Title: Discrete and Continuous Probability Distributions

Packages: gridExtra, knitr

Discrete Variable

I am sending this report to my senior manager who asked me to analyses how much our customers visiting the desired links on our website - which is called conversion rate - and how can we improve the profit. I have with me the rate of conversion and their respective probability of occurring, also, how much profit we get from that conversion rate. In this way I can determine what audience to target to give more improved website experience so that more profit can be generated from them.

```
library(gridExtra)
## Warning: package 'gridExtra' was built under R version 3.1.3
Conversion_rate <- 1:10
probability <- c(.2,.1,.15,.24,.09,.07,.06,.06,0.02,.01)
profit <- c(23,65,76,77,84,90,98,100,105,200)

expected_value <- (probability * profit)


d <- data.frame(Conversion_rate,probability,profit,expected_value)
grid.table(d)
```

	Conversion_rate	probability	profit	expected_value
1	1	0.2	23	4.6
2	2	0.1	65	6.5
3	3	0.15	76	11.4
4	4	0.24	77	18.48
5	5	0.09	84	7.56
6	6	0.07	90	6.3
7	7	0.06	98	5.88
8	8	0.06	100	6
9	9	0.02	105	2.1
10	10	0.01	200	2

```
print(paste("mean is",mean(d$expected_value)))

## [1] "mean is 7.082"

print(paste("variance is",var(d$expected_value)))

## [1] "variance is 23.2401288888889"

print(paste("standard dev is",sd(d$expected_value)))

## [1] "standard dev is 4.82080168528938"

print(paste("lower limit is", mean(d$expected_value) - sd(d$expected_value)))

## [1] "lower limit is 2.26119831471062"

print(paste("upper limit is", mean(d$expected_value) + sd(d$expected_value)))

## [1] "upper limit is 11.9028016852894"
```

Observation: If we compare the upper value and lower value with the conversion rate we have, and since it was discrete value, we can see that conversion rate 3 to 10 lies in that category. By that we can determine that the 70% of conversion rate that amounts to some profit to our website (1 to 3), rest 30% does not amount much, and hence, we should not worry about users who have a conversion rate of 1 or 2 or 3.

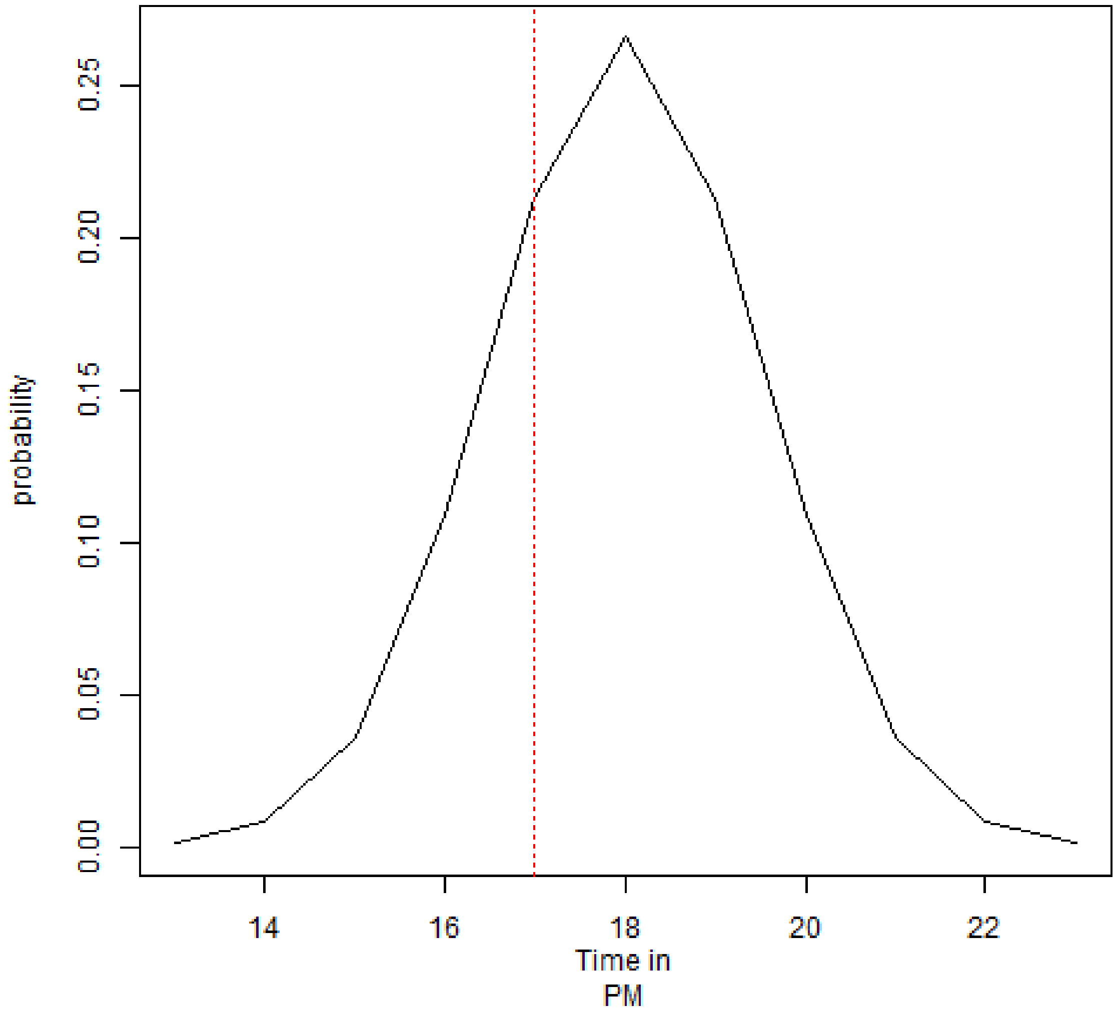
Continuous Variable

In the dataframe, we have time as a continuous variable. The probability that a flight will take off between that 1 hour timeframe at each interval is also given. Then a column represent the number of passengers that usually come to airport to have taken a flight. By doing this probability distribution we can determine as to at what time do we have the highest risk of losing a customer because he misses his flight.

```
time_PM <- seq(12,24,by=1)
probabilities <- c(0.000, 0.001, 0.008, 0.036 ,0.109, 0.213 ,0.266, 0.213
,0.109, 0.036, 0.008, 0.001,0.00)
Number_of_passengers <- c(00,80,70,69,56,45,30,42,49,53,67,78,00)
expected_value <- probabilities * Number_of_passengers
tab <- grid.table
join <- data.frame(time_PM,probabilities,Number_of_passengers,expected_value)
join <- join[-1,]
join <- join[-12,]
tab <- grid.table(join)
```

	time_PM	probabilities	Number_of_passengers	expected_value
2	13	0.001	80	0.08
3	14	0.008	70	0.56
4	15	0.036	69	2.484
5	16	0.109	56	6.104
6	17	0.213	45	9.585
7	18	0.266	30	7.98
8	19	0.213	42	8.946
9	20	0.109	49	5.341
10	21	0.036	53	1.908
11	22	0.008	67	0.536
12	23	0.001	78	0.078

```
print(paste("total expected value",sum(join$expected_value)))
## [1] "total expected value 43.602"
plot(join$time_PM,join$probabilities,type="l",xlab="Time in
PM",ylab="probability")
abline(v=17, col="red", lty=3)
```




```
prob <- sum(join$probabilities[which(time_PM< 17)])

print(paste("probability that a flight will take off before 5PM or 17hr is:",
round(prob,3)))

## [1] "probability that a flight will take off before 5PM or 17hr is: 0.367"

print(paste("mean is",mean(join$expected_value)))

## [1] "mean is 3.96381818181818"

print(paste("standard deviation is:",sd(join$expected_value)))

## [1] "standard deviation is: 3.72776525060744"
```

Observation: In this we can see that the probability that a plane will take off before 5PM is 0.367.