PRIME INTUIT
Finishing School

Effect of transformation on measure of centrality

Scale

Shift

Scale and Shift

Scaling : Eg Kilometers to Meters

Original distance: 52.32, 61.28, 71.28……
Scaled data : 52320, 61280, 71280……

Xnew = X * a

Xnew = X * a + c

F to C
5/ 9 a + - 160/9

PRIME INTUIT
Finishing School

Scale

Shift

Scale and Shift

$$\text{Mean} = \bar{x}_{new} = \bar{x} * a + C$$

$$\text{Xnew} = X * a + c$$

$$Mode_{new} = \text{Mode} * a + C$$

PRIME INTUIT
Finishing School

Summary

Mean

$$\bar{x} = \frac{1}{n}\sum_{i=0}^{n} x_i$$

Median

$$\frac{x_{n+1}}{2} \; or \; \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Mode

$Most\ frequently\ occuring\ element$

Mean is sensitive to outliers but median is not
Calculate the trimmed mean to avoid outliers
Mean is the center of gravity of the data
(almost always true)

Skewness:

Left Skewed: mean < median < mode
Right Skewed: mean > median > mode
Symmetric: mean = median = mode

Mean and Median can be approximately computed from histograms

Effect of Transformation:

Mean = $\bar{x}_{new} = \bar{x} * a + C$

$Median_{new}$ = median * a + c

$Mode_{new} = \mathbf{Mode} * \mathbf{a} + \mathbf{C}$

PRIME INTUIT
Finishing School

Introduction to Measures of Spread

PRIME INTUIT
Finishing School

Introduction to Descriptive statistics- Part 2

Descriptive Statistics

- ✓ Different types of data
- ✓ Different types of plots
- ✓ Measure of centrality and Spread

Probability Theory

- ✓ Sample Specs, events, axioms
- ✓ Discrete and continuous RVs
- ✓ Bernoulli, Uniform, Normal dist
- ✓ Sampling strategies

Inferential Statistics

- ✓ Interval Estimators
- ✓ Hypothesis testing (z-test, t-test)
- ✓ ANOVA, Chi-square test
- ✓ Linear Regression

PRIME INTUIT
Finishing School

Questions we are trying to answer

What are percentiles?

What are some frequently used percentiles?

How do you compute percentile rank of a value?

What is the effect of transformation on percentiles?

What are the different measures of spread?

What is the effect of transformation measures of spread?

What are box plots and how to use them to visualize some measures of centrality and spread?

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

Suppose you scored 45 out of 100 on a test, how would you rate your performance? Good or bad?

Is it bad? (because you scored less then 50%)

But ……

What if the questions were really hard?

What if the time provided was insufficient?

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

Suppose you scored 45 out of 100 on a test, Out of 100 students, Only 2 scored greater than 45. How would you rate your performance?

Does it look good now?

Yes it does ……

You can proudly say you lie in the top 98 percentile of your class
(the score of 98% of students was less then or equal to your score)

Introduction to measure of Spread - Percentiles

Another Example:

A university conducts a written test for 25 students and decides to call those students for an interview whose score is more then 70 percentile

44,43,37,68,55,46,19,59,34,46,51,62,47,52,44,28,36,56,65,
60,55,66,54,48,62

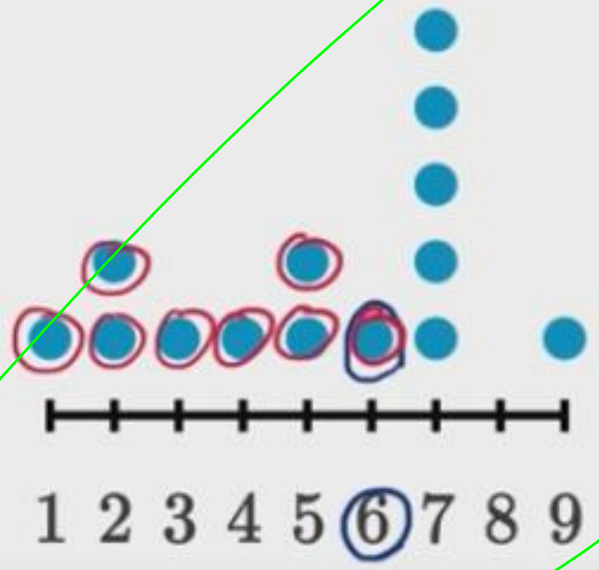
Can you identify which students will be called for the interview?

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

not related to this

The dot plot shows the number of hours of daily driving time for 14 school bus drivers. Each dot represents a driver.



Percentile → ① % of the data $\frac{7}{14} = 50\%$ that is below the amount in question

② % of the data that is at or below the amount

$\frac{8}{14} = \frac{4}{7}$

$$7\overline{)4.000} = 0.555$$

$-35\downarrow$

50

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

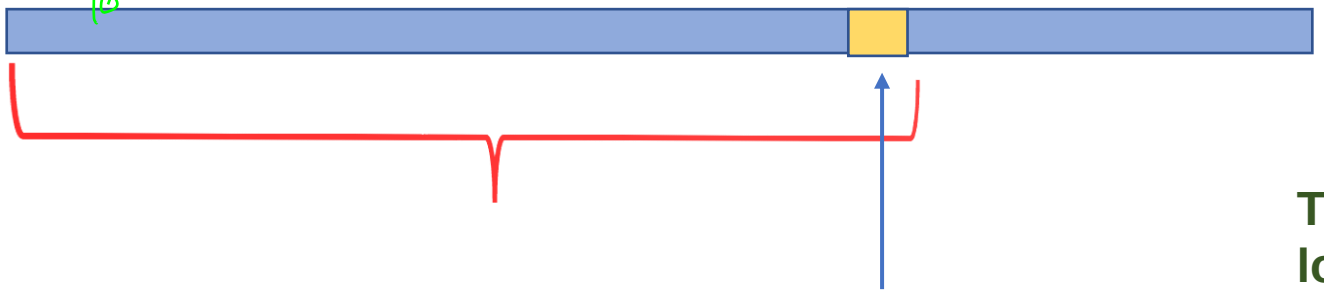
$$L_p = \frac{P}{100}(n+1)$$

$$= \frac{70}{100}(25+1) = 18.2$$

Percentiles:

25 students (sorted scores)

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55, 55, 56, 59, 60, 62, 62, 65, 66, 68



70% of the values in the data

70th Percentile

The 70th Percentile lies at location 18.2

The p Percentile of a sample is a value such that p Percentage of the values in the data are less than or equal to this value

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

Percentiles:

$$L_p = \frac{P}{100}(n+1)$$

$$= \frac{70}{100}(25+1) = 18.2$$

25 students (sorted scores)

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55, 55	56	59	60, 62, 62, 65, 66, 68

17 elements

18th 19th last 6 elements

18.2 is between 18 and 19 and is closer

Where is the position 18.2?

70th percentile should be between 56 and 59, greater then 56 but closer to 56. 56 + 0. 2 * (59-56) = 56.6

(59 - 56) x 0.2 = 0.6

PRIME INTUIT
Finishing School

Procedure for computing the Percentiles

What is the overall Procedure?

Sort the data

Compute location of the Pth Percentile

Lp = p / 100 (n+1)

Compute the integer part Lp = ip
Compute the fraction part Lp = fp

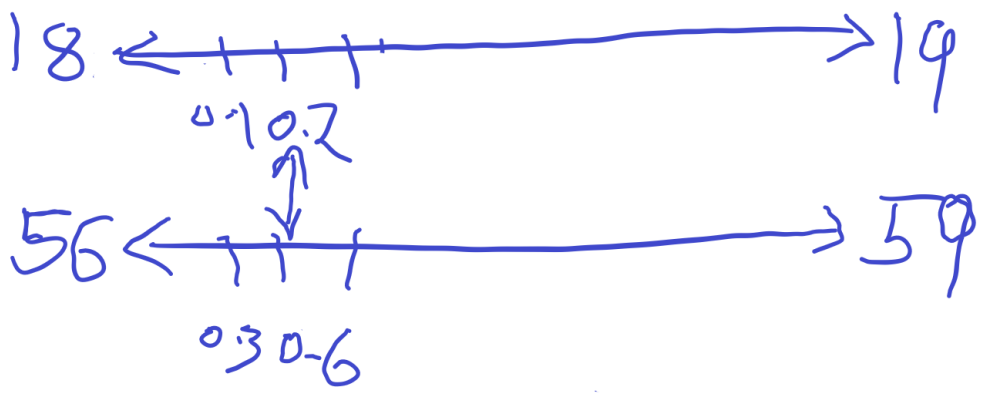
Compute the pth percentile as

$$y_p = x_i + f_p * (x_{i_p+1} - x_{i_p})$$

PRIME INTUIT
Finishing School

Procedure for computing the Percentiles

Some more information



$$x_{ip} = 56 + (0.2)(59 - 56)$$

$$3$$

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

Another Example: continued

A university conducts a written test for 25 students and decides to call those students for an interview whose score is more then 70 percentile

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55,
55, 56, 59, 60, 62, 62, 65, 66, 68

$$y_{70} = 56.6$$

The university will invite only those 7 students whose score was greater then 56.6

Introduction to measure of Spread - Percentiles

Another Example:

If the university decides to change its decision and now wants to invite only students who scored greater then 80 percentile, P = 80

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55, 55, 56, 59, 60, 62, 62, 65, 66, 68

$$L_p = \frac{P}{100}(n+1) = \frac{80}{100}(25+1) = 20.8$$

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55, 55, 56, 59	60	62, 65, 66, 68

$$y_{80} = x_{20} + 0.8 * (x_{21} - x_{20})$$

$$= 61.6$$

$$20 + 0.8\,(62 - 60)$$

The university will invite only those 5 students whose score was greater then 61.6

PRIME INTUIT
Finishing School

Introduction to measure of Spread - Percentiles

Another Example:

Suppose there were only 24 students and P = 80

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55,
55, 56, 59, 60, 62, 62, 65, 66

19, 28, 34, 36, 37, 43, 44, 44, 46, 46, 47, 48, 51, 52, 54, 55, 55, 56, 59	60	62, 65, 66, 68

$$L_p = \frac{P}{100}(n+1) = \frac{80}{100}(24+1) = 20$$

$$i_p = 20,\ f_p = 0,\ y_{80} = 60 + 0 \times (62-60)$$
$$= 60$$

The university will invite only those 4 students whose score was greater then 60

PRIME INTUIT
Finishing School

Alternative methods for computing the Percentiles

What is the standard Procedure?

Sort the data

Compute location of the Pth Percentile

Lp = p / 100 (n+1)

Compute the integer part Lp = ip : Yp = xi
Compute the fraction part Lp = fp

Compute the pth percentile as

$$y_p = x_i + f_p * (x_{i_p+1} - x_{i_p})$$

Alternative methods for computing the Percentiles

Alternative 1

Sort the data

Compute location of the Pth Percentile

Lp = p / 100 (n) - - - - Note: use of n instead of n+1

I integer part of Lp = ip

If Lp is an integer :

$$y_p = \frac{x_{Lp} + x_{Lp+1}}{2}$$

If Lp is not an integer :

$$y_p = x_{ip+1}$$