Inferential Statistics

Instructions:

Please share your answers filled inline in the word document. Submit code files wherever applicable.

Insights should be drawn from the plots about the data such as, is data normally distributed/not, outliers, measures like mean, median, mode, variance, std. deviation, etc.

Please ensure you update all the details:

Name: _Kavali Rakesh

Batch Id: <u>DSWDMCOM 180122</u>

Topic: Basic Statistics

Problem Statements:

Q1) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Sol:- Possible outcomes when we toss 3 coins is HHT,HTH,THH,HHH,TTH,THT,HTT,TTT

So, the probability of two heads and one tail are favourable outcomes/total outcomes = 3/8=0.375

Q2) Two Dice are rolled, find the probability that sum is

a) Equal to 1
b) Less than or equal to 4
c) Sum is divisible by 2 and 3

Sol :- a) Not possible as the outcome will be always greater than 1

b)(1,1),(1,2),(1,3)(2,1)(3,1)=5

Total possible outcomes=6^2=36

So,Less than or equal to 4 is 5/36=0.138

C) Favourable outcomes=(1,5) (3,3)(4,2)(5,1)(2,4)(6,6)

So,Sum Is divisible by 2& 3 are 6/36=0.167

Q3) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Sol :- n(s)=7c2 = 7*6/2*1=21, n(e)=5c2 = 5*4/2*1= 10

Probability is 10/21

Q4) Calculate the Expected number of candies for a randomly selected child:

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

i. Child A – probability of having 1 candy is 0.015

ii. Child B – probability of having 4 candies is 0.2

CHILD	Candies count	Probability
A	1	0.015
B	4	0.20
C	3	0.65
D	5	0.005
E	6	0.01
F	2	0.12

Sol :- Expected number of candies for a randomly selected child

= 1 * 0.015 + 4*0.20 + 3 *0.65 + 5*0.005 + 6 *0.01 + 2 * 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

Q5) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

- For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and comment about the values/ Draw some inferences.

Points	Score	Weigh
3.9	2.62	16.46
3.9	2.875	17.02
3.85	2.32	18.61
3.08	3.215	19.44
3.15	3.44	17.02
2.76	3.46	20.22
3.21	3.57	15.84
3.69	3.19	20
3.92	3.15	22.9
3.92	3.44	18.3
3.92	3.44	18.9
3.07	4.07	17.4
3.07	3.73	17.6
3.07	3.78	18
2.93	5.25	17.98
3	5.242	17.82
3.23	5.345	17.42
4.08	2.2	19.47
4.93	1.615	18.52
4.22	1.835	19.9
3.7	2.465	20.01
2.76	3.52	16.87
3.15	3.435	17.3
3.73	3.84	15.41
3.08	3.845	17.05

Dataset: Refer to Hands-on Material in LMS - Data Types EDA assignment snap shot of dataset is given data se

Sol :- Mean ,Median and Mode are approximately equal ..so we can infer that data is normally distributed

Q6) Calculate Expected Value for the problem below

    a) The weights (X) of patients at a clinic (in pounds), are
    108, 110, 123, 134, 135, 145, 167, 187, 199

    Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Sol :- Exp(x)=sum of (P(x)*X)

       P(x)=1/9

    Exp(x)=1/9(108+110+123+134+145+167+187+199)

       =1/9(1308)

       =145.333

Q7) Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2
Hint: [Use a plot which shows the data distribution, skewness along with the outliers; also use R/Python code to evaluate measures of centrality and spread]

Name of company	Measure X

Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%
Warner-Lambert	35.00%

Q8) AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that "could happen." Suppose that one in 200 long-distance telephone calls is misdirected.

What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)
Hint: [Using Probability formula evaluate the probability of one call being wrong out of five attempted calls]
Sol :- Probability of telephone call misdirecting   P=1/200

Not misdirecting(q)=1-1/200=199/200

n=5

Atleast one in five attempted calls reaches the wrong number is $1-P(x)=1-\,^nC_x p^x q^{n-x}$

=1-(199/200)^5

=0.02475

Q9) Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

X	P(x)
-2,000	0.1

-1,000	0.1
0	0.2
1000	0.2
2000	0.3
3000	0.1

(i)   What is the most likely monetary outcome of the business venture?

      Hint: [The outcome is most likely the expected returns of the venture]

Sol :- The probability (0.3) is more for 2000 $ as compared to others,

Therefore, most likely monetary outcome of the business venture = 2000$

(ii)   Is the venture likely to be successful? Explain.

      Hint: [Probability of % of venture being a successful one]

Sol :- Long term average = \sum{P(xi)*Xi} = (-2000*0.1) +(-1000*0.1) +(0) +(1000*0.2) +(2000*0.3) +(3000*0.1) = 800$

As the long-term average gives positive numbers the Business venture likely to be successful.

(iii)   What is the long-term average earning of business ventures of this kind? Explain.

      Hint: [Here, the expected returns to the venture is considered as the the required average]

Sol :- Long term average = \sum{P(xi)*Xi} = (-2000*0.1) +(-1000*0.1) +(0) +(1000*0.2) +(2000*0.3) +(3000*0.1) = 800$

- Means on an average Return will be 800 $

(iv)   What is the good measure of the risk involved in a venture of this kind? Compute this measure.

      Hint: [Risk here stems from the possible variability in the expected returns, therefore, name the risk measure for this venture]

Sol :- Risk involved in a venture

$$Var(X) = E(X^2) - \{E(X)\}^2$$

$$= 2800000 - 800^2$$

$$= \mathbf{2160000} \ (\text{Quite High})$$

$$SD = \sqrt{Var} \approx \mathbf{\$ \ 1470}$$

As **Variability is Quite high** hence **Risk is high**

Hints:

For each assignment, the solution should be submitted in the below format

1. Research and Perform all possible steps for obtaining solution.

2. For Statistics calculations, explanation of the solutions should be documented detail along with codes. Use the same word document to fill in your explanation

   Must follow these guidelines:

   2.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the

        calculation stepwise

   2.2. For True/False Questions, or short answer type questions explanation is must.

   2.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) the data

        distribution to be attached

3. All the codes (executable programs) should execute without errors

4. Code modularization should be followed

5. Each line of code should have comments explaining the logic and why you are using that function

Grading Guidelines:

Note: 1. An Assignment submission is considered complete only when successful executable code(s), and documentation explaining the applied solution and results are provided. Failing to submit either of them will be considered an invalid submission and will not be considered for evaluation.

2. Assignments submitted after the deadline date will affect your grades.

Grading:

Ans	Date			Ans	Date

Correct	On time	A	100		
80% & above	On time	B	85	Correct	Late
50% & above	On time	C	75	80% & above	Late
50% & below	On time	D	65	50% & above	Late
		E	55	50% & below	
Copied/No Submission		F	45		

- **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline date

- **Grade B: (>= 80 and < 90):**
  - When assignments are submitted on time but less than 80% of questions asked in assignments are completed. (or)
  - All assignments were submitted, however, after the given deadline

- **Grade C: (>= 70 and < 80):**
  - When assignments are submitted on time but less than 50% of questions asked in assignments are completed. (or)
  - Less than 80% of questions asked in assignments are submitted after the deadline

- **Grade D: (>= 60 and < 70):** Assignments submitted after the Deadline and with 50% or less of questions

- **Grade E: (>= 50 and < 60):**
  - Less than 30% of questions asked in the assignments are submitted after the deadline (OR)
  - Less than 30% of questions asked in the assignments are submitted before deadline

Grade F: (< 50): Copied submission or No submission