The Mean: Mathematical Notation

$$\textit{Population Mean} = \frac{\textit{Sum of all data values}}{\textit{Number of Data values}} = \mu$$

$$\textit{Sample Mean} = \frac{\textit{Sum of all SAMPLE values}}{\textit{Total number of all SAMPLE values}} = \bar{x}$$

$$\mu = \frac{\sum x}{N} \qquad \bar{x} = \frac{\sum x}{n}$$

The Mean: Mathematical Notation

Say we have a group of data values,

$x_1 = 75 \qquad x_2 = 77 \qquad x_3 = 80 \qquad x_4 = 73$

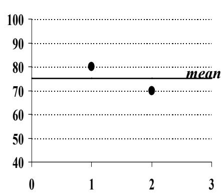
then $\sum x$ *means* $75 + 77 + 80 + 73$

and $\sum x^2$ *means* $(75)^2 + (77)^2 + (80)^2 + (73)^2$

and $\frac{\sum x}{N} = \frac{75 + 77 + 80 + 73}{4} = 76.25 = \mu =$ *the Mean*

and $\sum (x-\mu)$ *means* $(75-\mu) + (77-\mu) + (80-\mu) + (73-\mu)$

$$\textit{Population Mean} = \mu = \frac{\textit{sum of all data values}}{\textit{total number of data values}}$$



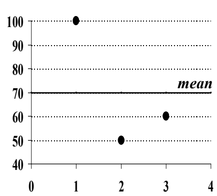
$MEAN = \mu = \frac{x_1 + x_2}{n} = \frac{80 + 70}{2} = 75$

Distances from the mean...

$(x_1 - \mu) = 80 - 75 = +5$

$(x_2 - \mu) = 70 - 75 = -5$

In general, $\sum (x-\mu) = (x_1 - \mu) + (x_2 - \mu) = 0$



$MEAN = \mu = \frac{x_1 + x_2 + x_3}{n} = \frac{100 + 50 + 60}{3} = 70$

Distances from the mean...

$(x_1 - \mu) = 100 - 70 = +30$

$(x_2 - \mu) = 50 - 70 = -20$

$(x_3 - \mu) = 60 - 70 = -10$

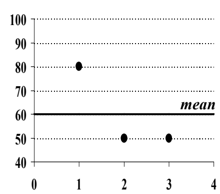
In general, $\sum (x-\mu) = (x_1 - \mu) + (x_2 - \mu) + (x_3 - \mu) = 0$

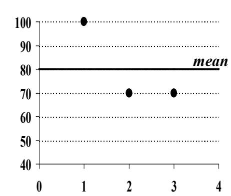
$$\mu = \frac{x_1 + x_2 + x_3 + x_4 + ... + x_N}{N} = \frac{\sum x}{N}$$

Again! The sum of the distances from the mean is zero.

Effect on Mean (μ) of adding a constant to each data value

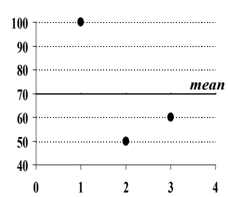
add 20 to each data value...
and the mean increases by that value

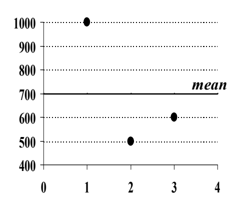




Effect on Mean (μ) of multiplying each value by a constant

multiply each data value by 10 ...
multiplies the mean by that value





The Median

Definition: The middle value of a sorted set of data.

For an odd number of data values,

8 7 9 9 9 6 10 7 5

…sort the values

5 6 7 7 8 9 9 9 10

Middle Value = Median

For an even number of data values,

8 7 9 9 9 10 10 7 8 5

…sort the values

5 7 7 8 8 | 9 9 9 10 10

$$= \frac{8+9}{2} = 8.5$$

The Mode

Definition: The data value in a distribution which occurs most frequently.

In this example,

5 6 7 7 8 (9 9 9) 10

the Mode is 9, because it occurs the most frequently.

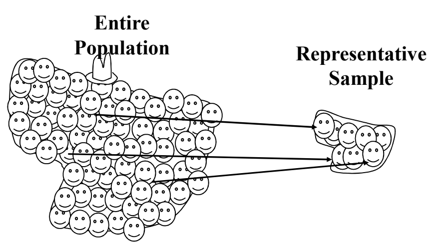
In this example,

5 6 (7 7 7) 8 (9 9 9) 10

there are 2 Modes, 7 and 9, because they both occur the most frequently. The distribution is bimodal.

If more than 2 values occur most frequently, the distribution is nonmodal.

Population Mean v. Sample Mean



$$\bar{x} = \frac{\sum x}{n}$$

$$\mu = \frac{\sum x}{N}$$

$\bar{x}$ = *mean value of sample*

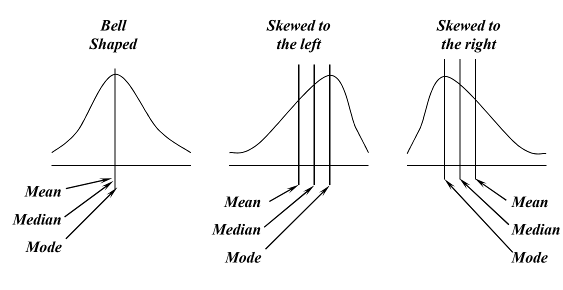
n = *number of data values in sample*

μ = *mean value of population*

N = *number of data values in population*

$\bar{x}$ is usually used to estimate μ

The Shapes of Distributions: Mean, Median and Mode



The mean is the most affected by extreme values

The median and the mode are the most resistant to outliers.

Problem Pg 146/19

For the distribution 6, 2, 4, 4, 5, 3, 5, 3, 1, 7, find:

a) the mean

c) the mode

e) $\sum x^2$

g) $\sum (x-4)^2$

i) $\sum x^2 - (\bar{x})^2$

Lucy and Joe's Diner was up for sale. They told a potential buyer that they sold an average of 100 lunches per day. Considering the data below, was this a fair statement?

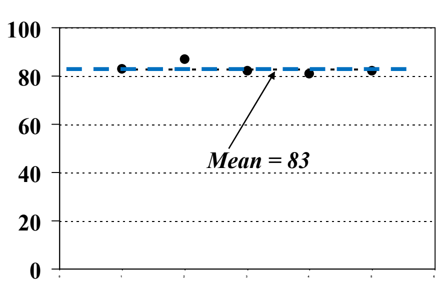
Number of lunches sold daily over the past month:

76	78	195	84	83	79
253	79	90	77	83	92
75	83	87	76	79	97
90	87	325	86	85	78
79	79	86	77	79	75

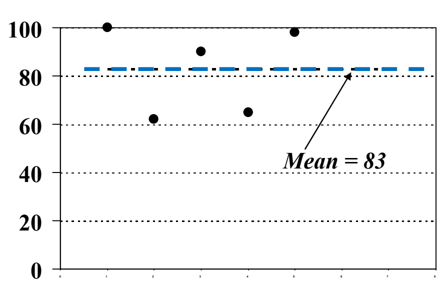
What was the true average number of lunches served per day?

Deviation from the Mean

Class 1: Exam Grades 83, 87, 82, 81 and 82

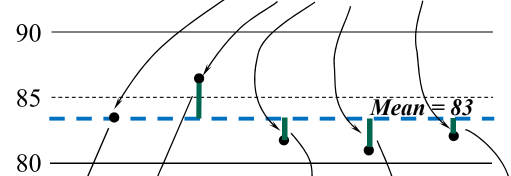


Class 2: Exam Grades 100, 62, 90, 65 and 98



Deviation from the Mean

Class 1: Exam Grades 83, 87, 82, 81 and 82



Compute the Mean of the <u>Squared</u> Deviations from the Mean

$$(83-83)^2 + (87-83)^2 + (82-83)^2 + (81-83)^2 + (82-83)^2$$

$$\frac{0+16+1+4+1}{5} = 4.4$$

This is the Mean of the <u>Squared</u> Deviations

…and is called the <u>Variance</u>.

We take the square root of the <u>Variance</u>

…and call it the <u>Standard Deviation</u> (σ)

$$\sigma = \sqrt{4.4} = 2.098$$

Compute the Mean of the Squared Deviation from the Mean

Class 1:

Mean Sq Dev = $\dfrac{\sum (x-\mu)^2}{N}$

Mean Sq Dev = $\dfrac{(83\text{-}83)^2+(87\text{-}83)^2+(82\text{-}83)^2+(81\text{-}83)^2+(82\text{-}83)^2}{5}$

Mean Sq Dev = $\dfrac{(0)^2+(4)^2+(-1)^2+(-2)^2+(-1)^2}{5} = \dfrac{22}{5} = 4.4$

Std Dev = σ = 2.096

Class 2:

Mean Sq Dev = $\dfrac{\sum (x-\mu)^2}{N}$

Mean Sq Dev = $\dfrac{(100\text{-}83)^2+(62\text{-}83)^2+(90\text{-}83)^2+(65\text{-}83)^2+(98\text{-}83)^2}{5}$

Mean Sq Dev = $\dfrac{(17)^2+(-21)^2+(7)^2+(-18)^2+(-15)^2}{5} = \dfrac{1328}{5} = 265.6$

Std Dev = σ = 16.297

Variance and Standard Deviation

For a Population:

$\dfrac{\sum (x - \mu)^2}{N}$ = *Variance* = σ^2

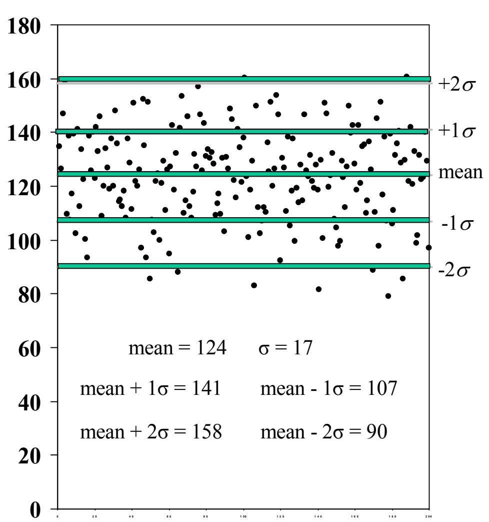
$\sqrt{\dfrac{\sum (x - \mu)^2}{N}} = \sqrt{Variance}$ = *Standard Deviation* = σ

For a Sample:

$\dfrac{\sum (x - \bar{x})^2}{n-1}$ = *Sample Variance* = s^2

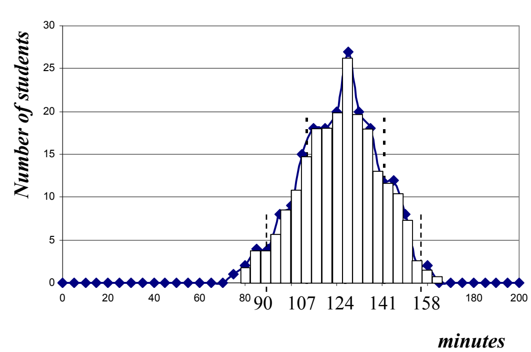
$\sqrt{\dfrac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{Sample\ Variance}$ = *Sample Standard Deviation* = s

Number of Minutes to Complete an Exam Given to 200 Students



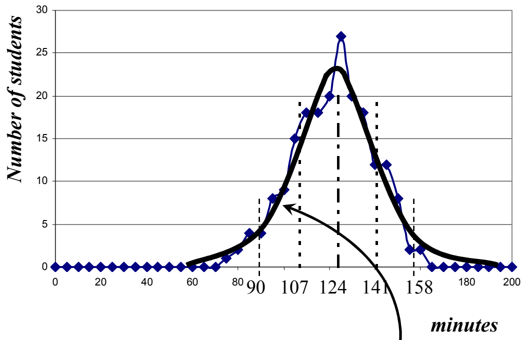
Number of Minutes to Complete an Exam Given to 200 Students

Shown the times to complete as a distribution



Number of Minutes to Complete an Exam Given to 200 Students

Shown as a distribution

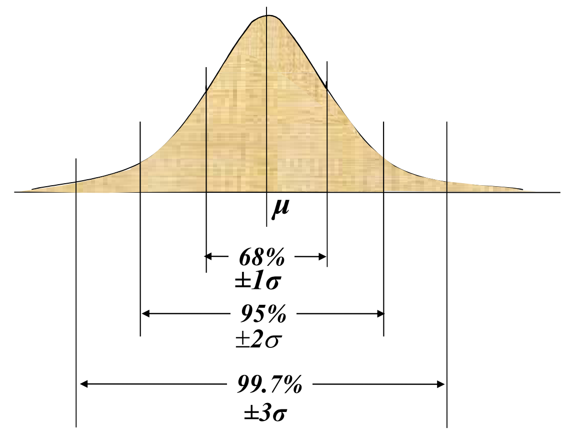


The distribution of grades
is approximated by a Normal Curve

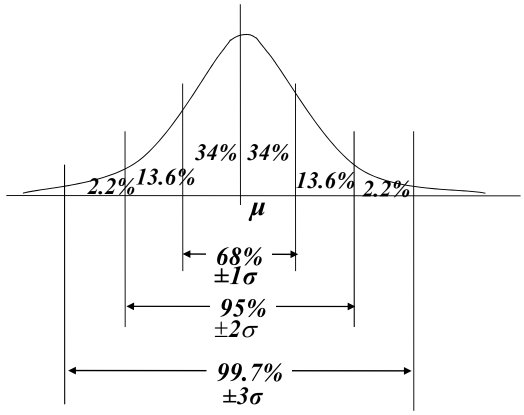
For a Bell-Shaped Distribution

The Empirical Rule States

68% of data values fall within $1\ \sigma$ of μ
95% of data values fall within $2\ \sigma$ of μ
99.7% of data values fall within $3\ \sigma$ of μ



Areas Under the Normal Distribution
if the total area under the curve is 100%

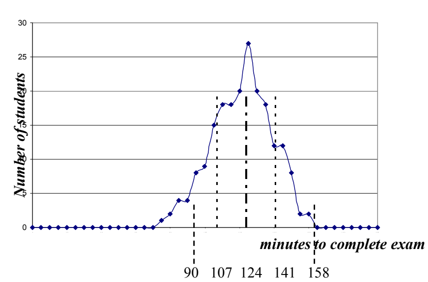


Use the calculator to find the other areas under the normal curve, say between -1σ and $+2\sigma$.

1. [2nd] [DISTR]
2. Select ⟶ 2: normalcdf
3. normalcdf(-1,2) [ENTER] ⟶ .8185946784

or approximately 81.9%

Number of Minutes to Complete an Exam Given to 200 Students



Chebyshev's Theorem :

Regardless of the shape of the distribution, the proportion of data values that fall within $\pm k$ standard deviations of the mean, is at least

$$1 - \frac{1}{k^2} \quad \text{...where } k > 1$$

For k = 2, at least $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4}$

3/4 of the data values will fall between ± 2x17 of 124

For k = 3, at least $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9}$

8/9ths of the data values will fall between ± 3x17 of 124

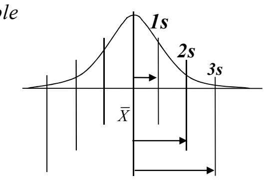
Using the Calculator to Determine Mean and Std Deviation
Using 1-Var Stats

STAT --> 1: EDIT...
enter data in L1 (or other L)
STAT --> CALC --> 1-var Stats
2nd --> L1 --> ENTER

Calculator Display

$\bar{x}$ = *mean*
$\sum x$ = *sum of data values*
$\sum x^2$ = *sum of squares of data values*
s_x = *Sample Standard Deviation*
σ_x = *Population Standard Deviation*
n = *number of data values*
minX = smallest data value
Q1 = 1st quartile
Med = median
Q3 = 3rd quartile
maxX = largest data value

For the Sample



$\pm 1S$ Contains 68% of the data points
$\pm 2S$ " 95% "
$\pm 3S$ " 99.7% "

Estimate s from the RANGE

Range: Difference between the largest and smallest data value

Sample Size	Estimated s
5	s = RANGE/2.5
10	s = RANGE/3
25	s = RANGE/4
100	s = RANGE/6

Coefficient of variation = $\frac{s}{\bar{x}} \times 100$

Review:

1. Mean, Median, Mode
2. Std Deviation characterizes the SPREAD of a distribution of data about the mean

Population:	***Sample***
Mean	***Mean***
$\mu = \frac{\sum x}{N}$	$\bar{x} = \frac{\sum x}{n}$
Std Dev	***Sample Std Dev***
$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

When all you've got is a sample…
$\bar{x}$ is the best estimate of the population mean, μ.
s is the best estimate of the population std dev, σ.

Empirical Rule: (for Bell Shaped distribution)

68% ***within*** $\pm 1\sigma$
95% ***within*** $\pm 2\sigma$
99.7% ***within*** $\pm 3\sigma$

Chebyshev's Theorem: (for any shape distribution)

Within k Std Deviations of the mean,
there are at least $(1-\frac{1}{k^2})$ of the data values. (For k > 1)

Coefficient of Variation = $\frac{\sigma}{\mu} \times 100$

Candidate	**Typing*	**Writing*	**Steno*	**WP*Law*	
Ms Libel	*7.9*	*7.9*	*7.8*	*7.7*	*8.8*
Ms Sue	*8.6*	*8.1*	*7.9*	*7.5*	*8.0*
Mr Will	*7.8*	*8.4*	*8.0*	*7.4*	*8.5*

Problem: Section3.2 Pg 146/49

For the distribution: 42, 30, 24, 48, 60, 36, 54

a) Find $\bar{x}$ and s.

b) Compute $\bar{x}$ + 1s.

c) Find the data values that are within one standard deviation above the mean.

d) Compute $\bar{x}$ - 1s.

e) Find the data values that are within one standard deviation below the mean.

f) Compute $\bar{x}$ - 2s and $\bar{x}$ + 2s.

g) What percent of the data values are within 1 standard deviation from the mean?

h) What percent of the data values are within 2 standard deviations from the mean?

Steven, a college professor, computed the mean and sample standard deviation of the final exam he administered to his day and evening statistics classes.

Class	**Mean Grade**	**Standard Deviation**
Day	**66**	**16**
Evening	**66**	**2**

a) In which class would you expect to find the lowest grade on this test? The highest grade? Explain.

b.) If Steven randomly selects a student from each one of these classes, from which class would Steven have a better chance of selecting a student with a grade closer to the mean grade?

---------------Expenditures------------------

DRUG COMPANIES		*COSMETIC COMPANIES*	
R&D	*Marketing*	*R&D*	*Marketing*
2.64	*1.74*	*1.17*	*3.75*
3.79	*2.59*	*2.38*	*4.69*
1.84	*1.97*	*0.87*	*2.75*
2.36	*3.64*	*1.24*	*3.15*
3.59	*2.18*	*1.75*	*3.56*
4.61	*2.64*	*0.75*	*2.94*
3.17	*2.55*	*1.29*	*4.75*
2.89	*1.90*	*1.67*	*3.87*

a) Compute the mean and standard deviation for each type of company and each type of expenditure.

b.) What conclusions can you draw from the data?