7.3 Sampling Distribution of Sample Mean

GOALS:
1. Understand the distribution of the sample mean.
2. Understand that using the distribution of the sample mean with sufficiently large sample sizes enables us to use parametric statistics for distributions that are not normal.

Study Ch. 7.3, # 63-71

d, e use calculator

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

What Do We Know so Far?

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

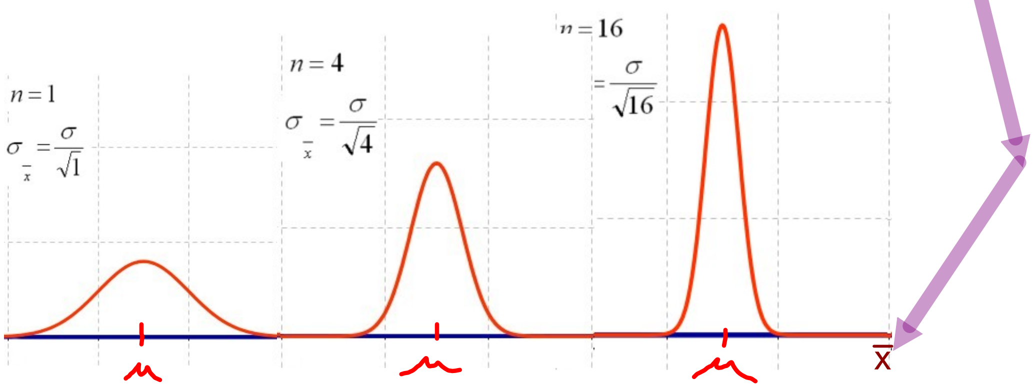
If there are a number of sample means of a particular sample size, *n*, what is the distribution of $\bar{x}$?

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean = **Distribution of $\bar{x}$**

Population **normally distributed** - easy, already done.

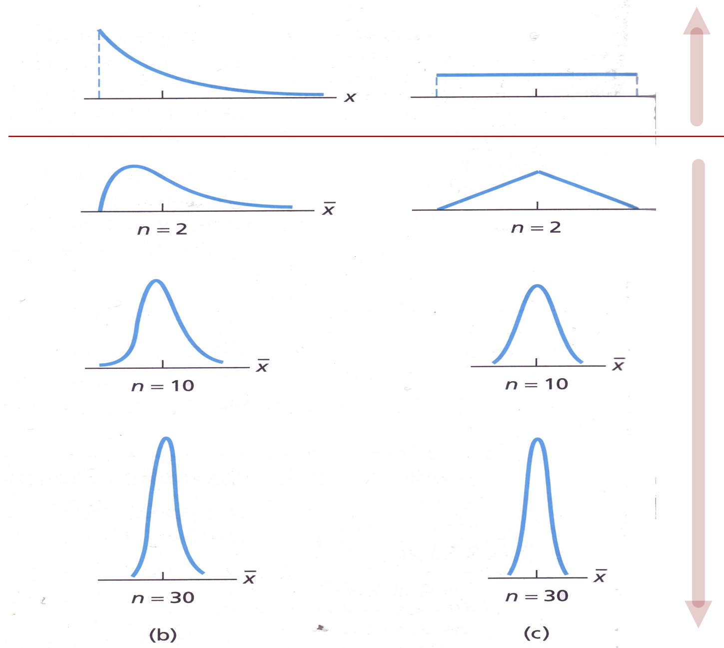


What about when the
Population is NOT normally distributed ?

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean



x **NOT** normal

$\bar{x}$ As n increases, distribution approaches normal

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

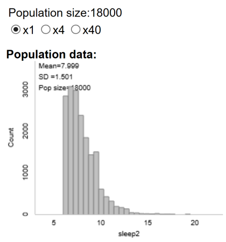
Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

http://www.rossmanchance.com/applets/OneSample.html

SETUP
1. Click the url:
2. Select POP2. This is skewed right.

POPULATION:
1. Examine the histogram.
2. The population size, N = 18000. This population consists of values from 6 to 19 (x-axis). It is right skewed. Note the mean (7.999) and standard deviation (1.501)



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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

http://www.rossmanchance.com/applets/OneSample.html

SAMPLING THE POPULATION: n = 5

7. Select the box next to Show Sampling Options at the top of the page.
8. Start by entering Number of samples: 1 and Sample size: 5
9. Be sure that you can see the histogram, and click "Draw Samples"
10. Two graphs show the sampling.
    The middle graph is a dot plot of the sample items (from the population).
    The right graph shows the mean of the sample.
11. Repeat step 9 by clicking "Draw Samples" again. Watch the graphs.
12. Continue the sampling. Watch the histogram of sample means grow.
13. Finally change the Number of Samples to 1000 and "Draw Samples"
    Notice that the histogram is less right skewed than the population, but it is still skewed, and not a normal distribution.
14. Compare the mean and standard deviation of the population to the mean and standard deviation of the sample means.

μ = ______________ μ_{xbar} = ______________

σ = ______________ σ_{xbar} = ______________

eg: 8.030,0.670
will be different each time

$\sigma_{\bar{x}} = \dfrac{1.501}{\sqrt{5}}$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

http://www.rossmanchance.com/applets/OneSample.html

REPEAT SAMPLING for n = 10 and for n = 30
1. Click the RESET button.
2. Change the sample size to n = 10 and repeat above.
3. Change the sample size to n = 30 and repeat above.

For 1000 samples for each

eg: 8.030, 0.670 will be different each time	eg: 7.992, 0.458 will be different each time	eg: 8.001, 0.284 will be different each time
$\sigma_{\bar{x}} = \frac{1.501}{\sqrt{5}}$ $= 0.671$	$\sigma_{\bar{x}} = \frac{1.501}{\sqrt{10}}$ ≈ 0.475	$\sigma_{\bar{x}} = \frac{1.501}{\sqrt{30}}$ $= 0.274$

For 10000 samples for n=30, get 0.275

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

Statistics Home Page | Class Notes | Homework

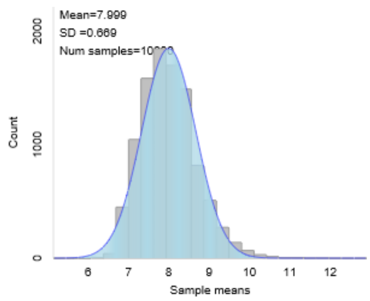
---

http://www.rossmanchance.com/applets/OneSample.html

Rossmanchance output for Right Skewed Population

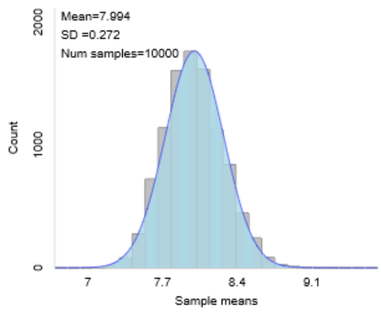
n=5, 10000 samples

Statistic: ◉ Mean ○ Median ○ *t*-statistic



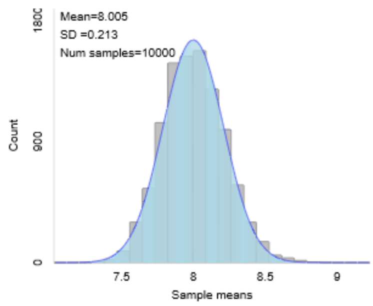
n=30, 10000 samples

Statistic: ◉ Mean ○ Median ○ *t*-statistic



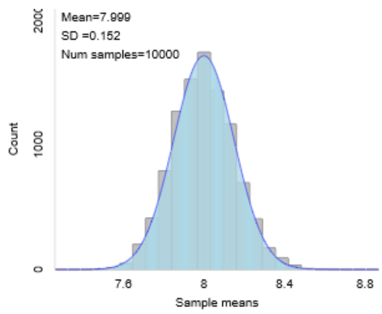
n=50, 10000 samples

Statistic: ◉ Mean ○ Median ○ *t*-statistic



n=100, 10000 samples

Statistic: ◉ Mean ○ Median ○ *t*-statistic



As n increases, the distribution of $\bar{x}$ is closer to normal distribution

7.3 Sampling Distribution of Sample Mean

n.d. = normally distributed $\bar{x}$

G: μ = 35, σ = 42, **n.d.**

a) F: sampling distribution of $\bar{x}$, n = 9

b) Can you answer part (a) if the distribution of original variable is unknown?

c) Can you answer part (a) if the distribution of original variable is unknown and n=36? Why or why not?

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

n.d. = normally distributed $\bar{x}$

G: μ = 35, σ = 42

a) G: n.d. F: sampling distribution of $\bar{x}$, n = 9

n.d. $\mu_{\bar{x}} = \mu = 35$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{9}} = 14$

b) Can you answer part (a) if the distribution of the original variable is unknown?

No, even though can compute $\mu_{\bar{x}} = 35$, $\sigma_{\bar{x}} = 14$

But distr. of $\bar{x}$ not known, n = 9 < 30

∴ Do NOT know th **shape** of the distribution of $\bar{x}$ Can NOT use for predictions.

c) Can you answer part (a) if the distribution of the original variable is unknown and n=36? Why or why not?

n=36 : Yes. distrib $\bar{x}$ is ~ n.d even if x not normal

$\mu_{\bar{x}} = \mu = 35$ | $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{36}} = 7$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

G: μ , σ , n large [$\geq$ 30]

a) Identifiy the distribution of $\bar{x}$

b) Does your answer to part (a) depend on n being large?

c) If n < 30, can you still identify the $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

G: μ , σ , n large [$\geq$ 30]

a) Identifiy the distribution of $\bar{x}$

$\sim$ n.d. $\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

b) Does your answer to part (a) depend on n being large?

Yes. We are not given the distribution of the population variable, x. Only when $n \geq 30$ can we say that $\bar{x}$ is ~ n.d.

c) If n < 30, you can still compute the $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$ but it is meaningless because can NOT identify the distribution (shape).

Without a known distribution, cannot use area under curve to interpret as either % or probability

$\bar{x}$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

G: NYC marathon, n.d., μ = 61 min, σ = 9 min

a) F: sampling distribution of $\bar{x}$, n = 4

b) F: sampling distribution of $\bar{x}$, n = 9

c) Sketch the distributions in (a) and (b)

d) F: % of all sample of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

e) F: % of all sample of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

$\bar{x}$
$\bar{x}$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

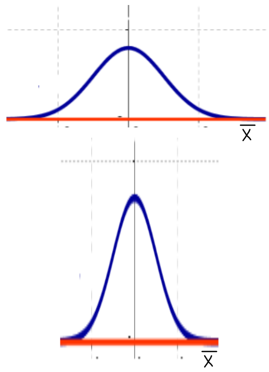
G: NYC marathon, n.d., μ = 61 min, σ = 9 min

a) F: sampling distribution of $\bar{x}$, n = 4

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{4}} = 4.5$ min

b) F: sampling distribution of $\bar{x}$, n = 9

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{9}} = 3.0$ min



c) Sketch the distributions in (a) and (b)

d) F: % of all samples of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

e) F: % of all samples of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

$\bar{x}$
$\bar{x}$

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

Statistics Home Page Class Notes Homework

7.3 Sampling Distribution of Sample Mean

G: NYC marathon, n.d., μ = 61 min, σ = 9 min

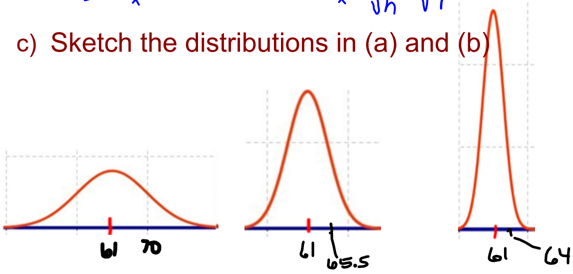
a) F: sampling distribution of $\bar{x}$, n = 4

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{4}} = 4.5$ min

b) F: sampling distribution of $\bar{x}$, n = 9

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{9}} = 3.0$ min

c) Sketch the distributions in (a) and (b)



Since n.d., can sketch distribution.

d) F: % of all sample of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

e) F: % of all sample of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

$\bar{x}$
$\bar{x}$

7.3 Sampling Distribution of Sample Mean

G: NYC marathon, n.d., μ = 61 min, σ = 9 min

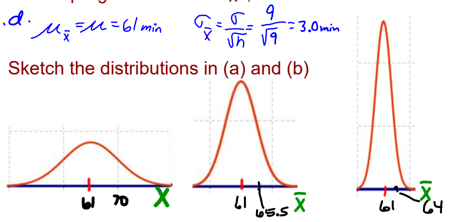
a) F: sampling distribution of $\bar{x}$, n = 4

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{4}} = 4.5$ min

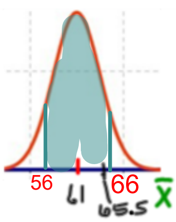
b) F: sampling distribution of $\bar{x}$, n = 9

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{9}} = 3.0$ min

c) Sketch the distributions in (a) and (b)

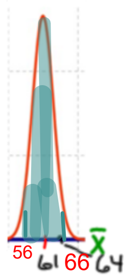


Since n.d., can find areas under the normal curve, and interpret as %.



d) F: % of all sample of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

normalcdf(____,____,____,____) = ________
or ________%



e) F: % of all sample of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

normalcdf(____,____,____,____) = ________
or ________%

$\bar{x}$ $\bar{x}$

7.3 Sampling Distribution of Sample Mean

G: NYC marathon, n.d., μ = 61 min, σ = 9 min

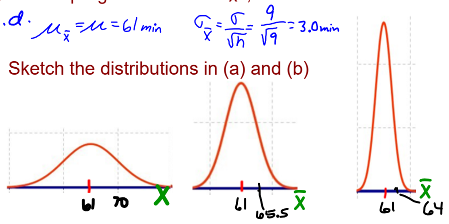
a) F: sampling distribution of $\bar{x}$, n = 4

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{4}} = 4.5$ min

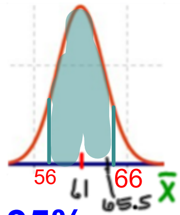
b) F: sampling distribution of $\bar{x}$, n = 9

n.d. $\mu_{\bar{x}} = \mu = 61$ min $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{9}} = 3.0$ min

c) Sketch the distributions in (a) and (b)



Since n.d., can find areas under the normal curve, and interpret as %.



d) F: % of all sample of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

normalcdf(56,66,61,4.5) = 0.7335 or 73.35% **(table:73.30%)**

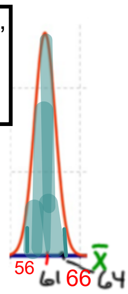
$z = \frac{x - \mu}{\sigma} = \frac{-5}{4.5} = -1.11$ $A_{-1.11} = 0.1335$ in tail

$\times 2 = 0.2670$ in 2 tails

$1 - 0.2670 = 0.7330$ in mid

Not required when using calculator.

When using the table, need to find z-score, then area in tail, then subtract from 1.



e) F: % of all sample of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

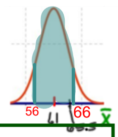
normalcdf(56,66,61,3) = 0.9044 or 90.44%
(table: 90.50%)

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

Statistics Home Page Class Notes Homework $\bar{x}$ $\bar{x}$

7.3 Sampling Distribution of Sample Mean

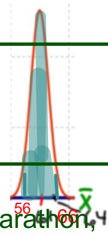
Since n.d., can find areas under the normal curve, and interpret as probability or %.



d) F: % of all sample of 4 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

normalcdf(56,66,61,4.5) = 0.7335 or 73.35% **(table:73.30%)**

INTERPRETATION:
When you take a random sample of **4** finishers in the NY Marathon, the **probability** that the mean of the sample is **within 5 minutes** of the population mean **is 0.7335**.
OR:
When you take a random sample of **4** finishers in the NY Marathon, **73.35%** of the sample means will be **within 5 minutes** of the population mean.



e) F: % of all sample of 9 finishers that finished within 5 min of μ = 61 min. Interpret, re: sampling error.

normalcdf(56,66,61,3) = 0.9044 or 90.44% **(table: 90.50%)**

INTERPRETATION:
When you take a random sample of **9** finishers in the NY Marathon, the **probability** that the mean of the sample is **within 5 minutes** of the population mean **is 0.9044**.
OR:
When you take a random sample of **9** finishers in the NY Marathon, **90.44%** of the sample means will be **within 5 minutes** of the population mean.

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

Statistics Home Page Class Notes Homework $\bar{x}$ $\bar{x}$