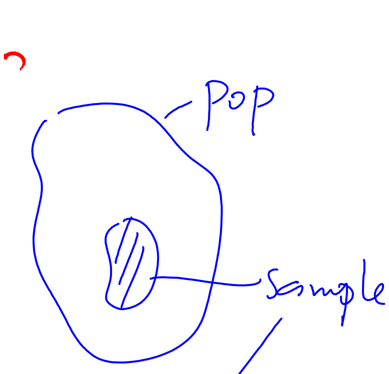
2013-10-16-C01

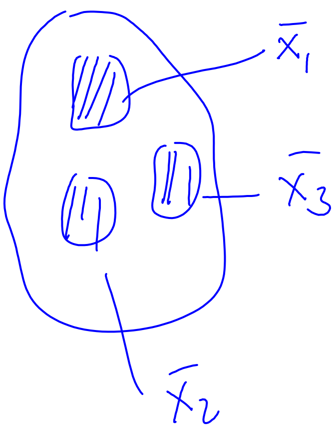
Wednesday, October 16, 2013
1:17 PM

Ch.6 Sampling distribution

Chs.1-5 : Descriptive
Chs.6 → : Inferential



summarized by "statistic", $\bar{x}$



Ex. GMAT-2003

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/GMAT-2003-Samples.xls

Count	GMAT-2003		Sample 1	Sample 2	Sample 3
1	580		580	620	620
2	610		610	620	660
3	660		660	610	650
4	660		660	640	680
5	650		650	610	690
6	650		650	640	640
7	690		690	620	650
8	690		690	640	670
9	620		620	610	640
10	600		600	680	680
11	650				
12	620				

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\GMAT-2003-Samples.xls>

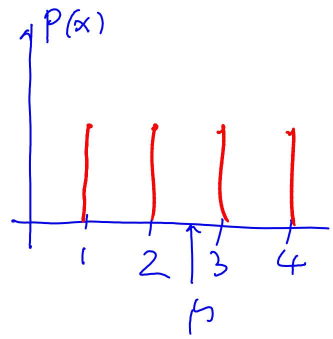
Descriptive statistics					
	GMAT-2003				
count	164				
mean	643.48				
minimum	560				
maximum	760				
range	200				
population variance	1,501.94				

minimum	560				
maximum	760				
range	200				
population variance	1,501.94				
population standard deviation	38.75				
Descriptive statistics					
	Sample 1	*Sample 2*	*Sample 3*		*Average of all three samples*
count	10	10	10		
mean	641.00	629.00	658.00		642.67
sample standard deviation	37.25	21.83	22.01		
sample variance	1,387.78	476.67	484.44		
minimum	580	610	620		
maximum	690	680	690		
range	110	70	70		

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\GMAT-2003-Samples.xls>

Ex. Prizes in a lottery (or, Stock returns)

Ball	Prize	# freq	$p(x)$
A	1	15	1/4
B	2	15	1/4
C	3	15	1/4
D	4	15	1/4
		60	



$$E(X) = \mu = 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\frac{1}{4} + 4\cdot\frac{1}{4} = 2.5$$

$$\sigma^2 = \sum (x-\mu)^2 p(x) = 1.25$$

$$\sigma = \sqrt{\sigma^2} = 1.12$$

(1) All possible samples of $n=2$

Samples

Sample 1	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

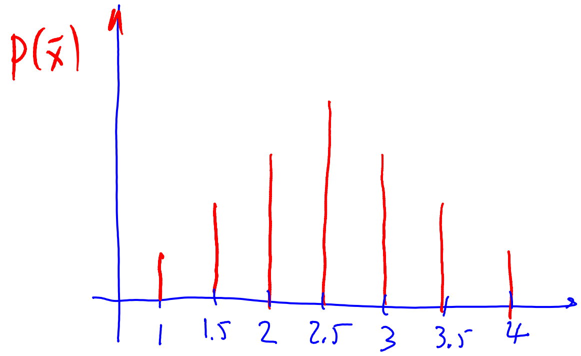
$\bar{x}$	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

Sample 1

1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

1	1	1.5	2	2.5
2	1.5	2	2.5	3
3	2	2.5	3	3.5
4	2.5	3	3.5	4

$\bar{x}$	Freq	$p(\bar{x})$
1	1	1/16
1.5	2	2/16
2	3	3/16
2.5	4	4/16
3	3	3/16
3.5	2	2/16
4	1	1/16



$$E(\bar{X}) = \mu_{\bar{X}} = 1\cdot\frac{1}{16} + 2\cdot\frac{2}{16} + \cdots + 1\cdot\frac{1}{16} = \boxed{2.5 = \mu}$$

$$\sigma_{\bar{X}}^2 = (1-2.5)^2\frac{1}{16} + \cdots + (4-2.5)^2\frac{1}{16} = 0.625$$

1.25	2	0.625
σ^2	n	$\sigma_{\bar{X}}^2$

$$\frac{1.25}{2} = 0.625$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Stocks-2013.xlsx

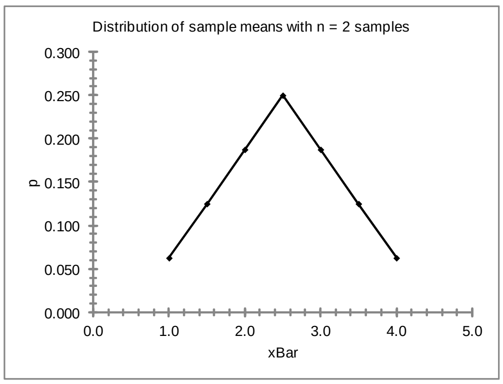
Stocks	% Return (x)	p(x)	x·p(x)	(x-μ)²	(x-μ)²·p(x)
A	1	0.25	0.25	2.25	0.5625
B	2	0.25	0.5	0.25	0.0625
C	3	0.25	0.75	0.25	0.0625
D	4	0.25	1	2.25	0.5625
			2.5		1.25
			μ		σ²

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Stocks-2013.xlsx>

n = 2 samples			xBar	xBar	Freq.	p	xBar·p	(xBar-μ)²	(xBar-μ)²·p

1	1		1.0	1.0	1	0.063	0.063	2.250	0.141
1	2		1.5	1.5	2	0.125	0.188	1.000	0.125
1	3		2.0	2.0	3	0.188	0.375	0.250	0.047
1	4		2.5	2.5	4	0.250	0.625	0.000	0.000
2	1		1.5	3.0	3	0.188	0.563	0.250	0.047
2	2		2.0	3.5	2	0.125	0.438	1.000	0.125
2	3		2.5	4.0	1	0.063	0.250	2.250	0.141
2	4		3.0						
3	1		2.0		16	1.000	2.500		0.625
3	2		2.5				E(Xbar)		σ²(Xbar)
3	3		3.0				Same as μ=2.5		Same as
3	4		3.5						σ²/2=1.25/2 =0.625
4	1		2.5						
4	2		3.0						
4	3		3.5						
4	4		4.0						

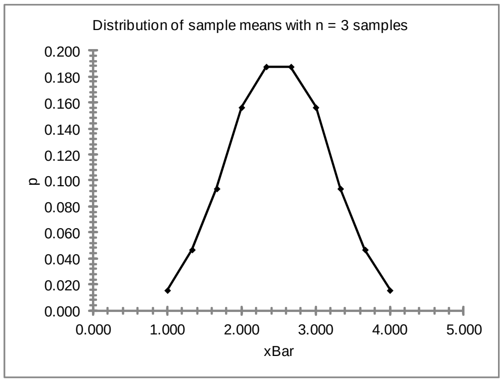
Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Stocks-2013.xlsx>



h2 (2) All Sampls of size n=3

n = 3 samples				xBar		xBar	Freq.	p	xBar·p	(xBar-μ)²	(xBar-μ)²·p
1	1	1		1.000		1.000	1	0.016	0.016	2.250	0.035
1	1	2		1.333		1.333	3	0.047	0.062	1.361	0.064
1	1	3		1.667		1.667	6	0.094	0.156	0.694	0.065
1	1	4		2.000		2.000	10	0.156	0.313	0.250	0.039
1	2	1		1.333		2.333	12	0.188	0.437	0.028	0.005
1	2	2		1.667		2.667	12	0.188	0.500	0.028	0.005
1	2	3		2.000		3.000	10	0.156	0.469	0.250	0.039
1	2	4		2.333		3.333	6	0.094	0.312	0.694	0.065
1	3	1		1.667		3.667	3	0.047	0.172	1.361	0.064
1	3	2		2.000		4.000	1	0.016	0.063	2.250	0.035
1	3	3		2.333							
1	3	4		2.667			64	1.000	2.500		0.417
1	4	1		2.000					E(Xbar)		σ²(Xbar)
1	4	2		2.333					Same as μ= 2.5		Same as
1	4	3		2.667							σ²/3= 1.25/3= 0.417
1	4	4		3.000							
2	1	1		1.333							
2	1	2		1.667							

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Stocks-2013.xlsx>



So, $E(\bar{X}) = \mu_{\bar{x}} = \mu$ always

$$\left.\begin{aligned} \sigma_{\bar{X}}^2 &= \frac{\sigma^2}{n} \\ \sigma_{\bar{X}} &= \frac{\sigma}{\sqrt{n}} \end{aligned}\right\} \text{ ~~differ~~ differ slightly for finite pop's}$$

In general

Populat. (N)

Infinite, or very large ($N \geq 20n$)

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Finite ($N < 20n$)

With replacement

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

w/o replacement

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

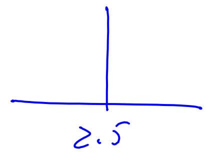
Ex. $N = 60$, $n = 4$: N vs. $20n$
$60 < 20 \cdot 4 = 80$ (finite)

So, $\sigma_{\bar{X}} = \frac{1.12}{\sqrt{4}}\sqrt{\frac{60-4}{60-1}} = \underbrace{(0.56)(0.974)}_{} = .545$

Ex If $N=60$, $n=60$

Finite $\sigma_{\bar{X}} = \frac{1.12}{\sqrt{60}} \sqrt{\frac{60-60}{60-1}} = 0$

$\mu_{\bar{X}} = \mu = 2.5$



Ex. Shape of the sample mean distrib

http://highered.mcgraw-hill.com/sites/0070000237/student_view0/visual_statistics.html

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/ComparisonOfSampleDistributions.wmf

b) Central Limit Theorem

If n is sufficiently large (≥ 30), then the sample mean $\bar{X}$ is approx. normal with

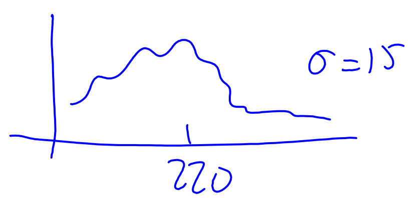
mean $\mu_{\bar{X}} = \mu$

st. dev. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

regardless of the shape of pop'n distrib

Ex. Speedboat engines (V6, Laser XRi) Mercury

Pop'n. $\mu = 220$ hp
$\sigma = 15$ hp



Potential buyer
$n=100$, Buy if $\bar{X} > 217$

$Pr(\bar{X} \leq 217) = ?$

$\bar{X}$ is normal

$$\mu_{\bar{X}} = \mu = 220$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$$

Standardize

$$\bar{X} \le 217$$

$$\underbrace{\frac{\bar{X} - 220}{1.5}} \le \frac{217 - 220}{1.5}$$

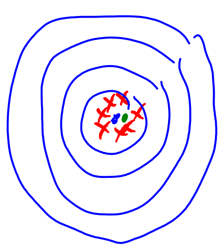
$$\Pr(Z \le -2) = 0.0228$$

no sale likelihood

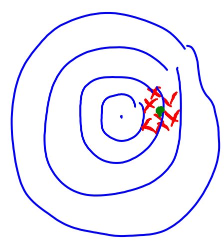
c) Unbiasedness & minimum variance estimation

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

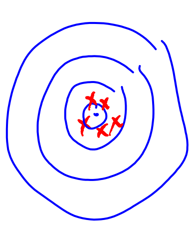
$$\sigma^2 \rightarrow s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

✓ 

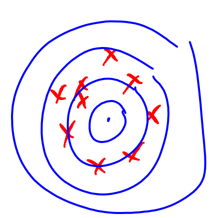
Unbiased
Correct aim



biased
aim incorrect

✓ 

small variance



large variance

Ex. $N=4$ people & their income

x_1: 1, x_2: 1, x_3: 3, x_4: 4

$$\mu = \frac{1+1+3+4}{4} = \frac{9}{4} = 2.25$$

Sample of size $n=3$

Estimation

NOT GOOD!

x_1	x_2	x_3	x_4	$\bar{x}$	Median
1	1	3		5/3	1
1	1		4	2	1
1		3	4	8/3	3
	1	3	4	8/3	3

$$\mu_{\bar{x}} = \frac{9}{4} = 2.25 = \mu \qquad \mu_{median} = \frac{8}{4} = 2 \neq \mu$$

unbiased ✓ biased

d) Distribution of sample proportion

Ex New Coke

Coca-Cola market share 60% 1945

24% 1983

CEO Roberto Goizueta

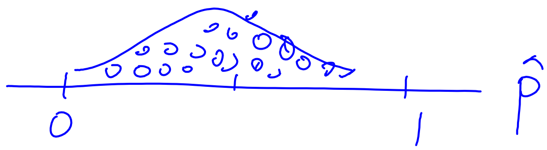
↓ "New Coke" → disaster

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

$n=100$, 40 would buy

$$\hat{p} = \frac{40}{100} = .4$$

$$\hat{p} = \frac{40}{100} = .4$$



Sample proportion $\hat{p}$ is normal, provided that
n is large $(np \geq 5,\ n(1-p) \geq 5)$

$$\left.\begin{array}{lll} \text{with mean} & & \mu_{\hat{p}} = p \\ & \text{variance} & \sigma^2_{\hat{p}} = \frac{1}{n} p(1-p) \\ & \text{s.d} & \sigma_{\hat{p}} = \sqrt{\frac{1}{n} p(1-p)} \end{array}\right\}$$

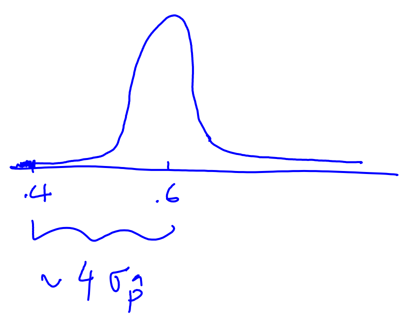
CEO thinks that $p = 0.6$

We found $\hat{p} = 0.4$ $(n = 100)$

$$Pr(\hat{p} \leq 0.4 \text{ given that } p = 0.6) = ?$$

$$\mu_{\hat{p}} = 0.6$$

$$\sigma_{\hat{p}} = \sqrt{\frac{1}{100}(.6)(.4)} = .049$$



$$\hat{p} \leq .4$$

$$\frac{\hat{p} - .6}{.049} \leq \frac{.4 - .6}{.049}$$

$Pr\left(Z \leq -4.08\right)$

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/SmallProb.xlsx

0.0000225178503892032

$\simeq 0.00002$