StateByState

Zongyan Wang

Saturday, December 12, 2015

This document gives an overall analyst to the b/g ratio over States

Load packages

```
require(data.table)
```

```
## Loading required package: data.table
```

```
## Warning: package 'data.table' was built under R version 3.1.3
```

```
require(agricolae)
```

```
## Loading required package: agricolae
```

```
## Warning: package 'agricolae' was built under R version 3.1.3
```

```
require(plyr)
```

```
## Loading required package: plyr
```

```
## Warning: package 'plyr' was built under R version 3.1.3
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
## Warning: package 'dplyr' was built under R version 3.1.3
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
## The following objects are masked from 'package:data.table':
##
##     between, last
##
## The following objects are masked from 'package:stats':
```

```
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Load data

```
partD_npi <- read.csv("F:/Academic/Stat 992/group project/prescription_fraud/analysis.py/NPI_bg.csv")
names(partD_npi)
```

```
##  [1] "b.g"                          "b.g.with.claim"
##  [3] "NPI"                          "NPPES_PROVIDER_LAST_ORG_NAME"
##  [5] "NPPES_PROVIDER_FIRST_NAME"    "NPPES_PROVIDER_CITY"
##  [7] "NPPES_PROVIDER_STATE"         "SPECIALTY_DESC"
##  [9] "DESCRIPTION_FLAG"             "BENE_COUNT"
## [11] "TOTAL_CLAIM_COUNT"            "TOTAL_DAY_SUPPLY"
## [13] "TOTAL_DRUG_COST"              "BENE_COUNT_GE65"
## [15] "BENE_COUNT_GE65_REDACT_FLAG"  "TOTAL_CLAIM_COUNT_GE65"
## [17] "GE65_REDACT_FLAG"             "TOTAL_DAY_SUPPLY_GE65"
## [19] "TOTAL_DRUG_COST_GE65"
```

ANOVA without counting claim

```
data <- cbind(partD_npi,trt=factor(partD_npi$NPPES_PROVIDER_STATE,labels=1:length(unique(partD_npi$NPPES
model <- aov(b.g ~ trt, data = data)
summary(model)
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## trt             60    234   3.903   55.26 <2e-16 ***
## Residuals   807959  57067   0.071
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model shows significant result to States ## Doing LSD and HSD test

```
lsd.out <- LSD.test(model, "trt", p.adj = "bonferroni")
hsd.out <- HSD.test(model,"trt", group = T)

#bar.group(lsd.out$groups,ylim=c(0,45),density=4,border="blue")
#bar.group(hsd.out$groups,ylim=c(0,45),density=4,border="blue")

#lsd.out$groups
#hsd.out$groups
```

match the group number with States with the data we use

```
id.group <- data %>%
  dplyr::select(NPPES_PROVIDER_STATE, trt) %>%
  unique() %>%
  arrange(trt)
groups <- as.data.frame(hsd.out$groups)
groups$trt <- as.integer(groups$trt)
id.group$trt <- as.integer(id.group$trt)
result <- plyr::join(groups, id.group)
```

```
## Joining by: trt
```

```
result
```

```
##    trt     means  M NPPES_PROVIDER_STATE
## 1   57 0.8296296  a                   WI
## 2    1 0.7536630  a                   AA
## 3   59 0.7460317  a                   WY
## 4    8 0.7435175  a                   AZ
## 5   41 0.7099463  a                   NY
## 6   58 0.7042467 ab                   WV
## 7   18 0.6931382 ab                   IA
## 8   50 0.6802200 ab                   TN
## 9   25 0.6795626 ab                   MA
## 10  22 0.6737069 ab                   KS
## 11  51 0.6691765 ab                   TX
## 12  20 0.6686648 ab                   IL
## 13  31 0.6632435 ab                   MP
## 14  10 0.6620653 ab                   CO
## 15  45 0.6607997 ab                   PA
## 16  19 0.6589100 ab                   ID
## 17  42 0.6584061 ab                   OH
## 18  33 0.6558891 ab                   MT
## 19   4 0.6534401 ab                   AL
## 20  53 0.6524948 ab                   VA
## 21  44 0.6508540 ab                   OR
## 22  30 0.6506284 ab                   MO
## 23   3 0.6492230 ab                   AK
## 24  52 0.6489724 ab                   UT
## 25  61 0.6479112 ab                   ZZ
## 26  36 0.6475830 ab                   NE
## 27  29 0.6462845 ab                   MN
## 28  40 0.6455697 ab                   NV
## 29   6 0.6453119 ab                   AR
## 30  35 0.6446090 ab                   ND
## 31  48 0.6442566 ab                   SC
## 32  21 0.6441525 ab                   IN
## 33   9 0.6422798 ab                   CA
## 34  13 0.6404267 ab                   DE
## 35  49 0.6398098 ab                   SD
## 36  37 0.6395605 ab                   NH
```

```
## 37  24 0.6385870 ab                  LA
## 38  16 0.6380884 ab                  GU
## 39  15 0.6362533 ab                  GA
## 40  27 0.6347261 ab                  ME
## 41   2 0.6345081 ab                  AE
## 42  39 0.6330565 ab                  NM
## 43  28 0.6327925 ab                  MI
## 44  60 0.6299534 ab                  XX
## 45  34 0.6294399 ab                  NC
## 46  43 0.6276517 ab                  OK
## 47  56 0.6271792 ab                  WA
## 48  11 0.6267211 ab                  CT
## 49  46 0.6261382 ab                  PR
## 50   5 0.6253142 ab                  AP
## 51  54 0.6246399 ab                  VI
## 52  32 0.6230888 ab                  MS
## 53   7 0.6223479 ab                  AS
## 54  14 0.6215336 ab                  FL
## 55  26 0.6209656 ab                  MD
## 56  55 0.6207932 ab                  VT
## 57  47 0.6204590 ab                  RI
## 58  38 0.6169197 ab                  NJ
## 59  23 0.6104840 ab                  KY
## 60  17 0.6103561 ab                  HI
## 61  12 0.5715370 ab                  DC
```

ANOVA with counting claim

```
data <- cbind(partD_npi,trt=factor(partD_npi$NPPES_PROVIDER_STATE,labels=1:length(unique(partD_npi$NPPES
model <- aov(b.g.with.claim ~ trt, data = data)
summary(model)
```

```
##                 Df    Sum Sq Mean Sq F value Pr(>F)
## trt             60 1.019e+08 1698278     134 <2e-16 ***
## Residuals   807959 1.024e+10   12675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model shows significant result to States ## Doing LSD and HSD test

```
lsd.out <- LSD.test(model, "trt", p.adj = "bonferroni")
hsd.out <- HSD.test(model,"trt", group = T)

#bar.group(lsd.out$groups,ylim=c(0,45),density=4,border="blue")
#bar.group(hsd.out$groups,ylim=c(0,45),density=4,border="blue")

#lsd.out$groups
#hsd.out$groups
```

match the group number with States with the data we use

```
id.group <- data %>%
  dplyr::select(NPPES_PROVIDER_STATE, trt) %>%
  unique() %>%
  arrange(trt)
groups <- as.data.frame(hsd.out$groups)
groups$trt <- as.integer(groups$trt)
id.group$trt <- as.integer(id.group$trt)
result <- plyr::join(groups, id.group)
```

```
## Joining by: trt
```

```
result
```

```
##    trt     means     M NPPES_PROVIDER_STATE
## 1   41 94.588324     a                   NY
## 2   34 79.634906     b                   NC
## 3   56 72.097695     c                   WA
## 4   26 67.739304    cd                   MD
## 5   16 64.588452    de                   GU
## 6   10 61.551558   def                   CO
## 7   46 61.181660    ef                   PR
## 8   54 60.247493   efg                   VI
## 9    6 59.412172    fg                   AR
## 10   7 58.030295    fg                   AS
## 11  24 56.585151   fgh                   LA
## 12  28 54.672548   ghi                   MI
## 13  43 54.398429  ghij                   OK
## 14  17 54.165338  ghij                   HI
## 15  15 53.394622  ghij                   GA
## 16  14 51.431418  hijk                   FL
## 17  30 50.739024 hijkl                   MO
## 18  38 49.895729  ijkl                   NJ
## 19  37 49.478959   jkl                   NH
## 20  35 48.827277  jklm                   ND
## 21  29 48.100536 jklmn                   MN
## 22  44 47.783472 jklmn                   OR
## 23  40 47.585188  klmn                   NV
## 24  42 46.637912  klmn                   OH
## 25  47 45.725775   lmn                   RI
## 26  49 44.721457   lmn                   SD
## 27  11 43.562404  lmno                   CT
## 28  61 43.001584   mno                   ZZ
## 29  27 42.754208  mnop                   ME
## 30  53 42.411970  mnop                   VA
## 31  39 42.214424  mnop                   NM
## 32  60 41.161856   nop                   XX
## 33   9 40.064134  nopq                   CA
## 34  13 39.604225  nopq                   DE
## 35  33 39.155338  nopq                   MT
## 36  21 38.944621  nopq                   IN
```

```
## 37  22 38.857150  nopq                KS
## 38  20 38.226426  nopq                IL
## 39  36 37.668456   opq                NE
## 40  51 37.251280  opqr                TX
## 41  52 37.223366  opqr                UT
## 42  32 36.700499  opqr                MS
## 43   5 34.830763 opqrs                AP
## 44  48 34.106327 opqrs                SC
## 45  18 33.723717  pqrs                IA
## 46  55 32.463806  pqrs                VT
## 47   2 31.966822   qrs                AE
## 48   3 31.129564   qrs                AK
## 49   8 30.883649  qrst                AZ
## 50  31 30.392160  qrst                MP
## 51  19 27.389842   rst                ID
## 52  50 27.241004   rst                TN
## 53  59 25.824322   rst                WY
## 54  25 22.888835   rst                MA
## 55  23 21.081185    st                KY
## 56  58 17.593594    st                WV
## 57   4 17.087016     t                AL
## 58  12  8.538531     t                DC
## 59  57  6.279915     t                WI
## 60   1  3.832884     t                AA
## 61  45  2.962611     t                PA
```