Mathematics 13 : Elementary Statistics

Unit 2 : Section 6

Lesson 29 :

The Central Limit Theorem

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X, and (3) any standard deviation σ_X, the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X, and (3) any standard deviation σ_X, the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

The sample size is n.

The random selection of one data value does not affect the random selection of another.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X, and (3) any standard deviation σ_X, the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

The results of the Central Limit Theorem apply to any population.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X, and (3) any standard deviation σ_X, the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

The normal probability distribution is used to approximate probabilities involving the sample mean $\bar{x}$.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X , and (3) any standard deviation σ_X , the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

As the sample size increases, the approximation becomes more accurate.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X , and (3) any standard deviation σ_X , the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

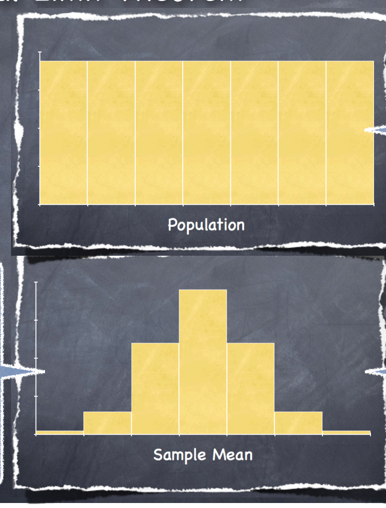
The mean of the sample mean is the same as the mean for the population from which the sample was collected.

The Central Limit Theorem

When a sample of n independent data values are randomly selected from a population with (1) any shaped distribution, (2) any mean μ_X , and (3) any standard deviation σ_X , the sampling distribution of the sample mean $\bar{x}$ has (1) an approximate normal distribution, (2) a mean $\mu_{\bar{x}} = \mu_X$, and (3) a standard deviation $\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$.

The standard deviation of the sample mean is the population standard deviation divided by the square root of the sample size.

The Central Limit Theorem

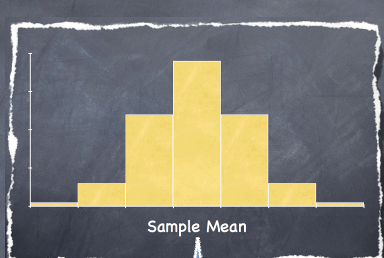


Regardless of the distribution of the population

the distribution of the sample mean has an approximate normal distribution

when the sample size is sufficiently large

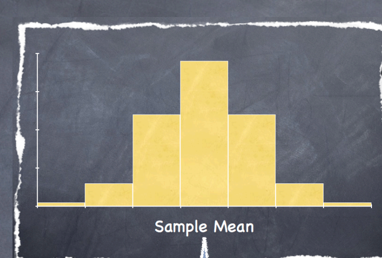
The Central Limit Theorem



The mean of the sampling distribution of the sample mean is equal to the mean of the population from which the sample was collected.

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

The Central Limit Theorem



The standard deviation of the sampling distribution of the sample mean is equal to the standard deviation of the population from which the sample was collected divided by the square root of the sample size.

$$\sigma_{\bar{x}} = \sigma_X / \sqrt{n}$$

The Central Limit Theorem

$$\mu_{\bar{x}} = E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{E(x_1 + x_2 + \cdots + x_n)}{n}$$

$$= \frac{E(x_1) + E(x_2) + \cdots + E(x_n)}{n}$$

$$= \frac{\mu_X + \mu_X + \cdots + \mu_X}{n} = \frac{n \cdot \mu_X}{n} = \mu_X$$

The Central Limit Theorem

The variance of a random variable X is the square of the standard deviation. It is denoted by

$$V(X) = \sigma_X^2$$

Consequently, the standard deviation of a random variable X is the square root of the variance.

$$\sigma_X = \sqrt{V(X)}$$

The Central Limit Theorem

$$\sigma_{\bar{x}} = \sqrt{V(\bar{x})} = \sqrt{V\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)}$$

$$= \sqrt{\frac{V(x_1 + x_2 + \cdots + x_n)}{n^2}}$$

$$= \sqrt{\frac{V(x_1) + V(x_2) + \cdots + V(x_n)}{n^2}}$$

$$= \sqrt{\frac{\sigma_X^2 + \sigma_X^2 + \cdots + \sigma_X^2}{n}} = \sqrt{\frac{n \cdot \sigma_X^2}{n^2}} = \sqrt{\frac{\sigma_X^2}{n}} = \frac{\sigma_X}{\sqrt{n}}$$

The Central Limit Theorem

So, under the conditions of the Central Limit Theorem, the normal probability distribution can be used to approximate probabilities involving the sample mean.

For the TI-84 calculator,

$$p_{L<\bar{x}<U} = normalcdf\left(L, U, \mu_X, \sigma_X/\sqrt{n}\right)$$

$$P_k = invNorm\left(k/100, \mu_{\bar{x}}, \sigma_X/\sqrt{n}\right)$$

Example 1

In the general population, IQ scores have a mean of 100 and a standard deviation of 15. In a classroom containing 32 students, what is the chance that the mean IQ score for these students is over 101?

> The results of the Central Limit Theorem can be applied to probability problems involving the sample mean.

Example 1

$$p_{\bar{x}>101} = normalcdf\left(L,U,\mu_X,\sigma_X/\sqrt{n}\right)$$
$$= normalcdf\left(101,9E99,100,15/\sqrt{32}\right)$$
$$\approx 0.353$$

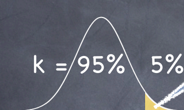
Thus, in a classroom containing 32 students, there is a 35.3% chance that the mean IQ score for these students is over 101.

Example 2

In the general population, IQ scores have a mean of 100 and a standard deviation of 15. In a classroom containing 32 students, what values of the mean IQ score for these students would be considered an unusually high mean IQ score?

> The results of the Central Limit Theorem can be applied to probability problems involving the sample mean.

Example 2



> unusually high mean IQ scores

$$P_{95} = invNorm\left(0.95,100,15/\sqrt{32}\right)$$
$$\approx 104.4$$

Thus, in a classroom containing 32 students, a mean IQ score of 104.4 or greater would be considered an unusually high mean IQ score for these students.

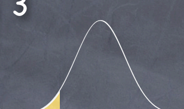
Example 3

The duration of human pregnancy follows a normal distribution with a mean of 274.5 days and a standard deviation of 13.6 days.

An obstetrician's practice is currently treating 25 pregnant women. Would it be unusual if the mean duration of pregnancy for these women were to be no more than 270 days?

> The results of the Central Limit Theorem can be applied to probability problems involving the sample mean.

Example 3



$$\mu_{\bar{x}} = \mu_X = 274.5$$
$$\sigma_{\bar{x}} = \sigma_X/\sqrt{n} = 13.6/\sqrt{25}$$

$$p_{\bar{x}\le 270} = normalcdf\left(0,270,274.5,13.6/\sqrt{25}\right)$$
$$\approx 0.049 \le 0.05$$

Thus, it would be unusual if the mean duration of pregnancy for these 25 women were to be no more than 270 days since the probability that this would occur by chance 0.049 is less than or equal to 0.05.

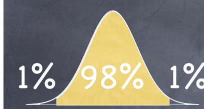
Example 4

The duration of human pregnancy follows a normal distribution with a mean of 274.5 days and a standard deviation of 13.6 days.

An obstetrician's practice is currently treating 25 pregnant women. Find the narrowest interval that has a 98% chance of containing the mean duration of pregnancy for these women.

> The results of the Central Limit Theorem can be applied to probability problems involving the sample mean.

Example 4



$\mu_{\bar{x}} = 274.5 \qquad \sigma_{\bar{x}} = 13.6/\sqrt{25}$

$\text{U} = invNorm\left(0.99, 274.5, 13.6/\sqrt{25}\right)$

≈ 280.8

$\text{L} = invNorm\left(0.01, 274.5, 13.6/\sqrt{25}\right)$

≈ 268.2

> The narrowest interval is the range of values from L to U symmetric about the mean.

Example 4

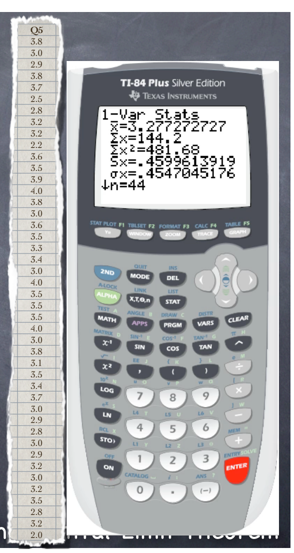
Thus, the narrowest interval that has a 98% chance of containing the mean duration of pregnancy for these 25 women ranges from 268.2 to 280.8 days.

Example 5

If six students were randomly selected from the Sierra College Elementary Statistics Student Survey, what is the probability that the mean GPA for this sample will be between 3.1 and 3.4?

Example 5

Q5. What is your college GPA? _______



$\mu_X \approx 3.277 \qquad \sigma_X \approx 0.455$

Example 5

$\mu_X \approx 3.277 \qquad \sigma_X \approx 0.455 \qquad n = 6$

$p_{3.1<\bar{x}<3.4} = normalcdf\left(3.1, 3.4, 3.277, 0.455/\sqrt{6}\right)$

≈ 0.576

If six students were randomly selected from the Sierra College Elementary Statistics Student Survey, the probability that the mean GPA for this sample will be between 3.1 and 3.4 is 57.6%.

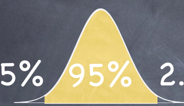
Example 6

If six students were randomly selected from the Sierra College Elementary Statistics Student Survey, what interval would contain the mean GPA for this sample 95% of the time?

Example 6

$$\mu_{\bar{x}} \approx 3.277$$

$$\sigma_{\bar{x}} \approx 0.455/\sqrt{6}$$



$$\text{U} = invNorm\left(0.975, 3.277, 0.455/\sqrt{6}\right) \approx 3.64$$

$$\text{L} = invNorm\left(0.025, 3.277, 0.455/\sqrt{6}\right) \approx 2.91$$

Example 6

If six students were randomly selected from the Sierra College Elementary Statistics Student Survey, the interval from 2.91 to 3.64 would contain the mean GPA for this sample 95% of the time.

Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

According to a study conducted at Purdue University, college students sleep an average of 7.65 hours per night with a standard deviation of 0.98 hours per night. These nightly sleep times of college students displayed an approximate bell-shaped distribution.

(a) For a randomly selected college class that contains 32 students, find the probability that the mean amount of sleep this sample of students had last night ranges from 7.5 to 8.5 hours.

(b) For a randomly selected college class that contains 32 students, find the values of the mean amount of sleep per night that would be considered an unusually high mean amount of sleep.

Exercise 2

The lengths of California Kingsnakes (*Lampropeltis getula californiae*) are normally distributed with a mean of 42.8 inches and a standard deviation of 9.7 inches.

Find the narrowest interval that has a 95% chance of containing the mean length for a random sample of 16 California Kingsnakes.

Exercise 3

The weights of adult male Rhesus monkeys (*Macaca mulatta*) are normally distributed with a mean of 7.71 kilograms and a standard deviation of 1.68 kilograms.

Would it be unusual if the mean weight for a random sample of 8 adult male Rhesus monkeys were no more than 7 kilograms?

Exercise 4

For Golden Retrievers, the number of puppies (X) born in one litter is approximated by the following probability distribution.

X	p_X
3	4.8%
4	6.5%
5	9.1%
6	12.0%
7	14.3%
8	15.2%
9	13.7%
10	10.3%
11	7.7%
12	6.4%

For a random sample of 12 Golden Retriever litters, find the narrowest interval that would accurately predict the mean number of puppies per litter 98% of the time.

Exercise 5

The number of M&M's (X) contained in one 8.2 ounce bag of Milk Chocolate Plain M&M's is approximated by the following probability distribution.

X	p_X
269	0.0014
270	0.0143
271	0.0637
272	0.1622
273	0.2580
274	0.2627
275	0.1672
276	0.0608
277	0.0097

For a random sample of 20 8.2 ounce bags of Milk Chocolate Plain M&M's, find the narrowest interval that would accurately predict the mean number of M&M's per bag 99% of the time.

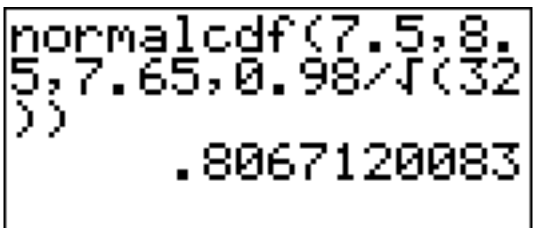
Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

According to a study conducted at Purdue University, college students sleep an average of 7.65 hours per night with a standard deviation of 0.98 hours per night. These nightly sleep times of college students displayed an approximate bell-shaped distribution.

(a) For a randomly selected college class that contains 32 students, find the probability that the mean amount of sleep this sample of students had last night ranges from 7.5 to 8.5 hours.

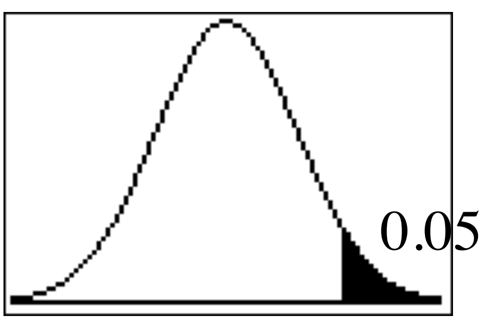
The results of the Central Limit Theorem can be applied since this problem involves the mean for a random sample of size n = 32.

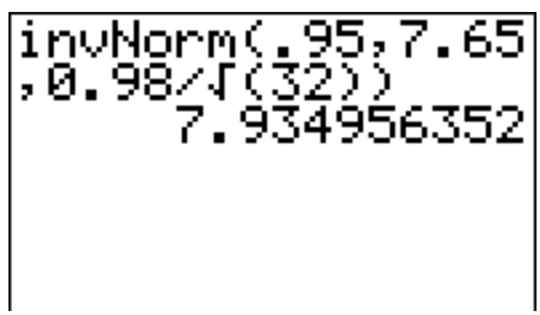


$$p_{7.5\text{ to }8.5} = p_{7.5\le \bar{X} \le 8.5} \approx 80.7\%$$

For a randomly selected college class that contains 32 students, the probability that the mean amount of sleep this sample of students had last night ranges from 7.5 to 8.5 hours is 80.7%.

(b) For a randomly selected college class that contains 32 students, find the values of the mean amount of sleep per night that would be considered an unusually high mean amount of sleep.





$$P_{95} \approx 7.93\ hours$$

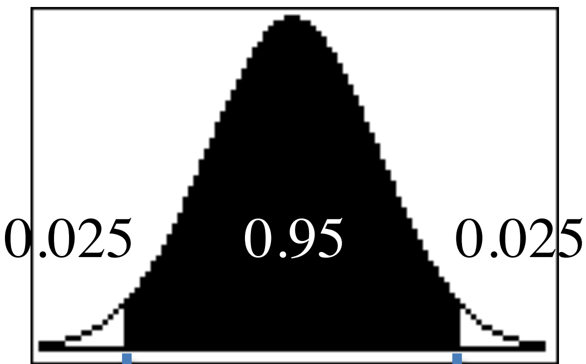
For a randomly selected college class that contains 32 students, the values of the mean amount of sleep per night that would be considered an unusually high mean amount of sleep would be 7.93 hours or more of sleep per night.

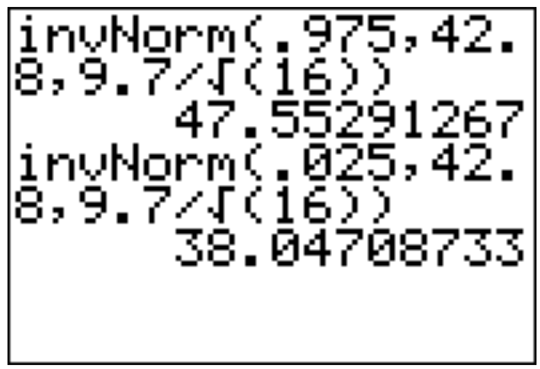
Exercise 2

The lengths of California Kingsnakes (*Lampropeltis getula californiae*) are normally distributed with a mean of 42.8 inches and a standard deviation of 9.7 inches.

Find the narrowest interval that has a 95% chance of containing the mean length for a random sample of 16 California Kingsnakes.

The results of the Central Limit Theorem can be applied since this problem involves the mean for a random sample of size n = 16.





$$U \approx 47.6 \; inches$$

$$L \approx 38.0 \; inches$$

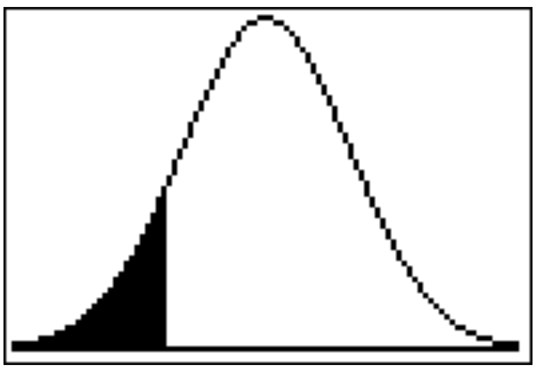
Thus, the narrowest interval that has a 95% chance of containing the mean length for a random sample of 16 California Kingsnakes ranges from a low of 38.0 inches to a high of 47.6 inches.

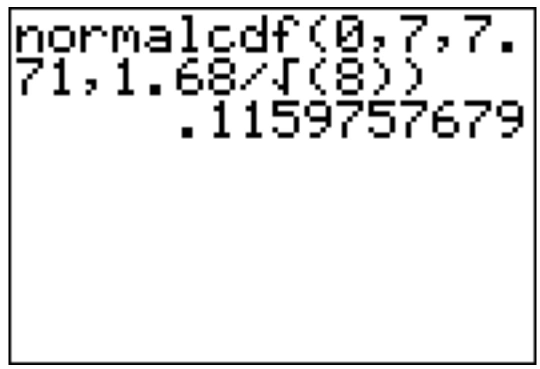
Exercise 3

The weights of adult male Rhesus monkeys (*Macaca mulatta*) are normally distributed with a mean of 7.71 kilograms and a standard deviation of 1.68 kilograms.

Would it be unusual if the mean weight for a random sample of 8 adult male Rhesus monkeys were no more than 7 kilograms?

The results of the Central Limit Theorem can be applied since this problem involves the mean for a random sample of size n = 8.





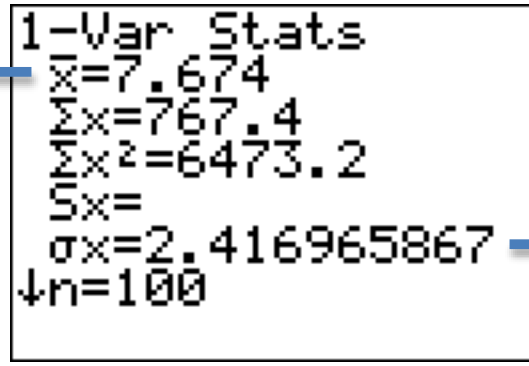
$$p_{\text{no more than 7}} = p_{\bar{X} \leq 7} \approx 0.116$$

No. It would not be unusual if the mean weight for a random sample of 8 adult male Rhesus monkeys were no more than 7 kilograms since the probability that this would occur 0.116 is not 0.05 or less.

Exercise 4

For Golden Retrievers, the number of puppies (X) born in one litter is approximated by the following probability distribution.

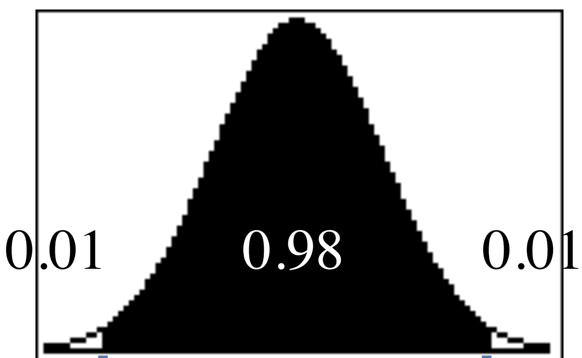
X	p_X
3	4.8%
4	6.5%
5	9.1%
6	12.0%
7	14.3%
8	15.2%
9	13.7%
10	10.3%
11	7.7%
12	6.4%

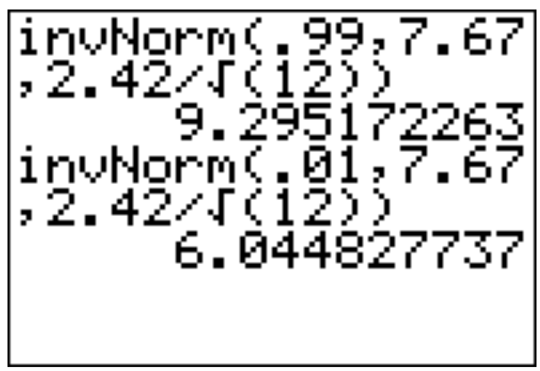


$\mu_X \approx 7.67$ $\sigma_X \approx 2.42$

The results of the Central Limit Theorem can be applied since this problem involves the mean for a random sample of size n = 12.

For a random sample of 12 Golden Retriever litters, find the narrowest interval that would accurately predict the mean number of puppies per litter 98% of the time.





$U \approx 9.3$ puppies per litter

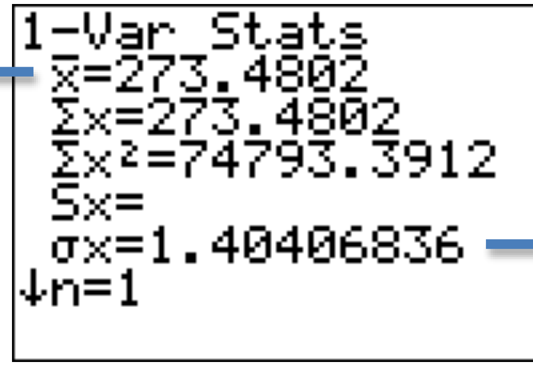
$L \approx 6.0$ puppies per litter

For a random sample of 12 Golden Retriever litters, the narrowest interval that would accurately predict the mean number of puppies per litter 98% of the time ranges from a low of 6.0 to a high of 9.3 puppies per litter.

Exercise 5

The number of M&M's (X) contained in one 8.2 ounce bag of Milk Chocolate Plain M&M's is approximated by the following probability distribution.

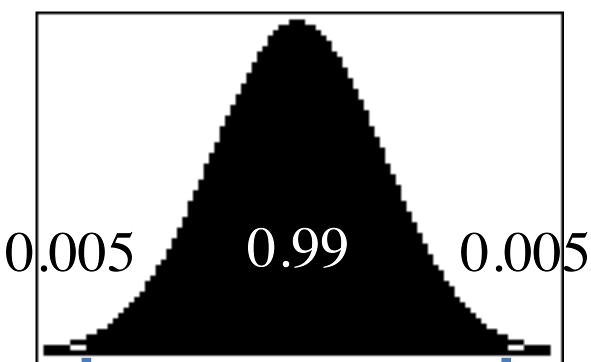
X	p_X
269	0.0014
270	0.0143
271	0.0637
272	0.1622
273	0.2580
274	0.2627
275	0.1672
276	0.0608
277	0.0097

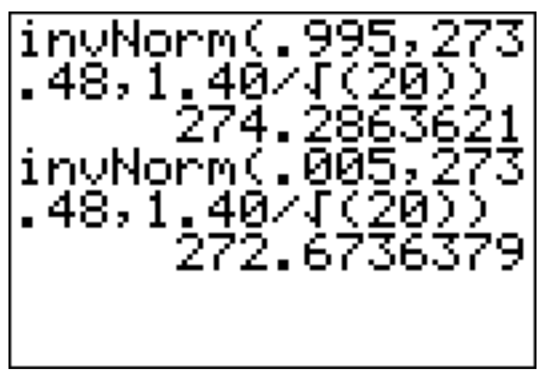


$\mu_X \approx 273.48$ $\sigma_X \approx 1.40$

The results of the Central Limit Theorem can be applied since this problem involves the mean for a random sample of size n = 20.

For a random sample of 20 8.2 ounce bags of Milk Chocolate Plain M&M's, find the narrowest interval that would accurately predict the mean number of M&M's per bag 99% of the time.





$U \approx 274.3$ M&M's per bag

$L \approx 272.7$ M&M's per bag

For a random sample of 20 8.2 ounce bags of Milk Chocolate Plain M&M's, the narrowest interval that would accurately predict the mean number of M&M's per bag 99% of the time ranges from a low of 272.7 to a high of 274.3 M&M's per bag.