**Dwight Howard Vs Kosta Koufos**

**Introduction:**

The aim of this assignment is to compare two basketball players (Dwight Howard and Kosta Koufos). Basketball is a game played between two teams of five players in which goals are scored by throwing a ball through a netted hoop fixed at each end of the court. Each player has a certain job to do and play in a different positive. Hence there are five different positions: point guard (PG), shooting guard (SG), small forward (SF), power forward (PF) and centre (C). In order to keep track of the score in a game, points are being used. Points can be accumulated by making field goals (two or three points) or free throws (one point). If a player makes a field goal from within the three point line, the player scores two points. If the player makes a field goal from beyond the three-point line, the player scores three points. The team that has recorded the most points at the end of a game is declared that game's winner. In part A of this assignment, I'm going to compare and contrast the two basketball players (Dwight Howard and KostaKoufos) to establish which one is best. Dwight Howard and KostaKoufos are two American professional basketball player plays at the centre of the field. In order to compare them, 20 matches of them were collected in 2014-2015 seasons as it can be seen below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Games played** | | | | |
| **Game/Players** | **Minutes played (KostaKoufos)** | **Points achieved (KostaKoufos)** | **Minutes played (Dwight Howard)** | **Points achieved (Dwight Howard)** |
| **1** | 24 | **8** | 27 | **12** |
| **2** | 14 | **4** | 28 | **19** |
| **3** | 15 | **9** | 22 | **16** |
| **4** | 17 | **8** | 21 | **14** |
| **5** | 16 | **10** | 17 | **11** |
| **6** | 19 | **4** | 33 | **28** |
| **7** | 19 | **6** | 36 | **13** |
| **8** | 16 | **2** | 39 | **13** |
| **9** | 14 | **4** | 39 | **18** |
| **10** | 18 | **2** | 39 | **22** |
| **11** | 24 | **10** | 35 | **24** |
| **12** | 14 | **4** | 38 | **14** |
| **13** | 13 | **4** | 18 | **7** |
| **14** | 13 | **2** | 32 | **20** |
| **15** | 17 | **2** | 40 | **20** |
| **16** | 15 | **6** | 32 | **16** |
| **17** | 34 | **6** | 26 | **7** |
| **18** | 13 | **4** | 40 | **19** |
| **19** | 16 | **9** | 33 | **14** |
| **20** | 17 | **0** | 34 | **14** |

In order to compare them, statistics is going to be used to help understand how each player is performing. Statistics is the practice of collecting, analyzing, organizing, interpreting and presenting data. We all use statistical data in our life and we sometimes don't even realize it as well. There are many different kinds of statistics that can be used to present data and the statistics that is going to be used in this assignment in order to compare the two players are central tendency (the mean, median, mode) , range, standard deviation, 5 number summary (box and whisker plot) and stem and leaf plot. Math Helper plus, Excel and the calculator is going to be used in order to find these statistical data.

**Central tendency** is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The **mean, median and mode** are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

Another way of comparing is by using **the measures of spread.** **Measures of spread describe how similar or varied the set of observed values are for a particular variable.** Measures of spread include the range, quartiles and the interquartile range, variance and standard deviation.

The **range** is the most obvious measure of dispersion and is the difference between the lowest and highest values in a dataset.

The **interquartile range** (**IQR**) is a measure of variability, based on dividing a data set into quartiles. Quartiles divide a rank-ordered data set into four equal parts. The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2, and Q3, respectively.

In statistics, the **standard deviation** is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A standard deviation close to 0 indicates that the data points tend to be very close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.

A **box and whisker plot** is a simple way of representing statistical data on a plot in which a rectangle is drawn to represent the second and third quartiles, usually with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines either side of the rectangle.

**Stem-and-leaf plots** are a method for showing the frequency with which certain classes of values occur.  It is also a device for presenting quantitative data in a graphical format.

As you can see below are the mean, median and mode for both players:

**Statics of Dwight Howard:**

**n = 20**

**mode = 14 ( f = 4 )**

**mean = 16.05**

**median = 15**

**range = 21**

**sd (sample) = 5.30615**

**var (sample) = 28.1553**

**sd (population) = 5.1718**

**var (population) = 26.7475**

**Five number summary:**

**min = 7**

**Q1 = 13**

**Q2 = 15**

**Q3 = 19.5**

**max = 28**

**interquartile range = 6.5**

**MODE = score with the highest frequency**

= 14 (which has frequency: 4)

**MEAN = sum of scores / number of scores**

= sum of 'fx' / sum of 'f'

= 321 / 20

= 16.05

**MEDIAN = middle score**

n = 20, which is EVEN

So the median is the average of scores

at positions: 20/2 = 10 and 10 + 1 = 11.

The score at position 10 is x = 14,

and the score at position 11 is x = 16.

So the median score = [14 + 16]/2

= 15

**RANGE = maximum score - minimum score**

= 28 - 7

= 21

**STANDARD DEVIATION**

sd (population) = sqrt[ (sum of fx²/n) - (sum of fx)/n)² ]

= sqrt[ (5687/20) - (321/20)² ]

= 5.1718

sd (sample) = sqrt[ (sum of fx² - {sum of fx}²/n) / (n-1)]

= sqrt[ (5687 - (321)²/20) / 19 ]

= 5.30615

**VARIANCE**

var (population) = (sum of fx²/n) - (sum of (fx)/n)²

= (5687/20) - (321/20)²

= 5.1718

var (sample) = (sum of fx² - {sum of fx}²/n) / (n-1)

= (5687 - (321)²/20) / 19

= 5.30615

**Static of KostaKoufos:**

**Statistics summary:**

**n = 20**

**mode = 4 ( f = 6 )**

**mean = 5.2**

**median = 4**

**range = 10**

**sd (sample) = 2.98417**

**var (sample) = 8.90526**

**sd (population) = 2.90861**

**var (population) = 8.46**

**Five number summary:**

**min = 0**

**Q1 = 3**

**Q2 = 4**

**Q3 = 8**

**max = 10**

**interquartile range = 5**

**MODE = score with the highest frequency**

= 4 (which has frequency: 6)

**MEAN = sum of scores / number of scores**

= sum of 'fx' / sum of 'f'

= 104 / 20

= 5.2

**MEDIAN = middle score**

n = 20, which is EVEN

So the median is the average of scores

at positions: 20/2 = 10 and 10 + 1 = 11.

The score at position 10 is x = 4,

and the score at position 11 is x = 4.

So the median score = [4 + 4]/2

= 4

**RANGE = maximum score - minimum score**

= 10 - 0

= 10

**STANDARD DEVIATION**

sd (population) = sqrt[ (sum of fx²/n) - (sum of fx)/n)² ]

= sqrt[ (710/20) - (104/20)² ]

= 2.90861

sd (sample) = sqrt[ (sum of fx² - {sum of fx}²/n) / (n-1)]

= sqrt[ (710 - (104)²/20) / 19 ]

= 2.98417

**VARIANCE**

var (population) = (sum of fx²/n) - (sum of (fx)/n)²

= (710/20) - (104/20)²

= 2.90861

var (sample) = (sum of fx² - {sum of fx}²/n) / (n-1)

= (710 - (104)²/20) / 19

= 2.98417

**Also the scores from both players can be seen in the graph below:**

As it can be seen above, Dwight Howard had better performance compare to Kosta Koufas as he was able to score more points. This shows that Dwight Howard was more helpful for his team compare to Kosta Joufas.

**Performance of the player:**

|  |  |  |  |
| --- | --- | --- | --- |
| Kostakoufos | Dwight Howard |  |  |
| 4 | 14 | mode | 1 |
| 5.2 | 16.05 | mean | 2 |
| 4 | 15 | median | 3 |
| 2.98417 | 5.30615 | sd | 4 |
| 10 | 21 | range | 5 |

Similarly, the graph above demonstrate same conclusion as the graph in the previous page. From this graph, it s conducted that Dwight Howard had higher mean compare to Kostakoufos. This means the average points that Dwight Howard achieved per game was higher than the average scores of Kostakoufos. Furthermore Dwight Howard had higher achieved 14 points in most of his games while Kostakoufos only achieved about 4 points in most of his games. Moreover Dwight Howard had higher Standard Deviation and range compare to Kostakoufos meaning his points are spread out over a wider range of values.

**Box and whisker plot:**

From the Box and whisker plot it can be deduced the better performance of Dwight Howard compare to Kosakoufos. First of all the box and whisker plot proves the fact that Dwight Howard achieved the highest score. Furthermore Dwight Howard higher middle 50% of data compared to Kostakoufos. Moreover, the lowest extreme value of Dwight Howard was nearly equivalent to upper half of Kostakoufos' data (upper quartile). Hence, Box and whisker plot clearly showed each players performance where Dwight Howard scored more points compare to Kosta Koufos.

**Stem and Leaf plot:**

|  |  |
| --- | --- |
| Kostakoufos | Dwight Howard |
| **STEM AND LEAF PLOT:**  **Unit = 1, Example: 1 | 2 represents 12**  0 | 0 2 2 2 2 4 4 4 4 4 4 6 6 6 8 8 9 9    1 | 0 0 | **STEM AND LEAF PLOT:**  **Unit = 1, Example: 1 | 2 represents 12**  0 | 7 7  1 | 1 2 3 3 4 4 4 4 6 6 8 9 9  2 | 0 0 2 4 8 |

From the stem and leaf plot it can be seen that Dwight Howard data points were negatively skewed while Kostakoufos data points were positively skewed. **Skewness** is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. Furthermore, for the stem and leaf plot it can be seen Dwight Howard scored more than 10 points in most of his matches while Kostakoufos scored less than 10 points in most of his matches.

**Strength:**

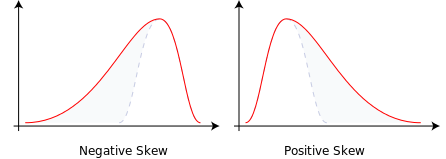
There was some strength of using this method. By using central tendency (mean, median & mode) and range, we could compare the two basketball players and to see which one is the best. It is also very easy to find central tendency as programs like math helper plus are able to find the answer very quick.

**Part B**

**Skewness**

In probability theory and statistics, **skewness** is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or even undefined.

The qualitative interpretation of the skew is complicated. For a unimodal distribution, negative skew indicates that the *tail* on the left side of the probability density function is longer or fatter than the right side – it does not distinguish these shapes. Conversely, positive skew indicates that the tail on the right side is longer or fatter than the left side. In cases where one tail is long but the other tail is fat, skewness does not obey a simple rule. For example, a zero value indicates that the tails on both sides of the mean balance out, which is the case for a symmetric distribution, but is also true for an asymmetric distribution where the asymmetries even out, such as one tail being long but thin, and the other being short but fat. Further, in multimodal distributions and discrete distributions, skewness is also difficult to interpret. Importantly, the skewness does not determine the relationship of mean and median.



### How to Calculate Skewness

1. Calculate the mean and standard deviation  
2. Subtract the mean from each raw score  
3. Raise each of these deviations from the mean to the third power and sum  
4. Calculate skewness, which is the sum of the deviations from the mean, raise to the third power, divided by number of cases minus 1, times the standard deviation raised to the third power

The [**mean, mode and median**](http://www.statisticshowto.com/how-to-find-the-mean-mode-and-median/) can be used to figure out if you have a positively or negatively skewed distribution.

* If the **mean** is greater than the **mode**, the distribution is **positively** skewed.
* If the **mean** is less than the **mode**, the distribution is **negatively** skewed.
* If the **mean** is greater than the **median**, the distribution is **positively** skewed.
* If the **mean** is less than the **median**, the distribution is **negatively** skewed.

**Five different datasets:**

|  |  |
| --- | --- |
| **Frequency** | **score** |
| 1 | 4 |
| 2 | 5 |
| 4 | 6 |
| 5 | 7 |
| 7 | 8 |
| 9 | 9 |
| 10 | 10 |
| 15 | 11 |
| 21 | 12 |
| 23 | 13 |
| 24 | 14 |
| 25 | 15 |
| 26 | 16 |
| 27 | 17 |
| 28 | 18 |
| 29 | 19 |
| 30 | 20 |
| 31 | 21 |
| 32 | 22 |
|  |  |



|  |  |
| --- | --- |
| Frequency | score |
| 1 | 19 |
| 2 | 20 |
| 4 | 21 |
| 5 | 22 |
| 7 | 23 |
| 9 | 18 |
| 10 | 17 |
| 15 | 16 |
| 21 | 15 |
| 23 | 14 |
| 24 | 13 |
| 25 | 12 |
| 26 | 11 |
| 27 | 10 |
| 28 | 9 |
| 29 | 8 |
| 30 | 7 |
| 31 | 6 |
| 32 | 5 |
|  |  |

|  |  |
| --- | --- |
| score | Frequency |
| 1 | 5 |
| 2 | 5 |
| 3 | 5 |
| 4 | 5 |
| 5 | 5 |
| 6 | 5 |
| 7 | 5 |
| 8 | 5 |
| 9 | 5 |
| 10 | 25 |
| 11 | 5 |
| 12 | 5 |
| 13 | 5 |
| 14 | 5 |
| 15 | 5 |
| 16 | 5 |
| 17 | 5 |
| 18 | 5 |
| 19 | 5 |
|  |  |





|  |  |
| --- | --- |
| score | Frequency |
| 1 | 3 |
| 2 | 5 |
| 3 | 7 |
| 4 | 9 |
| 5 | 11 |
| 6 | 13 |
| 7 | 15 |
| 8 | 17 |
| 9 | 19 |
| 10 | 21 |
| 11 | 19 |
| 12 | 17 |
| 13 | 15 |
| 14 | 13 |
| 15 | 11 |
| 16 | 9 |
| 17 | 7 |
| 18 | 5 |
| 19 | 3 |
|  |  |

|  |  |
| --- | --- |
| score | Frequency |
| 1 | 21 |
| 2 | 19 |
| 3 | 17 |
| 4 | 15 |
| 5 | 13 |
| 6 | 11 |
| 7 | 9 |
| 8 | 7 |
| 9 | 5 |
| 10 | 3 |
| 11 | 5 |
| 12 | 7 |
| 13 | 9 |
| 14 | 11 |
| 15 | 13 |
| 16 | 15 |
| 17 | 17 |
| 18 | 19 |
| 19 | 21 |
|  |  |

### 

### The Formula for Skewness Calculation

The term Skewness in Probability theory or Statistics, can be derived from the formula:

http://growingknowing.com/Images/skew.manual.jpg

### *Excel Formula*

http://growingknowing.com/Images/skew.excel.jpg

Where:

* **s** is the sample standard deviation.
* **Σ** means sum all the values.
* **x̄** represents a sample mean.