Descriptive Statistics

Making Stem Plot

&

Study of Measuring Positions

What is the **Stem Plot** of a sample?

It is a graphical summary for a set of **sorted numerical** sample. It shows a picture of the distribution of numerical values by displaying the actual values of the data.

How do we construct the **Stem Plot** of a sample?

- Sort the data from smallest to largest.
- Split each observation into a **stem** and a **leaf**.
- The **leaf** is the last digit in the observation.
- The **stem** is all digits that precede the leaf in the observation.

How do we display the **Stemplot** of a sample?

A typical stemplot graph clearly shows the stems and leaves separated by a vertical line with a key indicator. An example is given below.

Stemplot Sample Key: $5|4 = 54, 10|2 = 102$

Stem(tens)	Leaf(units)
5	4 5 6 9
6	
10	2

Example:

A sample of 40 exams in a math class was randomly taken. Scores are given below:

58	72	100	62	74	53	99	66	75	70
61	55	98	61	57	98	69	69	81	61
78	63	87	67	87	70	77	57	57	90
71	80	70	57	69	64	55	56	56	77

Construct the stemplot with key notation for this sample.

Solution:

We begin by sorting these exam scores in **ascending order** .

53	55	55	56	56	57	57	57	57	58
61	61	61	62	63	64	66	67	69	69
69	70	70	70	71	72	74	75	77	77
78	80	81	87	87	90	98	98	99	100

Now we identify the leaf of each observation which is the last digit and the rest of the digits are the stem. For the number 75, the leaf is 5 , and 7 is the stem.

Solution Continued:

We can use the sorted data to construct the **stemplot**.

Stemplot of Exam Scores Key: $5|3 = 53, 10|0 = 100$

Stem(tens)	Leaf(units)										
5	3	5	5	6	6	7	7	7	7	8	
6	1	1	1	2	3	4	6	7	9	9	9
7	0	0	0	1	2	4	5	7	7	8	
8	0	1	7	7							
9	0	8	8	9							
10	0										

What are the **Percentiles** of a sample?

Percentiles of a sorted sample are numbers that divide the data set into 100 groups. Each group contains 1% of the total sample. We use $P_1, P_2, P_3, \cdots, P_k, \cdots, P_{99}$ to denote percentiles.

What does P_{10} mean?

P_{10} of a sorted sample is the number that separates approximately the bottom 10% of the data from the top (100-10)%=90% of the data.

What does P_k mean?

P_k of a sorted sample is the number that separates approximately the bottom k% of the data from the top (100-k)% of the data.

Are there any special **Percentiles** ?

Here is a list of special **Percentiles** for any sample:

- $P_{25} = Q_1$
- $P_{50} = Q_2 = \widetilde{x} =$ Median
- $P_{75} = Q_3$

How do we find the location L for a given P_k?

Here are some steps that we need to take:

1. Sort the data from the smallest to the largest.
2. Compute $L = \frac{k}{100} \bullet n$ where n is the sample size.
3. When L is a whole number, then

$$P_k = \frac{Lth \text{ Value } + \text{ Next Value}}{2} \tag{1}$$

4. When L is a decimal number, then round it up to the next higher whole number and

$$P_k = Lth \text{ Value} \tag{2}$$

Example:

Consider these sorted exam scores below

52	53	53	54	56	57	57	57	57	58
61	61	61	62	63	64	66	67	69	69
69	70	70	70	71	72	74	75	77	77
78	80	81	87	87	90	98	98	99	100

Find

1. Find P_{10}.
2. Find P_{92}.

Solution:

We first need to compute the location for each percentile, then use the sorted data to locate the percentile.

1. For $P_{10} \rightarrow L = \frac{k}{100} \cdot n = \frac{10}{100} \cdot 40 = 4$

   Since the value of L is a whole number, we use equation (1).

   $$P_{10} = \frac{\text{4th value} + \text{5th value}}{2} = \frac{54 + 56}{2} = 55.$$

2. For $P_{92} \rightarrow L = \frac{k}{100} \cdot n = \frac{92}{100} \cdot 40 = 36.8$

   Since the value of L is a decimal number, we use equation (2).

   $P_{92} = \text{37th value} = 98.$

Example:

Consider these sorted exam scores below

58	59	60	61	65	67	70	72	75	78
80	81	84	85	63	86	87	88	90	91
93	93	94	95	97	99				

Find

1. Find the median.
2. Find Q_1.
3. Find Q_3.

Solution:

We use the fact that the median $= P_{50}, Q_1 = P_{25},$ and $Q_3 = P_{75}$.

1. For the median $\rightarrow L = \frac{k}{100} \cdot n = \frac{50}{100} \cdot 26 = 13$

   Since the value of L is a whole number, we use equation (1).

   $$\text{Median} = \frac{\text{13th value} + \text{14th value}}{2} = \frac{84 + 85}{2} = 84.5.$$

2. For $Q_1 \rightarrow L = \frac{k}{100} \cdot n = \frac{25}{100} \cdot 26 = 6.5$

   Since the value of L is a decimal number, we use equation (2).

   $Q_1 = \text{7th value} = 70.$

3. And $Q_3 = \text{20th value} = 91.$

How do we find the Percentile of a Data Value?

Here are some steps that we need to take:

1. Sort the data from the smallest to the largest.
2. Identify the sample size n.
3. Find B, the number of values are that are strictly below the data value in question.
4. Compute PR, the percentile ranking by using the formula below.

$$PR = \frac{B}{n} \bullet 100 \tag{3}$$

5. Always round PR to the nearest whole percent.

Example:

Consider these sorted exam scores below

41	44	45	45	47	48	49	49	50	54
58	59	60	61	65	67	70	72	75	78
80	81	84	85	63	86	87	88	90	91
93	93	94	95	97	99	103	107	110	112
115	118	119	120	120					

Find

1. Find the percentile ranking for 50.
2. Find the percentile ranking for 110.

Solution:

We first need to make sure that our data is sorted, which it is in our example, and determine the sample size, which is $n = 45$ in our example.

1. For the data element 50, there are $8 values$ strictly below it.

   $$PR = \frac{B}{n} \cdot 100 = \frac{8}{45} \cdot 100 \approx 17.7$$

   Rounding this result to the nearest whole percent, we get 18.

   So $P_{18} = 50$.

2. For the data element 110, there are $38 values$ strictly below it.

   $$PR = \frac{B}{n} \cdot 100 = \frac{38}{45} \cdot 100 \approx 84.4$$

   Rounding this result to the nearest whole percent, we get 84.

   So $P_{84} = 110$.


sta•tis•tics
[stuh-tis-tiks], n
1. the only science wherein two recognized experts, using exactly the same set of data, may come to completely opposite conclusions.