PH241 HW8

Kunal Mishra

3/14/2018

Question 1A

```
# Setting up Multilevel data
multilevelAlcoholConsumption = array(
    c(29, 75, 51, 45,
      386, 280, 87, 22),
    dim=c(4,2),
    dimnames=list( c("Alcohol Consumption - 0-39g/day", "40-79", "80-119", ">120"),
                   c("Esophegeal Cancer - Case", "Control")
    )
)
multilevelAlcoholConsumption
```

```
##                                 Esophegeal Cancer - Case Control
## Alcohol Consumption - 0-39g/day                       29     386
## 40-79                                                 75     280
## 80-119                                                51      87
## >120                                                  45      22
```

```
chisq.test(multilevelAlcoholConsumption,
           correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  multilevelAlcoholConsumption
## X-squared = 158.95, df = 3, p-value < 2.2e-16
```

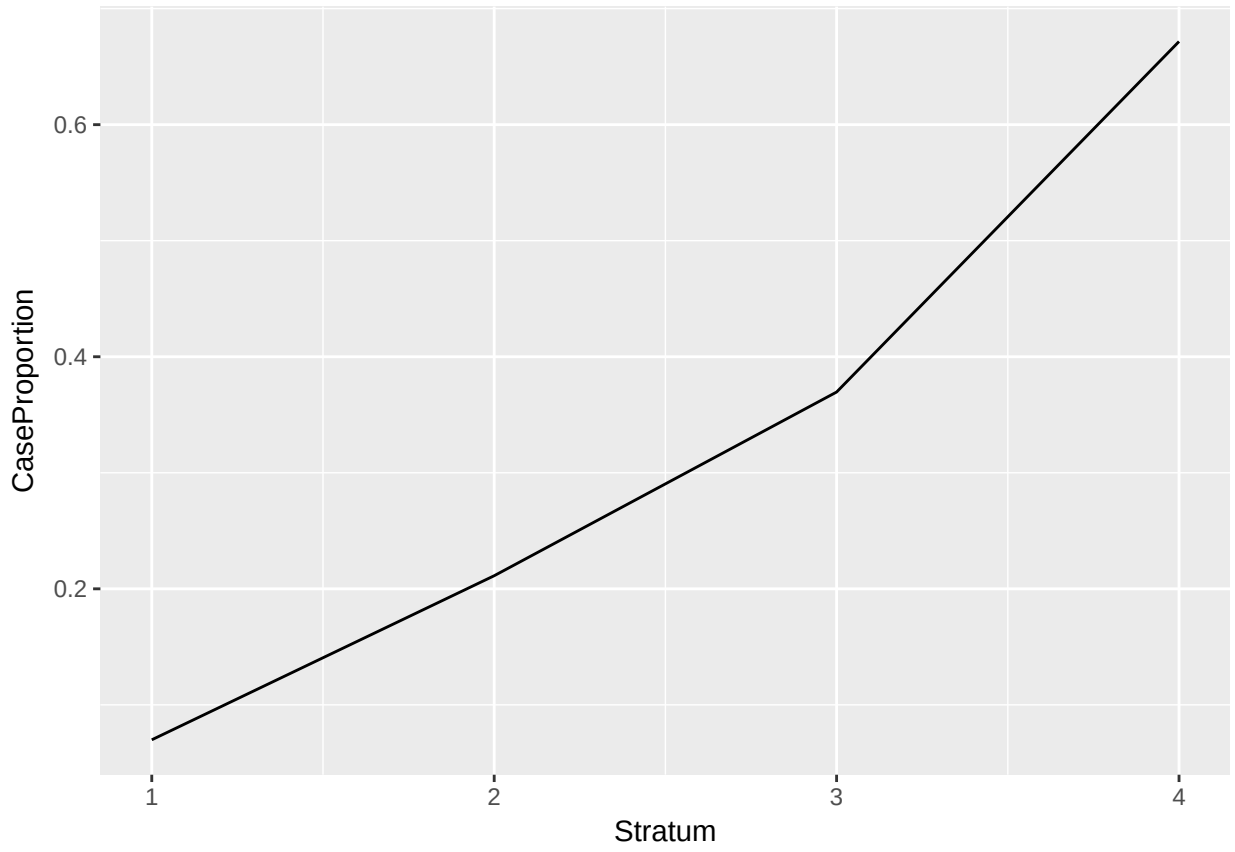
Question 1D

```
library(ggplot2)
library(data.table)

caseProportions = as.data.table(multilevelAlcoholConsumption)
colnames(caseProportions)[1] = "Case"
caseProportions[, CaseProportion := Case/(Case+Control)]
caseProportions
```

```
##    Case Control CaseProportion
## 1:   29     386     0.06987952
## 2:   75     280     0.21126761
## 3:   51      87     0.36956522
## 4:   45      22     0.67164179
```

```
ggplot(data=caseProportions[, .(CaseProportion)][, Stratum:=1:4],
       aes(x=Stratum, y=CaseProportion, group=1)) +
    geom_line()
```



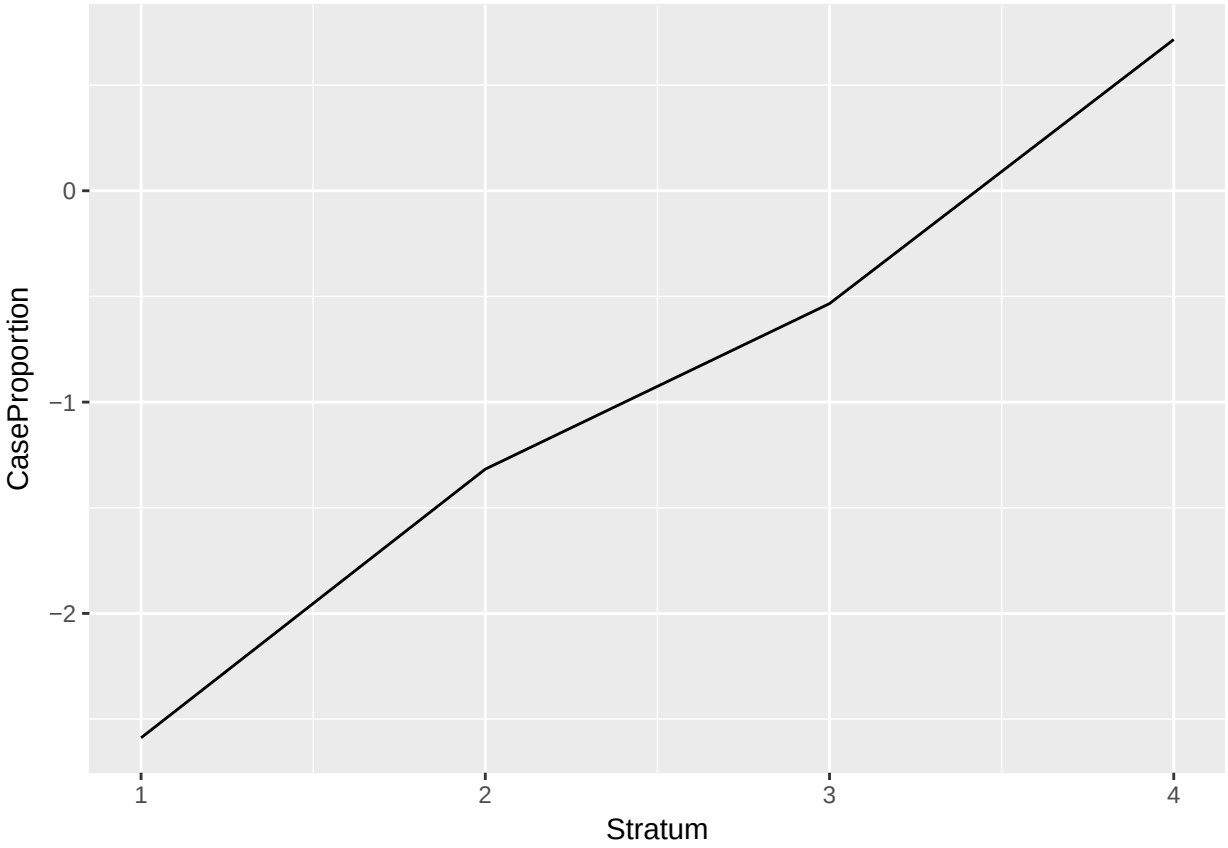
Question 1F

```
logCaseProportions = caseProportions[, .(CaseProportion=log(CaseProportion/(1-CaseProportion)))][, Strat

head(logCaseProportions)
```

```
##    CaseProportion Stratum
## 1:     -2.5885415       1
## 2:     -1.3173015       2
## 3:     -0.5340825       3
## 4:      0.7156200       4
```

```
ggplot(data=logCaseProportions,
       aes(x=Stratum, y=CaseProportion, group=1)) +
    geom_line()
```


0
−1
−2
CaseProportion
1
2
3
4
Stratum