Homework 1

Julia Van Dyke

January 7, 2015

1

```
sample(c(1:6),12,replace=TRUE)
```

```
##  [1] 1 3 2 6 3 4 3 2 3 5 3 6
```

2

```
pnorm(1)-pnorm(-1)
```

```
## [1] 0.6826895
```

```
pnorm(2)-pnorm(-2)
```

```
## [1] 0.9544997
```

```
pnorm(3)-pnorm(-3)
```

```
## [1] 0.9973002
```

3

```
1-pnorm(130,100,15)
```

```
## [1] 0.02275013
```

4

```
pexp(200,.0001)
```

```
## [1] 0.01980133
```

5

categorical: gender, class, smoker, selfhandedness, momhandedness, dadhandedness, region, oncampus, birthday, overtwenty

quantitative and discrete: childrank, numchildren, haircut, speedtickets, cds, cupscoffee,

quantitative and continuous: gpa, height, pulse, hourssleep, randomnum

6

```
ss = read.csv("http://www.calvin.edu/~scofield/data/csv/studentSurveyS04.csv")
prop.table(xtabs(~region, data=sample(ss,10,replace=TRUE)))
```

```
## region
##             Rural Suburban    Urban
##      0.0      0.3      0.5      0.2
```

```
prop.table(xtabs(~region, data=sample(ss,100,replace=TRUE)))
```

```
## region
##             Rural Suburban    Urban
##     0.00     0.17     0.75     0.08
```

```
prop.table(xtabs(~region, data=sample(ss,200,replace=TRUE)))
```

```
## region
##             Rural Suburban    Urban
##     0.00     0.16     0.66     0.18
```

```
prop.table(xtabs(~region, data=sample(ss,300,replace=TRUE)))
```

```
## region
##                   Rural    Suburban       Urban
## 0.003333333 0.206666667 0.643333333 0.146666667
```

```
prop.table(xtabs(~region,data=ss))
```

```
## region
##                   Rural    Suburban       Urban
## 0.003571429 0.178571429 0.678571429 0.139285714
```

as n increases, the distribution of the sample looks more like the distrbution of the population

7

```
cleanedSS = droplevels(subset(ss, selfhandedness!=""))
twiceCleanedSS = droplevels(subset(cleanedSS, momhandedness!=''))
prop.table(xtabs(~ momhandedness, data=twiceCleanedSS))
```

```
## momhandedness
##          L          R
## 0.05755396 0.94244604
```

8

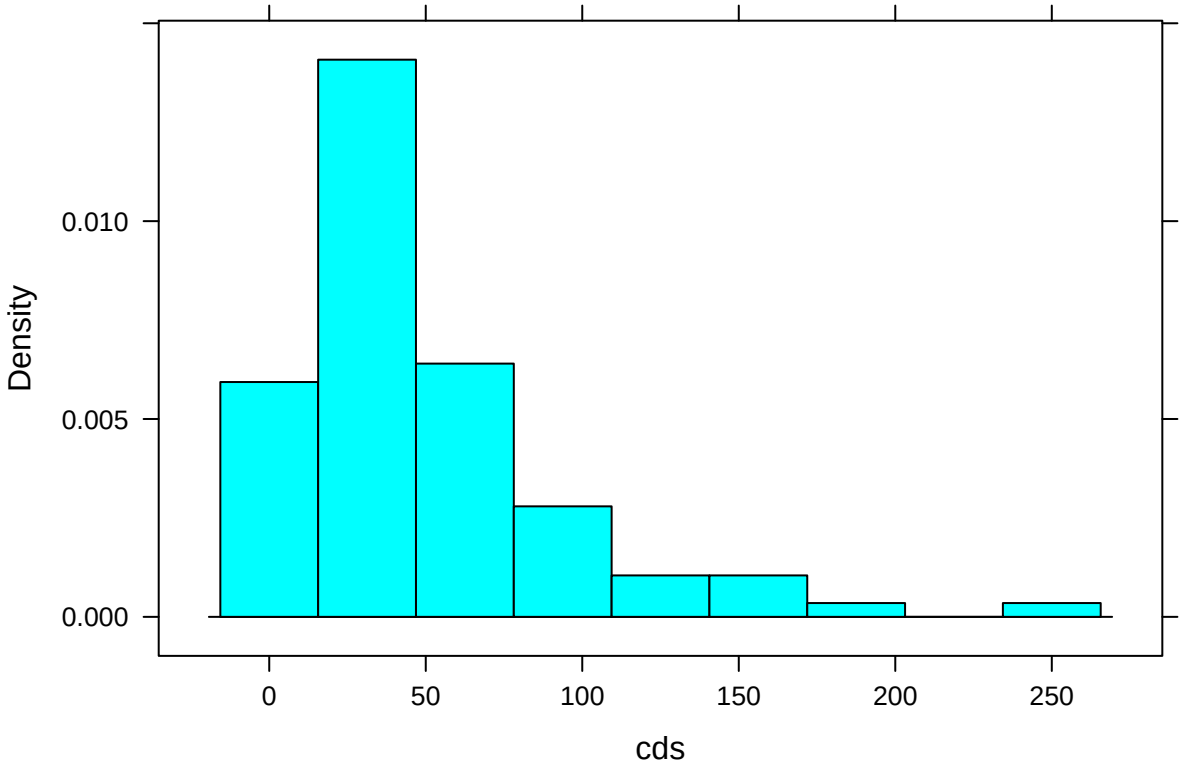
In most situations, the commands using twiceCleanedSS will be more useful, unless you are interested in the proportion of students who didn't answer the handedness questions. Using just xtabs is useful if you want the specific number of responses for each category, and adding prop.table is useful if you want to know the proportions of people who responded in each way. Adding margin=1 calculates the proportion of people who answered left and right for momhandedness, seperated by answers for selfhandedness. Adding margin=2 calculates the proportion of people who answered left and right for selfhandedness, seperated by answers for momhandedness. For a situation where momhandedness as the explanatory and selfhandedness as the response, using prop.table(xtabs(~selfhandedness + momhandedness, data=twiceCleanedSS), margin=2) makes the most sense.

9

momhandedness doesn't seem very useful for predicting selfhandedness. The proportion of righthanded and lefthanded people are only slightly different for righthanded and lefthanded moms.

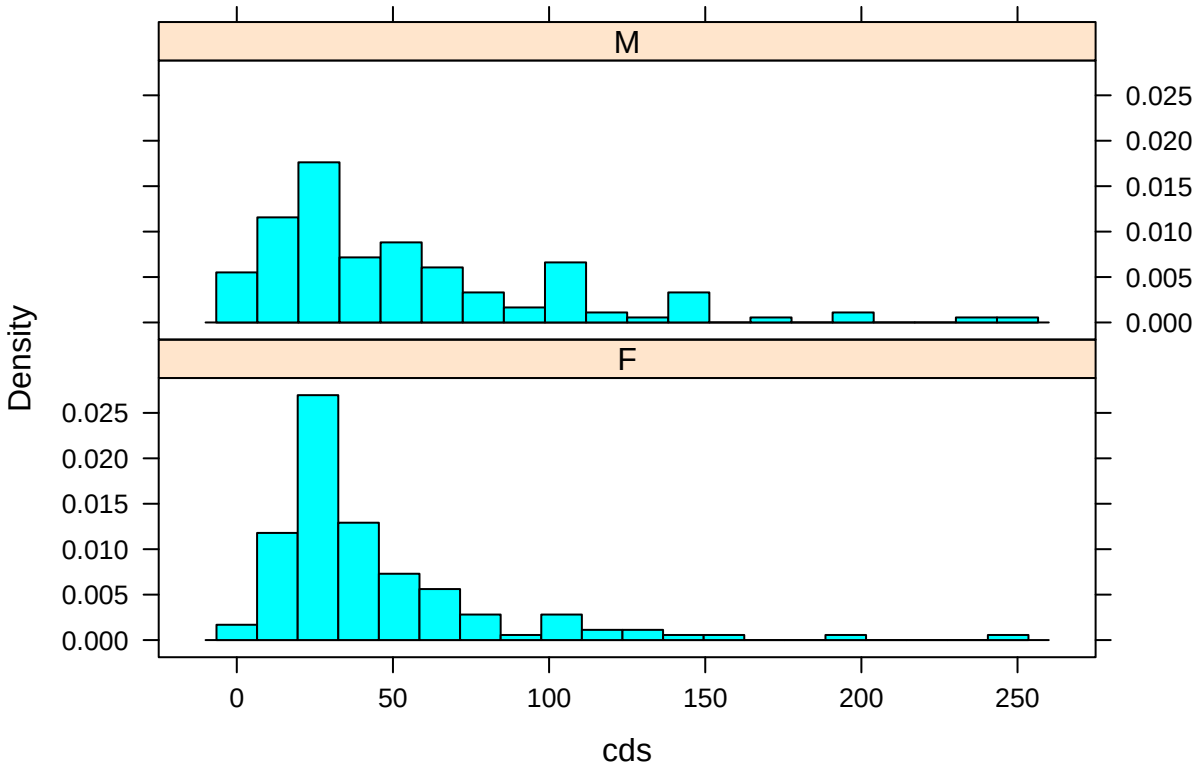
10

```
cleanedcd=droplevels(subset(ss,cds<400))
histogram(~cds, data=cleanedcd)
```



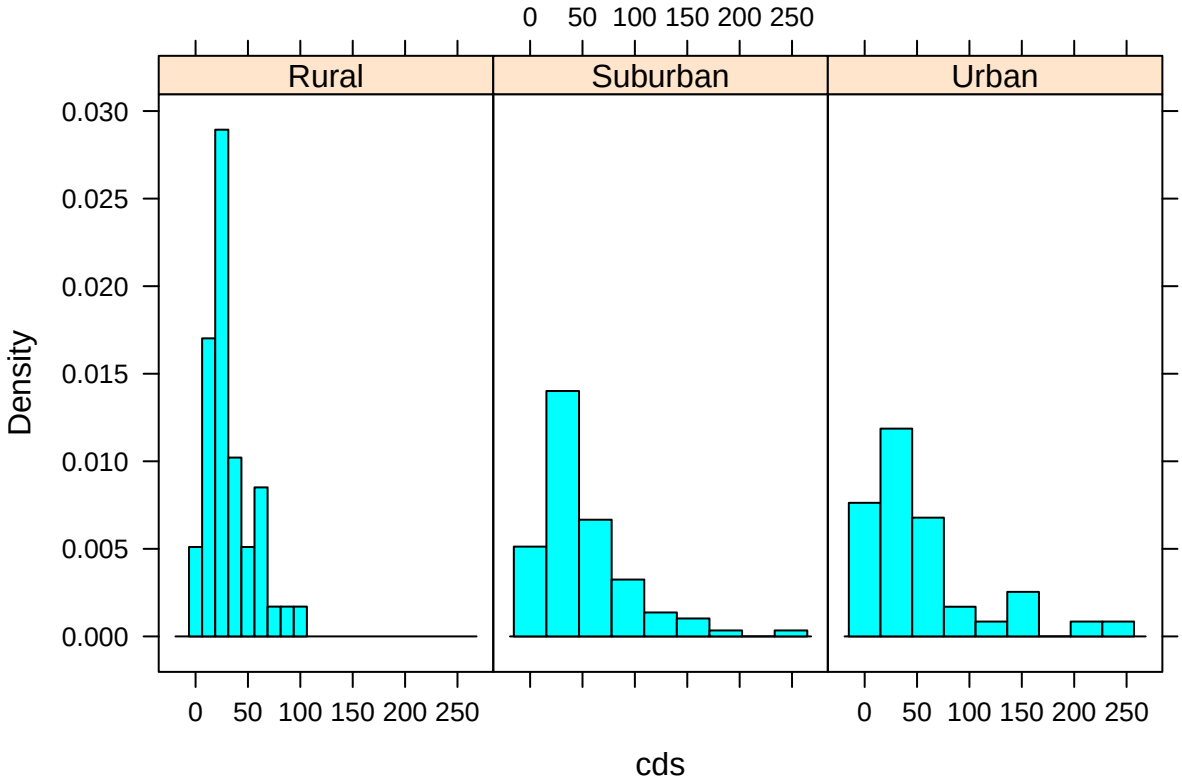
#11

```
histogram(~cds | gender, data=cleanedcd, n=20, layout=c(1,2))
```



#If there were no association, the histograms would be the same. These histograms aren't exactly the same, but they are very similar. #12

```
cleanedregion = droplevels(subset(cleanedcd, region!=""))
histogram(~cds | region, data=cleanedregion)
```



#Students coming from rural regions tend to have much fewer cds than students from suburban and urban regions. #13 #The data doesn't look like it has an association. The data is spread out everywhere, rather than following a line or a curve.