Bertelsmann Tech Scholarship - Data Track

9.18 UNC Introduction to Data Analysis & Programming

The median is the "middle" of the data =/= not average;
half of the data values < median < half are greater.

9.22 Find Median with Outlier

- Get median between 2 medians by get average.
- Median = best measure of central tendency in highly skewed distribution

9.24 Order Measures of Center 1

9.25 Order Measures of Center 2

- Symmetrical, Normal distribution. mean = median = mode

9.27 Udacians' Facebook Friends - Mean
=average(B2:B28) --> Formula for the mean of dataset.

9.28 Udacians' Facebook Friends - find Median

- Menu > Data > Sort range
  > Sort by Column B - A -> Z

9.29 Formula Location of Median -->

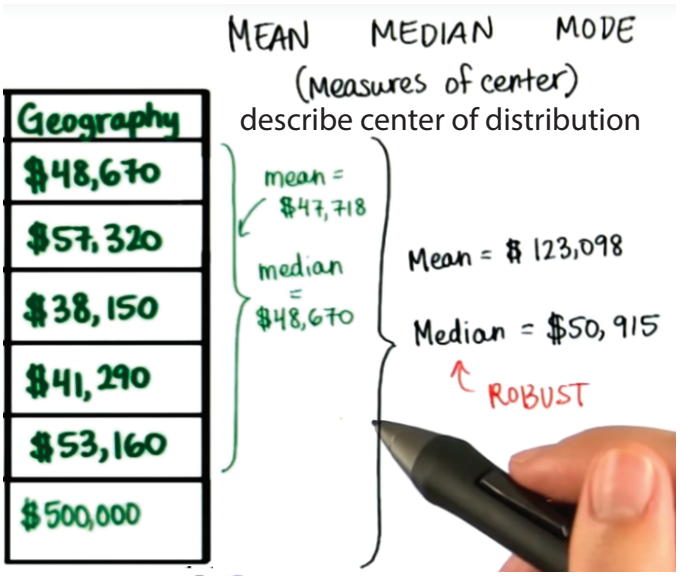
- Using symbol notation.

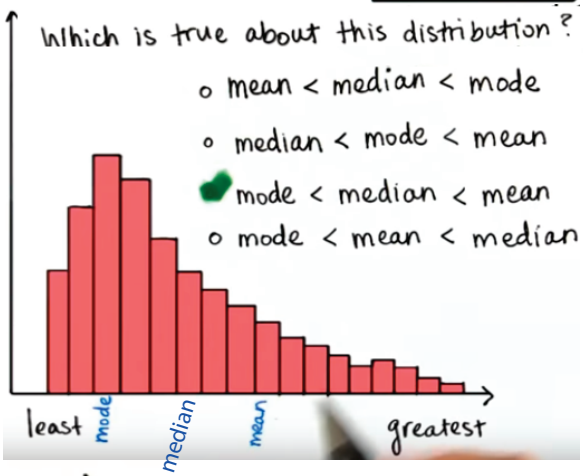
9.30 Wrap Up - Measures of Center
Here is a short doc outlining Mean, Median, and Mode.
http://tinyurl.com/measureOfCenter

LESSON 10

PS 3a: Central Tendency

10.4 Quiz 4. What Distribution?



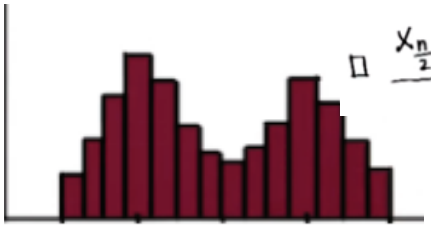


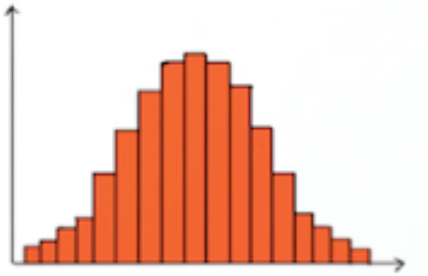
Rank	Value (x_i - i^{th} value)
1	x_1
2	x_2
3	x_3
⋮	⋮
13	x_{13}
14	x_{14} (median)
15	x_{15}
⋮	⋮
27	x_{27}

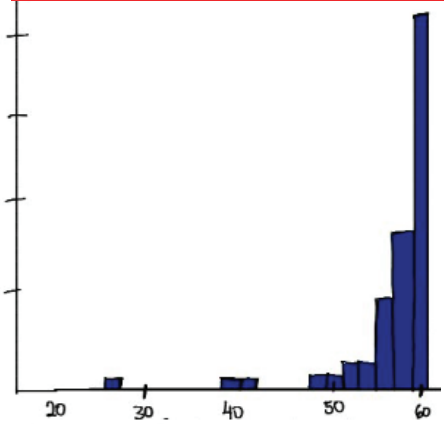
n even $\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$ n odd $x_{\frac{n+1}{2}}$

Measures of Center

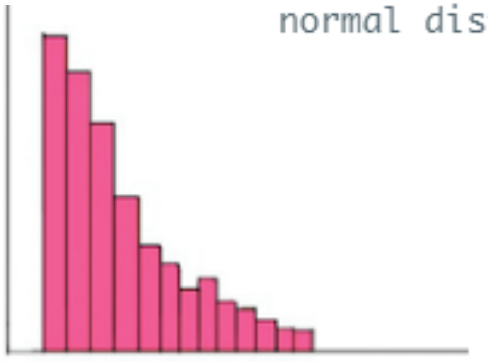
Has a simple equation	Will always change if any data value changes	Not affected by change in bin size	Not affected severely by outliers	Easy to find on a histogram
✓ $\frac{\sum x}{n}$	✓	✓		
Mean				
	✓	✓		highest frequency
Median				
			✓	✓
Mode				

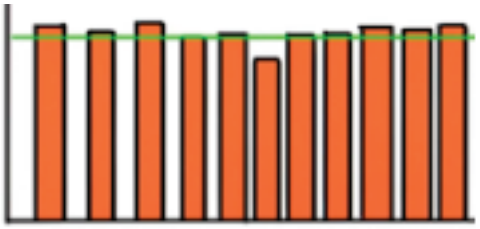

bimodal distribution


normal distribution:



negatively skewed.


positively skewed dist


uniform distribution:
Each value has the same frequency

Bertelsmann Tech Scholarship - Data Track

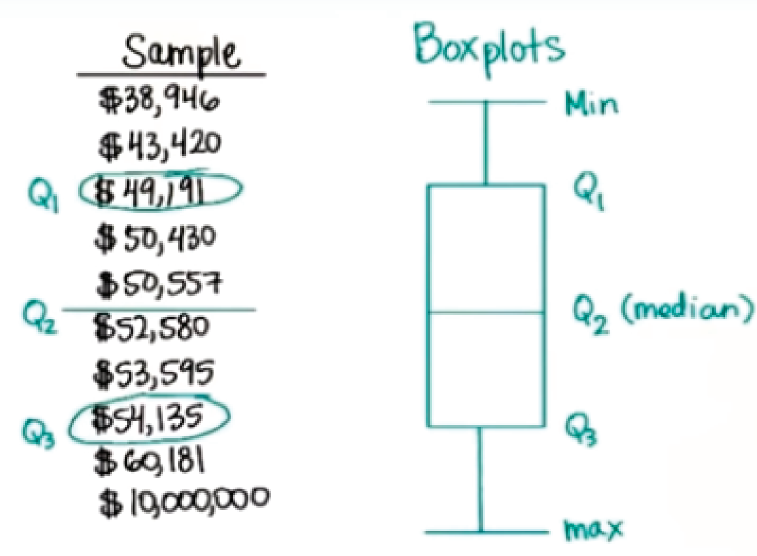
LESSON 11

PS 3b: Additional Practice (Optional)

LESSON 12

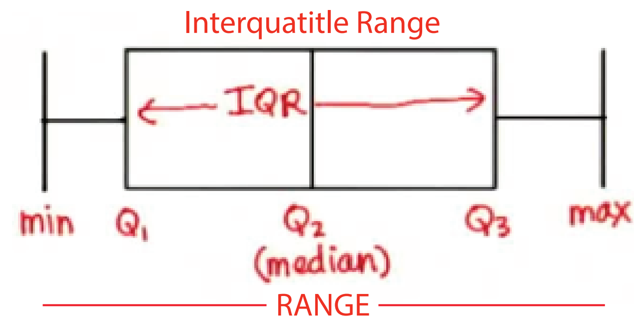
Variability

12.1 Social Networkers' Salaries
12.2 What's the Difference?
12.6 Mark Zuckerberg the Outlier
12.9 IQR
12.13 Match Boxplots
12.14 Mean Within IQR?
- We can visualize data w/ histogram & Boxplots
12.15 Problem with IQR
12.18 Deviation from Mean



Sample	Deviation from Mean $(x_i - \bar{x})$
$33,219	−19,574.80
$36,254	−16,539.80
$38,801	−13,992.80
$46,335	−6,458.80
$46,840	−5953.80
$47,596	−5197.80
$55,130	2336.20
$56,863	4069.20
$78,070	25,276.20
$88,830	36,036.20

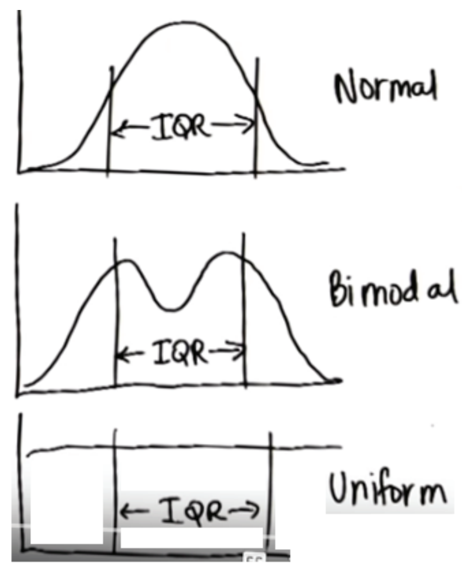
Mean $\bar{x} = \frac{\Sigma x}{10} = \$52,793.80$ Average deviation = 0 ⊙ $\frac{\Sigma(x_i - \bar{x})}{10}$



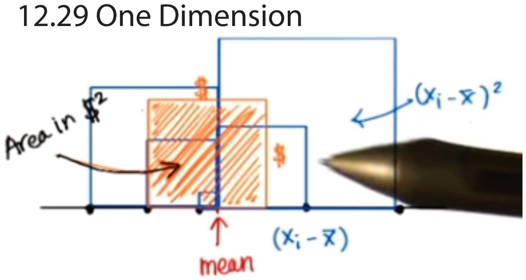
12.20 Equation for Average Deviation

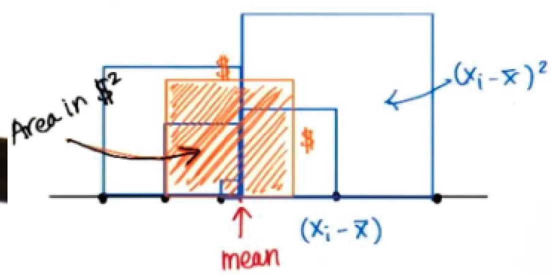
Average deviation = 0 ⊙ $\frac{\Sigma(x_i - \bar{x})}{10}$

Avg. squared de $= \frac{\Sigma(x_i - \bar{x})^2}{10} = 291,622,740$ VARIANCE



12.27 Average Squared Deviation = Variance
12.29 One Dimension



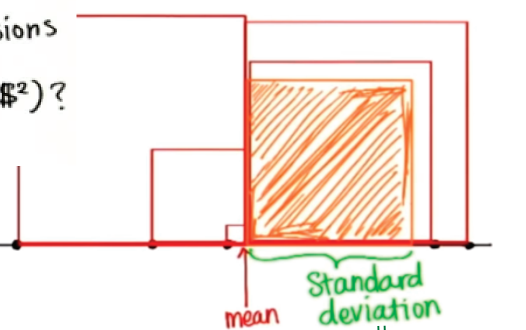


SS = Σ (area of each square)

$= \Sigma(x_i - \bar{x})^2$

Avg. squared deviation $= \frac{\Sigma(x_i - \bar{x})^2}{n}$

How do we convert the dimensions of the average square back to one dimension ($) from two ($²)?
⊙ Take the square root



12.31 Calculate Standard Deviation
square root of VARIANCE

$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{10}} = 17,077$

Avg. squared de $= \frac{\Sigma(x_i - \bar{x})^2}{10} = 291,622,740$ VARIANCE

square root of VARIANCE = σ (sigma)

Bertelsmann Tech Scholarship - Data Track

12.32 SD Social Networkers

12.32 SD Social Networkers

=A1-51511.1=constant =B1^2 = square

	A	B	C
1	38946	-12565.1	157881738.01
2	43420	-8091.1	65465899.21
3	49191	-2320.1	5382864.00999999
4	50430	-1081.1	1168777.21
5	50557	-954.099999999998	910306.809999997
6	52580	1068.9	1142547.21
7	53595	2083.9	4342639.21000001
8	54135	2623.9	6884851.21000001
9	60181	8669.9	75167166.01
10	62076	10564.9	111617[illegible]12.01
11			
12	515111		variance =
13	51511.1	<-- average	42996390.09
14	=sum(A1:A10)/10		standard deviation =
15	=average(c1:c10)=	=sqrt(c13)	6557.16326546777

12.33 SD in Words

12.34 Spreadsheet SD

12.35 Point of SD

12.38 Bessel's Correction

12.39 Clarifying Sample SD

What is a way to put the standard deviation in words?

12.33 SD in Words

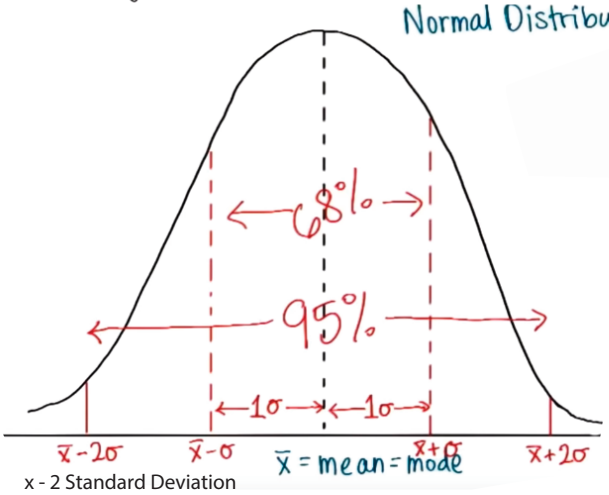
$\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$

- [x] Square root of average squared deviation ($(x_i-\bar{x})^2$, $x_i-\bar{x}$)
- [ ] (Average squared deviation) squared $\left(\frac{\Sigma(x_i-\bar{x})^2}{n}\right)^2$
- [ ] Sum of squared deviations $\Sigma(x_i-\bar{x})^2 = SS$
- [ ] Sum of (absolute deviations squared) = SS
- [x] Square root of ((sum of squared deviations) divided by n) $\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$

$$\sigma = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$$

What's so great about the standard deviation anyway?

Normal Distribution



x - 2 Standard Deviation

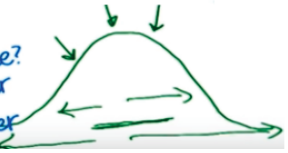
12.38 Bessel's Correction

Population
$\mu = 18.97$
$\sigma = 5.99$
$= 2.6$

Sample
$n = 9$

Bessel's Correction
Standard deviation = $\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$
Variance = $\frac{\Sigma(x_i-\bar{x})^2}{n-1}$

What will this do to the original standard deviation and variance?
- make them bigger
- make them smaller



Sample standard deviation

$$s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} \approx \sigma = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$$

(n=5) $\bar{x} = 3$

5	4
2	1
1	4
0	9
7	16

squared deviations

$\Sigma(x_i-\bar{x})^2 = 34$ (Sum of squares)

Standard deviation of sample $= \sqrt{\frac{34}{5}}$

sample standard dev. (s)

Standard deviation of population $\approx \sqrt{\frac{34}{4}}$

Bertelsmann Tech Scholarship - Data Track

LESSON 13

PS 4: Variability

L4-24 What Proportion? > L5-5 Range of Propo > L5-19 Prop over 60

13.7 Udacians' Facebook Friends (proportion)

- Proportion = SUM(n1:n2)/SUM(A1:A last)
  last 2 numbers of 13.6

13.8 Udacians' Facebook Friends (sample sd)

LESSON 14
Standardizing

14.1 Chess

14.2 USCF Distribution

- Propo data values

14.9 Continuous Distribution

- Theoritically Continuous Distribution, this model use relative frequencies. This equation help to calculate the propo of any 2 values on x axis

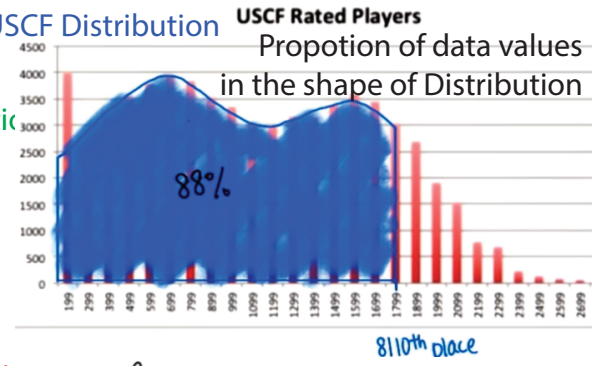
$$sv = \frac{ss}{n-1} \quad > ssd \sqrt{sv}$$

$$= 191477,35 > 437,58.$$

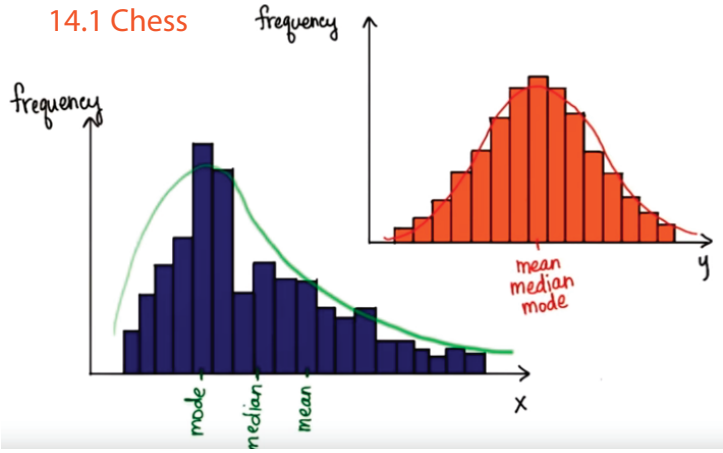
13.8 (sample sd)

14.2 USCF Distribution

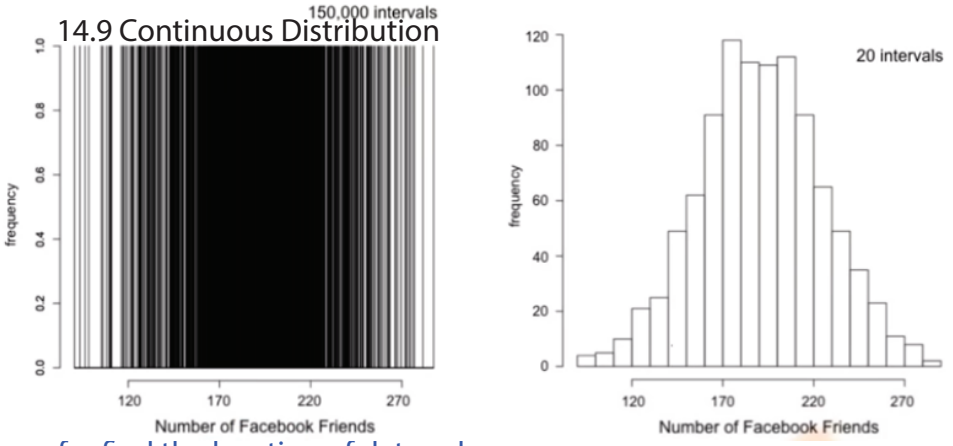
Introductio



14.1 Chess



14.9 Continuous Distribution

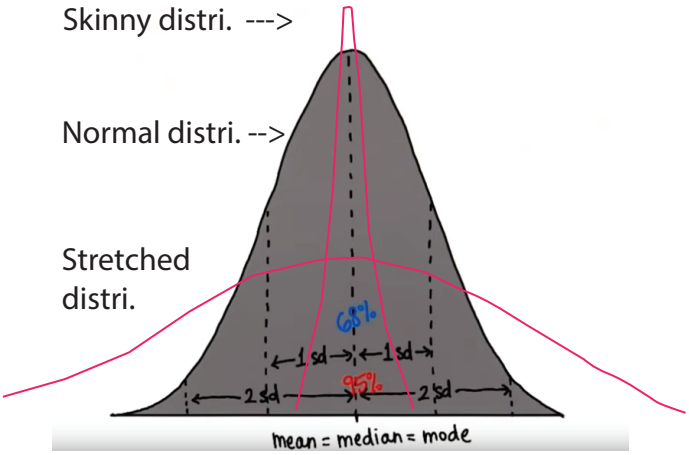


14.10 Theoretical Normal Distribution

Skinny distri. --->

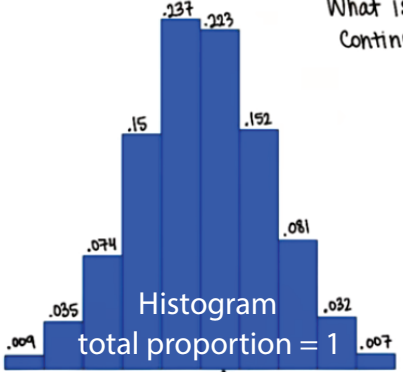
Normal distri. -->

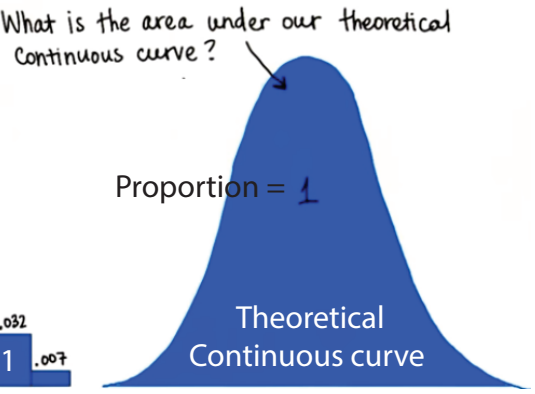
Stretched distri.



for find the location of data values

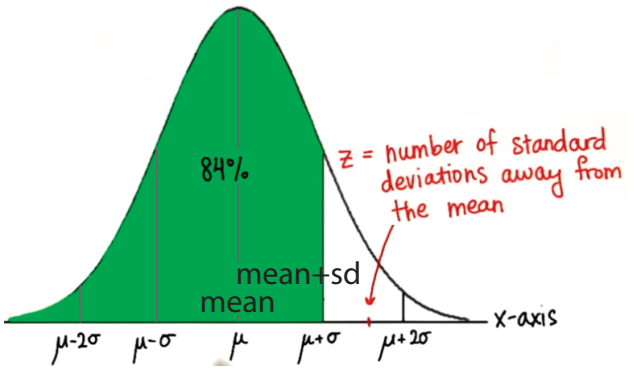
14.9 Continuous Distribution





14.11 Z

- Location of value on x-axis = SD = standard deviation

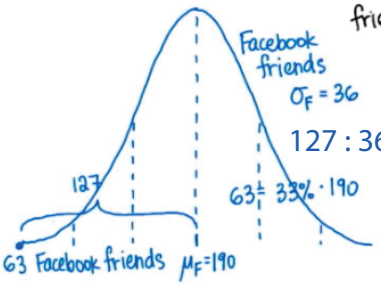


Bertelsmann Tech Scholarship - Data Track

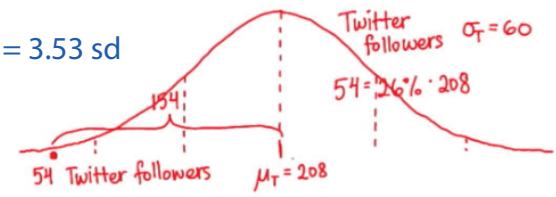
14.12 Unpopular

14.13 Katie - SDs Below

How many standard deviations is Andy's number of Twitter followers from the mean number of Facebook friends?

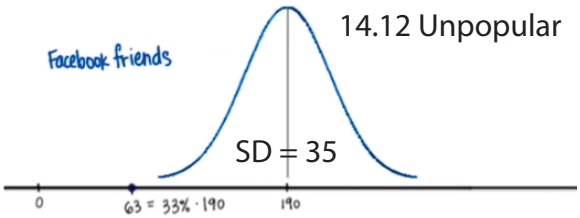


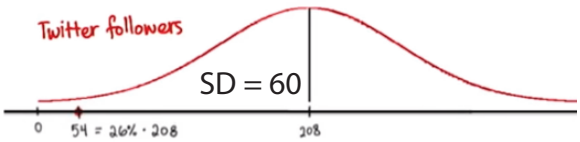
127 : 36 = 3.53 sd



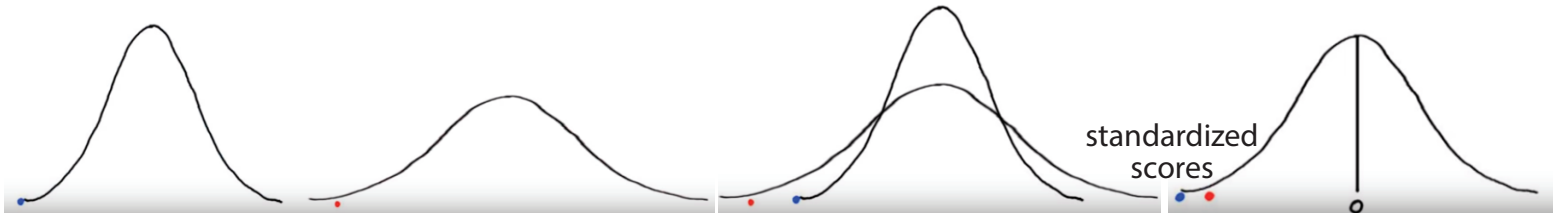
154 : 60 = 2.57 sd

14.12 Unpopular





14.15 Who's More Unpopular?



normal distribution stretched distribution standardizing distribution 3.53 sd 2.57 sd reference point

14.16 Formula for Number of SDs $\frac{(\mu - x)}{\sigma}$ $Z = \frac{(x - \mu)}{\sigma}$

14.17 Z-Score

14.19 Mean of Standardized Distribution

$\mu = 0$

If we standardize a distribution by converting every value to a z-score, what will be the new mean of this standardized distribution?

$\mu = 0$

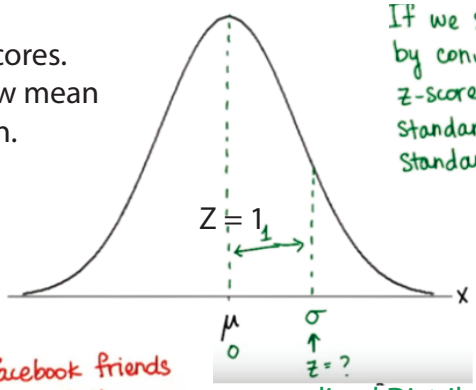
$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{\mu - \mu}{\sigma} = \frac{0}{\sigma} = 0$$

14.20 SD of Standardized Distribution = 1

- Standardized Distribution by converting values to Z scores.
- Calculate Z score by x substract the mean to get 0, now mean = 0. Then divide SD we change the shape of distribution.

14.22 Popular Chris



If we standardize a distribution by converting every value to a z-score, what will be the new standard deviation of this standardized distribution?

$$z = \frac{x - \mu}{\sigma} = \text{sigma}$$

$$\frac{\sigma - 0}{\sigma} = \frac{\sigma}{\sigma} = 1$$

normalized Distribution

normal distribution

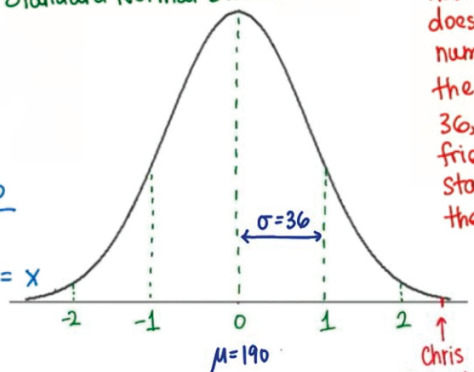
Standard Normal Distribution

$$z = \frac{x - \mu}{\sigma}$$

$$2.5 = \frac{x - 190}{36}$$

$$2.5(36) + 190 = x$$

$$280 = x$$



How many Facebook friends does Chris have if the mean number of friends is 190, the standard deviation is 36, and the number of friends Chris has is 2.5 standard deviations above the mean?

2.5 std. dev. = 36 × 2.5 = 90

190 + 90 = 280

Bertelsmann Tech Scholarship - Data

14.24 Convert to Popularity Score

LESSON 15

PS 5a: Standardizing

LESSON 16

PS 5b: Additional Practice (Optional)

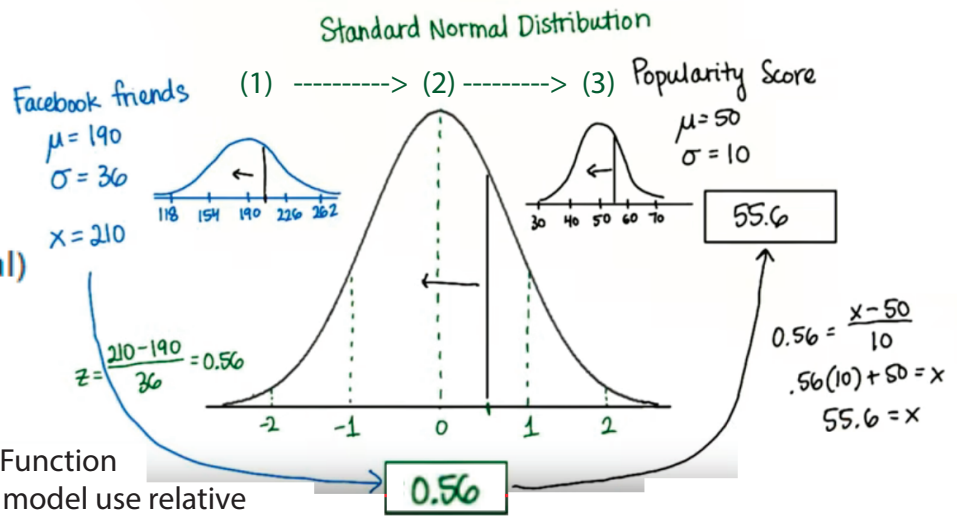
16.1 Quiz 7. Mean and SD

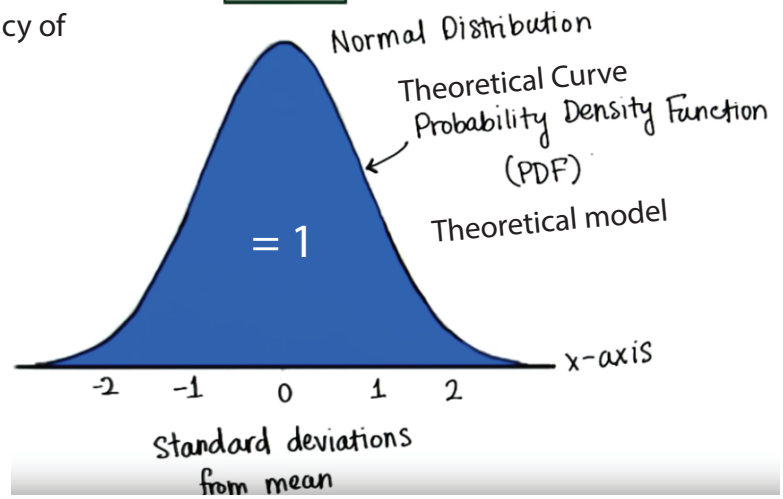
LESSON 17

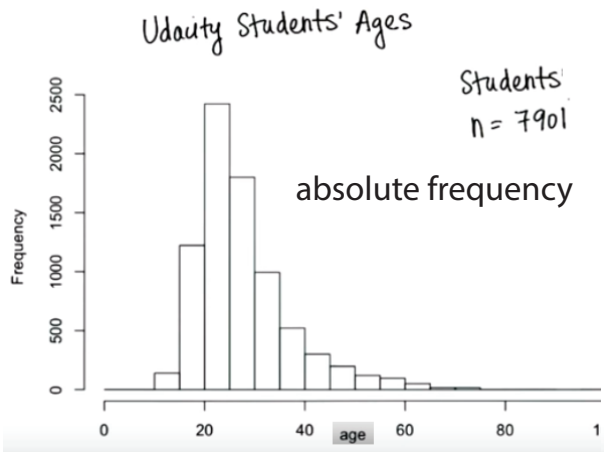
Normal Distribution

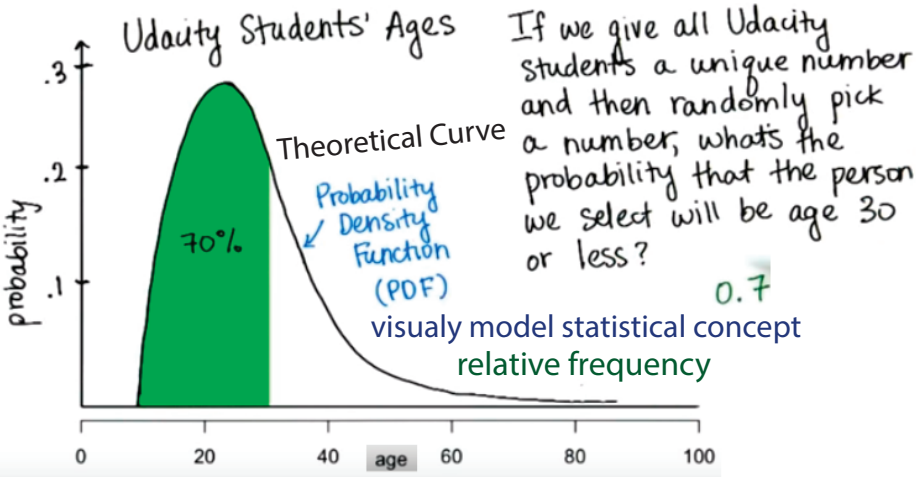
17.1 Intro to the PDF = Probability Density Function
- Theoritically Continuous Distribution, this model use relative frequencies. The curve shape model w/ relative frequency of distribution of data in term of proportions.

17.2 Probability









17.3 Get to Know the PDF