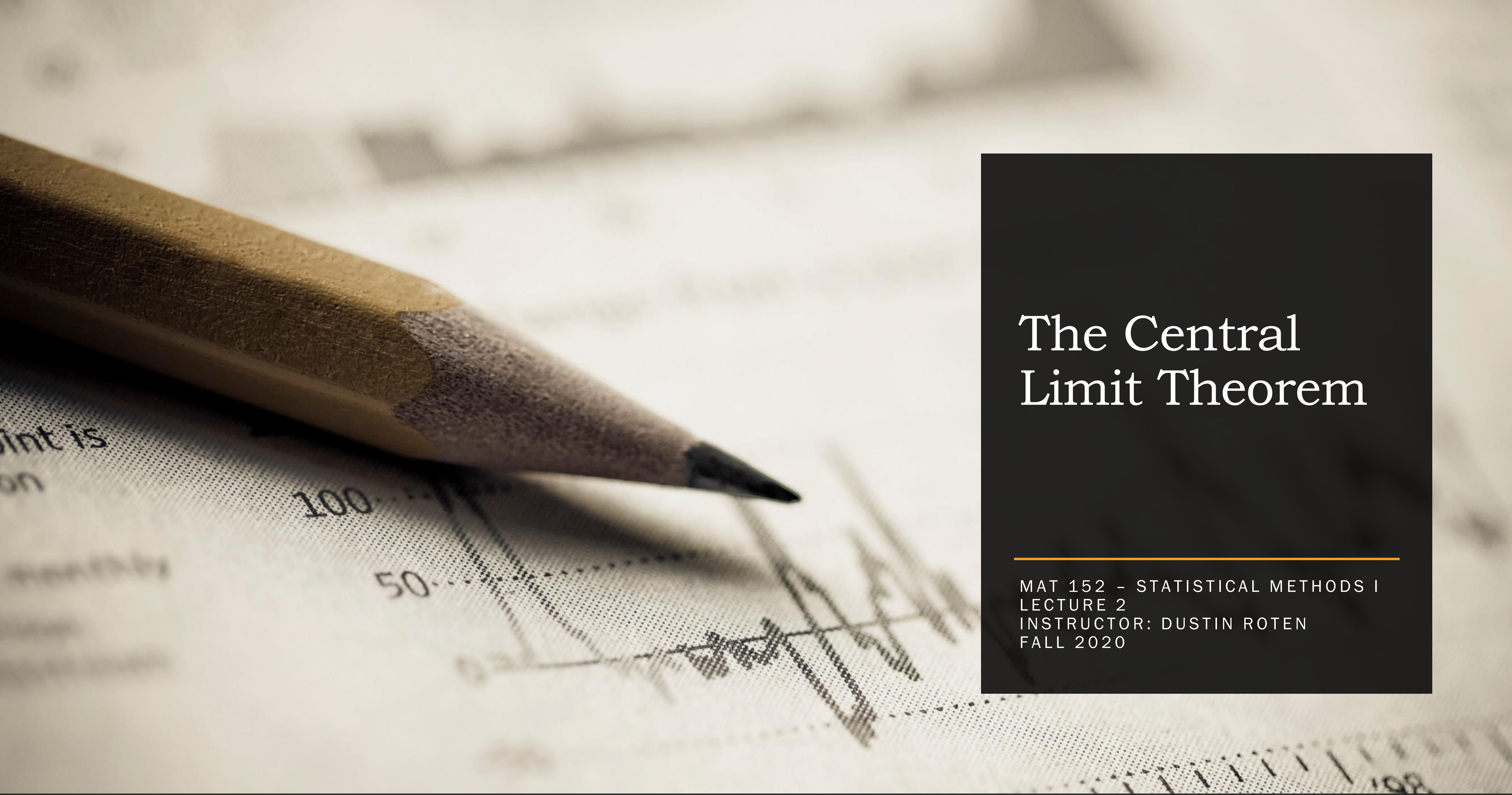


The Central Limit Theorem

MAT 152 – STATISTICAL METHODS I
LECTURE 2
INSTRUCTOR: DUSTIN ROTEN
FALL 2020

Review

Central Limit Theorem: The collection of sample means from ANY distribution will form their own normal distribution. $\bar{X}$ is the sampling distribution of the mean.

$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$.

The for large sample sizes, the "mean of the means" is the same as the population mean $(\mu_X = \mu_{\bar{X}})$

$\frac{\sigma_X}{\sqrt{n}}$ is the standard error. It states how far away, on average, the sample mean will be from the population mean.

$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)}$ (z-score for a mean)

Math Review

Distribution: $a(b+c) = ab + ac$

$\frac{x^a}{x^b} = x^{a-b}$

Equations: $1 + 2x = 4x - 3$

Must be balanced: $1 + 2x - 5 = 4x - 3 - 5$

Whatever is added to one side must be added to the other.

Multiplying only one side by one doesn't change the equation: $3x = 4y$

$(1) \cdot 3x = 4y$ is still the same as the original.

$$z = \frac{\text{distance between a value and the mean}}{\text{average distance from mean}} = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x-\mu}{\sigma}$$

Normal Distribution: $X \sim N(\mu, \sigma)$

Building New Mathematics from the Central Limit Theorem's Z-Scores

$$z = \frac{\bar{x}-\mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \frac{\frac{1}{n}\sum x-\mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = (1)\cdot\frac{\frac{1}{n}\sum x-\mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \left(\frac{n}{n}\right)\cdot\frac{\frac{1}{n}\sum x-\mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \frac{n\left(\frac{1}{n}\sum x-\mu_X\right)}{n\cdot\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \frac{\sum x-(n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$$

What does this new z-score mean?

$$z = \frac{\sum x-(n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$$

New mean: $(n)(\mu_X)$

New standard deviation: $(\sqrt{n})(\sigma_X)$

The Central Limit Theorem (for Sums)

If you keep drawing larger and larger samples and taking their sums, the sums form their own normal distribution which approaches a normal distribution as the sample size increases.

$$\sum X \sim N\left((n)(\mu_X), (\sqrt{n})(\sigma_X)\right)$$

Revisiting the previous dice example:

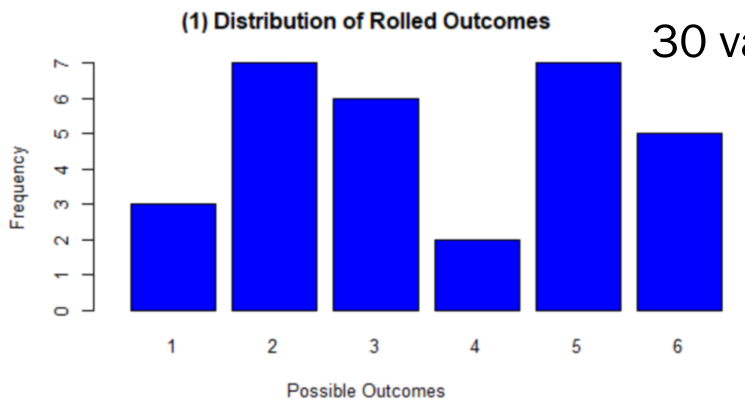
Suppose we collect samples of dice rolls?

This time, we will focus on sums, not means.

Example (Cont.)

What if we collect 1 sample of 30 rolls? (n=30)

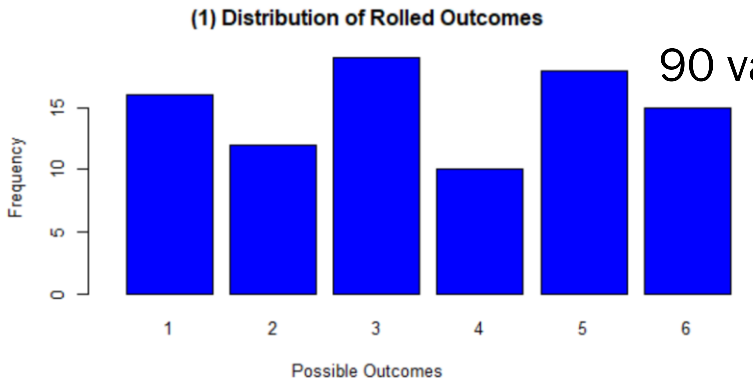
Sample	Values	
S1	3 4 3 6 2 3 3 1 6 2 2 1 5 5 6 2 5 1 6 4 2 2 3 5 2 3 5 5 5 6	108
	(30 total values)	



30 values

What if we collect 3 samples of 30 rolls? (n=30)

Sample	Values	Sample Mean
S1	1 2 2 6 6 6 1 6 5 4 3 5 1 3 2 5 1 5 3 1 2 6 3 3 5 4 5 3 4 2	108
S2	1 2 1 1 4 3 4 6 1 6 5 3 6 3 5 5 5 5 3 3 5 1 4 6 3 1 2 1 5 2	102
S3	1 6 2 2 2 3 3 3 1 2 4 1 6 3 6 4 5 6 6 5 3 4 5 5 5 4 6 3 1 3	110
	(90 total values)	Mean of the sums = 106.67 (close to 105)

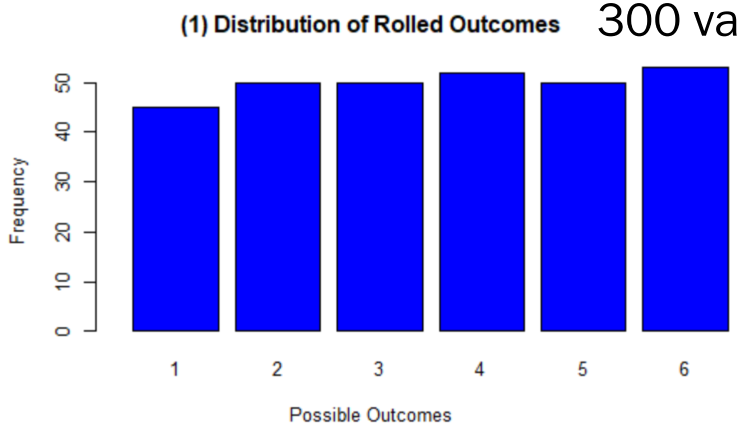


90 values

Example (Cont.)

What if we collect 10 samples of 30 rolls? (n=30)

Sample	Values	Sample Sums ($\sum x$)
S1	2 4 1 5 4 3 5 5 1 2 5 1 2 2 2 5 5 4 5 2 3 1 3 4 2 6 2 6 2 6	100
S2	4 4 4 2 1 3 6 1 2 6 3 3 2 5 6 6 6 6 3 6 4 3 6 6 6 6 1 4 3 5	123
S3	5 2 4 1 1 3 6 3 4 4 3 3 6 4 1 3 2 1 3 6 6 6 5 5 5 2 3 4 6 3	110
S4	4 1 3 2 2 4 6 2 3 4 1 1 2 3 4 4 3 6 5 3 4 3 6 3 6 2 3 5 2 6	103
S5	5 2 4 4 4 2 5 2 4 5 1 5 2 2 4 3 6 3 5 6 4 3 3 3 4 5 3 4 6 5	114
S6	1 6 2 6 5 1 3 4 6 5 6 5 5 1 3 4 4 6 1 1 6 1 4 3 6 2 6 3 2 2	110
S7	5 1 5 4 1 3 5 2 4 3 1 1 6 1 4 3 4 6 1 5 1 6 1 2 6 3 2 6 5 4	101
S8	6 3 5 5 1 5 4 5 3 1 3 4 2 3 1 6 6 1 4 6 2 5 3 6 4 4 3 3 2 4	110
S9	1 2 1 2 2 4 5 5 3 4 2 6 5 4 5 2 1 3 2 2 5 5 1 6 1 5 2 5 2 1	94
S10	6 4 1 6 1 5 1 5 4 5 2 6 2 4 5 6 5 1 1 4 2 6 5 1 4 3 3 2 4 2	106



300 values

Mean of the sums = 107.1 (close to 105)

Example (Cont.)

What if we collect 1000 samples of 30 rolls? (n=30)

In total how many rolls were there?

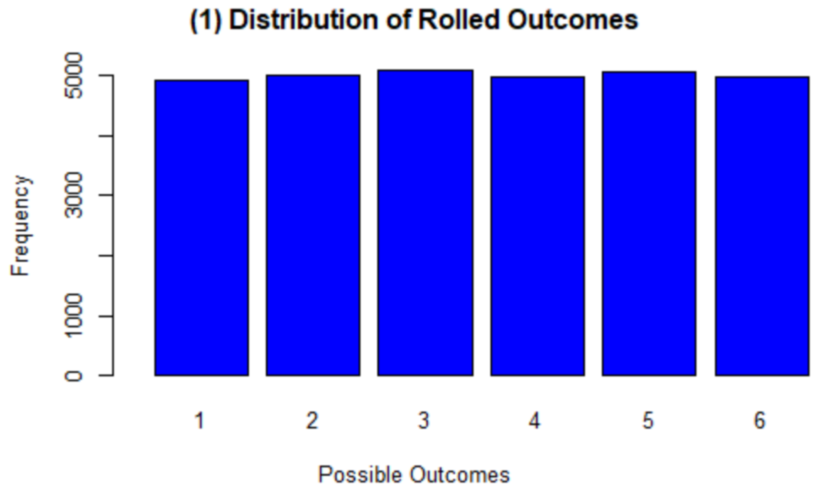
$1000 \cdot 30 = 30000$

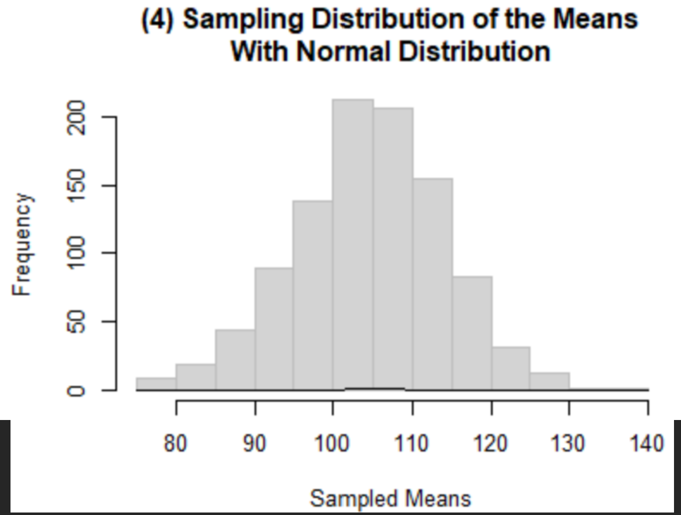
What is the mean of the sums?

Mean of $\sum X = (n)(\mu_X) = 30 \cdot 3.5 = 105$

What is the standard error of the sums?

Standard Deviation of $\sum X = (\sqrt{n})(\sigma_X) = \sqrt{30} \cdot 1.708 = 9.36$





Example (Cont.)

Given the previous scenario, what is the z-score of $\sum x = 107$?
(This is one sum value in our entire collection of 1000 sums)

$$z = \frac{\sum x - (n)(\mu_X)}{(\sqrt{n})(\sigma_X)} = \frac{107 - (30) \cdot (3.5)}{\sqrt{30} \cdot 1.708} = \frac{107 - 105}{9.36} = 0.214$$

Example

The distribution of final exam grades in a statistics course has a mean of 75 and a standard deviation of 10. Samples of 25 grades are taken from the gradebook.

X = a final grade in the class

$\bar{X}$ = the mean of a sample of size 25

Find the standard error of the mean.

$$\frac{\sigma_X}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Example (Cont.)

The distribution of final exam grades in a statistics course has a mean of 75 and a standard deviation of 10. Samples of 25 grades are taken from the gradebook.

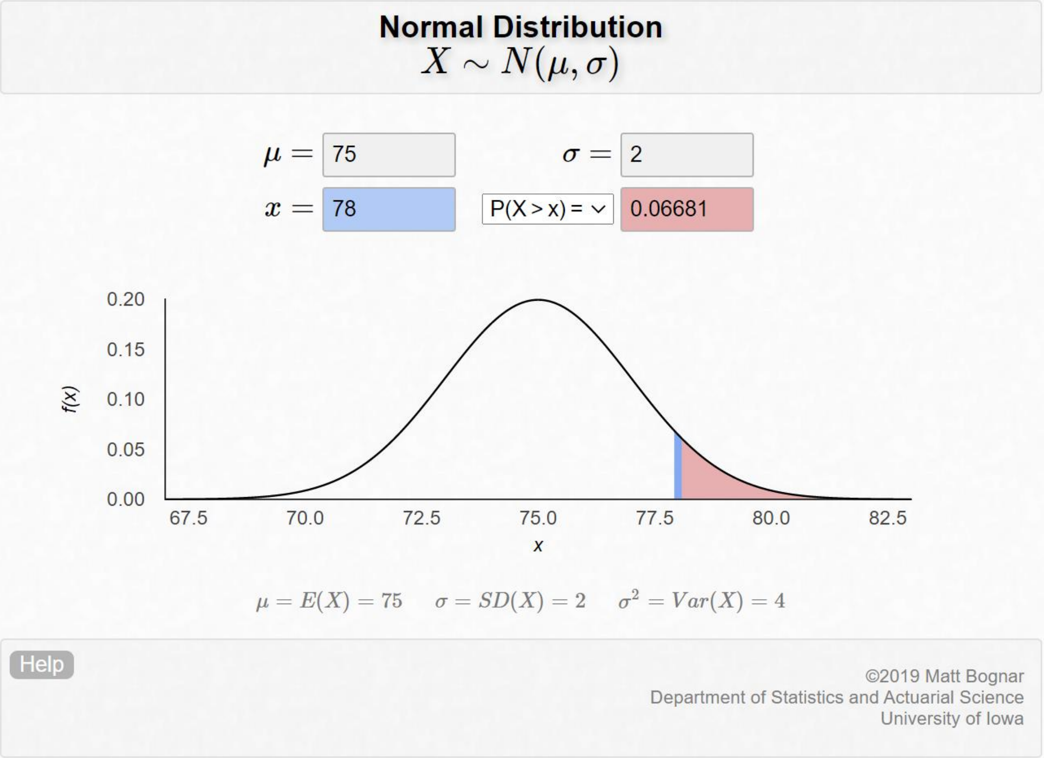
X = a final grade in the class

$\bar{X}$ = the mean of a sample of size 25

Find the probability that the sample mean is greater than 78.
$P(\bar{x} > 78) = ?$

$X \sim N(\mu_X, \sigma_X) = N(75,10)$

$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right) = N\left(75, \frac{10}{\sqrt{25}}\right) = N(75,2)$



Example (Cont.)

The distribution of final exam grades in a statistics course has a mean of 75 and a standard deviation of 10. Samples of 25 grades are taken from the gradebook.

X = a final grade in the class

$\bar{X}$ = the mean of a sample of size 25

Find the value two standard deviations (standard error) above the mean.

$$z_{+2} = \mu_X + z\left(\frac{\sigma_X}{\sqrt{n}}\right) = 75 + (2)\left(\frac{10}{\sqrt{25}}\right) = 75 + 2(2) = 79$$

Example (Cont.)

The distribution of final exam grades in a statistics course has a mean of 75 and a standard deviation of 10. Samples of 25 grades are taken from the gradebook.

X = a final grade in the class

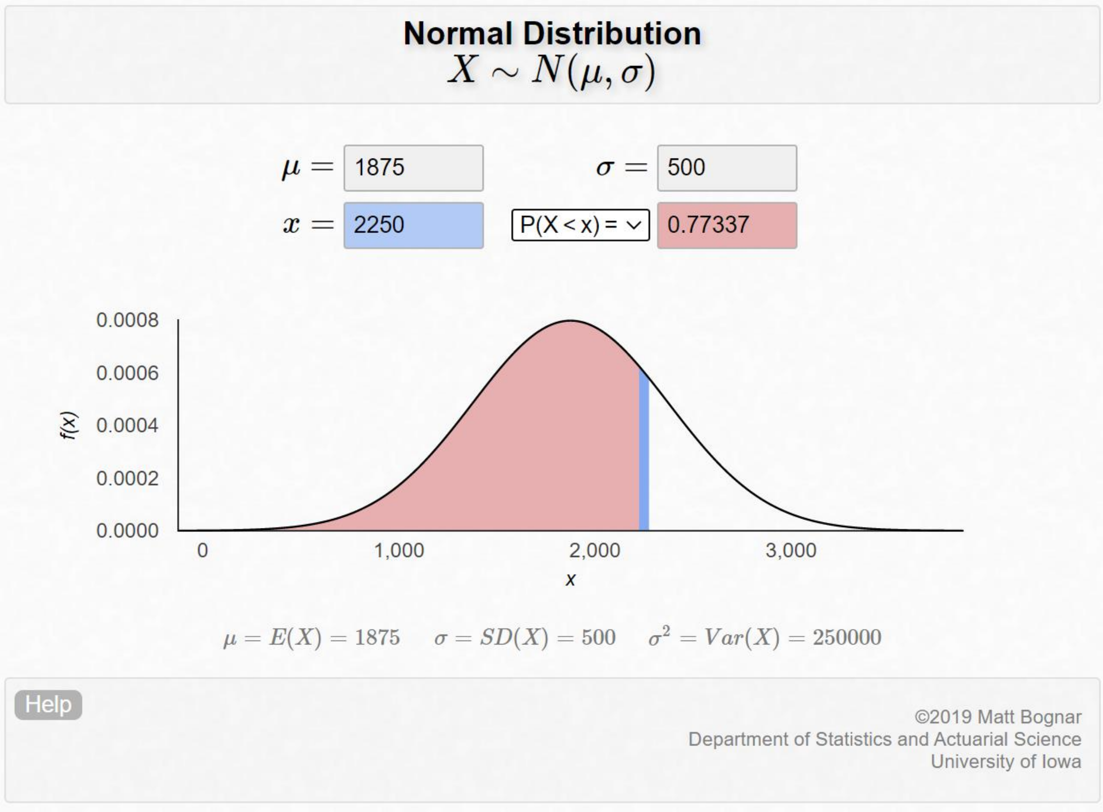
$\sum X$ = the sum of a sample of size 25

Find the probability that the sum of the sample mean is less than 2250.

$$\text{mean of } \sum X = (n)(\mu_X) = (25)(75) = 1875$$

$$\text{standard deviation of } \sum X = (\sqrt{n})(\sigma_X) = (\sqrt{25})(10) = 500$$

$$\sum X \sim N\left((n)(\mu_X), (\sqrt{n})(\sigma_X)\right) = N\left((25)(75), (\sqrt{25})(10)\right) = N(1875, 500)$$



Review

Central Limit Theorem for Sums: The collection of sample sums from ANY distribution will form their own normal distribution. $\bar{X}$ is the sampling distribution of the mean.

$$\sum X \sim N\left((n)(\mu_X), \left(\sqrt{n}\right)(\sigma_X)\right)$$

Mean = $(n)(\mu_X)$

Standard Deviation = $\left(\sqrt{n}\right)(\sigma_X)$

$z = \frac{\sum x - (n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$ (z-score for the random variable $\sum X$)