

Data 606: Chapter-2 Homework

Student Name : Sachid Deshmukh

Date : 09/19/2018

- GitHub Location for rmd file
 - GitHub Location for pdf file
 - RPub's location of published file
-

```
library(ggplot2)
library(VennDiagram)
```

```
## Warning: package 'VennDiagram' was built under R version 3.4.4
```

```
## Loading required package: grid
```

```
## Loading required package: futile.logger
```

```
## Warning: package 'futile.logger' was built under R version 3.4.4
```

1. 2.6 Dice rolls.

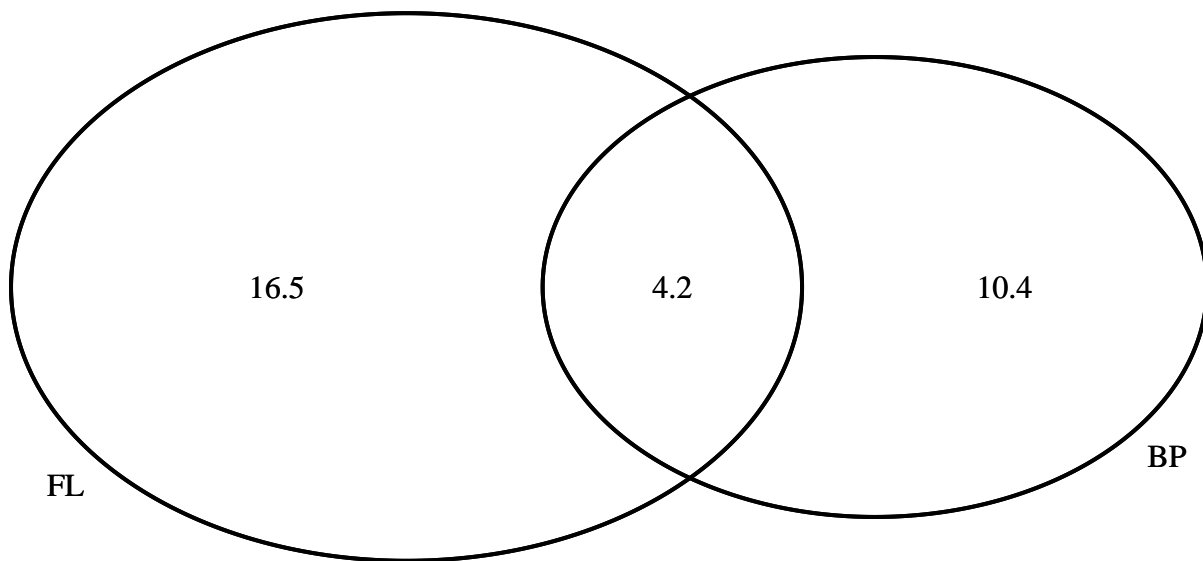
- Prob of getting sum of 1 is 0
- Prob of getting sum of 5 = $4/36$
- Prob of getting sum of 12 = $1/36$

2. 2.8 Poverty and language

- Living below poverty line and speaking a foreign language are not disjoint since there are 4.2% people who falls under both the categories
-

```
bp = 14.6
fl = 20.7
common = 4.2
```

```
diag <- draw.pairwise.venn(bp, fl, common, c("BP", "FL"))
grid.draw(diag)
```



- 10.4% of Americans only live below poverty line and only speaks english at home
- 31.1% Americans live below the poverty line or speak a foreign language at home
- 68.9% Americans live above the poverty line and only speaks english at home
- Since 4.2% population is common in below Americans living below poverty line and Americans speaking foreign language, these two events are not completely independent

3. 2.20 Assortative Mating

- $P(MB \cup FB) = P(MB) + P(FB) - P(\text{CommonB})$

```
(114/204 + 108/204) - 78/204
```

```
## [1] 0.7058824
```

- $P(FB | MB) = P(MB \text{ AND } FB)/P(MB)$

```
r{prob male blue} (78/204)/ (114/204)
```

- $P(FB | MBR) = P(FB \text{ AND } MBR)/P(MBR)$

```
(19/204)/(54/204)
```

```
## [1] 0.3518519
```

- $P(FB | MG) = P(FB \text{ AND } MG)/P(MG)$

```
(11/204)/(36/204)
```

```
## [1] 0.3055556
```

- If eyecolors of partners are independent then following equation would be true $P(MB \text{ ??? } FB) = P(MB) * P(FB)$. 78/204 is not equal to $(114/204) * (108/204)$. We can conclude that eyecolors of parters are

not independent

4. 2.30 Books on bookshelf

•

```
(28/95) * (59/94)
```

```
## [1] 0.1849944
```

•

```
(72/95) * (28/94)
```

```
## [1] 0.2257559
```

•

```
(72/95) * (28/95)
```

```
## [1] 0.2233795
```

- Since we are taking out only one book out of 95 total books, overall probability will not change drastically if we select random book with replacement or without replacement for scenario b and c.

5. 2.38 Baggage Fees

•

```
(0 * 0.54) + (0.34 * 25) + (0.12 * 60)
```

```
## [1] 15.7
```

•

```
sqrt(0.54*(0-15.7)^2 + 0.34*(25-15.7)^2 + 0.12*(60-15.7)^2)
```

```
## [1] 19.95019
```

- Avg Revenue per passenger is \$15.7. For 120 passenger expected revenue is

```
15.7*120
```

```
## [1] 1884
```

- Sample SD = 19.95. According to central limit theorem population SD = Sample SD * SQRT(n)

```
19.95 * sqrt(120)
```

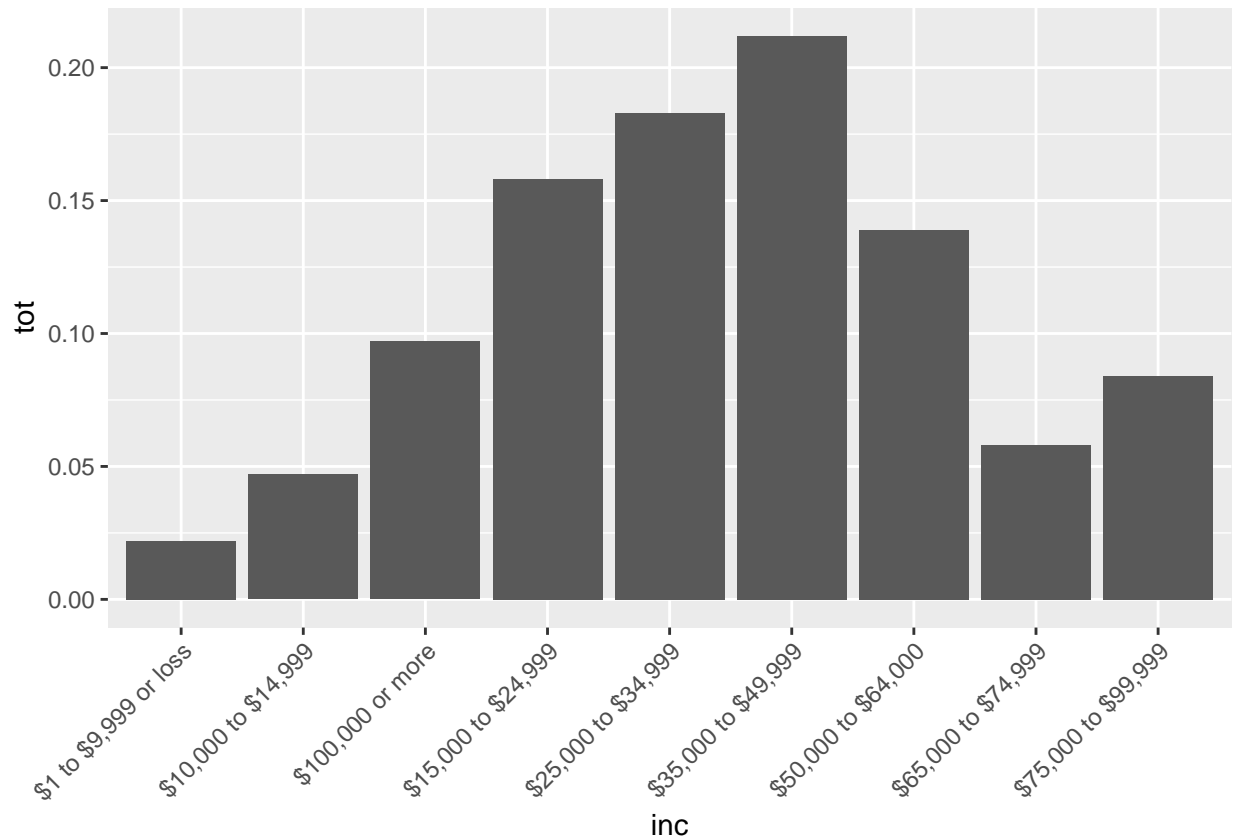
```
## [1] 218.5413
```

6. 2.44 Income and Gender

•

```
inc = c("$1 to $9,999 or loss",
        "$10,000 to $14,999",
        "$15,000 to $24,999",
        "$25,000 to $34,999",
        "$35,000 to $49,999",
        "$50,000 to $64,000",
        "$65,000 to $74,999",
        "$75,000 to $99,999",
        "$100,000 or more")
tot = c(.022, .047, .158, .183, .212, .139, .058, .084, 0.097)
inc_gender <- data.frame(inc, tot)
```

```
ggplot(inc_gender, aes(inc, tot)) + geom_bar(stat = "identity") + theme(axis.text.x = element_text(angl
```



Distribution is normal

-

```
sum(tot[1:5])
```

```
## [1] 0.622
```

- Assuming gender and income less than 50K per year are independent

```
sum(tot[1:5])*0.41
```

```
## [1] 0.25502
```

- Probability of any person making less than 50K an year is 0.62. However for female that probability is 0.71. This significant difference in probability indicates that those are dependent event. Probability of a person making less than 50K an year is gender dependent