**HUYNH DO**

**Module 2, Part 2/2 - Write a detailed summary explaining your findings of the z-score exercise**

1. **Continuous #1: Time at current job – Data Row: YEARSJOB**

**Frequency Data**

****

**Central Tendency data Key descriptive data**

** **

* **State the purpose of the exercise**

The exercise allows us to understand more about how the data is spread along the time of employees have spent at their current jobs.

The exercise also helps us to determine whether there are any outliers (unusual data), and patterns within the employees.

* **State how to calculate the statistical technique**

From the calculated z-scores data from the SPSS, these z-scores fall into absolute value fall in between [3-5.10]: 3.15, 3.2, 3.79, 4.6, especially **5.09** are over absolute 3, therefore, they are considered outliers.

How to calculate the z-scores:

1. **Mean**:

Add all data divide by N (Total entries)
2. **Standard Deviation**:

Calculate the average of squared deviations from the mean, then take the square root.
3. **z-scores**:
Subtract the mean from an individual value, divide by the standard deviation.


Where:

* **Xi**​ is the value of the data point.
* **Mean** = 7.73
* **Standard Deviation** = 9.28
* **f** is the frequency of the data point
* **σ** is the standard deviation.

For example the outlier has these z-scores are calculated as follow:

At 1 year: -0.725 = (1 – 7.73) / 9.28

At 55 years: 5.09375 = ( 55 - 7.73 ) / 9.28

* **Interpret the findings**

The combination of extreme positive and negative z-scores and a high range shows that the dataset has both very long-tenured and very short-tenured employees, with most people clustered closer to the mean. This finding also suggests that most people have much short-term compared to the average while quite a few employees spend decades with the company.

1. **Continuous #2: Number of hours usually work a week – Data View:HRS2**

**Frequency data**

****

**Central Tendency data Key descriptive data**

****

* **State the purpose of the exercise**

The purpose of calculating the z-scores is to have a better understanding of how each responded number of hours per week differs or how far off compared to the average in terms of stand deviation. Also, this will help us to find whether there’s the presence of any outliers.

* **State how to calculate the statistical technique**

Based on the frequency data table, below are technique and how numbers are calculated:

****

f = frequency of each data point

**∑** = (6×1)+(9×1)+(10×1)+(12×1)+(15×2)+(18×1)+(24×1)+(30×2)+(35×3)+(40×21)+(41×1)+(42×1)+(45×3)+(50×7)+(70×1)+(84×1)​ = 1836

N = 48

**Mean**  = 1836/48 = 38.25



For example the outlier has the z-score of 3.2127**,** which corresponds tothis data point **84 hours**, is calculated below:

 **3.2127** = ( 84 - 38.25 ) / 14.24

* **Interpret the findings**

Based on the z-scores table, only the 84 hours have a z-score of 3.21 which is greater than a typical absolute 3. Therefore, the number 3.21 is the outlier corresponding to 84 hours while other values do not exceed the typical threshold for being considered outliers.

The z-scores also highlight that most employees work within a reasonable range of the average hours (38.25), with only one stand-out outlier (84 hours). This indicates that overtime is not common (or rare) for hours per week.