07 March Assignment

May 2, 2023

[ ]:
```
Q1. What are the three measures of central tendency?
```

[ ]:
```
ANS -
```

[ ]:
```
A Measure of central tendency is a summary measure that attempts to describe a␣
 ↪whole set of data with a
single value that represents the middle or centre of its distribution.
 There are three main measure of central tendency.

1. Mean
2. Median
3. Mode
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
Q2. What is the difference between the mean, median, and mode? How are they␣
 ↪used to measure the
central tendency of a dataset?
```

[ ]:
```
ANS -
```

[ ]:
```
The Mean(average) of a data set is found by adding all numbers in the data set␣
 ↪and then dividing by the
numbers of values in the set.
The Median is the middle value when a data set is odered from least to greatest.
The Mode is the number that occurs most often in a dataset.
```

[ ]:
```
In a normal distribution , data is symmetrically distributed with no skew. most␣
 ↪values cluster around a central
region, with values tapering off as they go further away from the center. the␣
 ↪mean ,mode , and median are exactly the same
in a normal distribution
```

[ ]:
```
In skewed distribution , more values fall on one side of the center than the␣
 ↪other, and the mean,mode , and median all
```

```
differ from each other.
```

[ ]:

[ ]:

[ ]:
```
Q3. Measure the three measures of central tendency for the given height data:
[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]
```

[ ]:
```
ANS -
```

[2]:
```
height=[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.
↪5]
```

[3]:
```
import numpy as np
```

[4]:
```
np.mean(height)
```

[4]:
```
177.01875
```

[5]:
```
np.median(height)
```

[5]:
```
177.0
```

[7]:
```
from scipy import stats
```

[8]:
```
stats.mode(height)
```

```
/tmp/ipykernel_70/1690947930.py:1: FutureWarning: Unlike other reduction
functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically
preserves the axis it acts along. In SciPy 1.11.0, this behavior will change:
the default value of `keepdims` will become False, the `axis` over which the
statistic is taken will be eliminated, and the value None will no longer be
accepted. Set `keepdims` to True or False to avoid this warning.
  stats.mode(height)
```

[8]:
```
ModeResult(mode=array([177.]), count=array([3]))
```

[ ]:

[ ]:

[ ]:
```
Q4. Find the standard deviation for the given data:
[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]
```

[ ]:
```
ANS -
```

[9]:
```
import numpy as np
```

[10]:
```
data=[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]
```

[11]:
```
std=np.std(data)
std
```

[11]:
```
1.7885814036548633
```

[ ]:

[ ]:

[ ]:
```
Q5. How are measures of dispersion such as range, variance, and standard␣
 ↪deviation used to describe
the spread of a dataset? Provide an example.
```

[ ]:
```
ANS -
```

[ ]:
```
The range is the easiest dispersion of data or measure of variability. The␣
 ↪range can measure by subtracting the lowest
value from the massive number. the wide range indicates high variability,and␣
 ↪the small range specifies low variability in
the distribution.
```

[ ]:
```
Variance is a simple measure of dispersion . Variance measures how far each␣
 ↪number in the datset from the mean.To
compute varaince first , calculate the mean and squared deviations from a mean.
 ↪Observation near to mean value gets
higher value.
```

[ ]:
```
Standard deviation is a aquared root of the variance to get original values ,␣
 ↪low standard deviation indicate
data points close to mean.
```

[ ]:

[ ]:

[ ]:
```
Q6. What is a Venn diagram?
```

[ ]:
```
ANS -
```

[ ]:
```
A venn diagram representing mathematical or logical sets pictorially as circles␣
 ↪or closed curves within
```

```
an enclosing rectangle , common elements of the sets being represented by␣
 ↪intersections of the
circles.
A venn diagram uses overlapping circles or other shapes to illustrate the␣
 ↪logical relationships between two
or more sets of items.
```

[ ]:

[ ]:

[ ]:
```
Q8. What do you understand about skewness in data?
```

[ ]:
```
ANS -
```

[ ]:
```
Skewness is a measurment of the distortion of symmetrical distribution of␣
 ↪symmetry in a data set.
Skewness is demonstrated on a bell curve when data points are not distributed␣
 ↪symmetrically to the
left and right sides of the median on a bell curve.
```

[ ]:

[ ]:

[ ]:
```
Q9. If a data is right skewed then what will be the position of median with␣
 ↪respect to mean?
```

[ ]:
```
ANS -
```

[ ]:
```
Genarally , if the distribution of data is skewed to the left , the mean is␣
 ↪less than the median ,
which is often less than the mode. if the distribution of data is skewed to the␣
 ↪right, the mode is
often less than the median , which is less than the mean.
```

[ ]:

[ ]:

[ ]:
```
Q10. Explain the difference between covariance and correlation. How are these␣
 ↪measures used in
statistical analysis?
```

[ ]:
```
ANS -
```

[ ]:
```
Covariance is a statistical term that refers to a systematic relationship␣
 ↪between two random variable in which a chnage in the other
reflects a change in one variable.
The greater the number , the more reliant the relationship. positive covariance␣
 ↪denotes a direct relationship and is represented
by a positive number.A negative number , on the other hand , denotes negative␣
 ↪covariance , which indicates an inverse relationship
between the two variable.

IN statistics, correlation is a measure that determines the degree to which two␣
 ↪or more random variable move in sequence . when an
equivalent movement of naother variable in some way or another during the study␣
 ↪of two variables.the variable are said to be
correlated.
```

[ ]:
```
Covariance is an indicator of the extent to which 2 random variable are␣
 ↪dependent on each other . A higher number denotes
higher dependency.
The value of covariance lies in the range of - 00 and + 00.

correlation is a statistical measure that indicates how strongly two variable␣
 ↪are related.
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
Q11. What is the formula for calculating the sample mean? Provide an example␣
 ↪calculation for a
dataset.
```

[ ]:
```
ANS -
```

[ ]:
```
Its obtained by simply dividing the sum of all values in a dataset by the␣
 ↪number of values.
```

[ ]:
```
The general formula for calculating the sample mean is given by `x = ( xi ) / n.
 ↪ Here , x represnts the sample mean , xi refers
all X sample values and n stands for thr number of sample terms in the data set.
```

[ ]:
```
sum of terms = 60 + 56 + 76 + 45 + 34
```

[1]:
```
60 + 56 + 76 + 45 + 34
```

[1]:
```
271
```

[ ]:
```
Number of terms = 5
```

[3]:
```
mean = 271 / 5
```

[4]:
```
mean
```

[4]: 54.2

[ ]:

[ ]:

[ ]:
```
Q12. For a normal distribution data what is the relationship between its␣
↪measure of central tendency?
```

[ ]:
```
ANS -
```

[ ]:
```
Any normal distribution has a graph that is perfectly symmetric about a␣
↪vertical line through its peak.
Therefore , all measures of entral tendency (most commonly , the  mean , median␣
↪, mode)give the same answer
the x - value of the peak.
```

[ ]:

[ ]:

[ ]:
```
Q13. How is covariance different from correlation?
```

[ ]:
```
ANS -
```

[ ]:
```
Covariance and correlation are two terms that are opposed and are both used in␣
↪statistics and regression
analysis . Covariance shows you how the two variable differ , whereas␣
↪correlation shows you how the two
variable are related.
```

[ ]:

[ ]:

[ ]:
```
Q14. How do outliers affect measures of central tendency and dispersion?␣
↪Provide an example.
```

[ ]:
```
ANS -
```

[ ]:
```
Outliers are numbers in a data set that are vastly larger or smaller than the␣
 ↪other values in the set . Mean , median and mode
are measures of central tendency . Mean is the only measure of central tendency␣
 ↪that is always affected by an outlier.
```

[11]:
```
score = [0,34,54,67,87,98,99]
```

[13]:
```
import numpy as np
```

[14]:
```
mean = np.mean(score)
```

[15]:
```
mean
```

[15]:
```
62.714285714285715
```

[ ]:
```

```