2013-09-18-C01

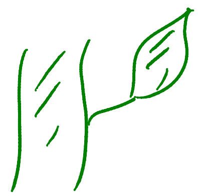
Wednesday, September 18, 2013
1:42 PM

Ch. 2 Descriptive Statistics

Ex. Charles Minard's graph 1812 war

http://en.wikipedia.org/wiki/Charles_Joseph_Minard

a) shape of distribution



· Stem & leaf diagram

173 : $\underset{\times 10}{17} \mid \underset{\times 1}{3}$

Ex. Test Scores

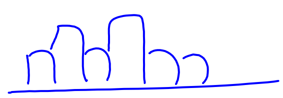
70, 72, 76, 80, 84, 84, 88, 90, 90, 94, 96, 98, 100, 100, 100

$n = 15$

Count	Stem (×10)	Leaf (×1)
3	7	0 2 6
4	8	0 4 4 8
5	9	0 0 4 6 8
3	10	0 0 0
15		

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/TestScoresMBA-StemLeaf.xls

Histogram



http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/TestScoresMBA-DifferentWidths.xls

Options for different class lengths.
Which one is best?

General rule. n: # measurements
K: # classes

K: # classes

L: class length

① Find smallest K such that

$$2^K > n$$

② $L = \dfrac{\text{max. meas} - \text{min. meas}}{K}$

$n=15$

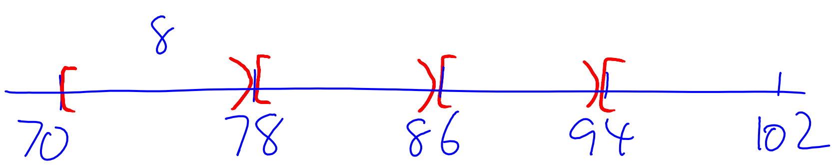
		2^K		$n=15$
×	$K=1$	2	$\not>$	15
×	$K=2$	$2^2=4$	$\not>$	15
×	$K=3$	$2^3=8$	$\not>$	15
	$K=4$	$2^4=16$	$>$	15

$$L = \frac{100-70}{4} = 7.5 \xrightarrow{} 8$$

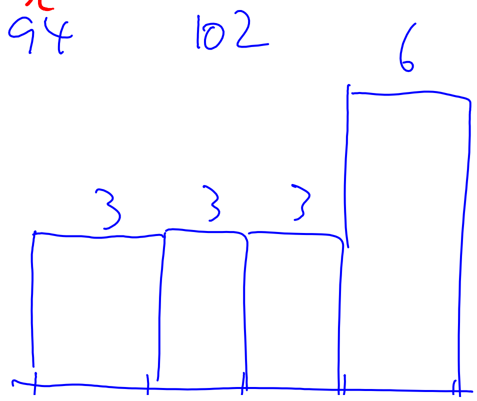
round up

∴ $K=4$ classes

$L=8$ class length

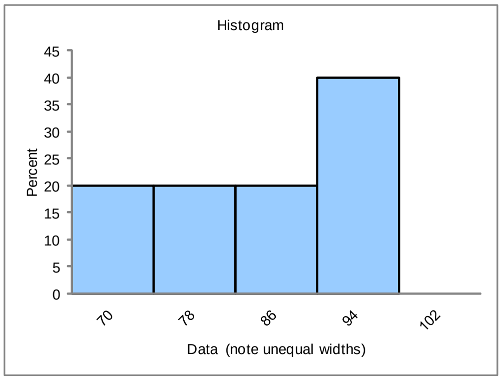


Class	Freq
70 to <78	3
78 to <86	3
86 to <94	3



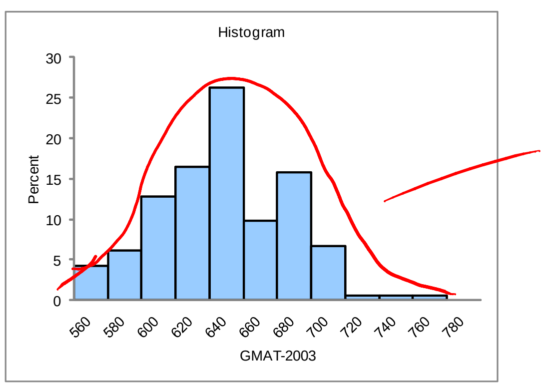
94 to <102 $\frac{6}{15}$

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/TestScoresMBA-Nice.xls

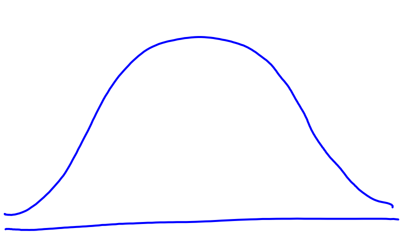


Ex. GMAT Scores

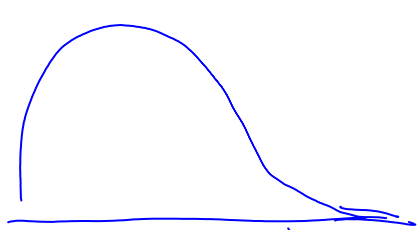
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/GMAT-03-04-05.xls



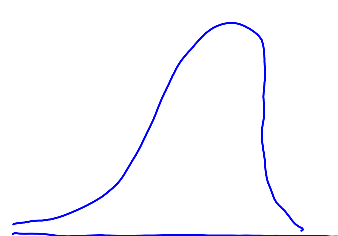
approx. normal



Symmetric



positively-skewed



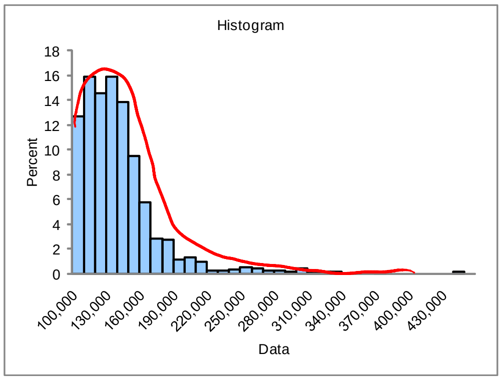
negatively-skewed

Ex. Pb. 2.8, p.37 unequal widths

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Pb28-p37.pdf

Ex. Mac salaries

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/McMaster Salary-2011-for-2010.xlsx



b) Mean, median + mode

popn parameters μ σ^2

sample statistics $\bar{x}$ s^2

· Mean

Ex. 15 test scores: 70, 72, 76, ..., 98, 100, 100, 100

$$\mu = \frac{70+72+\cdots+100}{15} = 88.13$$

$$\mu = \frac{X_1+\cdots+X_N}{N}$$

Sample of size n

$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$, $\bar{x} = \frac{70+72+80}{3} = 74$

↗ sample statistic : point estimate of μ

• Median M_d

Ex. 6 people in an office

	A	36	(\$1,000)
	B	42	
	C	30	
	D	40	
	E	45	
President	F	300	

$\mu = \frac{36 + \cdots + 300}{6} = \$82{,}167$

Order

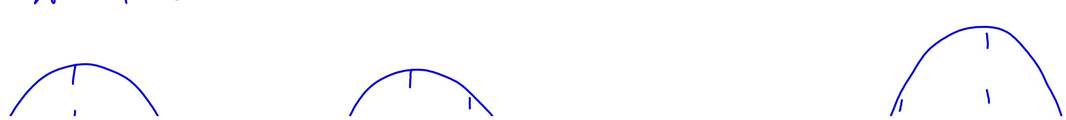
30
36
40 ⟵ $M_d = 41$
42
45
300

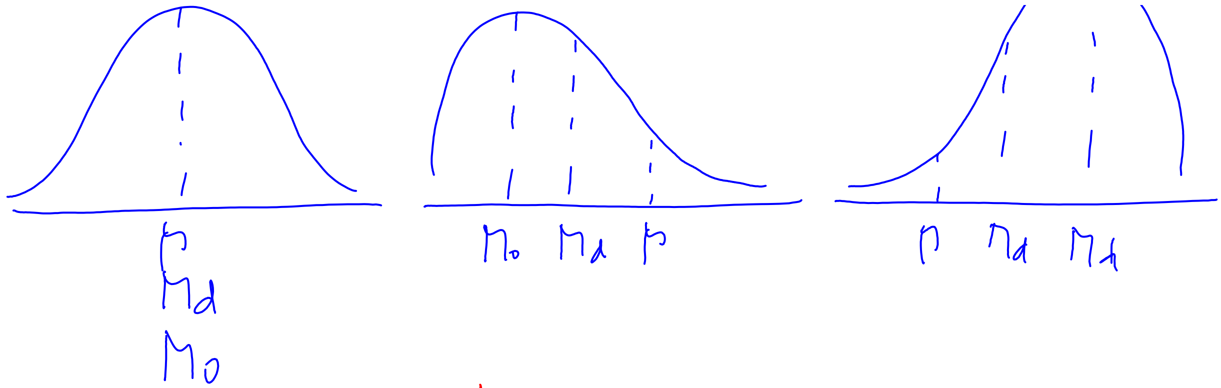
Median & mean supplement each other

Mode M_o : most frequently occurring value in data

Ex. Test scores

70 72 76 80 84
84 88 90 90 94 ↙ Mode
96 98 100 100 100





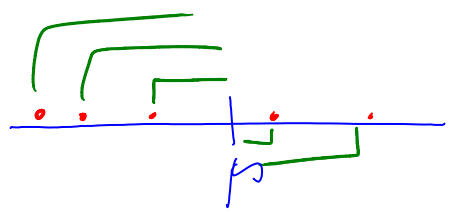
c) Range + variation measures

Ex. Three groups of salaries

		Range	Mean	Median	Mode
A	1, 3, 5, 7, 9	8	5	5	N/A
B	3, 5, 5, 5, 7	4	5	5	5
C	5, 5, 5, 5, 5	0	5	5	5
D	5 5 5 5 100	95			
E	5 60 70 80 100	95			

Range

Variance



N, μ

$$\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2}{N} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

A: 1, 3, 5, 7, 9 , $\mu = 5$

x	$x-\mu$	$(x-\mu)^2$
1	-4	16
3	-2	4
5	0	0
7	2	4
9	4	16
	0	40

$$\sigma^2 = \frac{40}{5} = 8$$

$$\sigma = \sqrt{\sigma^2} = 2.83$$

X	$X-\mu$	$(X-\mu)^2$
1	-4	16
3	-2	4
5	0	0
7	2	4
9	4	16
	0	40

$$\sigma^2 = \frac{40}{5} = 8$$

$$\sigma = \sqrt{\sigma^2} = 2.83$$

	μ	σ^2	σ
A	5	8	2.83
B	5	6	1.26
C	5	0	0

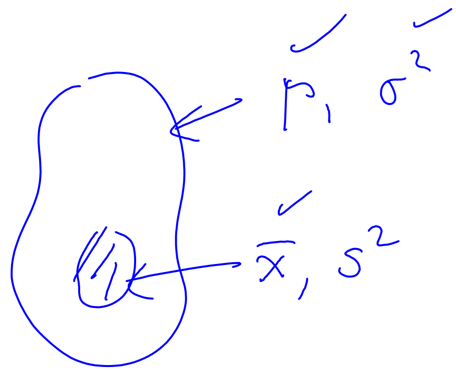
Q: Why do we use σ (not σ^2)

μ : \$

σ^2 : \$^{2}

σ : \$

What about Sample?



$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

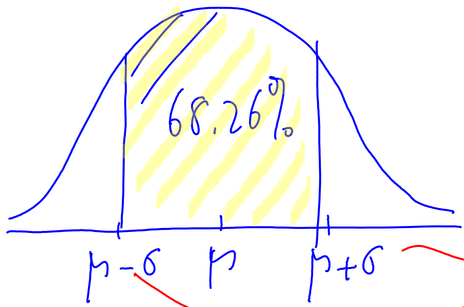
Pop σ^2

Sample s^2

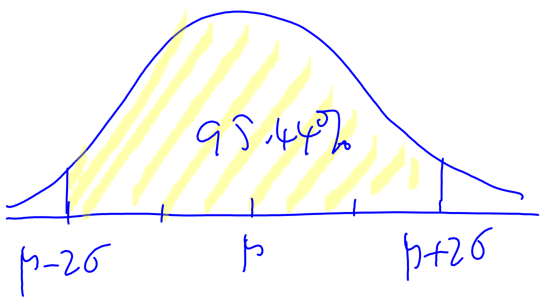
Empirical rule for "normal" pop'ns using μ, σ

Empirical rule for "normal" pop'ns using μ, σ

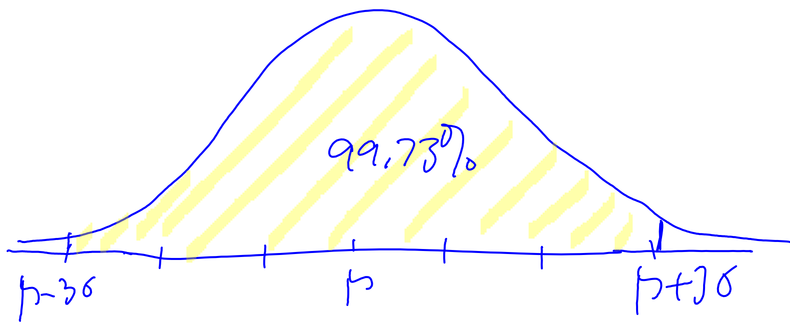
Tolerance interval



$(\mu \mp \sigma) = (\mu-\sigma, \mu+\sigma)$



$(\mu \mp 2\sigma)$

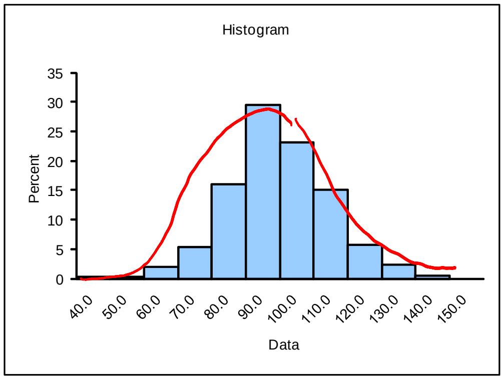


$(\mu \mp 3\sigma)$

Six-sigma 99.9997% (3 defects in a million)
(Quality Control)

Ex. IQ

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/IQScores-500.xls



What are z-scores?

(μ, σ)

μ $\quad x = \mu + \sigma$

$$x = \mu + \sigma$$
$$x - \mu = \sigma$$
$$\frac{x-\mu}{\sigma} = 1$$

μ $\quad x = \mu + 2\sigma$

$$x = \mu + 2\sigma$$
$$\frac{x-\mu}{\sigma} = 2$$

$x = \mu - 3\sigma$ $\quad \mu$

$$\frac{x-\mu}{\sigma} = -3$$

If $\quad x = \mu + z\sigma$

$\Updownarrow$

$\Rightarrow \quad \dfrac{x-\mu}{\sigma} = z$

(Jay-Z?)