Section 3 – 3B: Percentiles

A **Percentile** is **a number** in a **SORTED LIST**
that has **a given percentage of the data below it**.

If **n** represents the 70^{th} percentile then **70% of the data is less than n**
We use the notation P_{70} to represent the number that is the 70^{th} percentile

If **n** represents the 30^{th} percentile then **30% of the data is less than n**
We use the notation P_{30} to represent the number that is the 30^{th} percentile

Finding the k^{th} percentile P_k for a given value of x in a SORTED LIST

To find **what percentile corresponds to a given number x** on a **sorted list**, divide the number of values in the sorted list that are **less than x** by the total number of values in the list and multiple by 100. Then **round the answer to the nearest whole percentile.**

$$\text{the percentile of } x = p_k = \frac{\text{number of values less than x}}{\text{total number of values in list}} \bullet 100$$

(rounded to the nearest whole percentile)

x is the k^{th} percentile

$x = P_k$

Example 1

Consider the sorted list of 10 values below

2 4 5 7 8 12 15 19 20 25

What percent of the data is less than 20?

2 4 5 7 10 12 15 19 20 25

8 of the 10 numbers are less than 20

8 out of 10 is 80%

80% of the data is less than 20

we call 20 the 80^{th} percentile

$20 = P_{80}$

What percent of the data is less than 5?

2 4 5 7 10 12 15 19 20 25

3 of the 10 numbers are less than 7

3 out of 10 is 30%

30% of the data is less than 7

we call 7 the 30^{th} percentile

$7 = P_{30}$

Example 2

Consider the sorted list of 15 values below

2 4 5 7 8 10 11 13 15 16 17 18 20 21 24

Example 2A

What percent of the data is less than 16?

2 4 5 7 8 10 11 13 15 16 17 18 20 21 24

9 of the 15 numbers are less than 16

9 out of 15 is 60%

60% of the data is less than 16

we call 16 the 60^{th} percentile

$16 = P_{60}$

Example 2B

What percent of the data is less than 10?

2 4 5 7 8 10 11 13 15 16 17 18 20 21 24

5 of the 15 numbers aare less than 10

5 out of 15 is 33% (rounded to a whole percent)

33% of the data is less than 10

we call 10 the 33^{th} percentile

$10 = P_{33}$

Example 3

Finding the k^{th} percentile for a given value of x in a SORTED LIST

$$\text{the percentile of } x = k = \frac{\text{number of values less than } x}{\text{total number of values in list}} \bullet 100$$

(rounded to the nearest whole percentile)

x is the k^{th} percentile

$x = P_k$

The **sorted list** below represents a sample of 25 random numbers generated by an excel program.

(lowest)									
3	5	6	7	8	9	11	12	14	18
20	24	26	27	28	30	33	34	35	40
43	44	45	48	51	**(highest)**				

Example 3A

Find the percentile for the number 18

$$= \frac{\text{number of values less than } 18}{\text{total number of values in list}} \bullet 100$$

$$= \frac{9}{25} \bullet 100 = 36\%$$

$$18 = 36^{th} \text{ percentile} = P_{36}$$

Example 3B

Find the percentile for the number 27

$$= \frac{\text{number of values less than } 27}{\text{total number of values in list}} \bullet 100$$

$$= \frac{13}{25} \bullet 100 = 52\%$$

$$27 = 52^{th} \text{ percentile} = P_{52}$$

Example 3C

Find the percentile for the value 43

$$= \frac{\text{number of values less than } 43}{\text{total number of values in list}} \bullet 100$$

$$= \frac{20}{25} \bullet 100 = 80\%$$

$$43 = 80^{th} \text{ percentile} = P_{80}$$

Example 3D

Find the percentile for the value 51

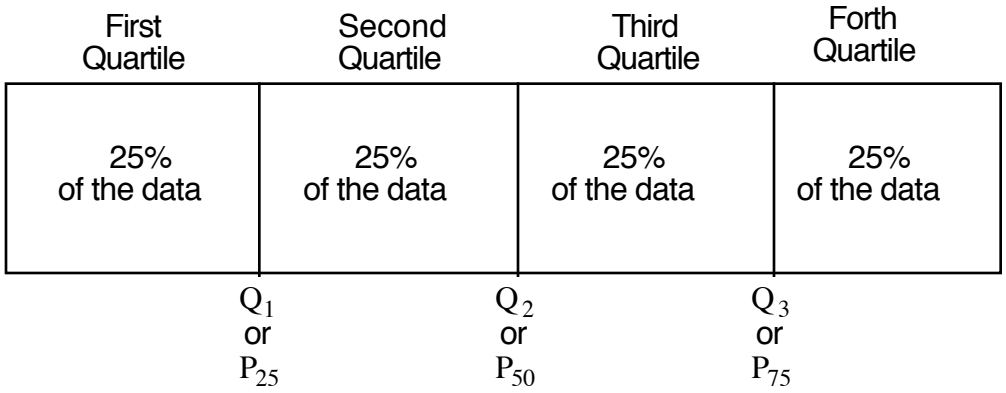
$$= \frac{\text{number of values less than } 51}{\text{total number of values in list}} \bullet 100$$

$$= \frac{24}{25} \bullet 100 = 96\%$$

$$51 = 96^{th} \text{ percentile} = P_{96}$$

Quartiles

It is common to **break the sorted data up into groups of 4.** Each group represents a quarter or 25% of the data. We call each of the groups that contain a quarter of the data a **Quartile**.



$Q_1 = P_{25}$ Q_1 is the 25 th percentile. **25% of the data is less than Q_1.**

$Q_2 = P_{50}$ Q_2 is the 50 th percentile. **50% of the data is less than Q_2.**

$Q_3 = P_{75}$ Q_3 is the 75 th percentile. **75% of the data is less than Q_3.**

Finding a specific P_k in a Sorted List

In this section you will be given a **specific P_k** and required to **find the number N in the sorted list that has k percent of the numbers below it.**

What is K

A) P_{60} = ???

$P_{60} = P_k$ so k = 60 **Find the number N on the sorted list that has 60% of the data below it.**

B) P_{35} **= ???**

$P_{35} = P_k$ so k = 35 **Find the number N on the sorted list that has 35% of the data below it.**

Finding the number **on the sorted list** that represents **P_k given a specific value of k**

It may seem strange but we **do not find the number in the list first**. We **find the LOCATION of the number first** and then **go to that location** and **find the number** that represents **P_k.**

A list sorted low to high

3 5 6 7 8 9 11 12 14 18

Example 1

If I knew that **P_{30}** was at location 4 then I would count **up 4 numbers from the bottom of the sorted list**. **The number 7** is at **location 4**. **7 is the 30th percentile.** $P_{30} = 11$

Check: 3 of the 10 numbers on the list are below 11. 3 out of 10 is 30%. $P_{30} = 11$

Note: If 3 out of 10 is 30% then the 4th location has 3 numbers below it and is t **P_{30}**.

Example 2

If I knew that **P_{60}** was a location 7 then I would count **up 7 numbers from the bottom of the sorted list. The number 11** is at **location 7**. **11 is the 60th percentile.** $P_{60} = 11$

Check: 6 of the 10 numbers on the list are below 11. 6 out of 10 is 60%. $P_{60} = 11$

Note: If 6 out of 10 is 60% then the 7th location has 6 numbers below it and is **P_{60}**.

You will be given a specific **P_k** and required to find a **number** on the sorted list **that has k percent of the numbers below it.** You must first find the **location of that number**. **We will call the location of the number L.**

The question is how do I find L?

Finding the location L of P_k in a SORTED LIST

To find P_k for a data set means, find a number N such that **k% of the data is less than N. There** is no standard formula or technique for finding the location of P_k.

This course requires that N be a member of the data set. Some books do not require the number to be in the sorted list.

3 4 **8** 9 10 11 12 13 14 18

The number 8 in the data set above has 2 of the 10 numbers on the list below it. 2 of 10 is 20% so the number 8 is P_{20}. If we do not specify that the answer for P_{20} must be a number on the list, any number above 4 up to and including 8 would work as an answer for P_{20}. 7 has 2 of the 10 numbers below it. So does 6 and 5 and 4,5 and 4.2. Any number, including decimal values, in the interval from $4 < x \le 8$ will have 2 numbers out of the 10 below them and serve as an answer for P_{20}.

Different textbooks, software packages, spreadsheets and calculators use different formulas and techniques to find P_{20}. Some of these use complicated techniques to arrive at several different decimal numbers between 4 and 8. It is very common to use the decimal value half way between 4 and 8. **As the data set increases in size the difference in the answers decreases.** It is in small data sets that the differences are more noticeable.

Many basic statistics books simply **use the lowest value IN THE SORTED LIST that has 20% of the data less than it.** We will use this simplified technique for this course and leave the more sophisticated methods for later courses.

How to find the location L **in a SORTED LIST** for a given **P_k**

The number at location L is the value of **P_k.** To find P_k we need **the location L** of where **P_k is located in the SORTED list**.

The value of P_k is **the number at the location L on the sorted list**

where L is found by

$$\text{Given : a specific } P_k$$

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

where n is the number of values in the set

and k is the k^{th} percentile

Find P_k in a SORTED LIST

Given : a specific P_k

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

where n is the number of values in the set

and k is the k^{th} percentile

Example 1

Find the k^{th} percentile for the **sorted list of 20 values**

3	5	6	7	8	9	11	12	14	18
20	24	26	27	28	30	33	34	35	40

Example 1A

Find **P_{40}**

The specific **P_k** is **P_{40} so K = 40**

there are 20 numbers in the list so **n = 20**

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$L = \text{the next integer} > \frac{40}{100} \bullet 20$$

$$L = \text{the next integer} > 8$$

$$L = 9$$

P_{40} is in the 9th location

$P_{40} = 14$

Example 1B

Find **P_{73}**

The specific **P_k** is **P_{73} so K = 73**

there are 20 numbers in the list so **n = 20**

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$L = \text{the next integer} > \frac{73}{100} \bullet 20$$

$$L = \text{the next integer} > 14.6$$

$$L = 15$$

P_{73} is in the 15th location

$P_{73} = 28$

Example 2

Find the k^{th} percentile for the **sorted list of 30 values**

3	5	6	7	8	9	11	12	14	18
20	24	26	27	28	30	33	34	35	40
43	44	45	48	51	55	57	58	59	61

Example 2A

Find P_{65}

$k = 65 \quad n = 30$

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$L = \text{the next integer} > \frac{65}{100} \bullet 30$$

$$L = \text{the next integer} > 19.5$$

$$L = 20$$

P_{65} is in the 20th location

$P_{65} = 40$

Example 2B

Find $Q_3 = P_{75}$

$k = 75 \quad n = 30$

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$L = \text{the next integer} > \frac{75}{100} \bullet 30$$

$$L = \text{the next integer} > 22.5$$

$$L = 23$$

P_{75} is in the 23rd location

$Q_3 = P_{75} = 45$

Example 3

Find the k th percentile for the sorted list of 40 values

2	6	9	10	11	13	15	16	17	18
20	24	26	27	28	30	33	34	35	40
41	42	43	46	47	50	51	53	58	50
60	63	64	68	69	71	72	73	75	79

Example 3A

Find P_{22}

$k = 22 \quad n = 40$

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$L = \text{the next integer} > \frac{22}{100} \bullet 40$$

$$L = \text{the next integer} > 8.8$$

$$L = 9$$

P_{22} is in the 9th location

$P_{22} = 17$

Example 3B

Find $Q_1 = P_{25}$

$k = 25 \quad n = 40$

$$L = \text{the next integer} > \frac{k}{100} \bullet n$$

$$`L = \text{the next integer} > \frac{25}{100} \bullet 40$$

$$L = \text{the next integer} > 10$$

$$L = 11$$

Q_1 is in the 11th location

$Q1 = 20$

