MSDA Bridge Week 3 – Data Science Assignment
Author: Arun Kumar

2.34

Outcome	Win (X)($)	P(X)	$X*P(X=x_i)$	$(X-E(X))^2$	$P(X)*(X-E(X))^2$
Red Card	0	26/52 =0.5	0	$(0 – 4.3269)^2 = 18.72$	9.36
Spade	5	13/52 = 0.25	1.25	$(5 – 4.3269)^2 = 0.45$	0.1125
Club	10	13/52 = 0.25	2.5	$(10 – 4.3269)^2 = 32.18$	8.045
Ace Club	30	1/52 = 0.0192	0.5769	$(30 – 4.3269)^2 = 659.10$	12.65
			E(X) = $4.3269		

Variance = $V(X) = \Sigma P(X)*(X – E(X))^2 =$ $ 30.17
Standard Deviation = SD(X) = Sqrt(V(X)) = $5.49

I will not play this game as the average expected win is around $4 but the wins have a standard deviation of $5.49, which means I don't have a chance to make a profit. On the other hand I can loose lot of money as the losses per game can go up to $5.49.

2.40: Baggage Fees

X = Revenue per passenger
(a)

	No Bag	One Bag	Two Bag	
Cost(X in $)	0	25	60	
P(X)	0.54	0.34	0.12	
X*P(X)	0	8.5	7.2	E(X) = $ 15.7
X – E(X)	-15.7	9.3	44.3	
$(X – E(X))^2$	246.49	86.49	1962.49	

Variance = $V(X) = \Sigma P(X)*(X – E(X))^2 =$ $ 2295.47
Standard Deviation = SD(X) = Sqrt(V(X)) = $47.91

(b)

Expected revenue for 120 passengers = 120 * E(X) = $1884
Standard deviation for 120 passengers = 120 * SD(X) =
Sqrt(120*V(X)) = $524.84

(c)
Assumption for the above calculation is that a passenger check-in baggage is independent of other passenger's choice.

2.42: Textbooks and Mario Kart video game

X = Item1 = Text book
Y = Item 2 = Mario Kart video game for Nintendo wii

Given information:
E(X) = $110 , SD(X) = $4
E(Y) = $38, SD(Y) = $5

(a)
Profit = 1*Y + (-1)*X = Y – X
E(Y – X) = E(Y) – E(X) = 38 – 110 = - $ 72 (Loss)

To calculate the standard deviation of the profit/loss:
Var(Y) = $5^2 = 25$
Var(X) = $4^2 = 16$

Var(Profit/Loss) = Var(Y-X) = Var(Y) – Var(X) = 25 – 16 = $9
SD(Y – X) = Sqrt(Var(Y – X)) = $3

So,
Expected loss = $72
Expected standard deviation of loss = $3

(b)
Expected money earned from the sale of text book= E(X) = $110

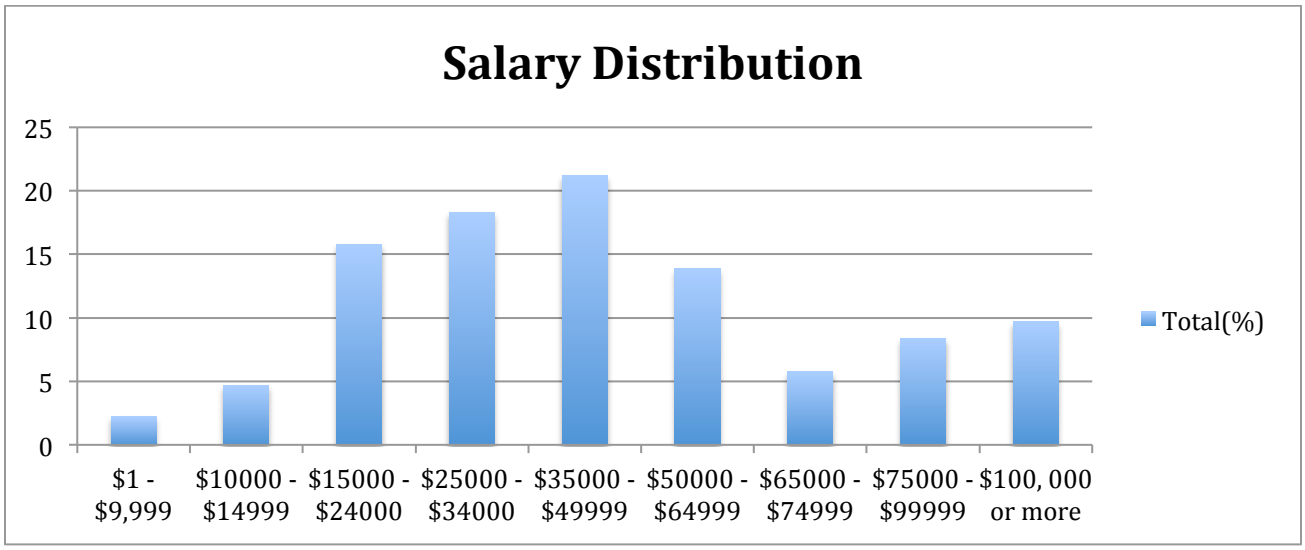
As 10% is earned for the sale –
Expected money earned of the sale = 10% * E(X) = 10% * 110 = $11.

Standard deviation of the expected money earned =
10% * Standard deviation of the money earned of sale = 10 % * SD(X) =
$ 0.4

2.46: Salary distribution

Income	Total(%)
$1 - $9,999	2.2
$10000 - $14999	4.7
$15000 - $24000	15.8
$25000 - $34000	18.3
$35000 - $49999	21.2
$50000 - $64999	13.9
$65000 - $74999	5.8
$75000 - $99999	8.4
$100, 000 or more	9.7

Below chart is created in excel:



(a)

Based on the distribution, it can be inferred that more people earn between $35,000 - $49,999.
The distribution seems to be a normal distribution with the pivot at $35,000 - $49,999 salary range.

(b)
P(Person makes less than $50,000/year) = 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2 % = 0.62

(c)
P(Female and earns less than $50,000) =
P(Female) * P(Earns less than $50,000) = 0.41 * 0.62 = 0.2542 = 25.42 %

The above calculation is on an assumption that the salary earned by person is independent of the gender.

(d)

The above assumption is not true.
Note: Many reports/study shows that salary earned by female co-worker are less than the male co-worker.
One citing: http://time.com/3836977/un-women-wages-and-careers/