Thursday, September 19, 2013
10:13 AM

Ch. 2 Descriptive Statistics

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

C. Minard

a) Shape of a distribution

- Stem + leaf

Stem (×10)	Leaf (×1)
17	3

173 ⟹

Ex. Test scores

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

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

- Dot plots + histograms

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/TestScores MBA-DotPlot.xls

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

width (length) : 1, 4, 10, 20

Which is best?

n: # data points
K: # classes
L: class length

① Find smallest K such that $2^K > n$. Then,

② → $L = \frac{\max - \min}{K}$

$n = 15$

2^K vs. $n = 15$

$K = 1$ $\quad 2^1 = 2 \not> 15$

$\times$ 2

$\times$ 3

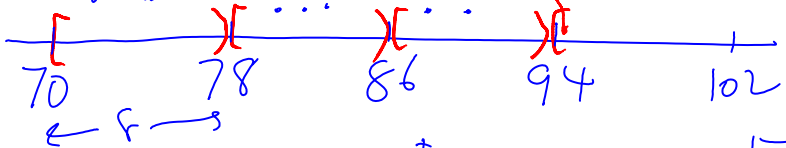
$K = 4$ $\quad 2^4 = 16 > 15$

$L = \frac{100 - 70}{4} = 7.5 \longrightarrow 8$ round up to

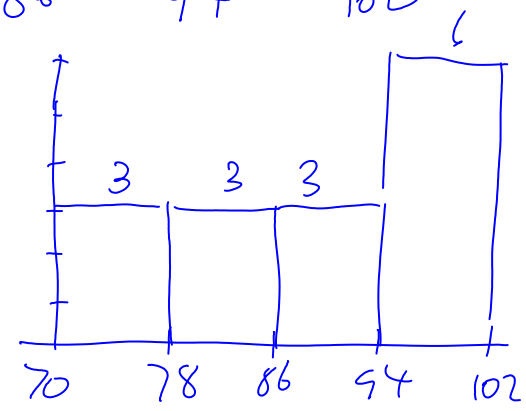
$K = 4, \quad L = 8$

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

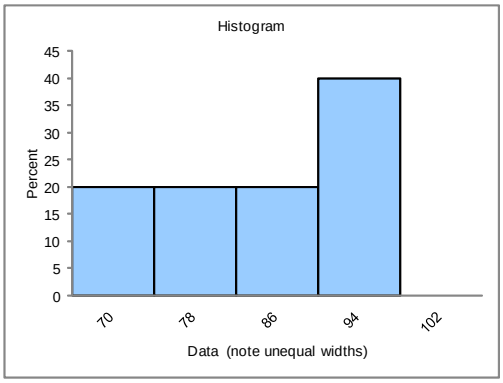
Next?



Class	Freq
70 to <78	3
78 " <86	3
86 " <94	3
94 " <102	6

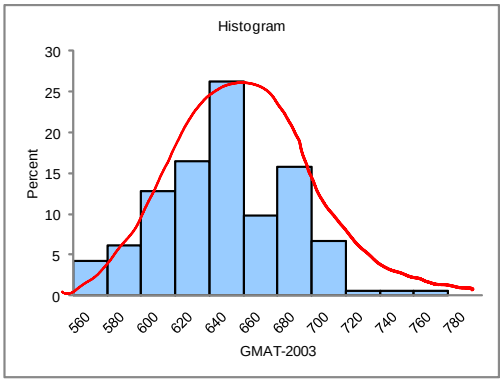


http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/TestScores MBA-Nice.xls

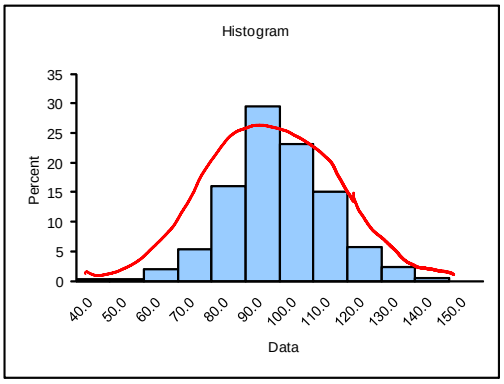


Ex. GMAT Scores 2003-2004-2005

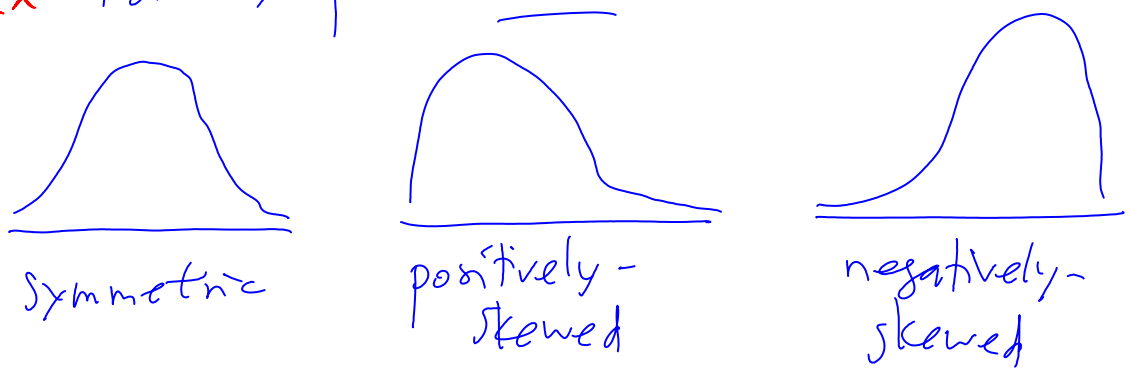
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/GMAT-03-04-05.xls



Ex IQ



Ex Pb. 2.8, p. 37



Ex Mac salaries

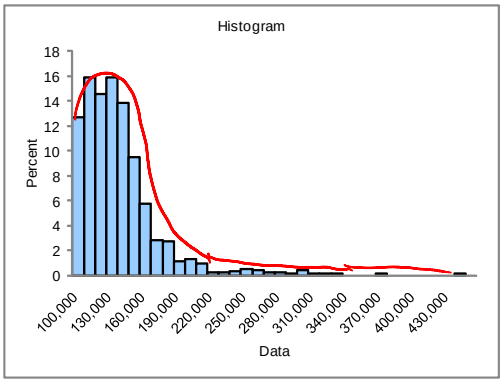
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/McMasterSalary-2011-for-2010.xlsx

count	928
mean	140,885.0597
sample standard deviation	35,939.6578
sample variance	1,291,659,004.8693
minimum	100032.24
maximum	448977.4
range	348945.16

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\McMasterSalary-2011-for-2010.xlsx>

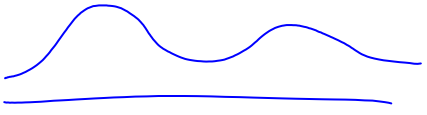
1st quartile	117,646.0025
median	134,052.8300
3rd quartile	151,663.5975
interquartile range	34,017.5950
mode	124,382.6300

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\McMasterSalary-2011-for-2010.xlsx>

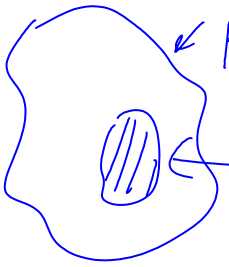


Positively skewed!

Heights



b) Mean, median, mode



pop'n parameter μ σ^2

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

- Mean (μ or $\bar{x}$)

$N=15$, 70, 72, 76, 80, 84; 84; 88, 90, 90, 94; 96, 98, 100, 100, 100

Pasted from <http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/ch-02.html>

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

$$\mu = \frac{X_1+\cdots+X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Sample of $n=3$

$$\bar{x} = \frac{70+72+80}{3} = 74$$

$$\bar{x} = \frac{x_1+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

sample statistic for population mean μ

- Median M_d

Ex. 6 people

Typist Sec	A	36	($1,000)
	B	42	
	C	30	

Sec
B 42
C 30
D 40
E 45
President F 300

$\mu = \$62,167$

Rearrange
30
36
40 ← 41 = M_d
42
45
300

Median M_d: value above & below which lie an equal # of measurements

Ex. Mac $\mu = 141$
$M_d = 134$

Mode M_o: most frequently occurring value in data

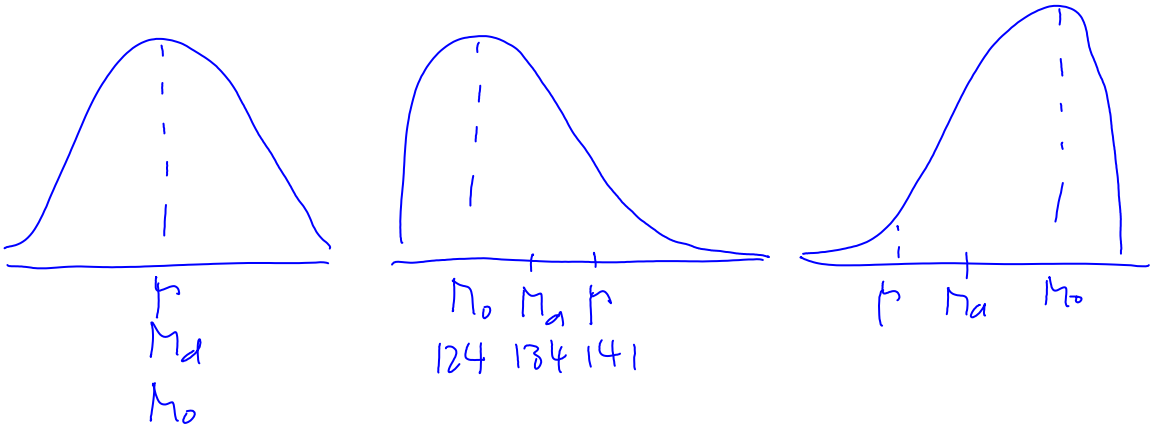
Ex. Test Scores

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

$M_o = 100$

Pasted from <http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/ch-02.html>

Ex. Mac $M_o = 124$



c) Range + Variance

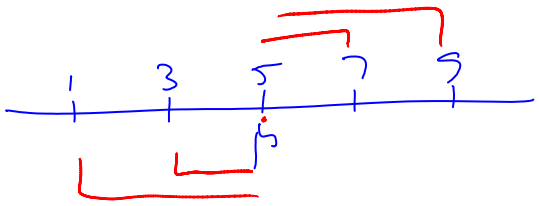
Ex Three groups

		μ	M_d	M_o	Range
A	1, 3, 5, 7, 9	5	5	N/A	8
B	3, 5, 5, 5, 7	5	5	5	4
C	5, 5, 5, 5, 5	5	5	5	0
D	5 5 5 5 100				95
E	5 60 70 80 100				95

Variance σ^2

$$\sigma^2 = \frac{(X_1-\mu)^2+(X_2-\mu)^2+\cdots+(X_N-\mu)^2}{N} = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$$

A



$\mu = 5$

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

(var)

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

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

(std. deviation)

	μ	σ^2	σ
A	5	8	2.83
B	5	1.6	1.26
C	5	0	0

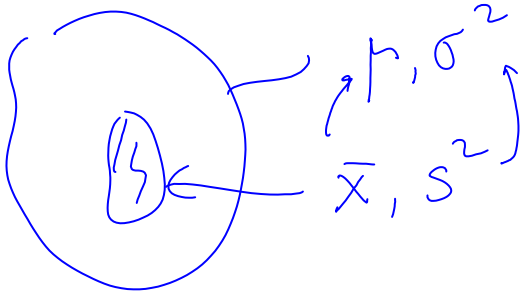
X	X-μ	(X-μ)²
1	-4	16
3	-2	4
5	0	0
7	2	4
9	4	16
	Σ=0	Σ=40

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

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

(std. deviation

	μ	σ^2	σ
A	5	8	2.83
B	5	1.6	1.26
C	5	0	0
	$	$^2	$



$$\mu = \frac{1}{N}\sum x_i$$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$\sigma^2 = \frac{1}{N}\sum (x_i - \mu)^2$$

Sample var

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

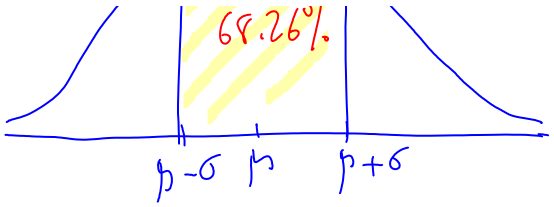
Pop σ^2

Sample

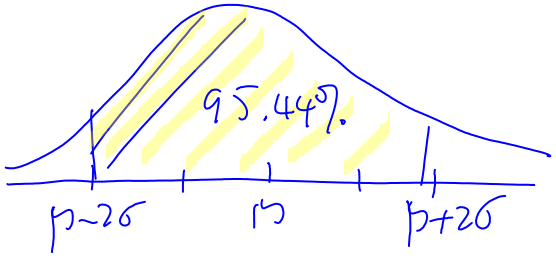
Empirical rule for "normal" pop'n μ, σ

Tolerance interval

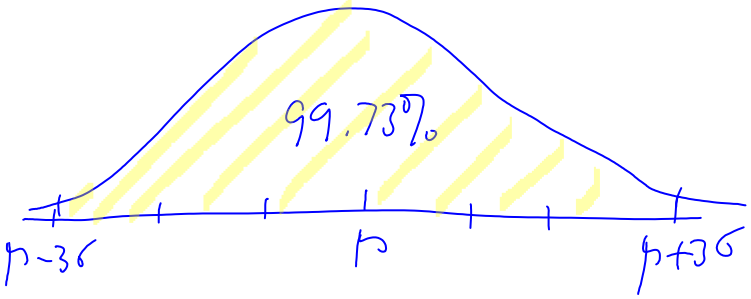
$(\mu \mp \sigma) - (\mu - \sigma \;\; \mu + \sigma)$



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



$(\mu \mp 2\sigma)$

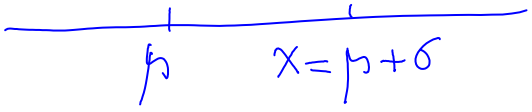


$(\mu \mp 3\sigma)$

Six-sigma (6σ) 99.9997%

(3 defects out of 1 million)

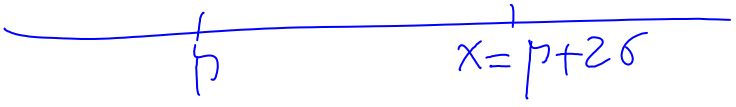
What z-score?



$x = \mu + \sigma$

$x - \mu = \sigma$

$\frac{x-\mu}{\sigma} = (1)$



$x = \mu + 2\sigma$

$\frac{x-\mu}{\sigma} = (2)$