

# The Central Limit Theorem for Sums

By:

OpenStaxCollege

Suppose  $X$  is a random variable with a distribution that may be **known or unknown** (it can be any distribution) and suppose:

1.  $\mu_X$  = the mean of  $X$
2.  $\sigma_X$  = the standard deviation of  $X$

If you draw random samples of size  $n$ , then as  $n$  increases, the random variable  $\Sigma X$  consisting of sums tends to be normally distributed and  $\Sigma X \sim N((n)(\mu_X), (\sqrt{n})(\sigma_X))$ .

**The** central limit theorem for sums says that if you keep drawing larger and larger samples and taking their sums, the sums form their own normal distribution (the sampling distribution), which approaches a normal distribution as the sample size increases. **The normal distribution has a mean equal to the original mean multiplied by the sample size and a standard deviation equal to the original standard deviation multiplied by the square root of the sample size.**

The random variable  $\Sigma X$  has the following  $z$ -score associated with it:

1.  $\Sigma x$  is one sum.
2.  $z = \frac{\Sigma x - (n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$ 
  1.  $(n)(\mu_X)$  = the mean of  $\Sigma X$
  2.  $(\sqrt{n})(\sigma_X)$  = standard deviation of  $\Sigma X$

To find probabilities for sums on the calculator, follow these steps.

2<sup>nd</sup> DISTR

2:normalcdf

normalcdf(lower value of the area, upper value of the area,  $(n)$ (mean),  $(\sqrt{n})$ (standard deviation))

## The Central Limit Theorem for Sums

where:

- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- *sample size* =  $n$

An unknown distribution has a mean of 90 and a standard deviation of 15. A sample of size 80 is drawn randomly from the population.

1. Find the probability that the sum of the 80 values (or the total of the 80 values) is more than 7,500.
2. Find the sum that is 1.5 standard deviations above the mean of the sums.

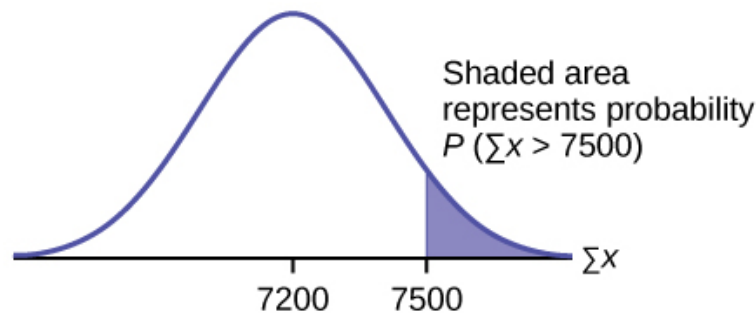
Let  $X$  = one value from the original unknown population. The probability question asks you to find a probability for **the sum (or total of) 80 values**.

$\Sigma X$  = the sum or total of 80 values. Since  $\mu_X = 90$ ,  $\sigma_X = 15$ , and  $n = 80$ ,  $\Sigma X \sim N((80)(90), (\sqrt{80})(15))$

- mean of the sums =  $(n)(\mu_X) = (80)(90) = 7,200$
- standard deviation of the sums =  $(\sqrt{n})(\sigma_X) = (\sqrt{80})(15)$
- sum of 80 values =  $\Sigma x = 7,500$

a. Find  $P(\Sigma x > 7,500)$

$$P(\Sigma x > 7,500) = 0.0127$$



`normalcdf(lower value, upper value, mean of sums, stdev of sums)`

The parameter list is abbreviated(lower, upper,  $(n)(\mu_X)$ ,  $(\sqrt{n})(\sigma_X)$ )

$$\text{normalcdf}(7500, 1E99, (80)(90), (\sqrt{80})(15)) = 0.0127$$

Reminder

## The Central Limit Theorem for Sums

$$1E99 = 10^{99}.$$

Press the EE key for E.

b. Find  $\Sigma x$  where  $z = 1.5$ .

$$\Sigma x = (n)(\mu_X) + (z)(\sqrt{n})(\sigma_X) = (80)(90) + (1.5)(\sqrt{80})(15) = 7,401.2$$

Try It

An unknown distribution has a mean of 45 and a standard deviation of eight. A sample size of 50 is drawn randomly from the population. Find the probability that the sum of the 50 values is more than 2,400.

0.0040

To find percentiles for sums on the calculator, follow these steps.

2<sup>nd</sup> DIStR

3:invNorm

$k = \text{invNorm}(\text{area to the left of } k, (n)(\text{mean}), (\sqrt{n})(\text{standard deviation}))$

where:

- $k$  is the  $k^{\text{th}}$  percentile
- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- *sample size* =  $n$

In a recent study reported Oct. 29, 2012 on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. The sample of size is 50.

1. What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?
2. Find the probability that the sum of the ages is between 1,500 and 1,800 years.
3. Find the 80<sup>th</sup> percentile for the sum of the 50 ages.
  1.  $\mu_{\Sigma x} = n\mu_x = 50(34) = 1,700$  and  $\sigma_{\Sigma x} = \sqrt{n}\sigma_x = (\sqrt{50})(15) = 106.01$   
The distribution is normal for sums by the central limit theorem.
  2.  $P(1500 < \Sigma x < 1800) = \text{normalcdf}(1,500, 1,800, (50)(34), (\sqrt{50})(15)) = 0.7974$
  3. Let  $k =$  the 80<sup>th</sup> percentile.  
 $k = \text{invNorm}(0.80, (50)(34), (\sqrt{50})(15)) = 1,789.3$

Try It

In a recent study reported Oct.29, 2012 on the Flurry Blog, the mean age of tablet users is 35 years. Suppose the standard deviation is ten years. The sample size is 39.

1. What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?
2. Find the probability that the sum of the ages is between 1,400 and 1,500 years.
3. Find the 90<sup>th</sup> percentile for the sum of the 39 ages.
  1.  $\mu_{\Sigma x} = n\mu_x = 1,365$  and  $\sigma_{\Sigma x} = \sqrt{n}\sigma_x = 62.4$   
The distribution is normal for sums by the central limit theorem.
  2.  $P(1400 < \Sigma x < 1500) = \text{normalcdf}(1400,1500,(39)(35),(\sqrt{39})(10)) = 0.2723$
  3. Let  $k =$  the 90<sup>th</sup> percentile.  
 $k = \text{invNorm}(0.90,(39)(35),(\sqrt{39})(10)) = 1445.0$

The mean number of minutes for app engagement by a tablet user is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample of size 70.

1. What are the mean and standard deviation for the sums?
2. Find the 95<sup>th</sup> percentile for the sum of the sample. Interpret this value in a complete sentence.
3. Find the probability that the sum of the sample is at least ten hours.
  1.  $\mu_{\Sigma x} = n\mu_x = 70(8.2) = 574$  minutes and  $\sigma_{\Sigma x} = (\sqrt{n})(\sigma_x) = (\sqrt{70})(1) = 8.37$  minutes
  2. Let  $k =$  the 95<sup>th</sup> percentile.  
 $k = \text{invNorm}(0.95,(70)(8.2),(\sqrt{70})(1)) = 587.76$  minutes  
Ninety five percent of the app engagement times are at most 587.76 minutes.
  3. ten hours = 600 minutes  
 $P(\Sigma x \geq 600) = \text{normalcdf}(600, E99, (70)(8.2), (\sqrt{70})(1)) = 0.0009$

The mean number of minutes for app engagement by a table use is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample size of 70.

1. What is the probability that the sum of the sample is between seven hours and ten hours? What does this mean in context of the problem?
2. Find the 84<sup>th</sup> and 16<sup>th</sup> percentiles for the sum of the sample. Interpret these values in context.
  1. 7 hours = 420 minutes  
10 hours = 600 minutes  
normalcdf  
 $P(420 \leq \Sigma x \leq 600) = \text{normalcdf}(420, 600, (70)(8.2), \sqrt{70}(1)) = 0.9991$

This means that for this sample sums there is a 99.9% chance that the sums of usage minutes will be between 420 minutes and 600 minutes.

2.  $invNorm(0.84, (70)(8.2), \sqrt{70}(1)) = 582.32$

$invNorm(0.16, (70)(8.2), \sqrt{70}(1)) = 565.68$

Since 84% of the app engagement times are at most 582.32 minutes and 16% of the app engagement times are at most 565.68 minutes, we may state that 68% of the app engagement times are between 565.68 minutes and 582.32 minutes.

## References

Farago, Peter. "The Truth About Cats and Dogs: Smartphone vs Tablet Usage Differences." The Flurry Blog, 2013. Posted October 29, 2012. Available online at <http://blog.flurry.com> (accessed May 17, 2013).

## Chapter Review

The central limit theorem tells us that for a population with any distribution, the distribution of the sums for the sample means approaches a normal distribution as the sample size increases. In other words, if the sample size is large enough, the distribution of the sums can be approximated by a normal distribution even if the original population is not normally distributed. Additionally, if the original population has a mean of  $\mu_X$  and a standard deviation of  $\sigma_x$ , the mean of the sums is  $n\mu_x$  and the standard deviation is  $(\sqrt{n})(\sigma_x)$  where  $n$  is the sample size.

## Formula Review

The Central Limit Theorem for Sums:  $\sum X \sim N[(n)(\mu_x), (\sqrt{n})(\sigma_x)]$

Mean for Sums ( $\sum X$ ):  $(n)(\mu_x)$

The Central Limit Theorem for Sums  $z$ -score and standard deviation for sums:  
 $z$  for the sample mean =  $\frac{\sum x - (n)(\mu_x)}{(\sqrt{n})(\sigma_x)}$

Standard deviation for Sums ( $\sum X$ ):  $(\sqrt{n})(\sigma_x)$

*Use the following information to answer the next four exercises:* An unknown distribution has a mean of 80 and a standard deviation of 12. A sample size of 95 is drawn randomly from the population.

Find the probability that the sum of the 95 values is greater than 7,650.

0.3345

## The Central Limit Theorem for Sums

Find the probability that the sum of the 95 values is less than 7,400.

Find the sum that is two standard deviations above the mean of the sums.

7,833.92

Find the sum that is 1.5 standard deviations below the mean of the sums.

*Use the following information to answer the next five exercises:* The distribution of results from a cholesterol test has a mean of 180 and a standard deviation of 20. A sample size of 40 is drawn randomly.

Find the probability that the sum of the 40 values is greater than 7,500.

0.0089

Find the probability that the sum of the 40 values is less than 7,000.

Find the sum that is one standard deviation above the mean of the sums.

7,326.49

Find the sum that is 1.5 standard deviations below the mean of the sums.

Find the percentage of sums between 1.5 standard deviations below the mean of the sums and one standard deviation above the mean of the sums.

77.45%

*Use the following information to answer the next six exercises:* A researcher measures the amount of sugar in several cans of the same soda. The mean is 39.01 with a standard deviation of 0.5. The researcher randomly selects a sample of 100.

Find the probability that the sum of the 100 values is greater than 3,910.

Find the probability that the sum of the 100 values is less than 3,900.

0.4207

Find the probability that the sum of the 100 values falls between the numbers you found in [\[link\]](#) and [\[link\]](#).

## The Central Limit Theorem for Sums

Find the sum with a  $z$ -score of  $-2.5$ .

3,888.5

Find the sum with a  $z$ -score of  $0.5$ .

Find the probability that the sums will fall between the  $z$ -scores  $-2$  and  $1$ .

0.8186

*Use the following information to answer the next four exercise:* An unknown distribution has a mean  $12$  and a standard deviation of one. A sample size of  $25$  is taken. Let  $X$  = the object of interest.

What is the mean of  $\Sigma X$ ?

What is the standard deviation of  $\Sigma X$ ?

5

What is  $P(\Sigma x = 290)$ ?

What is  $P(\Sigma x > 290)$ ?

0.9772

True or False: only the sums of normal distributions are also normal distributions.

In order for the sums of a distribution to approach a normal distribution, what must be true?

The sample size,  $n$ , gets larger.

What three things must you know about a distribution to find the probability of sums?

An unknown distribution has a mean of  $25$  and a standard deviation of six. Let  $X$  = one object from this distribution. What is the sample size if the standard deviation of  $\Sigma X$  is  $42$ ?

49

An unknown distribution has a mean of  $19$  and a standard deviation of  $20$ . Let  $X$  = the object of interest. What is the sample size if the mean of  $\Sigma X$  is  $15,200$ ?

Use the following information to answer the next three exercises. A market researcher analyzes how many electronics devices customers buy in a single purchase. The distribution has a mean of three with a standard deviation of 0.7. She samples 400 customers.

What is the  $z$ -score for  $\Sigma x = 840$ ?

26.00

What is the  $z$ -score for  $\Sigma x = 1,186$ ?

What is  $P(\Sigma x < 1,186)$ ?

0.1587

Use the following information to answer the next three exercises: An unknown distribution has a mean of 100, a standard deviation of 100, and a sample size of 100. Let  $X$  = one object of interest.

What is the mean of  $\Sigma X$ ?

What is the standard deviation of  $\Sigma X$ ?

1,000

What is  $P(\Sigma x > 9,000)$ ?

## Homework

Which of the following is NOT TRUE about the theoretical distribution of sums?

1. The mean, median and mode are equal.
2. The area under the curve is one.
3. The curve never touches the  $x$ -axis.
4. The curve is skewed to the right.

Suppose that the duration of a particular type of criminal trial is known to have a mean of 21 days and a standard deviation of seven days. We randomly sample nine trials.

1. In words,  $\Sigma X =$  \_\_\_\_\_
2.  $\Sigma X \sim$  \_\_\_\_\_ (\_\_\_\_\_, \_\_\_\_\_)
3. Find the probability that the total length of the nine trials is at least 225 days.



## The Central Limit Theorem for Sums

4. Ninety percent of the total of nine of these types of trials will last at least how long?
  1. the total length of time for nine criminal trials
  2.  $N(189, 21)$
  3. 0.0432
  4. 162.09; ninety percent of the total nine trials of this type will last 162 days or more.

Suppose that the weight of open boxes of cereal in a home with children is uniformly distributed from two to six pounds with a mean of four pounds and standard deviation of 1.1547. We randomly survey 64 homes with children.

1. In words,  $X =$  \_\_\_\_\_
2. The distribution is \_\_\_\_\_.
3. In words,  $\Sigma X =$  \_\_\_\_\_
4.  $\Sigma X \sim$  \_\_\_\_\_ (\_\_\_\_\_, \_\_\_\_\_)
5. Find the probability that the total weight of open boxes is less than 250 pounds.
6. Find the 35<sup>th</sup> percentile for the total weight of open boxes of cereal.

Salaries for teachers in a particular elementary school district are normally distributed with a mean of \$44,000 and a standard deviation of \$6,500. We randomly survey ten teachers from that district.

1. In words,  $X =$  \_\_\_\_\_
2.  $X \sim$  \_\_\_\_\_ (\_\_\_\_\_, \_\_\_\_\_)
3. In words,  $\Sigma X =$  \_\_\_\_\_
4.  $\Sigma X \sim$  \_\_\_\_\_ (\_\_\_\_\_, \_\_\_\_\_)
5. Find the probability that the teachers earn a total of over \$400,000.
6. Find the 90<sup>th</sup> percentile for an individual teacher's salary.
7. Find the 90<sup>th</sup> percentile for the sum of ten teachers' salary.
8. If we surveyed 70 teachers instead of ten, graphically, how would that change the distribution in part d?
9. If each of the 70 teachers received a \$3,000 raise, graphically, how would that change the distribution in part b?
  1.  $X =$  the salary of one elementary school teacher in the district
  2.  $X \sim N(44,000, 6,500)$
  3.  $\Sigma X \sim$  sum of the salaries of ten elementary school teachers in the sample
  4.  $\Sigma X \sim N(44000, 20554.80)$
  5. 0.9742
  6. \$52,330.09
  7. 466,342.04
  8. Sampling 70 teachers instead of ten would cause the distribution to be more spread out. It would be a more symmetrical normal curve.

## The Central Limit Theorem for Sums

9. If every teacher received a \$3,000 raise, the distribution of  $X$  would shift to the right by \$3,000. In other words, it would have a mean of \$47,000.