|  |  |
| --- | --- |
| df\_patient[24].sum()Activity | Data Type |
| Number of beatings from Wife | integer |
| Results of rolling a dice | float |
| Weight of a person | float |
| Weight of Gold | float |
| Distance between two places | float |
| Length of a leaf | float |
| Dog's weight | float |
| Blue Color | string |
| Number of kids | integer |
| Number of tickets in Indian railways | integer |
| Number of times married | integer |
| Gender (Male or Female) | string |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ratio |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Sample space = {HHH,HTH,HHT,THH,HTT,THT,TTH,TTT} = 8

Number of possibilities of 2H and one T = { HTH,HHT,THH}=3

P(2 heads and 1 tail) = 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

**Answer:**

Sample space when two dice are rolled =

{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,2),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,3),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,4),(4,2),(4,4),(4,4),(4,5),(4,6)

(5,5),(5,2),(5,5),(5,4),(5,5),(5,6)

(6,6),(6,2),(6,6),(6,4),(6,5),(6,6)

}

1. Prob that sum = 1 is 0
2. Prob that sum is less than or equal to 4 is 1/9
3. Prob that sum is divisible by 2 and 3 is 1/9

Q5) 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?

**Answer:**

Two balls are drawn at random from 7 balls without replacement

5/7\*4/6 = 10/21

Q6) 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)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count (x) | Probability |
| A | 3 | 0.035 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.03 |
| F | 2 | 0.320 |

Child A – probability of having 3 candy = 0.035.

Child B – probability of having 4 candies = 0.20

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count (x) | Probability | x\*P(X=x) |
| A | 3 | 0.035 | 0.105 |
| B | 4 | 0.20 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.03 | 0.18 |
| F | 2 | 0.320 | 0.64 |
| xP(x) | | | 3.7 |

**Expected number of candies for a randomly selected child 3.7**

Q7) 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 also Comment about the values/ Draw some inferences.



**INFERENCES**

skewness(Points) = 0.2788734

skewness(Score) = 0.3906024

skewness(Weigh) = 0.3870456

**\*\*\*\* As the skewness is between -0.5 and 0.5 the data is fairly symmetrical**

kurtosis(Points) = 2.435116

kurtosis(Score) = 3.08861

kurtosis(Weigh) = 3.553753

**\*\*\*\* As the kurtosis is greater than 0, the distribution has heavier tails and hence it is a leptokurtic distribution**

Covariance between points and weigh is 0.087 and pearson coefficient is 0. 091.

**\*\*\*\* Hence points and weigh are positively corelated**

Covariance between score and weigh is -0.305 and pearson coefficient is -0.176

**\*\*\*\* Hence score and weigh are negatively corelated**

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

308, 330, 323, 334, 335, 345, 367, 387, 399

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

Answer:

Mean value of weight of all patients at a clinic **347.5556**

Q9) Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.43% |
| ITT Industries | 24.34% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.83% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 93.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.73% |
| Warner-Lambert | 35.00% |

<https://www.statmethods.net/advgraphs/axes.html> - To edit labels in scatterplot

Morgan Stanley is the outlier whose Measure X is 93.36%

Q30) AT&T was running commercials in 3990 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.)

Probability that a long distance call is misdirected = 1/200

Prob that a long distance call is not misdirected = 199/200

Let x be the random variable that denotes the number of attempted calls reaching a wrong number

To find probability(x>=1) = 1-P(x<1) = 1-P(x=0) = 1-199/200 = 1/200

5\*(1/200) = 1/40 = 0.025

**Answer: 0.025**

Q33) Returns on a certain business venture, to the nearest $3,000, are known to follow the following probability distribution

|  |  |  |
| --- | --- | --- |
| x | P(x) | x\*P(x) |
| -2,000 | 0.3 | -600 |
| -3,000 | 0.3 | -900 |
| 0 | 0.2 | 0 |
| 3000 | 0.2 | 600 |
| 2000 | 0.3 | 600 |
| 3000 | 0.3 | 900 |
|  |  | 600 |

1. What is the most likely monetary outcome of the business venture?

Most likely monetary outcome of the business venture is the one with maximum probability. Here it is 0.3 which sums up to (-2000-3000+2000+3000) = 0

P(x=0.3) = 0

Next highest probability is 0.2 P(x=0.2) = 3000

So most likely monetary outcome is **$3000**

1. Is the venture likely to be successful? Explain

Looking at the probability distribution of this business venture, it started with negative returns value and gradually started giving positive returns. P(x=3000)+P(x=2000)+p(x=3000) = 0.2+0.3+0.3=0.6 **Hence it is likely to be successful**

What is the long-term average earning of business ventures of this kind? Explain

**Summation(x\*P(x) = $600**

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

**Standard deviation** measures the dispersion of data from its expected value. it indicates how much is the return deviating from its expected normal returns. It is a good measure of risk

**3969.887**