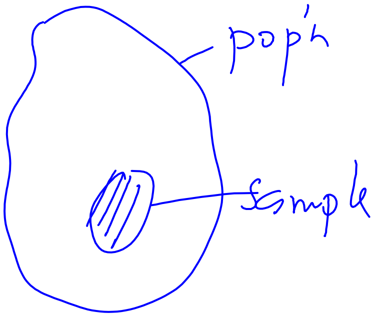
2013-10-17-C02

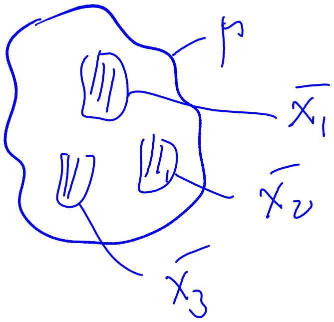
Thursday, October 17, 2013
10:27 AM

Ch.6 Sampling Distribution

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



a) Distribution of sample mean



Ex. GMAT-2007

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				

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

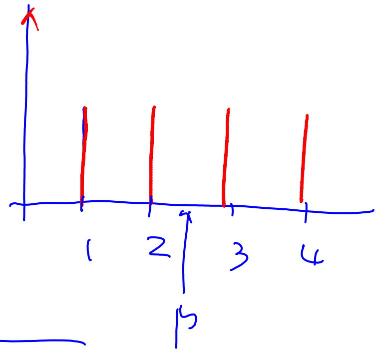
Descriptive statistics					
	GMAT-2003				
count	164				
mean	643.48				
minimum	560				
maximum	760				
range	200				
population variance	1,501.94				
population standard deviation	38.75				
Descriptive statistics					
	Sample 1	*Sample 2*	*Sample 3*		*Average of all three samples*

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. Lottery Prizes

	Balls	$ Prize (x)	p(x)
●	A	1	1/4
●	B	2	1/4
●	C	3	1/4
●	D	4	1/4



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

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

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

① All possible sampls of n=2 (avg prize)

Sample 1 \ Sample 2	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

Prizes

$\bar{x}$	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

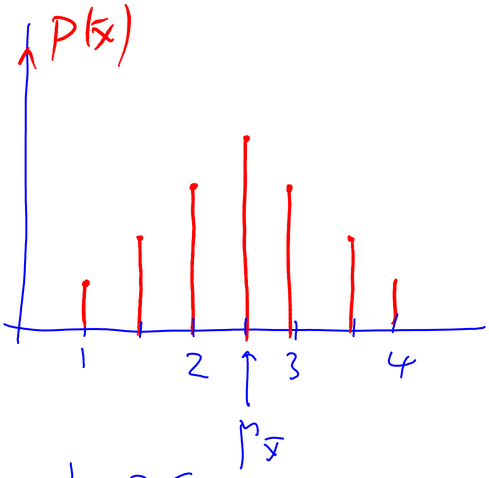
$\bar{x}$	Freq	$p(\bar{x})$
1	1	1/16
1.5	2	2/16
2	3	3/16
2.5	4	4/16
3	3	3/16
3.5	2	2/16

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



$n=2$

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

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

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

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

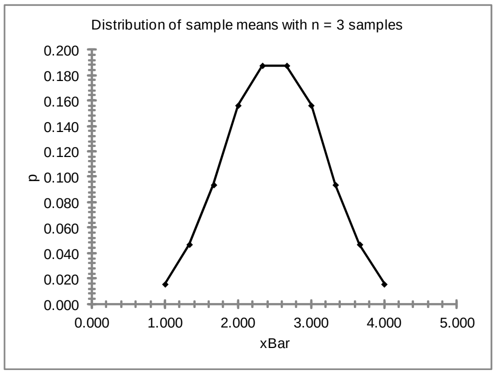
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Stocks-2013.xlsx

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>

② size n=3



n = 3 samples				xBar		xBar	Freq.	p	xBar·p	$(xBar-\mu)^2$	$(xBar-\mu)^2\cdot p$
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							$\sigma^2/3$= 1.25/3= 0.417
1	4	4		3.000							

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

Population (N)

Infinite, or
very large
(N ≥ 20n)

Finite
(N < 20n)

with replacement

without replacement

n – n

with replacement without replacement

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

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

Ex. $N = 60$, $n = 4$ N vs. 20n

$60 < 20 \cdot 4 = 80$

So, $\sigma_{\bar{x}} = \frac{1.12}{\sqrt{4}}\sqrt{\frac{60-4}{60-1}} = (0.56)(.974) \doteq .545$

Ex. $N = 60$
$n = 60$

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

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 ($n \geq 30$), then the sample mean $\bar{X}$ is always approximately normal with

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

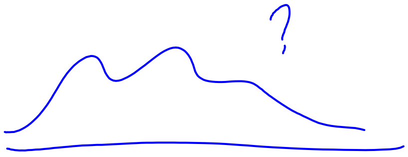
st. dev. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

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

regardless of the shape of pop'n distrib.

Ex. Speedboat engines (V6, Laser XRi) Mercury

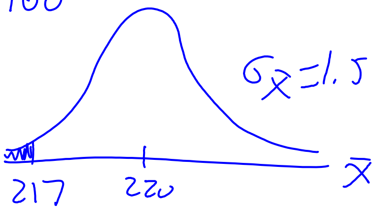
Engineers: $\mu = 220$ HP
$\sigma = 15$ HP



Customer tests $n=100 \Rightarrow$ will buy if $\bar X > 217$

$Pr(\text{no sale}) = Pr(\bar X \le 217) = ?$

$$\underbrace{\mu_{\bar X} = 220 \quad , \quad \sigma_{\bar X} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5}_{\text{normal}}$$



$$\bar X \le 217$$

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

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

Good news!

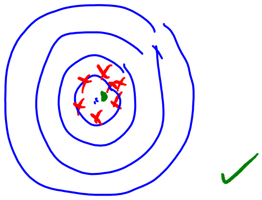
c) Unbiasedness & minimum variance estimation

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

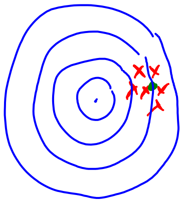
$\sigma^2 \longrightarrow s^2 = \frac{1}{n-1} \sum (x_i - \bar x)^2$

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

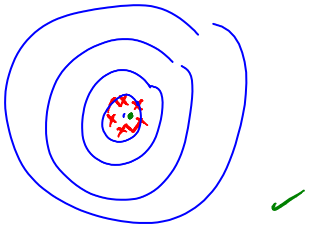
$$\sigma^2 \longrightarrow s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$



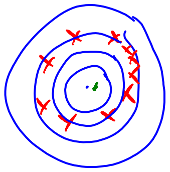
unbiased



biased



small var



large variance

Ex. N=4 & incomes

Pop $\begin{cases} x_1 = 1 \\ x_2 = 1 \\ x_3 = 3 \\ x_4 = 4 \end{cases}$ $\mu = \frac{9}{4} = 2.25$

n=3 Possible samples

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

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

d) Estimation of sample proportion

d) Estimation of sample proportion

Ex. "New Coke"

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

The Coca-Cola Company introduced the "New Coke" in April 1985 and stopped the production of the original coke. This turned out to be a poor decision as the sales plummeted. The company then decided to return to the original formula in June 1985. At that time the "Old Coke" became the "Coke Classic." Here's the sequence of events resulting in several name changes:

- New product's name changes: New Coke (April 1985) > Coca-Cola (July 1985) > Coca Cola II (1990) > Stopped production (2002).
- "Original" product's name changes: Coke became Coke Classic in July 1985 > Coke.

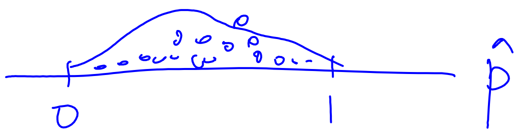
Pasted from <http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/ch-06.html>

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

CEO believes that $p = .6$

$n = 100$ → Say, 40 would buy New Coke

So, sample prop. $\hat{p} = \frac{40}{100} = .4$



Fact: If n is large ($np \geq 5$, $n(1-p) \geq 5$), then $\hat{p}$ is appr. normal with

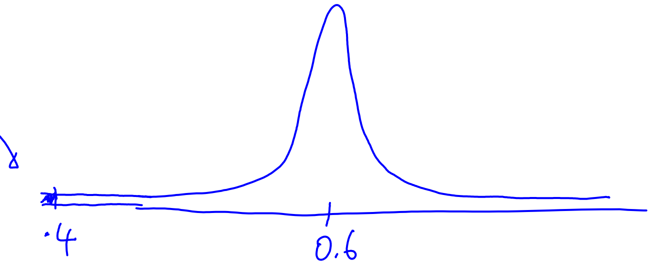
with mean $\mu_{\hat{p}} = p$ (true proport.)

var $\sigma^2_{\hat{p}} = \frac{1}{n} p(1-p)$

s.d. $\sigma_{\hat{p}} = \sqrt{\frac{1}{n} p(1-p)}$

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

$\mu_{\hat{p}} = 0.6,\qquad \sigma_{\hat{p}} = \sqrt{\frac{1}{100}(.6)(.4)} = .049$



$\Pr(\hat{p} \leq .4)$

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

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

$$\Downarrow$$

$$\Pr(Z \leq -4.08) = ?$$

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

0.0000225178503892032