

# Sample questions for quiz

Data Set

Total population 1, 9, 4, 3, 7, 3

## A. Central Tendencies

Using the data set above determine the following-

1. Mean
2. median
3. mode

## B. Variance - What is the variance of the total population

1, 9, 4, 3, 7, 3

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Fill in the chart below

$i$	$x_i$	$\mu$	$x_i - \mu$	$(x_i - \mu)^2$
$x_1$				
$x_2$		4.5		
$x_3$	4			
$x_4$				
$x_5$				
$x_6$			-1.5	

$\sigma^2 =$

## How do I get the answers? Notes

Data Set    1 , 9 , 4 , 3 , 7 , 3  
                  $x_1$     $x_2$     $x_3$     $x_4$     $x_5$     $x_6$

the order only matters when determining the median

**Central Tendencies** = a number most representative of all of the numbers

1. mean → add up all of the numbers then divide by the total number (of numbers)

$$1 + 9 + 4 + 3 + 7 + 3 = 27$$

$$27/6 = \boxed{4.5}$$

2. median → put the numbers in order and find the middle. If there are an even number of data points add the middle two and divide by two. 1, 3, 3, 4, 7, 9

$$(3+4)/2 = \boxed{3.5}$$

3. mode → The number that occurs most frequently

1, 9, 4, 3, 7, 3

$$= \boxed{3}$$

- why does a good researcher look at all three results?
- which central tendency is most typically used for business reports?

our data set 1, 9, 4, 3, 7, 3

formulas for determining the mean of a total population (N) and a sample population (n)

$\mu$  = population mean

$\bar{x}$  = Sample population mean

Population Mean

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

*N* - total number in the entire population

this sign means sum or add

tells you to start with the first data point and go until you reach N (the last) point or total.

$$\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} = \frac{1 + 9 + 4 + 3 + 7 + 3}{6} = 4.5$$

Sample mean

The sample is missing one of the 3's.

The sample population data set is 1, 9, 4, 3, 7

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

*n* - Total elements in sample (5)

sum of elements in the sample

$$\frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{1 + 9 + 4 + 3 + 7}{5} = 4.8$$

Taking these formulas we can determine dispersion. Dispersion = how close or far away the numbers/points are to the mean, \*

\* This is why we took the "hard" or long method to finding the mean version of central tendency.

on to variance ... (4)

we will only look at the variance of a total population.

$\sigma^2 = \text{variance} \Rightarrow$  the average squared number each of the data points is from the mean (central tendency).

formula

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$N$  ← number of elements in the total population  
 $x_i$  ← each point value  
 $\mu$  ← minus the mean  
 $\left. \begin{matrix} \text{each point value} \\ \text{minus the mean} \end{matrix} \right\}$  then square the answer  
 $i=1$  ← the points in the population. This simply tells you to start with the first point and go all the way to  $N$ .

Let's break down the formula with our population

1, 9, 4, 3, 7, 3

$i$	$x_i$	$\mu$	$x_i - \mu$	$(x_i - \mu)^2$
$x_1$	1	4.5	$1 - 4.5 = -3.5$	12.25
$x_2$	9	4.5	$9 - 4.5 = 4.5$	20.25
$x_3$	4	4.5	$-0.5$	.25
$x_4$	3	4.5	$-1.5$	2.25
$x_5$	7	4.5	$2.5$	6.25
$x_6$	3	4.5	$-1.5$	2.25

→ This is  $-3.5^2$  or  $-3.5 * -3.5$

→  $4.5^2$  or  $4.5 * 4.5$

↑  
mean from prior page

$x_i - \mu =$  data point minus the average of the population

$$\sigma^2 = \frac{\sum_{i=1}^6 (x_i - \mu)^2}{6}$$

sum of all numbers divided by  $N$   
 $\frac{12.25 + 20.25 + .25 + 2.25 + 6.25 + 2.25}{6}$

$$\frac{43.5}{6} = 7.25 \rightarrow \text{the average squared distance from the mean}$$

$$\sigma^2 = 7.25$$

Lastly = the standard deviation

$$\sigma = \sqrt{\sigma^2} \quad \text{or simply stated}$$

The standard deviation is the square root of the variance.

So what is the standard deviation of our data set?

$$\sigma = \sqrt{7.25} = 2.69^*$$

\* Since we cannot all do square roots in our heads, you will not have to do the square root (figure the standard deviation) for the quiz.

-BUT-

You do need to know why the square root of the variance (aka the standard deviation) is used in data reports, or - why aren't reports given using variance alone?