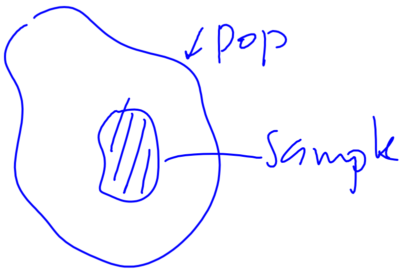
2013-10-15-EC01

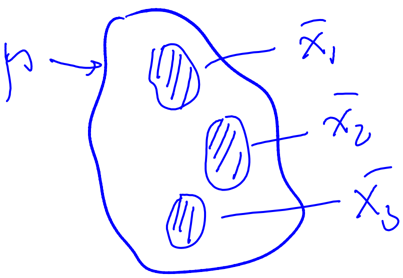
Tuesday, October 15, 2013
5:46 PM

Ch.6. Sampling Distributions

Chs. 1–5 Descriptive
Chs. 6 → Inferential



a) Distribution of sample mean



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				
13	650				
14	640				
15	610				
16	690				

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

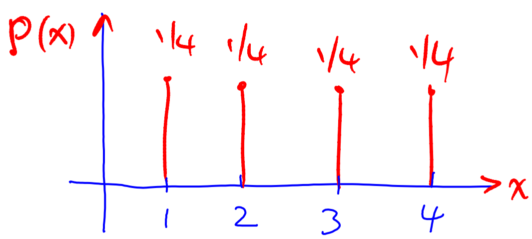
Descriptive statistics					
	GMAT-2003				
count	164				

mean	643.48				
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-1.xls>

Ex. four different types of stocks

Stocks	% return (x)	# stocks	p(x)
A	1	15	1/4
B	2	15	1/4
C	3	15	1/4
D	4	15	1/4
		60	



Single sample (n=1)

$$\mu = \sum x p(x) = 2.5$$

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

$$\sigma = \sqrt{1.25} = 1.12$$

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

Sample 2

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

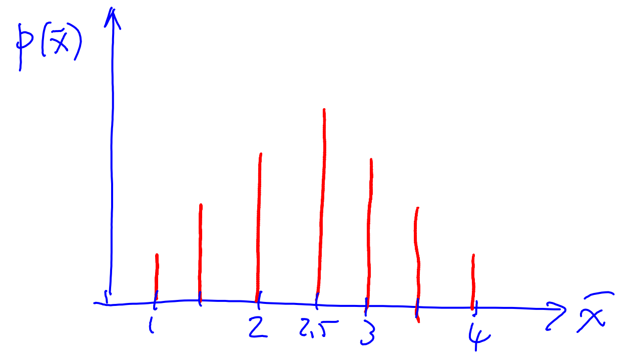
$\bar{x}$	1	2	3	4
1	1	1.5	2	2.5
2	1.5	2	2.5	3

Sample 1

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

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
	16	



$$E(\bar{X}) = \mu_{\bar{x}} = 1 \cdot \frac{1}{16} + 1.5\frac{2}{16} + \cdots + 4\frac{1}{16} = 2.5$$

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

1.25 0.625 2

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{2}$$

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

✓

(2) n = 3 case

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

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)		σ^2(Xbar)
1	4	2		2.333					Same as μ= 2.5		Same as
1	4	3		2.667							σ^2/3= 1.25/3= 0.417
1	4	4		3.000							
2	1	1		1.333							
2	1	2		1.667							
2	1	3		2.000							
2	1	4		2.333							
2	2	1		1.667							

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

General result with sample size n

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

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

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

(almost) always true

In general,

Population (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. Stocks

$N = 60, \; n = 4$ $\qquad$ N vs. $20n$

$60 < 20 \cdot 4 = 80 \rightarrow$ Use finite pop n form

So, $\sigma_{\bar{x}} = \frac{1.12}{\sqrt{4}} \sqrt{\frac{60-4}{60-1}} = (0.56)\sqrt{\frac{56}{59}} =$

$= 0.56 \times 0.9742 = .54$

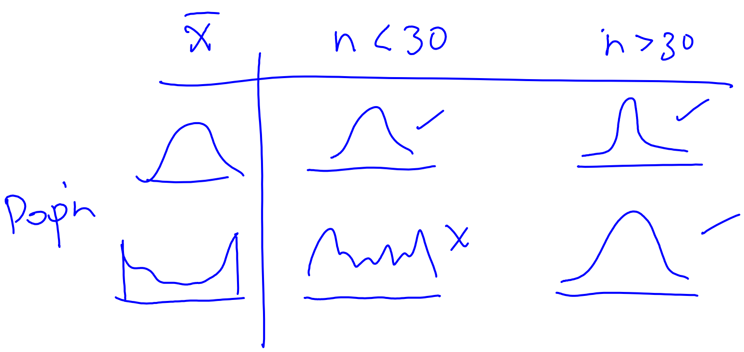
Ex. $N = 60, \quad n = 60$

$\sigma_{\bar{x}} = \frac{1.12}{\sqrt{60}} \sqrt{\frac{60-60}{60-1}} = 0$ $\qquad$ 100% sure about pop'n mean

Ex. Shape of sample mean $(\bar{x})$ distribution

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

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



b) Central limit theorem: If n is sufficiently large $(n > 30)$, then the sample mean $\bar{x}$ is approximately normal with mean

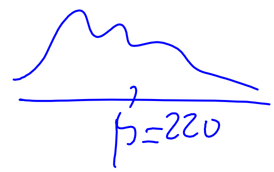
$\mu_{\bar{x}} = \mu$

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

regardless of the shape of pop'n distrib.

Ex. Speedboat engines by Mercury (V6, Laser XRi)

Eng'rs believe that $\mu = 220$ hp
$\sigma = 15$ hp

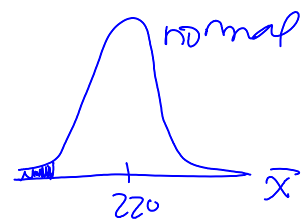


Potential buyer
$n = 100$ If $\bar{X} \leq 217$, won't buy

$Pr(\bar{X} \leq 217$ given that $\mu = 220)$

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

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



Standard

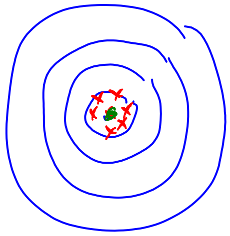
$\bar{X} \leq 217$

$\underbrace{\frac{\bar{X} - 220}{1.5}}_{Z} \leq \frac{217 - 220}{1.5}$
$\qquad\qquad \| $

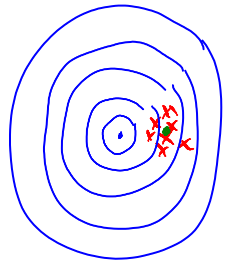
$Pr(Z \leq -2) = 0.022$
(Tables)

c) Unbiasedness & minimum variance estimation

$\left.\begin{array}{l} \mu \rightarrow \bar{x} \\ \sigma^2 \rightarrow s^2 \end{array}\right\}$ Correct?

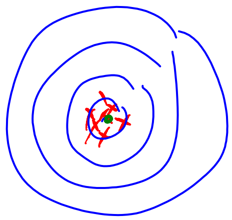


unbiased

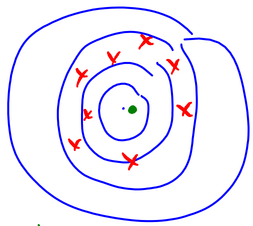


biased

2h



Small variance



Large variance

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

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

Ex: Error committed if median is used to estimate μ

$N = 4$ $\{1, 1, 3, 4\}$

$x_1 = 1$
$x_2 = 1$
$x_3 = 3$
$x_4 = 4$

True mean

$\mu = \frac{9}{4} = 2.25$

$n = 3$

Possible Samples	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

	x_1	x_2	x_3	x_4	$\bar{x}$	Median
Possible Samples	1	1	3		5/3	1
	1	1		4	2	1
	1		3	4	8/3	3
		1	3	4	8/3	3
					$\mu_{\bar{x}} = \frac{9}{4} = 2.25 = \mu$	mean of medians $= \frac{8}{4} = 2 \neq \mu$

d) Distribution of sample proportion

Ex. New Coke

Coca Cola comp. 1886

1945 : 60% market share

1983 : 24% " "

CEO Roberto Goizueta 1980

1985 : "New Coke"

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

New Coke : Corn Syrup

Old " , Sugar Cane

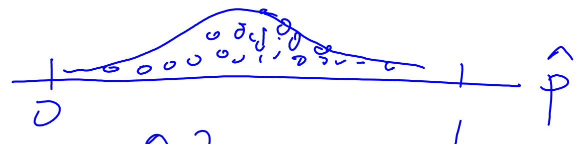
http://www.youtube.com/watch?v=W6t7deaplgY

Ask, say, $n = 100$

Say, 40 would buy New Coke

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

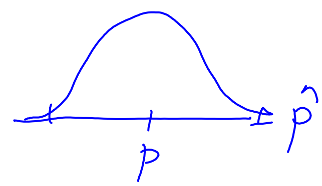
other samples



Q: What's distrib. of $\hat{p}$?

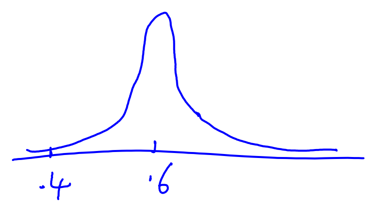
A. If n is large ($np \geq 5$, $n(1-p) \geq 5$), then

- approx. normal
- has mean $\mu_{\hat{p}} = p$
- has variance $\sigma^2_{\hat{p}} = \frac{1}{n} p(1-p)$
- " s.d. $\sigma_{\hat{p}} = \sqrt{\frac{1}{n} p(1-p)}$



$\mu_{\hat{p}} = p = 0.6$, $n = 100$

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



But, we found $\hat{p} = \frac{40}{100} = .4$: Could the CEO be wrong?

$Pr(\hat{p} \leq .4$ given that CEO thinks that $p = .6) = ?$

$$Pr(\hat{p} \leq .4) = Pr\left(\frac{\hat{p} - .6}{.049} \leq \frac{.4 - .6}{.049}\right)$$

$$= Pr(Z \leq -4.08) = .00025$$

Normal distribution		
		P(lower)
		2.25E-05

If CEO was right, it's almost impossible to see what we saw to see what we saw

($\hat{p}$ =.4)