2013-09-17-EC01

Tuesday, September 17, 2013
6:35 PM

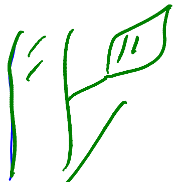
Ch.2 Descriptive Statistics

Ex. Charles Minard

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

a) Shape of a distribution

· Stem & leaf diagram



173

$\overset{\times 10}{17} \mid \overset{\times 1}{3}$ ← last digit

Ex. Test scores in an MBA elective course

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

n=15 Scores

Count	Stem (x10)	Leaf (x1)
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

Dot plot & histogram

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

4 different cases → which is best?

General rule. n: # measurements

K: # classes
L: class length

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

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

In our case, $n = 15$

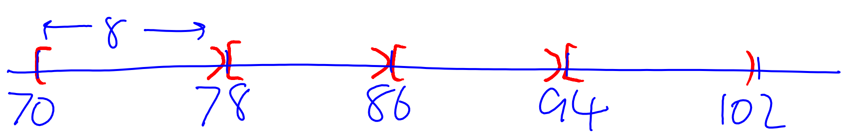
✗	$K=1$:	$2^1 = 2$	$\not>$	15	
✗	$K=2$:	$2^2 = 4$	$\not>$	15	
✗	$K=3$:	$2^3 = 8$	$\not>$	15	
	$\boxed{K=4}$:	$2^4 = 16$	$>$	15	✓

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

$$\therefore \boxed{K=4, \quad L=8}$$

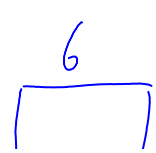
What's next

Range: $70 \rightarrow 100$

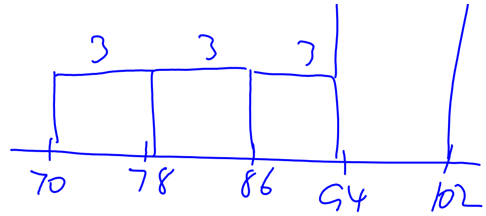


What to do with 94? Make cells disjointed

Class	Freq
70 to <78	3

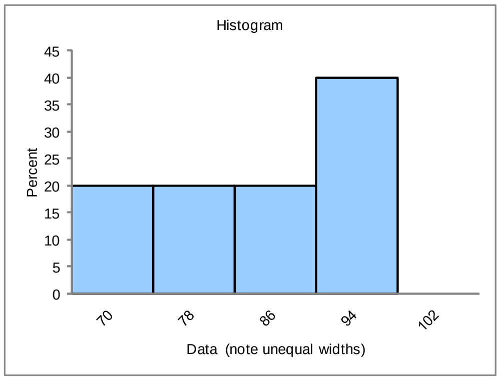


78 to <86 3
86 to <94 3
94 to <102 6
15



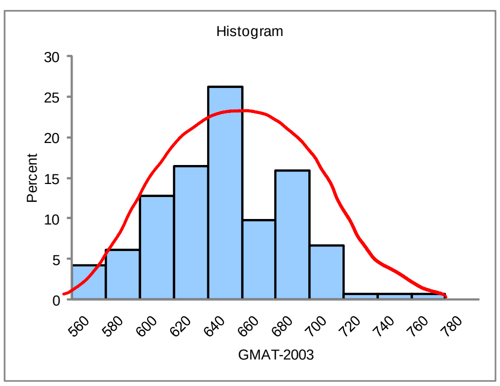
Ideal sol'n w/ Megastat

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



Ex GMAT 2003, 2004, 2005

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

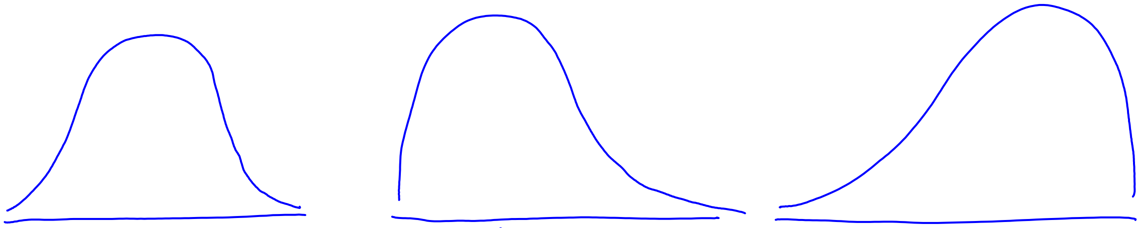


2003
L = 20
K = 11

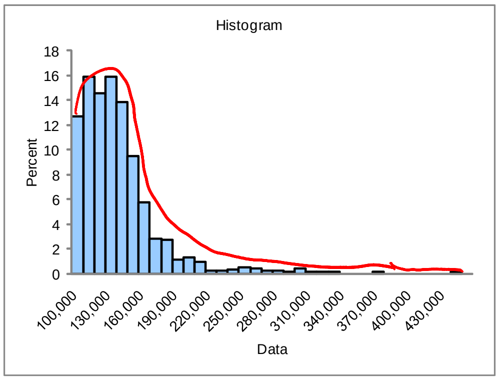
Ex. Pb. 2.8, p.37

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

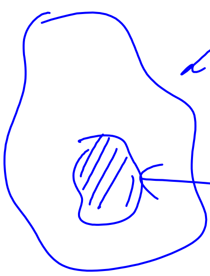
Skewness (+, -)



Symmetric | positively skewed | negatively skewed



b) Mean, median & mode (Central tendency)



popn (parameter) μ σ^2

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

- Mean

Ex. Test Scores

$N = 15$: 70, 72, 76, . . . , 98, 100, 100, 100

μ: popn mean (avg)

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

$$\mu = \frac{X_1 + \cdots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$ (Σ: sigma summation

$$\mu = \frac{X_1 + \cdots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$ (Σ : sigma summation notation)

Sample of $n=3$: 70, 72, 80

$\bar{x}$: sample mean

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

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

"sample statistic" : point estimate of the pop'n mean

· Median M_d

Ex. 6 people in an office

			Order
	A	36 (1,000)	30
	B	42	36
	C	30	40 ← $M_d = 41$
	D	40	42
	E	45	45
Pres.	F	300	300

$\mu = \$82,167$

M_d : value above + below which lie an equal # of measurements (50th percentile)

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
1st quartile	117,646.0025
median	134,052.8300

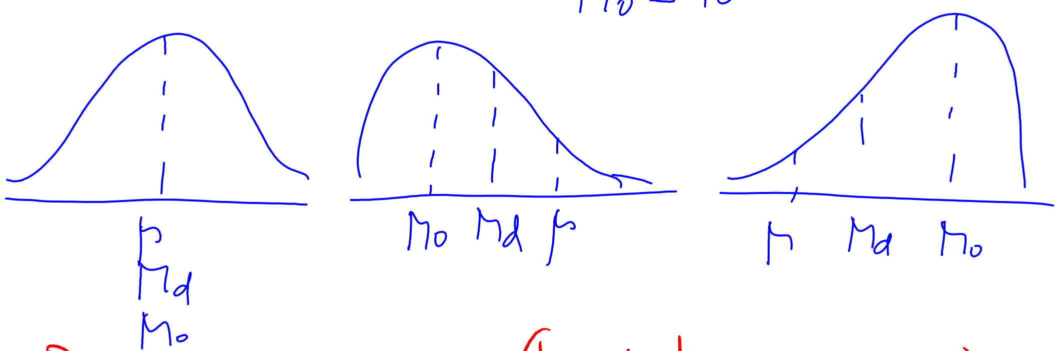
Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\McMasterSalary-2011-for-2010.xlsx>

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

$\mu = 88.13$
$M_d = 90$
$M_o = 100$



c) Range + variance (Variation measures)

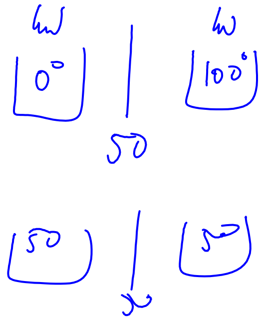
Ex Three groups of workers

$1,000			Mean	Median	Mode
	A	1, 3, 5, 7, 9	5	5	N/A
	B	3, 5, 5, 5, 7	5	5	5
	C	5, 5, 5, 5, 5	5	5	5

Range: A : 8
B : 4
C : 0

		Range
D:	5, 5, 5, 5, 100	95
E:	5, 60, 70, 80, 100	95

Variance



Hi variance

no "

N, μ

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

Set 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
	$\Sigma = 0$	$\Sigma = 40$

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

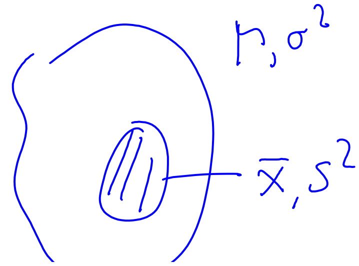
$\sigma = \sqrt{\sigma^2}$

standard deviation

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

μ : \$ units
σ^2 : $\$^2$ "
σ : \$ "

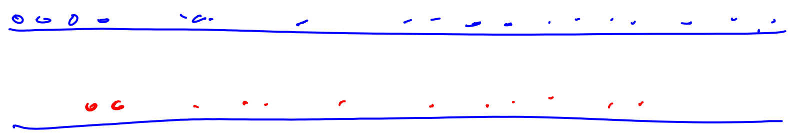
Sample variance



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

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

Pop



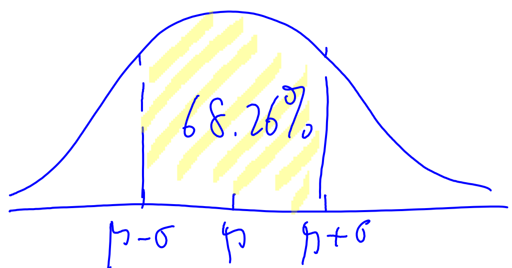
Samp

σ^2

s^2

Empirical rule for "normal" populations (using μ, σ)

Tolerance interval



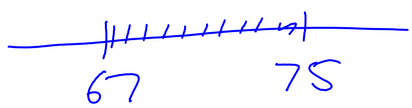
$(\mu \mp \sigma] = (\mu-\sigma, \mu+\sigma]$

$(\mu \mp 2\sigma)$

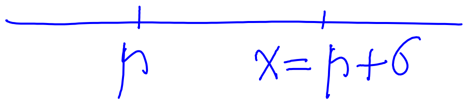
$(\mu \mp 3\sigma)$

6-sigma 99.9997% 3 defects in a million
(Quality control)

Coffee temp.



What are z-scores?



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

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

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