Central Limit Theorem: Using the C.L.T. (modified R. Bloom)*

Roberta Bloom

Based on *Central Limit Theorem: Using the Central Limit Theorem*† by

Barbara Illowsky, Ph.D.

Susan Dean

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0‡

Abstract

This module has examples illustrating use of the Central Limit Theorem is used. This revision of the original module in the collection Collaborative Statistics by S. Dean and Dr. B. Illowsky includes examples only for the CLT for means, and omits their material for the CLT for sums. The second example in this section has been changed to correct errors in the earlier versions of this module.

It is important to understand when to use the **CLT**. Use the CLT for means or averages when you are asked to find the probability for a sample average or mean, or when working with percentiles for sample averages. (If you are being asked to find the probability or percentile of a sum or total, use the CLT for sums.)

NOTE: If you are being asked to find the probability of an **individual** value, do **not** use the CLT. **Use the distribution of its random variable.**

1 Law of Large Numbers

The Law of Large Numbers says that if you take samples of larger and larger size from any population, then the mean $\overline{x}$ of the sample gets closer and closer to μ. From the Central Limit Theorem, we know that as n gets larger and larger, the sample averages follow a normal distribution. The larger n gets, the smaller the standard deviation gets. (Remember that the standard deviation for $\overline{X}$ is $\frac{\sigma}{\sqrt{n}}$.) This means that the sample mean $\overline{x}$ must be close to the population mean μ. We can say that μ is the value that the sample averages approach as n gets larger. The Central Limit Theorem illustrates the Law of Large Numbers.

Example 1

A study involving stress is done on a college campus among the students. **The stress scores follow a continuous uniform distribution** with the lowest stress score equal to 1 and the highest equal to 5. Using a sample of 75 students, find:

*Version 1.4: Dec 16, 2008 12:32 pm +0000

†http://cnx.org/content/m16958/1.6/

‡http://creativecommons.org/licenses/by/2.0/

a. The probability that the **average stress score** for the 75 students is less than 2.
b. The 90th percentile for the **average stress score** for the 75 students.

Let X = the stress score for one individual student

The individual stress scores follow a continuous uniform distribution, $X \sim U(1,5)$ where $a = 1$ and $b = 5$ (See the chapter on Continuous Random Variables).

$\mu_X = \frac{a+b}{2} = \frac{1+5}{2} = 3$

$\sigma_X = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(5-1)^2}{12}} = 1.15$

Problems a and b ask you to find a probability or a percentile for an **average** or **mean**. The sample size, n, is equal to 75.

Let $\overline{X}$ = the average stress score for the 75 students.

For the average stress score, use the CLT which tells us that $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$\overline{X} \sim N\left(3, \frac{1.15}{\sqrt{75}}\right)$ where n = 75.

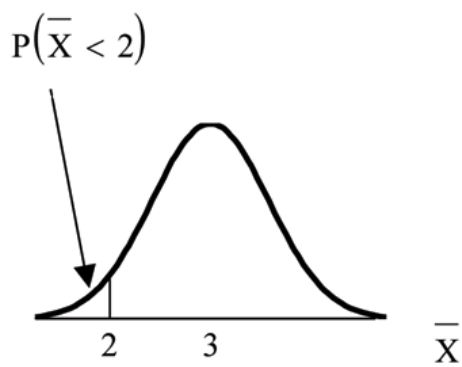
Problem 1

Find $P\left(\overline{X} < 2\right)$. Draw the graph.

Solution

$P\left(\overline{X} < 2\right) = 0$

The probability that the average stress score is less than 2 is about 0.



`normalcdf` $\left(1, 2, 3, \frac{1.15}{\sqrt{75}}\right) = 0$

NOTE: The smallest stress score is 1. Therefore, the smallest average for 75 stress scores is 1.

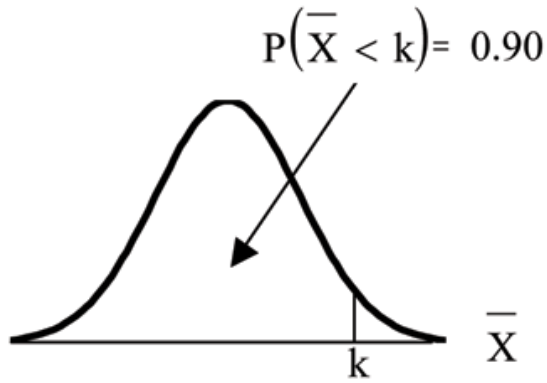
Problem 2

Find the 90th percentile for the sample average of 75 stress scores. Draw a graph.

Solution

Let k = the 90th precentile. Find k where $P\left(\overline{X} < k\right) = 0.90$.

$k = 3.17$ using `invNorm`$\left(.90, 3, \frac{1.15}{\sqrt{75}}\right) = 3.17$



The 90th percentile for the sample average of 75 scores is about 3.17. This means that 90% of all the averages of samples of 75 stress scores are at most 3.17 and 10% of the sample averages are at least 3.17 .

Example 2

Suppose that a market research analyst for a cell phone company conducts a study of their customers who exceed the time allowance included on their basic cell phone contract; the analyst finds that for those people who exceed the time included in their basic contract, the **excess time used** follows an **exponential distribution** with a mean of 22 minutes.

Consider a random sample of 80 customers who exceed the time allowance included in their basic cell phone contract.

Let X = the excess time used by one INDIVIDUAL cell phone customer who exceeds his contracted time allowance.

$X \sim \text{Exp}\left(\frac{1}{22}\right)$ From Chapter 5, we know that $\mu = 22$ and $\sigma = 22$.

Let $\overline{X}$ = the AVERAGE excess time used by a sample of $n =$ 80 customers who exceed their contracted time allowance.

$\overline{X} \sim N\left(22, \frac{22}{\sqrt{80}}\right)$ by the CLT for Sample Means or Averages

Problem 1

Using the CLT to find Probability:

a. Find the probability that the average excess time used by the 80 customers in the sample is longer than 20 minutes. This is asking us to find $P\left(\overline{X} > 20\right)$ Draw the graph.

b. Suppose that one customer who exceeds the time limit for his cell phone contract is randomly selected. Find the probability that this individual customer's excess time is longer than 20 minutes. This is asking us to find $P\left(X > 20\right)$

c. Explain why the probabilities in (a) and (b) are different.

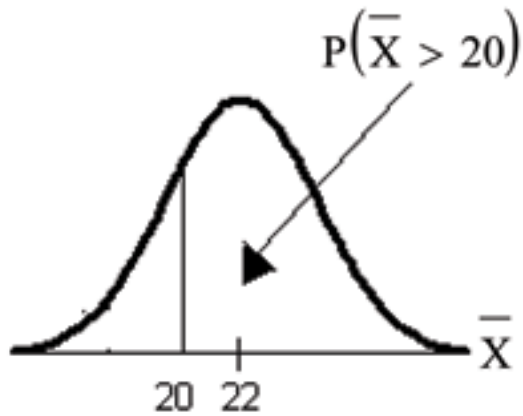
Solution

Part a.

Find: $P\left(\overline{X} > 20\right)$

$P\left(\overline{X} > 20\right) = 0.7919$ using `normalcdf`$\left(20, 1\text{E}99, 22, \frac{22}{\sqrt{80}}\right)$

The probability is 0.7919 that the average excess time used is more than 20 minutes, for a sample of 80 customers who exceed their contracted time allowance.



NOTE: 1E99 = 10^{99} and$-$1E99 = -10^{99}. Press the EE key for E. Or just use 10^99 instead of 1E99.

Part b.

Find P(X>20) . Remember to use the exponential distribution for an **individual: X∼Exp(1/22)**.

P(X>20) = e^(–(1/22)*20) or e^(–.04545*20) = 0.4029

Part c. Explain why the probabilities in (a) and (b) are different.

$P(X > 20) = 0.4029$ but $P\left(\overline{X} > 20\right) = 0.7919$

The probabilities are not equal because we use different distributions to calculate the probability for individuals and for averages.

When asked to find the probability of an individual value, use the stated distribution of its random variable; do not use the CLT. Use the CLT with the normal distribution when you are being asked to find the probability for an average.

Problem 2

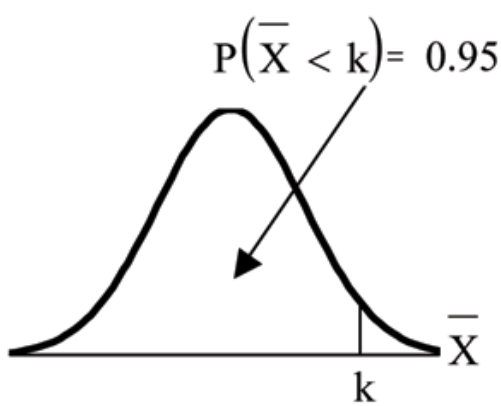
Using the CLT to find Percentiles:

Find the 95th percentile for the **sample average excess time** for samples of 80 customers who exceed their basic contract time allowances. Draw a graph.

Solution

Let k = the 95th percentile. Find k where $P\left(\overline{X} < k\right) = 0.95$

$k = 26.0$ using $\texttt{invNorm}\left(.95, 22, \frac{22}{\sqrt{80}}\right) = 26.0$



The 95th percentile for the **sample average excess time used** is about 26.0 minutes for random samples of 80 customers who exceed their contractual allowed time.

95% of such samples would have averages under 26 minutes; only 5% of such samples would have averages above 26 minutes.

Glossary

Definition 1: Average
A number that describes the central tendency of the data. There are a number of specialized averages, including the arithmetic mean, weighted mean, median, mode, and geometric mean.

Definition 2: Central Limit Theorem
Given a random variable (RV) with known mean μ and known variance σ^2, we are sampling with size n and we are interested in two new RV - sample mean, $\overset{\Psi}{X}$,and sample sum,ΣX. If the size n of the sample is sufficiently large, then $\overset{\Psi}{X} \sim N\left(\mathrm{n}\mu, \frac{\sigma^2}{n}\right)$ and $\Sigma X \sim N\left(\mathrm{n}\mu, n, \sigma^2\right)$. In words, if the size n of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distribution regardless of the shape of the population. And even more, the mean of the sampling distribution will equal the population mean and mean of sampling sums will equal n times the population mean. The standard deviation of the distribution of the sample means, $\frac{\sigma}{\sqrt{n}}$, is called standard error of the mean.

Definition 3: Exponential Distribution
Continuous random variable (RV) that appears when we are interested in intervals of time between some random events, for example, the length of time between emergency arrivals at a hospital. Notation: $X\tilde{}\mathrm{Exp}\left(m\right)$; the mean is $\mu = \frac{1}{m}$, and the variance is $\sigma^2 = \frac{1}{m^2}$, the probability density function is $f\left(x\right) = \mathrm{me}^{-\mathrm{mx}}, x \geq 0$and cumulative distribution is $P\left(X \leq x\right) = 1 - e^{-\mathrm{mx}}$.

Definition 4: Mean
A number to measure the central tendency (average), shortening from arithmetic mean. By definition, the mean for a sample (usually denoted by $\overset{\Psi}{X}$) is $\overset{\Psi}{X} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}}$, and the mean for a population (usually denoted by m) is $m = \frac{\text{Sum of all values in the population}}{\text{Number of values in the population}}$.