Assignment - 1

Creating the data

```r
year = seq(1979,2008,1)
accidents = c(4,0,1,1,1,3,1,1,2,1,6,0,3,2,0,1,5,2,1,6,0,2,2,2,3,3,2,3,2,0)

data = data.frame(accidents = accidents)
rownames(data) = year
```

1)

```r
probability_11million = data["2006",]/11000000
print(probability_11million)
```

```
[1] 2.727273e-07
```

The number of accidents recorded in 2006 is 3. Given that there were 11000000 flights in 2006, the probability that it involved in an accident is 3/11000000 which comes up to 2.727273e-07.

2)

```r
p = 0.00000001
n = 11000000

#for loop to calculate the binomial distribution values for x = 0 to x = 12.
x = c()
for(i in 0:12){
```

```r
  x = c(x,choose(n,i) * p^i * (1-p)^(n-i))
}
df = data.frame("X" = seq(0,12,1), "Probabilities" = x)
knitr::kable(df)
```

X	Probabilities
0	0.8958341
1	0.0985418
2	0.0054198
3	0.0001987
4	0.0000055
5	0.0000001
6	0.0000000
7	0.0000000
8	0.0000000
9	0.0000000
10	0.0000000
11	0.0000000
12	0.0000000

The above table displays the binomial distribution from x = 0 to x = 12 for p = 0.00000001 and n = 11000000.

3)

```r
#i - 4 fatal accidents
print(x[4])
```

```
[1] 0.0001987258
```

```r
#ii - 10 fatal accidents
print(x[10])
```

```
[1] 5.820997e-15
```

```r
#iii - between 1 - 5 inclusive
print(sum(x[1:5]))
```

```
[1] 0.9999999
```

i. Probability for 4 fatal accidents : 0.000198.

ii. Probability for 10 fatal accidents : 5.82997e-15.

iii. Probability for 1 - 5 fatal accidents : 0.999999.

4)

```r
p = 0.00000002
n = 11000000

x = c()
for(i in 0:12){
  x = c(x,choose(n,i) * p^i * (1-p)^(n-i))
}
df = data.frame("X" = seq(0,12,1), "Probabilities" = x)
knitr::kable(df)
```

X	Probabilities
0	0.8025188
1	0.1765541
2	0.0194210
3	0.0014242
4	0.0000783
5	0.0000034
6	0.0000001
7	0.0000000
8	0.0000000
9	0.0000000
10	0.0000000
11	0.0000000
12	0.0000000

The above table displays the binomial distribution from x = 0 to x = 12 for p = 0.00000002 and n = 11000000.

5)

A binomial distribution might not be the best model for this particular scenario i.e. to determine number of accidents. The binomial distribution assumption:

- The number of trials(n) is fixed.
- Each trial should have two outcomes (Success/Failure).
- Each trial should be independent of each other.
- Each trial should have the same fixed probability.

Modeling number of accidents with this model violates some of these assumptions.

- The total number of potential accidents (n) might not be fixed.
- The probability of each accident might not be the same in a real world scenario.
- Each accident might not be independent of each other. There might be accidents caused due to other accidents.

6)

It is said that a person would have to take flights daily for 64000 years before dying in an accident. Lets calculate the probability. 64000 years is 64000 x 365 days which is equal to 23360000 days. This means the person is taking a flight daily for 23360000 days which is 23360000 flights in total. So the probability to for one accident in 23360000 flights is 1/23360000 which is calculated below.

```
years = 64000
days = 64000 * 365
probability_death = 1/days
difference = probability_11million - probability_death
cat("Probability =", probability_death," , ","Difference = ",difference)
```

```
Probability = 4.280822e-08  ,  Difference =  2.299191e-07
```

This comes out to around 4.28022e-08. According to the data given to us for 2006, only 3 in 11,000,000 flights end up in an accident. We calculated that probability to around 2.72727e-07. The difference in probabilities can be accounted by the fact that not all accidents might end up in death. But the difference here is low enough to justify the statement.