|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

**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 | Ordinal |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

When three coins are tossed the total number of possible outcomes =23 = 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

Possible outcomes of getting two heads and one tail =3(HHT, HTH, THH )

So, number of favorable outcomes=3

P (Two heads and One Tail) = Number of favorable outcomes/total number of outcomes

Probability= 3/8=0.375

**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 2and 3.**

When two dice are rolled, possible outcomes are=62=36

a) The favorable outcome (sum is equal to 1) =0(i.e. not possible that sum always exceed to 1)

Probability =0/36=0

b) The favorable outcomes (sum less than or equal4) = 6

Combinations= (1, 1), (1, 2), (2, 1), (1,3),(2,2),(3,1)

Probability=6/36=1/6

c)The favorable outcomes (sum is divisible by 2 &3) =6

Combinations= (1,5), (2,4),(3,3),(4,2),(5,1),(6,6)

Probability=6/36=1/6

**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?**

In case of finding probability that:

First ball is either red / green, and the second ball is also either red / green.

Probability that first ball is red / green = (2+3)/7 = 5/7

While a second ball is being drawn, the count of “red + green” is four.

Probability that second ball is red / green = 4/6 = 2/3

Required probability = (5/7) \* (2/3) = 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 | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

=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

= 3.09

**Expected number of candies for a randomly selected child  = 3.09**

**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**.

**Use Q7.csv file**

[**https://colab.research.google.com/drive/1\_Z1wwquScCMlETY71kxSafys7oy1Scla#scrollTo=IaOXgqBhHSxN**](https://colab.research.google.com/drive/1_Z1wwquScCMlETY71kxSafys7oy1Scla#scrollTo=IaOXgqBhHSxN)

**Q8) Calculate Expected Value for the problem below**

1. **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?**

Expected Value = ∑ (probability \* Value)

 ∑ P(x). E(x)

There are 9 patients.

Probability of selecting each patient = 1/9

Ex 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) 1/9 1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value = (1/9) (108) + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

[**https://colab.research.google.com/drive/1NFlLKkkY54avQ9T5Q-VGrwro0pBfgdpW**](https://colab.research.google.com/drive/1NFlLKkkY54avQ9T5Q-VGrwro0pBfgdpW)

**SP and Weight (WT)**

**Use Q9\_b.csv**

[**https://colab.research.google.com/drive/1pQOPcxCpC42dB6pvffCQ\_NOzfBXdBp2-#scrollTo=x7Mgr\_W7-f5N**](https://colab.research.google.com/drive/1pQOPcxCpC42dB6pvffCQ_NOzfBXdBp2-#scrollTo=x7Mgr_W7-f5N)

**Q10) Draw inferences about the following boxplot & histogram.**

From Histogram We can say that maximum data concentrated towards left. We can see data skewed towards right & tail on right. It’s positive skewness data. It has outliers on the higher side.



Box plot has outliers on maximum side.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

The given information-

* Sample mean of x̅=200
* Sample **standard deviation** of s=30
* Sample **size**of n=2000

The confidence **interval** is:

The **interval** is:

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%5Cpm%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D

* In which t is the critical value for the two-tailed confidence interval l.

For 94% confidence interval T value is =1.8819

= 200 +1.8819=201.26

= 200 -1.8819=198.73

So, the confidence interval range is 198.73 to 201.27.

For 96 % confidence interval T value is =2.055

= 200 +2.055=201.37

= 200 -2.055=198.62

So, confidence interval range is 198.62 to 201.37

For 96 % confidence interval T value is =2.328

= 200 +2.328=201.56

= 200 -2.328=198.43

So confidence interval range is 198.43 to 201.56

**Q12) Below are the scores obtained by a student in tests.**

**Find mean, median,variance,standard deviation.**

**Mean**

σ=

Mean=(34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18

Mean=738/18

Mean=41

Median= (40+41)/2

Median=40.5

Variance:

(Sigma)2= µ)2/n

Variance= [(34-41)2+(36-41)2+(36-41)2+(38-41)2+(38-41)2+(39-41)2+(39-41)2+(40-41)2+(40-41)2+0+0+0+0+(42-41)2+(42-41)2+(45-41)2+(49-41)2+(56-41)]/17

Variance= (49+25+25+9+9+4+4+1+1+1+1+16+64+225)/17

Variance=434/17

Variance=25.529

Std=square root of Variance

Standard deviation=√25.529

Standard deviation= 5.0526

**2)what can we say about the student marks?**

Mean>median it’s positive skewed dada

**Q13) What is the nature of skewness when mean, median of data are equal?**

When mean & median of data are equal then there is no skewness in data. its symmetric distribution data.

Q14) What is the nature of skewness when mean >median?

When mean>median, shape of distribution of data skewed to the right & maximum part of data is concentrated towards left. There is a long tail stretching to the right. It’s positive skewed data.

**Q15) what is the nature of skewness when median > mean?**

When median>mean, shape of distribution of data skewed to the left & maximum part of data is concentrated towards right. There is a long tail stretching to the left. It’s negative skewed data.

**Q16) What does positive kurtosis value indicate for a data?**

Positive kurtosis indicates Higher peak & taller (heavy) tails than normal frequency distribution.

**Q17) What does negative kurtosis value indicate for a data?**

negative kurtosis indicates flatter, wider, less peak & thin tails than normal frequency distribution.

**Q18) Answer the below questions using the below boxplot visualization.**



**What can we say about the distribution of data?**

This is Box plot distribution of data.

**What is the nature of skewness of the data?**

The nature of skewness of data is negative skewness (left skewed data).

**What will be the IQR of the data (approximately)?**

IQR=Upper Quartile (UQ)-Lower Quartile (LQ)

IQR=18-10

IQR=8

**Q19) Comment on the below Boxplot visualizations?**



Both boxplots don’t have any outliers. Median of both plots looks same in middle of 250&275.That means 262.5.Data distributed equally on both side of median.no skewness in data.

Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Q 20) Calculate probability from the given dataset for the below cases**

**Data \_set: Cars.csv**

**Calculate the probability of MPG of Cars for the below cases.**

**MPG<- Cars$MPG**

* 1. **P(MPG>38)**
  2. **P(MPG<40)**

**c. P (20<MPG<50)**

<https://colab.research.google.com/drive/14t2-8ogHO7MOgDCzJVlaJG8SBYDyZI44>

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

<https://colab.research.google.com/drive/1RKwtswew0jsFj_QOqxRf5LZF6B-5qA8Y>

1. **Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

<https://colab.research.google.com/drive/1gF_3_FPnUq-xfpI8HG2sVvkwKxQbnAQB>

**Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval**

Z-score at 90% confidence interval is 1.6448.

Z-score at 94% confidence interval is 1.8807

Z-score at 90% confidence interval is 0.8416

[**https://colab.research.google.com/drive/1JT2AqkFBSCgsR30JsqVCxhy1BBhCyYl1**](https://colab.research.google.com/drive/1JT2AqkFBSCgsR30JsqVCxhy1BBhCyYl1)

**Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25**

#t value for 95%confidence interval

stats.t.ppf(0.975,df=24)

2.0638985616280205

#t value for 96%confidence interva

stats.t.ppf(0.98,df=24)

2.1715446760080677

#t value for 99%confidence interva

stats.t.ppf(0.995,df=24)

2.796939504772804

[**https://colab.research.google.com/drive/1JT2AqkFBSCgsR30JsqVCxhy1BBhCyYl1**](https://colab.research.google.com/drive/1JT2AqkFBSCgsR30JsqVCxhy1BBhCyYl1)

**Q 24) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.**

**Hint:**

**rcode 🡪pt(tscore,df)**

**df 🡪 degrees of freedom**

µ (population mean) =270

x̅ (sample mean) =260

s (sample std) =90

n (sample size) =18

T-test

Tvalue

260-270/(90/Ѵ18)

10/21.22

-.471

<https://colab.research.google.com/drive/18y8cASgrrj7W76q0jT5DyE7WOnCSWjyC#scrollTo=MfsieXpuL4Rw>