
attach(data)
fit.m1<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
summary(fit.m1)
```

```
##
## Call:
## vglm(formula = cbind(Region_NC, Region_S, Region_W, Region_NE) ~
##     PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
##         Physicians + `Hospital Beds` + Crimes + `High school graduates` +
##         `Bachelor degrees` + `Below poverty level` + Unemployement +
##         `Percapita income` + `Personal income`, family = multinomial,
##     data = data)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept):1             -1.648e+01  6.143e+00      NA       NA
## (Intercept):2              1.156e+01  5.168e+00   2.237 0.025289 *
## (Intercept):3             -2.238e+01  7.333e+00      NA       NA
## PopulationDensity:1       -2.057e-04  1.785e-04  -1.152 0.249146
## PopulationDensity:2       -3.648e-04  1.320e-04  -2.763 0.005725 **
## PopulationDensity:3       -6.680e-04  2.215e-04  -3.015 0.002566 **
## `POP aged 18-34`:1        -4.554e-02  7.379e-02  -0.617 0.537148
## `POP aged 18-34`:2        -9.320e-02  7.014e-02  -1.329 0.183952
## `POP aged 18-34`:3        -2.251e-01  8.426e-02  -2.671 0.007561 **
## `POP 65 or older`:1       -1.593e-01  5.952e-02  -2.676 0.007455 **
## `POP 65 or older`:2       -1.926e-02  4.902e-02  -0.393 0.694426
## `POP 65 or older`:3       -1.240e-01  7.027e-02  -1.764 0.077717 .
## Physicians:1              -4.138e-04  5.403e-04  -0.766 0.443729
## Physicians:2              -3.281e-04  4.937e-04  -0.665 0.506350
## Physicians:3               1.560e-03  6.917e-04   2.255 0.024102 *
## `Hospital Beds`:1          5.060e-04  3.790e-04   1.335 0.181873
## `Hospital Beds`:2         -6.693e-04  3.468e-04  -1.930 0.053592 .
## `Hospital Beds`:3         -2.904e-03  5.985e-04  -4.852 1.22e-06 ***
## Crimes:1                  -1.753e-05  1.507e-05  -1.163 0.244739
## Crimes:2                   1.764e-05  9.418e-06      NA       NA
## Crimes:3                   4.468e-05  1.420e-05   3.147 0.001648 **
## `High school graduates`:1  2.815e-01  6.056e-02   4.647 3.37e-06 ***
## `High school graduates`:2 -1.108e-01  5.062e-02  -2.188 0.028690 *
## `High school graduates`:3  3.537e-01  7.068e-02   5.005 5.60e-07 ***
## `Bachelor degrees`:1      -2.068e-01  6.545e-02  -3.159 0.001581 **
## `Bachelor degrees`:2       1.485e-01  5.774e-02   2.572 0.010101 *
## `Bachelor degrees`:3      -5.721e-02  6.740e-02  -0.849 0.395971
## `Below poverty level`:1    4.418e-01  9.789e-02   4.513 6.40e-06 ***
## `Below poverty level`:2    5.404e-01  9.252e-02   5.841 5.20e-09 ***
## `Below poverty level`:3    8.063e-01  1.131e-01   7.129 1.01e-12 ***
## Unemployement:1           -3.718e-01  1.110e-01  -3.350 0.000807 ***
## Unemployement:2           -6.615e-01  1.044e-01  -6.338 2.33e-10 ***
## Unemployement:3           -1.641e-01  1.088e-01  -1.508 0.131563
## `Percapita income`:1       6.544e-05  1.025e-04   0.638 0.523395
## `Percapita income`:2      -1.187e-04  9.389e-05  -1.265 0.205980
## `Percapita income`:3      -8.305e-05  1.116e-04  -0.744 0.456829
## `Personal income`:1        3.617e-05  7.263e-05   0.498 0.618480
## `Personal income`:2        8.897e-05  6.397e-05   1.391 0.164240
## `Personal income`:3        1.625e-04  7.739e-05   2.099 0.035785 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Names of linear predictors: log(mu[,1]/mu[,4]), log(mu[,2]/mu[,4]),
## log(mu[,3]/mu[,4])
##
## Residual deviance: 801.8632 on 1281 degrees of freedom
##
## Log-likelihood: -400.9316 on 1281 degrees of freedom
##
## Number of Fisher scoring iterations: 6
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1', '(Intercept):3', '`Hospital Beds`:3', 'Crimes:2'
##
##
## Reference group is level  4  of the response
```

Interpretation of variables:

1. For each unit increase in Population density of a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.999794321154794 controlling for other variables.
2. For each unit increase in Population density of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.99963526653143 controlling for other variables.
3. For each unit increase in Population density of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.999332223062329 controlling for other variables.
4. For each unit increase in POP aged 18-34 of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.955481382552257 controlling for other variables.
5. For each unit increase in POP aged 18-34 of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.911011279818736 controlling for other variables.
6. For each unit increase in POP aged 18-34 of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.798436371129949 controlling for other variables.
7. For each unit increase in POP 65 or older of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.852740498442439 controlling for other variables.
8. For each unit increase in POP 65 or older of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.980924288769611 controlling for other variables.
9. For each unit increase in POP 65 or older of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.883379840882751 controlling for other variables.

10. For each unit increase in Physicians of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.999586285603412 controlling for other variables.
11. For each unit increase in Physicians of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.999671953818919 controlling for other variables.
12. For each unit increase in Physicians of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 1.00156121743298 controlling for other variables.
13. For each unit increase in Hospital Beds of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 1.0005061280396 controlling for other variables.
14. For each unit increase in Hospital Beds of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.999330923931283 controlling for other variables.
15. For each unit increase in Hospital Beds of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.997100212529285 controlling for other variables.
16. For each unit increase in Crimes of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.99998247015365 controlling for other variables.
17. For each unit increase in Crimes of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 1.00001764015559 controlling for other variables.
18. For each unit increase in Crimes of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 1.00004468099817 controlling for other variables.
19. For each unit increase in High school graduates of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 1.32511599632152 controlling for other variables.
20. For each unit increase in High school graduates of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.895117754578785 controlling for other variables.
21. For each unit increase in High school graduates of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 1.42432782403152 controlling for other variables.
22. For each unit increase in Bachelor degrees of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.813182270179046 controlling for other variables.
23. For each unit increase in Bachelor degrees of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 1.16009279777443 controlling for other variables.

24. For each unit increase in Bachelor degrees of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.944395725438798 controlling for other variables.
25. For each unit increase in Below poverty level of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 1.55550460840026 controlling for other variables.
26. For each unit increase in Below poverty level of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 1.71669340222859 controlling for other variables.
27. For each unit increase in Below poverty level of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 2.23960609509626 controlling for other variables.
28. For each unit increase in Unemployment of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 0.689492127160731 controlling for other variables.
29. For each unit increase in Unemployment of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.516076638657093 controlling for other variables.
30. For each unit increase in Unemployment of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.848657151921587 controlling for other variables.
31. For each unit increase in Percapita income of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 1.00006544214124 controlling for other variables.
32. For each unit increase in Percapita income of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 0.999881307044566 controlling for other variables.
33. For each unit increase in Percapita income of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 0.999916953448556 controlling for other variables.
34. For each unit increase in Personal income of a county in a county in NC region wrt to county in NE region, the cumulative odds of county in the NE region being more popular than the county in NC region changes multiplicatively by a factor of 1.00003617065414 controlling for other variables.
35. For each unit increase in Personal income of a county in S region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in S region changes multiplicatively by a factor of 1.00008897395795 controlling for other variables.
36. For each unit increase in Personal income of a county in W region wrt to NE region, the cumulative odds of county in the NE region being more popular than the county in W region changes multiplicatively by a factor of 1.00016251320384 controlling for other variables.

Liklihood Ratio tests:

1.Without PopulationDensity

```
fit.m2<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~  `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m2 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ `POP aged 18-34` +
##     `POP 65 or older` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     815.44 -3  -13.572 0.003549 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Population Density is not a significant variable
Ha: Population Density is a significant variable

Since p-value is less than $\alpha = 0.01$, we reject Ho.
Therefore, Population Density is a significant variable.

2.Without POP aged 18-34

```
fit.m3<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity +  `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m3 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP 65 or older` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     811.36 -3   -9.499  0.02334 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: POP aged 18-34 is not a significant variable
Ha: POP aged 18-34 is a significant variable

Since p-value is greater than α = 0.01, we fail to reject Ho.
Therefore, POP aged 18-34 is not a significant variable.

3.Without POP 65 or older

```
fit.m4<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34`  +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m4 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     812.17 -3  -10.309  0.01611 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: POP 65 or older is not a significant variable
Ha: POP 65 or older is a significant variable

Since p-value is greater than α = 0.01, we fail to reject Ho.
Therefore, POP 65 or older is not a significant variable.

4.Without Physicians

```
fit.m5<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
                `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degrees` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m5 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     812.10 -3  -10.236  0.01666 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Physicians is not a significant variable
Ha: Physicians is a significant variable

Since p-value is greater than α = 0.01, we fail to reject Ho.
Therefore, Physicians is not a significant variable.

5.Without Hospital Beds

```
fit.m6<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians  + Crimes + `High school graduates` + `Bachelor degrees` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m6 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1281     801.86
## 2      1284     860.65 -3  -58.787 1.068e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Hospital Beds is not a significant variable
Ha: Hospital Beds is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, Hospital Beds is a significant variable.

6.Without Crimes

```
fit.m7<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds`  + `High school graduates` + `Bachelor degrees` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m7 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1281     801.86
## 2      1284     820.34 -3  -18.473 0.0003514 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Crimes is not a significant variable
Ha: Crimes is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, Crimes is a significant variable.

7.Without High school graduates

```
fit.m8<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes +  `Bachelor degrees` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m8 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `Bachelor degrees` + `Below poverty level` + Unemployement +
##     `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1281     801.86
## 2      1284     903.10 -3  -101.24 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: High school graduates is not a significant variable
Ha: High school graduates is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, High school graduates is a significant variable.

8.Without Bachelor degrees

```
fit.m9<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = multinomial, data=data)
anova(fit.m1 , fit.m9 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1281     801.86
## 2      1284     844.33 -3  -42.467 3.193e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Bachelor degrees is not a significant variable
Ha: Bachelor degrees is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, Bachelor degrees is a significant variable.

9.Without Below poverty level

```
fit.m10<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
              ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
                Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
                 Unemployement + `Percapita income` + `Personal income` ,
              family = multinomial, data=data)
anova(fit.m1 , fit.m10 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + Unemployement +
##     `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     871.59 -3  -69.724 4.89e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Below poverty level is not a significant variable
Ha: Below poverty level is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, Below poverty level is a significant variable.

10.Without Unemployement

```
fit.m11<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
              ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
                Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
                `Below poverty level`  + `Percapita income` + `Personal income` ,
              family = multinomial, data=data)
anova(fit.m1 , fit.m11 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     `Percapita income` + `Personal income`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1281     801.86
## 2      1284     860.17 -3  -58.302 1.355e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ho: Unemployement is not a significant variable
Ha: Unemployement level is a significant variable

Since p-value is less than α = 0.01, we reject Ho.
Therefore, Unemployement is a significant variable.

11.Without Percapita income

```
fit.m12<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
              ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
                Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
                `Below poverty level` + Unemployement +  `Personal income` ,
              family = multinomial, data=data)
anova(fit.m1 , fit.m12 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Personal income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     805.84 -3   -3.977    0.264
```

Ho: Percapita income is not a significant variable
Ha: Percapita income level is a significant variable

Since p-value is greater than α = 0.01, we fail to reject Ho.
Therefore, Percapita income is a not significant variable.

12.Without Personal income

```
fit.m13<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
              ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
                Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
                `Below poverty level` + Unemployement + `Percapita income`  ,
              family = multinomial, data=data)
anova(fit.m1 , fit.m13 , type = 1)
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`
## Model 2: cbind(Region_NC, Region_S, Region_W, Region_NE) ~ PopulationDensity +
##     `POP aged 18-34` + `POP 65 or older` + Physicians + `Hospital Beds` +
##     Crimes + `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income`
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1281     801.86
## 2      1284     806.83 -3  -4.9673   0.1742
```

Ho: Personal income is not a significant variable
Ha: Personal income level is a significant variable

Since p-value is greater than α = 0.01, we fail to reject Ho.
Therefore, Personal income is not a significant variable.

Predictors those should be retained are:
Population Density
Hospital Beds

Crimes
High school graduates
Bachelor degrees
Below poverty level
Unemployement

Binary logistic regressions

```
fit.b.1<-glm(Region_NC
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = binomial(link = "logit"), data=data)
summary(fit.b.1)
```

```
##
## Call:
## glm(formula = Region_NC ~ PopulationDensity + `POP aged 18-34` +
##     `POP 65 or older` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`, family = binomial(link = "logi
t"),
##     data = data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.57530  -0.70266  -0.42632  -0.01544   2.81765
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -1.861e+01  4.748e+00  -3.920 8.87e-05 ***
## PopulationDensity       -1.041e-04  1.642e-04  -0.634   0.5260
## `POP aged 18-34`         7.972e-02  5.296e-02   1.505   0.1323
## `POP 65 or older`       -1.002e-01  5.005e-02  -2.001   0.0454 *
## Physicians              -6.018e-04  4.584e-04  -1.313   0.1893
## `Hospital Beds`          1.369e-03  3.178e-04   4.307 1.65e-05 ***
## Crimes                  -2.982e-05  1.404e-05  -2.125   0.0336 *
## `High school graduates`  2.494e-01  4.567e-02   5.460 4.76e-08 ***
## `Bachelor degrees`      -2.812e-01  5.283e-02  -5.322 1.02e-07 ***
## `Below poverty level`   -1.304e-02  6.249e-02  -0.209   0.8347
## Unemployement           -7.637e-02  8.873e-02  -0.861   0.3894
## `Percapita income`       1.602e-04  8.251e-05   1.942   0.0522 .
## `Personal income`       -4.624e-05  6.194e-05  -0.746   0.4554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 490.41  on 439  degrees of freedom
## Residual deviance: 392.20  on 427  degrees of freedom
## AIC: 418.2
##
## Number of Fisher Scoring iterations: 6
```

```
fit.b.2<-glm(Region_S
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = binomial(link = "logit"), data=data)
summary(fit.b.2)
```

```
##
## Call:
## glm(formula = Region_S ~ PopulationDensity + `POP aged 18-34` +
##     `POP 65 or older` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`, family = binomial(link = "logi
t"),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4348  -0.7439  -0.4668   0.6622   2.4728
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)              1.843e+01  4.032e+00   4.571 4.86e-06 ***
## PopulationDensity       -2.801e-04  1.014e-04  -2.761  0.00576 **
## `POP aged 18-34`         1.249e-02  4.695e-02   0.266  0.79017
## `POP 65 or older`        5.599e-02  4.020e-02   1.393  0.16375
## Physicians              -5.582e-04  3.763e-04  -1.484  0.13793
## `Hospital Beds`         -9.512e-06  2.294e-04  -0.041  0.96693
## Crimes                   1.600e-05  6.657e-06   2.404  0.01623 *
## `High school graduates` -2.602e-01  3.983e-02  -6.534 6.41e-11 ***
## `Bachelor degrees`       1.818e-01  4.201e-02   4.327 1.51e-05 ***
## `Below poverty level`    1.065e-01  5.408e-02   1.969  0.04900 *
## Unemployement           -4.698e-01  8.128e-02  -5.780 7.49e-09 ***
## `Percapita income`      -7.075e-05  7.265e-05  -0.974  0.33012
## `Personal income`       -1.134e-05  4.757e-05  -0.238  0.81167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 567.24  on 439  degrees of freedom
## Residual deviance: 418.72  on 427  degrees of freedom
## AIC: 444.72
##
## Number of Fisher Scoring iterations: 5
```

```
fit.b.3<-glm(Region_W
             ~ PopulationDensity + `POP aged 18-34` + `POP 65 or older` +
               Physicians + `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degre
es` +
               `Below poverty level` + Unemployement + `Percapita income` + `Personal income` ,
             family = binomial(link = "logit"), data=data)
summary(fit.b.3)
```

```
##
## Call:
## glm(formula = Region_W ~ PopulationDensity + `POP aged 18-34` +
##     `POP 65 or older` + Physicians + `Hospital Beds` + Crimes +
##     `High school graduates` + `Bachelor degrees` + `Below poverty level` +
##     Unemployement + `Percapita income` + `Personal income`, family = binomial(link = "logi
t"),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6781  -0.5213  -0.2964  -0.0849   4.0703
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -2.303e+01  5.881e+00  -3.915 9.03e-05 ***
## PopulationDensity       -5.259e-04  1.867e-04  -2.817 0.004854 **
## `POP aged 18-34`        -1.425e-01  5.859e-02  -2.433 0.014993 *
## `POP 65 or older`       -3.851e-02  5.591e-02  -0.689 0.490946
## Physicians               1.659e-03  5.987e-04   2.770 0.005600 **
## `Hospital Beds`         -2.495e-03  5.289e-04  -4.717 2.39e-06 ***
## Crimes                   4.010e-05  1.176e-05   3.409 0.000652 ***
## `High school graduates`  3.100e-01  5.589e-02   5.546 2.93e-08 ***
## `Bachelor degrees`      -7.326e-02  4.966e-02  -1.475 0.140153
## `Below poverty level`    3.101e-01  7.324e-02   4.234 2.30e-05 ***
## Unemployement            2.397e-01  8.200e-02   2.923 0.003468 **
## `Percapita income`      -6.304e-05  8.903e-05  -0.708 0.478877
## `Personal income`        1.044e-04  5.885e-05   1.774 0.076035 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 408.08  on 439  degrees of freedom
## Residual deviance: 282.47  on 427  degrees of freedom
## AIC: 308.47
##
## Number of Fisher Scoring iterations: 7
```

Comparison of slope coefficients wrt to baseline model

Variable Baseline Binary
PopulationDensity - NC region -0.00021 -0.00010
PopulationDensity - S region -0.00036 -0.00028
PopulationDensity - W region -0.00067 -0.00053
POP aged 8-34 - NC region -0.04554 0.07972
POP aged 8-34 - S region -0.09320 0.01249
POP aged 8-34 - W region -0.22510 -0.14250
POP 65 or older - NC region -0.15930 -0.10020
POP 65 or older - S region -0.01926 0.05599
POP 65 or older - W region -0.12400 -0.03851
Physicians - NC region -0.00041 -0.00060

Physicians - S region -0.00033 -0.00056
Physicians - W region 0.00156 0.00166
Hospital Beds - NC region 0.00051 0.00137
Hospital Beds - S region -0.00067 -0.00001
Hospital Beds - W region -0.00290 -0.00250
Crimes - NC region -0.00002 -0.00003
Crimes - S region 0.00002 0.00002
Crimes - W region 0.00004 0.00004
High school graduates-NC region 0.28150 0.24940
High school graduates-S region -0.11080 -0.26020
High school graduates-W region 0.35370 0.31000
Bachelor degrees - NC region -0.20680 -0.01304
Bachelor degrees - S region 0.14850 0.18180
Bachelor degrees - W region -0.05721 -0.07326
Below poverty level-NC region 0.44180 -0.01558
Below poverty level-S region 0.54040 0.10650
Below poverty level-W region 0.80630 0.31010
Unemployement - NC region -0.37180 -0.07637
Unemployement - S region -0.66150 -0.46980
Unemployement - W region -0.16410 0.23970
Percapita income - NC region 0.00007 0.00016
Percapita income - S region -0.00012 -0.00007
Percapita income - W region -0.00008 -0.00006
Personal income - NC region 0.00004 -0.000046
Personal income - S region 0.00009 -0.00001
Personal income - W region 0.00016 0.00010

optimal Baseline Category Logit model

```
fit.final<-vglm(cbind(Region_NC,Region_S,Region_W,Region_NE)
             ~ PopulationDensity +
              `Hospital Beds` + Crimes + `High school graduates` + `Bachelor degrees` +
               `Below poverty level` + Unemployement ,
             family = multinomial, data=data)
summary(fit.final)
```

```
##
## Call:
## vglm(formula = cbind(Region_NC, Region_S, Region_W, Region_NE) ~
##     PopulationDensity + `Hospital Beds` + Crimes + `High school graduates` +
##         `Bachelor degrees` + `Below poverty level` + Unemployement,
##     family = multinomial, data = data)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept):1             -1.918e+01  4.283e+00  -4.477 7.56e-06 ***
## (Intercept):2              6.527e+00  3.639e+00   1.794 0.072878 .
## (Intercept):3             -2.786e+01  5.197e+00  -5.362 8.24e-08 ***
## PopulationDensity:1       -2.276e-04  1.530e-04  -1.488 0.136794
## PopulationDensity:2       -4.567e-04  1.333e-04  -3.426 0.000613 ***
## PopulationDensity:3       -8.309e-04  2.070e-04  -4.013 5.99e-05 ***
## `Hospital Beds`:1          2.503e-04  2.367e-04   1.058 0.290126
## `Hospital Beds`:2         -7.096e-04  2.375e-04  -2.988 0.002810 **
## `Hospital Beds`:3         -1.145e-03  3.036e-04  -3.773 0.000161 ***
## Crimes:1                  -8.281e-06  1.132e-05  -0.732 0.464369
## Crimes:2                   3.022e-05  1.131e-05   2.672 0.007532 **
## Crimes:3                   5.693e-05  1.369e-05   4.158 3.21e-05 ***
## `High school graduates`:1  2.849e-01  5.497e-02   5.183 2.18e-07 ***
## `High school graduates`:2 -9.109e-02  4.723e-02  -1.929 0.053756 .
## `High school graduates`:3  3.076e-01  6.211e-02   4.952 7.35e-07 ***
## `Bachelor degrees`:1      -1.717e-01  4.106e-02  -4.181 2.90e-05 ***
## `Bachelor degrees`:2       8.725e-02  3.458e-02   2.524 0.011617 *
## `Bachelor degrees`:3      -1.777e-02  3.800e-02  -0.468 0.640021
## `Below poverty level`:1    3.912e-01  7.167e-02   5.458 4.81e-08 ***
## `Below poverty level`:2    5.067e-01  6.977e-02   7.263 3.77e-13 ***
## `Below poverty level`:3    6.325e-01  8.108e-02   7.800 6.18e-15 ***
## Unemployement:1           -3.525e-01  1.018e-01  -3.463 0.000535 ***
## Unemployement:2           -6.460e-01  9.667e-02  -6.683 2.34e-11 ***
## Unemployement:3           -7.633e-02  1.004e-01  -0.760 0.447086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,4]), log(mu[,2]/mu[,4]),
## log(mu[,3]/mu[,4])
##
## Residual deviance: 867.2641 on 1296 degrees of freedom
##
## Log-likelihood: -433.632 on 1296 degrees of freedom
##
## Number of Fisher scoring iterations: 6
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1', '(Intercept):3'
##
##
## Reference group is level  4  of the response
```