STAT30280 Inference for Data Analytics - Assignment 1

Daniel Williams 21203054

17/02/2022

Exercise 1a

Show that the conditional probability of a survivor being male and the conditional probability of a survivor being female are approx 0.5.

```
no_male <- 387
yes_male <- 75
no_female <- 89
yes_female <- 76

male_total <- no_male + yes_male
female_total <- no_female + yes_female
yes_total <- yes_male + yes_female
no_total <- no_male + no_female
total <- no_male + yes_male + no_female + yes_female
```

Conditional probability P(A|B) = P(AB) / P(B)
We use this with A = male, B = survivor

```
Prob_AB <- yes_male / total
Prob_B <- yes_total / total
Prob_A_given_B <- Prob_AB / Prob_B
Prob_A_given_B
```

```
## [1] 0.4966887
```

```
# Result is 0.49668 ~ 0.5
```

Use the same formula with P(C|D) = P(CD) / P(D)
We use this with c = Female, D = survivor

```
Prob_CD <- yes_female / total
Prob_D <- yes_total / total
Prob_C_given_D <- Prob_CD / Prob_D
Prob_C_given_D
```

```
## [1] 0.5033113
```

```
# Result is 0.503 ~ 0.5
```

Exercise 1b

conditional probability of a male being a survivor
A = male, B = survivor

```
Prob_A <- male_total / total
Prob_B_given_A <- Prob_AB / Prob_A
Prob_B_given_A
```

```
## [1] 0.1623377
```

```
# Result is 0.16
```

conditional probability of a female being a survivor
C = female, D = survivor

```
Prob_C <- female_total / total
Prob_D_given_C <- Prob_CD / Prob_C
Prob_D_given_C
```

```
## [1] 0.4606061
```

```
# Result is 0.46
```

Exercise 1c

Prob being male, given survivor ~ 0.5
Prob being female, given survivor ~ 0.5
Prob being a survivor, given male ~ 0.16
Prob being a survivor, given female ~ 0.46

If we know a person survived, they are equally likely to be male or female
Survivors were equally split between male & female
Males were less likely to survive than females
Females had a 46% survival rate
Males had a 16% survival rate
Its key to note P(A|B) does not equal P(B|A)

Exercise 2a

n = number of rolls of the dice
X = number of times you roll a 5 P(X) = 1/6
Y = number of times you roll a 6 P(Y) = 1/6
let Z = X + Y , P(Z) = 2/6 , follows the binomial

```
Prob_Z <- 2/6
Var_Z_over_n <- Prob_Z * (1 - Prob_Z)

# Variance of Z = 0.222n = 8n/36
```

Exercise 2b

Var(X+Y) = Var(X) + Var(Y) + 2 Cov(X,Y)
We know Var (X+Y) from the above question
Var(X) = np(1-p) = n(1/6)(1 - 1/6)
Var(X) = 5n/36
Var(Y) = 5n/36 as it has equal p
Plugging into the top equation
8n/36 = 5n/36 + 5n/36 + 2 Cov(X,Y)
Cov(X,Y) = -n/36

Exercise 2c

Correlation (X,Y) = Cov(X,Y) / (stdev(x) * stdev(y))
stdev(x) = (5n/36)^0.5
stdev(y) = (5n/36)^0.5
Correlation (X,Y) = (-n/36) / (5n/36)
Correlation (X,Y) = - 1 / 5

Exercise 2d

n is number of repeat experiments
size is number of dice

```
set.seed(32)
dice_output <- rmultinom(n=10000, size=100, rep(1,6))
x <- dice_output[5,]
y <- dice_output[6,]
# not printing x & y due to their size
```

Exercise 2e

```
Z <- x + y
var(Z)
```

```
## [1] 22.80735
```

```
# Result is 22.81, expected 8/36 * 100 = 22.22, yes close.
cov(x,y)
```

```
## [1] -2.718637
```

```
# Result is -2.72, expected -100/36 = -2.77, yes close.
```

Exercise 3a

for the purposes of this answer let lambda = n
integral (int) between 0 and infinity

E[X] = int (x n exp(-n x) dx)
Integrate by parts
E[X] = n [-x exp(-nx)/n] + (1/n) int(exp(-nx) dx)
Take the first part between limits 0 and infinity
Evaluates to zero
E[X] = n [(1/n) int(exp(-nx) dx)]
E[X] = n [-(1/n) (1/n) exp(-nx)]
Evaluate between limits of 0 and infinity
Exp(-infinity) = 0
Exp(0) = 1
E[X] = n [1/n^2]
E[X] = 1/n
Here we used lambda = n, therefore proof Expectation is 1 / lambda.

Exercise 3b

```
set.seed(32)
G <- 10000  #number of repeat experiments
results <- rep(NA, G)  #setting up my empty vector to store my results

#for loop
for (k in 1:G) {
                results[k] <- 1 / mean(rexp(100,rate=5))}

mean(results)
```

```
## [1] 5.044403
```

```
# Result is 5.044, bias of 0.044
# Expected bias is 5/99 = 0.05
# Validates expected result.
```