7.1 Sampling Error

GOALS:

1. Consider mean and standard deviation of all the possible samples of a population.
2. Consider the distribution of all the sample means.
3. Understand that the sample mean is not expected to be exactly the same as the population mean.
4. Understand how the sample size relates to the sample standard deviation.
5. Understand how the sample size relates to the reliability of the sample mean as an approximation for the population mean.

Study Ch. 7.1, # 1-7, ~11

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:50 PM

7.1 Sampling Error

1. Take a random sample of 5 from the following height data.
2. Compute $\bar{x}$ and s.

STUDENT HEIGHTS, MULTIPLE YEARS

Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt
1	77	11	71	21	63	31	64	41	65	51	67	61	60	71	64
2	72	12	76	22	59	32	63	42	61	52	62	62	67	72	70
3	66	13	67	23	69	33	64	43	66	53	67	63	68	73	72
4	65	14	61	24	67	34	64	44	71	54	66	64	66	74	68
5	62	15	68	25	70	35	69	45	66	55	71	65	72	75	67
6	63	16	63	26	64	36	70	46	66	56	71	66	73	76	70
7	69	17	66	27	66	37	66	47	67	57	59	67	63	77	69
8	61	18	74	28	70	38	73	48	70	58	60	68	64	78	65
9	70	19	69	29	64	39	67	49	70	59	60	69	66	79	61
10	68	20	71	30	71	40	65	50	71	60	64	70	64	80	72

Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt	Pers	Hgt
81	64	91	71	101	62	111	59	121	66	131	70	141	62	151	72
82	61	92	59	102	67	112	64	122	62	132	68	142	68	152	65
83	66	93	62	103	69	113	66	123	67	133	69	143	67	153	69
84	70	94	74	104	62	114	72	124	61	134	75	144	68	154	71
85	72	95	67	105	64	115	71	125	71	135	67	145	65	155	62
86	64	96	71	106	67	116	64	126	63	136	67	146	62	156	68
87	65	97	63	107	68	117	75	127	63	137	67	147	70	157	67
88	62	98	62	108	67	118	78	128	62	138	63	148	68		
89	68	99	67	109	67	119	61	129	70	139	69	149	67		
90	67	100	73	110	66	120	64	130	70	140	69	150	63		

66.8 Mean 3.97 stdev

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

1. Take a random sample from the following height data.
2. Compute $\bar{x}$ and s.

1. We need to consider all the possible samples of size n from the population.
2. A population of N=157 is too large to look at every possible sample of n = 5.
$_{157}C_5$ = 745,395,651 possibilities

Examples

sample, n=5		sample	n=10	sample	n=15
149	67	126	63	2	72
143	67	150	63	5	62
24	67	35	69	9	70
81	64	59	60	20	71
64	66	2	72	33	64
	66.2	147	70	44	71
	1.30	17	66	48	70
		1	77	49	70
		87	65	67	63
		135	67	98	62
			67.2	114	72
			4.98	132	68
				153	69
				154	71
				155	62
					67.8
					3.968

$\bar{x}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

Investigate how samples relate to populations.
Use a contrived small population.

Print this page to use with the next several slides.

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

μ = ___

Samples of 2

Sample	Mean $\bar{x}$

Samples of 3

Sample	Mean $\bar{x}$

Samples of 4

Sample	Mean $\bar{x}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

http://espn.go.com/mlb/team/stats/batting/_/name/nyy/new-york-yankees

Consider this a population: **Find** μ and σ

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$

$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$

$\mu_{\bar{x}}$

$\sigma_{\bar{x}}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

Why $\sigma_{\bar{x}}$ instead of $s_{\bar{x}}$?

We have every possible sample of n=2, so together they form a population. Use $\sigma_{\bar{x}}$

$\mu_{\bar{x}}$
$\sigma_{\bar{x}}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

CS

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

Samples of 3

Sample	Mean $\bar{x}$

We have every possible sample of n=2, so together they form a population. ∴ Use $\sigma_{\bar{x}}$

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 10.65$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$

$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

$\mu_{\bar{x}} = 279.4$

$\sigma_{\bar{x}} = 10.65$

Samples of 3

Sample	Mean $\bar{x}$
CJR	291.7
CJG	287.0
CJS	286.3
CRG	280.0
CRS	279.3
CGS	274.7
JRG	278.3
JRS	277.7
JGS	273.0
RGS	266.0

$\mu_{\bar{x}} = 279.4$

$\sigma_{\bar{x}} = 7.10$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

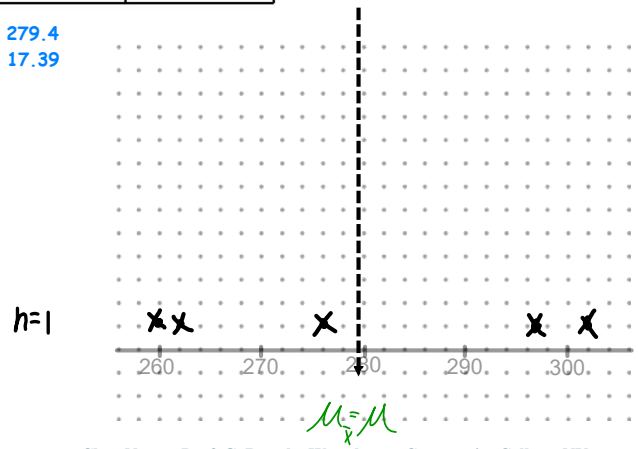
7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$

$\sigma = 17.39$



Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

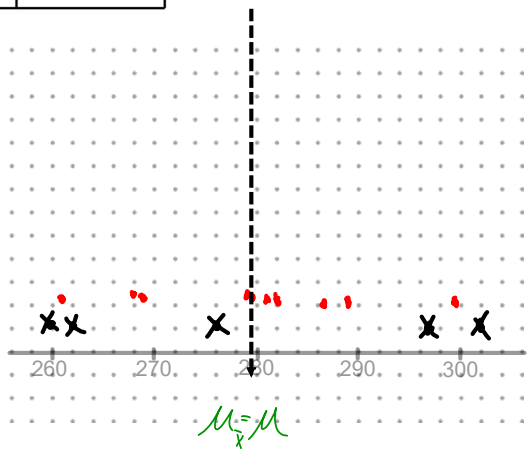
$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 10.65$

h=2
h=1



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

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

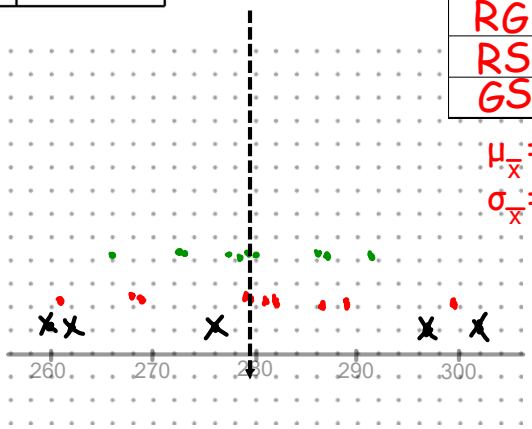
$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 10.65$

Samples of 3

Sample	Mean $\bar{x}$
CJR	291.7
CJG	287.0
CJS	286.3
CRG	280.0
CRS	279.3
CGS	274.7
JRG	278.3
JRS	277.7
JGS	273.0
RGS	266.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 7.10$

h=3
h=2
h=1



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

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 10.65$

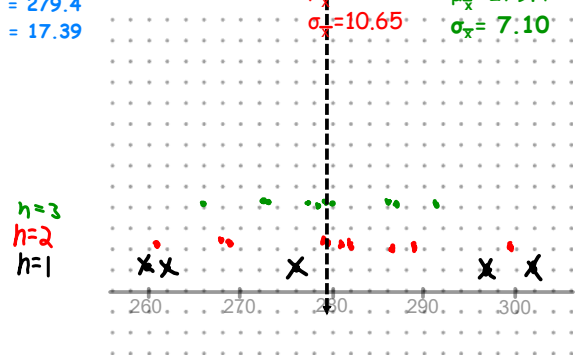
Samples of 3

Sample	Mean $\bar{x}$
CJR	291.7
CJG	287.0
CJS	286.3
CRG	280.0
CRS	279.3
CGS	274.7
JRG	278.3
JRS	277.7
JGS	273.0
RGS	266.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 7.10$

Samples of 4

Sample	Mean $\bar{x}$
CJRG	



Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

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

Oct 4-3:49 PM

7.1 Sampling Error **Distributions using DOTPLOTS**

NY Yankees, 2011 Stats

	Player	Batting Average
C	Robinson Cano	302
J	Derek Jeter	297
R	Alex Rodriguez	276
G	Curtis Granderson	262
S	Nick Swisher	260

$\mu = 279.4$
$\sigma = 17.39$

Samples of 2

Sample	Mean $\bar{x}$
CJ	299.5
CR	289.0
CG	282.0
CS	281.0
JR	286.5
JG	279.5
JS	278.5
RG	269.0
RS	268.0
GS	261.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 10.65$

Samples of 3

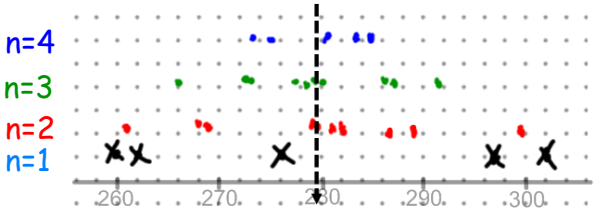
Sample	Mean $\bar{x}$
CJR	291.7
CJG	287.0
CJS	286.3
CRG	280.0
CRS	279.3
CGS	274.7
JRG	278.3
JRS	277.7
JGS	273.0
RGS	266.0

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 7.10$

Samples of 4

Sample	Mean $\bar{x}$
CJRG	284.25
CJRS	283.75
CJGS	280.25
CRGS	275.00
JRGS	273.75

$\mu_{\bar{x}} = 279.4$
$\sigma_{\bar{x}} = 4.35$



$\sigma_{\bar{x}} = 4.35$
$\sigma_{\bar{x}} = 7.10$
$\sigma_{\bar{x}} = 10.65$
$\sigma = 17.39$

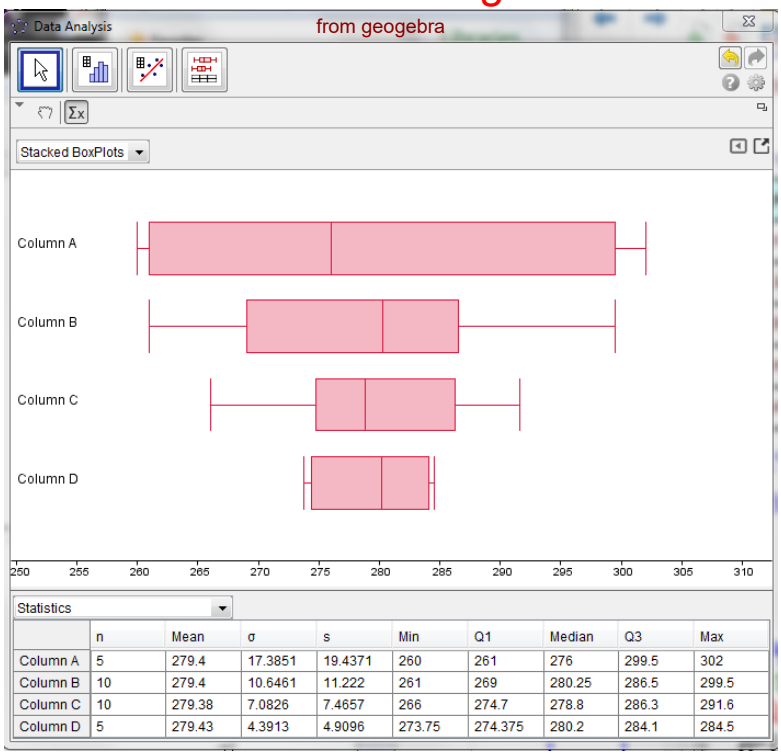
Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error

Distributions using BOXPLOTS



Statistics

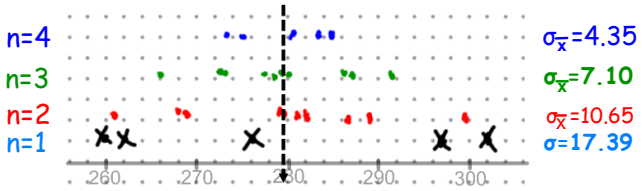
	n	Mean	σ	s	Min	Q1	Median	Q3	Max
Column A	5	279.4	17.3851	19.4371	260	261	276	299.5	302
Column B	10	279.4	10.6461	11.222	261	269	280.25	286.5	299.5
Column C	10	279.38	7.0826	7.4657	266	274.7	278.8	286.3	291.6
Column D	5	279.43	4.3913	4.9096	273.75	274.375	280.2	284.1	284.5

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.1 Sampling Error



As n increases,
the sample means cluster closer to true population mean

When we find the mean of the
sample means
from all possible samples of a given size n,
the mean of the sample means = population mean.

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

$\bar{x}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.2 Mean and Standard Dev of Sample Mean

GOALS:
1. Understand how the sample mean relates to the population mean.
2. Understand how the sample standard deviation relates to the population standard deviation.
3. Define the standard deviation of all the sample means as the STANDARD ERROR.

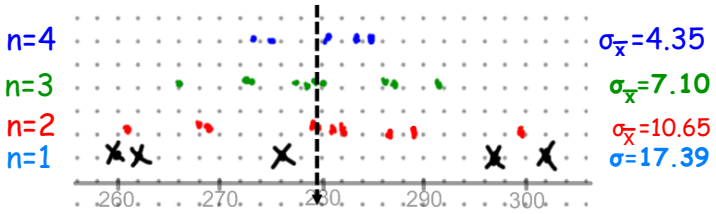
Study Ch. 7.2, # ~41, 47, 49

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.2 Mean and Standard Dev of Sample Mean



Mean of the Sample Mean

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

How does the Standard Deviation of the Sample Mean relate to the population standard dev.?

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

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.2 Mean and Standard Dev of Sample Mean

Sampling without Replacement

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

Sampling with Replacement

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

If n is small relative to N ($n \leq 0.05N$) then there is little difference between with and without replacement. So use simpler formula.

Statistics Home Page Class Notes Homework

Feb 20-11:41 AM

7.2 Mean and Standard Dev of Sample Mean

Mean of the Sample Mean

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

Standard Deviation of the Sample Mean

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

Standard Error (of the Mean)

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

Oct 4-3:49 PM

7.2 Mean and Standard Dev of Sample Mean

G: mean age of self-employed in the U.S. is 46.6 yrs, the stdev is 10.8 yrs
F: a) identify the population the variable
b) n = 100, F: mean, stdev of $\bar{x}$
c) repeat, n=175

Statistics Home Page Class Notes Homework

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Oct 4-3:49 PM

7.2 Mean and Standard Dev of Sample Mean

G: mean age of self-employed in the U.S. is 46.6 yrs, the stdev is 10.8 yrs
F: a) identify the population the variable
b) n = 100, F: mean, stdev of $\bar{x}$
c) repeat, n=175

a) Population: self-employed in US; variable: age

b) $\mu_{\bar{x}} = \mu = 46.6$ yrs. $n = 100$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10.8}{\sqrt{100}} = 1.08$ yrs

c) $n = 175$

$\mu_{\bar{x}} = \mu = 46.6$ yrs

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10.8}{\sqrt{175}} = \frac{10.8}{13.229} = 0.816$

Statistics Home Page Class Notes Homework

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Oct 4-3:49 PM