

Course Information

- [Schedule and Homework](#)
- [Course Requirements](#)
- [Instructor's Information](#)
- [Links](#)
- [Main Page](#)

Topics

1. [Introduction](#)
2. [Summarizing Data: listing and grouping](#)
3. [Summarizing Data: Statistical Descriptions](#)
4. [Possibilities and Probabilities](#)
5. [Some rules of probability](#)
6. [Probability Distributions](#)
7. [Normal Distributions](#)
8. [Sampling Distributions](#)
9. [Problems of Estimation](#)
10. [Hypothesis Tests](#)
11. [Tests based on count data](#)
12. [Regression](#)
13. [Non-parametric Tests](#)

Projects

1. [Stock return distribution](#)

- [Forum](#)

[edit](#)   [sec](#)   [src](#)   [prt](#)

[Site Manager](#)

[Fold](#)

|  |
|--|
| <b>Table of Contents</b>               |
| <a href="#">Sampling Distributions</a> |
| <a href="#">Central Limit Theorem</a>  |

# Chapter 8 Sampling Distributions

## Sampling Distributions

Suppose that without doing any research or having any knowledge of the business, you naively decide to open a pizzeria. On the first day you sell 9 pizzas, and on the next three days you sell 4,4, and 3 pizzas respectively. After the third day, you are broke and forced to close the business.

Leaving the sad story aside, consider our population data. Namely,  $N=4$  for 4 says, and  $x_1=9$ ,  $x_2=4a$ ,  $x_3=4b$  and  $x_4=3$ , where we just denote the two fours as  $4a$  and  $4b$  to indicate that they are from different days.

Let us now list all possible samples of size 2 (with replacement), namely:

3-3, 3-4a, 3-4b, 3-9, 4a-3,4a-4a, 4a-4b, 4a-9, 4b-3,4b-4a, 4b-4b, 4b-9, 9-3,9-4a,9-4b, 9-9. For these 16 samples, let us consider, the mean, median, standard deviation (sd), variance (var) and proportion of odd numbers in the sample. Calculating these values for the 16 samples we obtain:

| Sample  | Mean | Median | range | variance | standard deviation | proportion odd | probability |
|---|------|--------|-------|----------|--------------------|----------------|-------------|
| 3 3   | 3    | 3      | 0     | 0        | 0                  | 1              | 1/16        |
| 3 4   | 3.5  | 3.5    | 1     | 0.5      | 0.71               | 0.5            | 1/16        |
| 3 4   | 3.5  | 3.5    | 1     | 0.5      | 0.71               | 0.5            | 1/16        |
| 3 9   | 6    | 6      | 6     | 18       | 4.24               | 1              | 1/16        |
| 4 3   | 3.5  | 3.5    | 1     | 0.5      | 0.71               | 0.5            | 1/16        |
| 4 4   | 4    | 4      | 0     | 0        | 0.00               | 0              | 1/16        |
| 4 4   | 4    | 4      | 0     | 0        | 0.00               | 0              | 1/16        |
| 4 9   | 6.5  | 6.5    | 5     | 12.5     | 3.54               | 0.5            | 1/16        |
| 4 3   | 3.5  | 3.5    | 1     | 0.5      | 0.71               | 0.5            | 1/16        |
| 4 4   | 4    | 4      | 0     | 0        | 0.00               | 0              | 1/16        |
| 4 4   | 4    | 4      | 0     | 0        | 0.00               | 0              | 1/16        |
| 4 9   | 6.5  | 6.5    | 5     | 12.5     | 3.54               | 0.5            | 1/16        |
| 9 3   | 6    | 6      | 6     | 18       | 4.24               | 1              | 1/16        |
| 9 4   | 6.5  | 6.5    | 5     | 12.5     | 3.54               | 0.5            | 1/16        |
| 9 4   | 6.5  | 6.5    | 5     | 12.5     | 3.54               | 0.5            | 1/16        |
| 9 9   | 9    | 9      | 0     | 0        | 0.00               | 1              | 1/16        |
| Mean of statistic values                          | 5.0  | 5.0    | 2.3   | 5.5      | 1.6                | 0.50           |             |
| Population parameter                              | 5    | 4      | 6     | 5.5      | 2.35               | 0.5            |             |
| Does sample statistic target population parameter | Yes  | No     | No    | Yes      | No                 | Yes            |             |

Excel file available [here](#).

## Central Limit Theorem

The central limit theorem (CLT) is a fundamental theorem in statistics that allows us to have inferential statistics. Namely, suppose that we have a random procedure that follows some random variable with a given distribution, a mean  $\mu$  and standard deviation  $\sigma$ . Assume also

that you can repeat the experiment as many times as desired, and then you have random variables

$$X_1, X_2, X_3, \dots, X_n, \dots \quad (1)$$

that we call **i.i.d.r.v's**, that is *independent, identically distributed random variables*.

Then, the sample average is given by the sum of the first  $n$  such random variables divided by  $n$ :

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \quad (2)$$

The CLT states that as  $n \rightarrow \infty$  (becomes large), the distribution of the sample average  $\bar{X}_n$  becomes closer and closer to a normal distribution with mean  $\mu$  and diminishing standard deviation  $\sigma/\sqrt{n}$ .

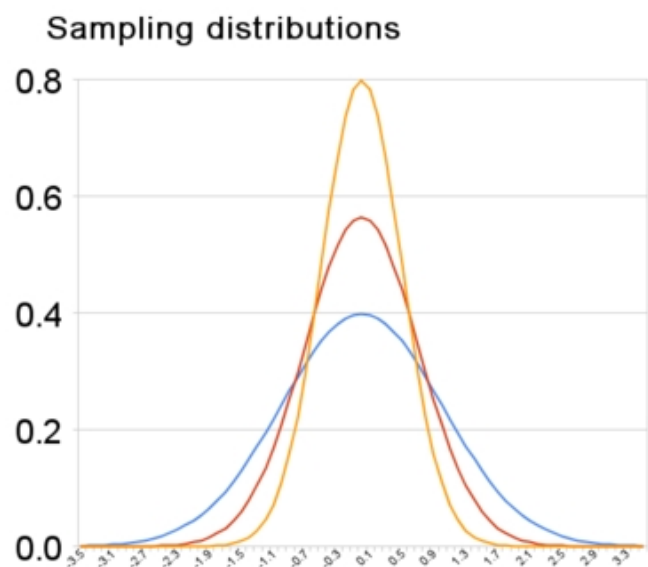
This can be denoted as follows:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow Z \sim N(0, 1) \quad (3)$$

The importance of this theorem cannot be stressed enough, suffice to say that the **result is independent** of the underlying distribution of the random variable  $X$ .

Notice that the mean remains unchanged and as  $n$  increases, the standard deviation of  $\bar{X}_n$  diminishes as  $1/\sqrt{n}$ .

As the graph below shows, this means that the weight gets concentrated on the center around the mean.



In the graph above, if the blue line represents the case of the original sample size  $n$ , the red is then  $2n$  and the yellow is when the sample is 4 times its original size.

An excellent way to see this theorem in action can be seen by using the applet that can be found at [http://www.ruf.rice.edu/~lane/stat\\_sim/sampling\\_dist/](http://www.ruf.rice.edu/~lane/stat_sim/sampling_dist/)

Click on begin in the top right of the site to get the Java applet running. Set the distribution on the top to be uniform  $U(0,32)$ , and in the 3rd and 4th panel the mean and  $N=5$  and  $N=20$  respectively. Click in both panels on the normal fit button. Click on animate in the second panel once or twice, then on the number of repetitions 5, 1000, and 10000, and observe how the mean  $\bar{X}_n$  becomes closer and closer to a normal distribution. Also notice that for  $n=5$ , the standard deviation is about twice as large than for  $n=20$ , as the number of repetitions increases.

page revision: 17, last edited: 31 Mar 2009, 02:51 (3577 days ago)

[Edit](#)   [Tags](#)   [History](#)   [Files](#)   [Print](#)   [Site tools](#)   [+ Options](#)

Powered by [Wikidot.com](http://www.wikidot.com)

[Help](#) | [Terms of Service](#) | [Privacy](#) | [Report a bug](#) | [Flag as objectionable](#)

Unless otherwise stated, the content of this page is licensed under [Creative Commons Attribution-ShareAlike 3.0 License](#)