Marianne Trisha B. Raroque

BS in Psychology – 2B

1. Give a short definition of the following concepts and cite an example of each:

**Qualitative data** - It describes qualities or characteristics which expressed in terms of language rather than numerical values. It is collected using questionnaires, interviews, or observation, and frequently appears in narrative form.

Example: Observational notes: descriptive records of observations made during a research study, such as behaviors or interactions.

**Quantitative data** – it refers to any information that can be quantified, counted or measured, and given a numerical value.

Example: Financial data: numerical records of monetary transactions, including revenue, expenses, profit, or losses.

**Nominal** - the data can only be categorized.

Example: Eye color (Blue/Brown/Green)

**Ordinal** - the data can be categorized and ranked.

Example: Socioeconomic status (Lower class/Middle class/Upper class)

**Interval** - the data can be categorized and ranked, and evenly spaced.

Example: Years (e.g., 1990, 2010, 2020)

**Ratio** - the data can be categorized, ranked, evenly spaced and has a natural zero.

Example: Height in centimeters or inches.

1. How do we get the mean, median, and mode for (a) ungrouped data (b) grouped data?

**(a) ungrouped data:**

Mean for ungrouped data:

To calculate the mean for ungrouped data, you follow these steps:

* Add up all the individual values in the dataset.
* Count the total number of values in the dataset.
* Divide the sum of all the values by the total number of values.

The formula for the mean (μ) is:

μ = (x₁ + x₂ + x₃ + ... + xn) / n

where x₁, x₂, x₃, ..., xn represent the individual values, and n is the total number of values in the dataset.

Median for ungrouped data:

To find the median for ungrouped data, you can follow these steps:

* Collect the data: Gather all the individual values of your data set.
* Sort the data: Arrange the values in ascending order from smallest to largest.
* Determine the number of data points: Count the total number of values in your data set. Let's denote this as "n".
* Find the middle position: If "n" is an odd number, the median is the value at the center position. If "n" is an even number, the median is the average of the two middle values.
* For odd "n": The median is the value at position (n + 1)/2.
* For even "n": The median is the average of the values at positions n/2 and (n/2) + 1.
* Find the median: Locate the value at the determined position based on the steps above. This value represents the median of the ungrouped data set.

Mode for ungrouped data:

Here's a step-by-step process to find the mode for ungrouped data:

* Arrange the data in ascending or descending order.
* Count the frequency of each unique value in the dataset.
* Identify the value(s) with the highest frequency. These value(s) represent the mode(s) for the ungrouped data.

If there is a single value that occurs most frequently, then the dataset has a single mode. However, if multiple values have the same highest frequency, the dataset is said to be multimodal, and there are multiple modes.

It's worth noting that in some cases, ungrouped data may not have a mode if no value appears more than once in the dataset.

Remember that the mode is just one of several measures of central tendency, along with the mean and median, that can provide insights about a dataset.

**(b) grouped data**

For grouped data, the formula for finding the mean is as follows:

Where: = i th frequency

= i th class mark

i = 1,…, n

n = total number of observations

The formula of finding the median of grouped data is given by:

where: = Lower class boundary of the median class

= cumulative frequency below the median class

i = class size

= frequency of the median class

To find the mode of a grouped data, the formula below is applied.

where: = lower class boundary of the modal class

= difference between the frequency of the modal class and frequency above it.

= difference between the frequency of the modal class and frequency below it.

3. In grouped data, what do we mean: (a) class limit (b) lower and upper class limit (c) lower and upper class boundary (d) class width (e) median class and (f) modal class?

In grouped data, the following terms are commonly used:

**(a) Class limit:** In grouped data, the class limit refers to the smallest and largest values that can be included in a particular class. It defines the lower and upper boundaries of a class. The lower class limit is the smallest value that can be included in the class, while the upper class limit is the largest value that can be included.

**(b) Lower and upper class limit:** As mentioned above, the lower class limit is the smallest value that can be included in a class, while the upper class limit is the largest value that can be included. These limits define the range of values for each class.

**(c) Lower and upper class boundary:** The lower class boundary and upper class boundary are values that lie midway between the lower and upper class limits of a class. They are calculated by subtracting or adding half of the class width to the lower and upper class limits, respectively. Class boundaries are useful for ensuring that data points are properly allocated to classes when constructing frequency distributions.

**(d) Class width:** The class width is the difference between the upper or lower class limits of adjacent classes. It represents the range of values that are included in a class. Class width is determined by dividing the range of the data by the desired number of classes and rounding up to the nearest whole number.

**(e) Median class:** The median class is the class that contains the median value of the data. The median value is the middle value when the data is arranged in ascending or descending order. The median class is important in statistical analysis as it helps in determining the median and quartiles of grouped data.

**(f) Modal class:** The modal class refers to the class with the highest frequency in a frequency distribution. It represents the value or range of values that occur most frequently in the data. The modal class is useful for identifying the mode, which is the most common value or values in a dataset.

These terms are commonly used when organizing and analyzing data in grouped form, where data values are grouped into intervals or classes to simplify the presentation and analysis of large datasets.

4. What do we refer to as outliers?

* Outlier is referred to as a single data point that significantly deviates from the average value of a set of statistics. Outliers can also be exceptions that are outside of specific population samples.

5. Construct an ungrouped frequency distribution table of the following data on students’ exam scores in a 30 point exam. Then find the mean, median, mode, and standard deviation for the given data set. You may use MS Excel or any spreadsheet application if you want.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 19 | 12 | 15 | 10 | 18 | 15 | 10 | 25 | 13 |
| 16 | 16 | 15 | 16 | 27 | 16 | 23 | 11 | 17 | 12 |
| 21 | 11 | 13 | 21 | 29 | 15 | 24 | 12 | 21 | 12 |
| 11 | 17 | 24 | 12 | 23 | 26 | 15 | 11 | 14 | 13 |
| 19 | 13 | 18 | 20 | 11 | 11 | 12 | 18 | 12 | 16 |

**Ungrouped frequency distribution table:**

|  |  |  |
| --- | --- | --- |
| Students exam scores | Frequency | fx |
| 10 | 2 | 20 |
| 11 | 7 | 77 |
| 12 | 7 | 84 |
| 13 | 4 | 52 |
| 14 | 1 | 14 |
| 15 | 5 | 75 |
| 16 | 5 | 80 |
| 17 | 2 | 34 |
| 18 | 3 | 54 |
| 19 | 2 | 38 |
| 20 | 1 | 20 |
| 21 | 3 | 63 |
| 22 | 0 | 0 |
| 23 | 2 | 46 |
| 24 | 2 | 48 |
| 25 | 1 | 25 |
| 26 | 1 | 26 |
| 27 | 1 | 27 |
| 28 | 0 | 0 |
| 29 | 1 | 29 |
| Total: 50 |  |  |

**Mean:** 16.24

**Median:** 19.5

**Mode:** 11 and 12

**Standard Deviation:** 4.99

6. Using the same data above, construct another frequency distribution this time using grouped data. Then find the mean, median, mode, and standard deviation. You may use MS Excel or any spreadsheet application if you want.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | frequency | midpoint | f•m | <cf |  |  | 2 | 2 |
| 28-30 | 1 | 29 | 29 | 50 | 16.1 | 12.9 | 166.41 | 166.41 |
| 25-27 | 3 | 26 | 78 | 49 | 16.1 | 9.9 | 98.01 | 294.03 |
| 22-24 | 4 | 23 | 92 | 46 | 16.1 | 6.9 | 47.61 | 190.44 |
| 19-21 | 6 | 20 | 120 | 42 | 16.1 | 3.9 | 15.21 | 91.26 |
| 16-18 | 10 | 17 | 170 | 36 | 16.1 | 0.9 | 0.81 | 8.1 |
| 13-15 | 10 | 14 | 140 | 26 | 16.1 | -2.1 | 4.41 | 44.1 |
| 10-12 | 16 | 11 | 176 | 16 | 16.1 | 5.1 | 26.01 | 416.16 |

**Mean:** 16.5

**Median:** 17

**Mode:** 14.2

**Standard Deviation:** 4.76

7. Compare the mean, median, mode, and standard deviation values obtained from treating the data as ungrouped and grouped. What can you conclude from this?

* The mean, median, mode, and standard deviation in grouped and ungrouped data varies from each other.

8. Using MS Excel or any spreadsheet application, construct a frequency polygon for both the ungrouped and grouped data. Compare the resulting graphs. (Attach print screen of the graphs)

