Chapter 5
RANDOM SAMPLE

NGUYỄN THỊ THU THỦY (1)

SCHOOL OF APPLIED MATHEMATICS AND INFORMATICS
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

HANOI – 2020

(1)Email: thuy.nguyenthithu2@hust.edu.vn

5.1.1 Populations and Samples

Definition 1 (Population)

A population consists of the totality of the observations with which we are concerned.

Size

The number of observations in the population is defined to be the size of the population N.

5.1.1 Populations and Samples

Definition 1 (Population)

A population consists of the totality of the observations with which we are concerned.

Size

The number of observations in the population is defined to be the size of the population N.

5.1.1 Populations and Samples

Example 1

1. If there are 600 students in the school whom we classified according to blood type, we say that we have a population of size $N = 600$.
2. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake are examples of populations with finite size. In each case, the total number of observations is a finite number.
3. The observations obtained by measuring the atmospheric pressure every day, from the past on into the future, or all measurements of the depth of a lake, from any conceivable position, are examples of populations whose sizes are infinite.

5.1.1 Populations and Samples

Example 1

1. If there are 600 students in the school whom we classified according to blood type, we say that we have a population of size $N = 600$.
2. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake are examples of populations with finite size. In each case, the total number of observations is a finite number.
3. The observations obtained by measuring the atmospheric pressure every day, from the past on into the future, or all measurements of the depth of a lake, from any conceivable position, are examples of populations whose sizes are infinite.

5.1.1 Populations and Samples

Example 1

1. If there are 600 students in the school whom we classified according to blood type, we say that we have a population of size $N = 600$.
2. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake are examples of populations with finite size. In each case, the total number of observations is a finite number.
3. The observations obtained by measuring the atmospheric pressure every day, from the past on into the future, or all measurements of the depth of a lake, from any conceivable position, are examples of populations whose sizes are infinite.

5.1.1 Populations and Samples

Definition 2 (Sample)

A sample is a subset of a population.

Size

The number of observations in the sample is defined to be the size of the sample n.

5.1.2 Random Sample

Definition 3 (Random sample)

Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each having the same probability distribution $F_X(x)$. Define $(X_1, X_2, \ldots, X_n)$ to be a random sample of size n from the population $F_X(x)$ and write its joint probability distribution as

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n).$$

5.1.2 Random Sample

Example 2

If one makes a random selection of $n = 8$ storage batteries from a manufacturing process that has maintained the same specification throughout and records the length of life for each battery, with the first measurement x_1 being a value of X_1, the second measurement x_2 a value of X_2, and so forth, then $x_1, x_2, \dots, x_8$ are the values of the random sample $X_1, X_2, \dots, X_8$. If we assume the population of battery lives to be normal, the possible values of any X_i, $i = 1, 2, \dots, 8$, will be precisely the same as those in the original population, and hence X_i has the same identical normal distribution as X.

5.2 Some Important Statistics

Definition 4 (Statistic)

Any function of the random variables constituting a random sample is called a statistic.

5.2.1 Sample Mean. Median. Mode

Definition 5 (Sample mean, median, and mode)

1. Sample mean:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Note that the statistic $\overline{X}$ assumes the value $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ when X_1 assumes the value x_1, X_2 assumes the value x_2, and so forth.

2. Sample median:

$$\hat{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

3. The sample mode is the value of the sample that occurs most often.

5.2.1 Sample Mean. Median. Mode

Definition 5 (Sample mean, median, and mode)

1. Sample mean:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Note that the statistic $\overline{X}$ assumes the value $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ when X_1 assumes the value x_1, X_2 assumes the value x_2, and so forth.

2. Sample median:

$$\hat{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

3. The sample mode is the value of the sample that occurs most often.

5.2.1 Sample Mean. Median. Mode

Definition 5 (Sample mean, median, and mode)

1. Sample mean:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Note that the statistic $\overline{X}$ assumes the value $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ when X_1 assumes the value x_1, X_2 assumes the value x_2, and so forth.

2. Sample median:

$$\hat{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

3. The sample mode is the value of the sample that occurs most often.

5.2.1 Sample Mean. Median. Mode

Example 3

Suppose a data set consists of the following observations:

$$0.32 \quad 0.53 \quad 0.28 \quad 0.37 \quad 0.47 \quad 0.43 \quad 0.36 \quad 0.42 \quad 0.38 \quad 0.43.$$

The sample mode is 0.43, since this value occurs more than any other value.

5.2.2 Sample Variance. Standard Deviation. Range

Definition 6 (Sample variance, standard deviation, and range)

1. Sample variance:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

The computed value of S^2 for a given sample is denoted by s^2.

2. Sample standard deviation:

$$\boxed{S = \sqrt{S^2}}$$

3. Let $X_{\max}$ denote the largest of the X_i values and $X_{\min}$ the smallest. Sample range:

$$\boxed{\mathbf{R} = X_{\max} - X_{\min}}$$

5.2.2 Sample Variance. Standard Deviation. Range

Definition 6 (Sample variance, standard deviation, and range)

1. Sample variance:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

The computed value of S^2 for a given sample is denoted by s^2.

2. Sample standard deviation:

$$\boxed{S = \sqrt{S^2}}$$

3. Let $X_{\max}$ denote the largest of the X_i values and $X_{\min}$ the smallest. Sample range:

$$\boxed{\mathbf{R} = X_{\max} - X_{\min}}$$

5.2.2 Sample Variance. Standard Deviation. Range

Definition 6 (Sample variance, standard deviation, and range)

1. Sample variance:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

The computed value of S^2 for a given sample is denoted by s^2.

2. Sample standard deviation:

$$\boxed{S = \sqrt{S^2}}$$

3. Let $X_{\max}$ denote the largest of the X_i values and $X_{\min}$ the smallest. Sample range:

$$\boxed{\mathbf{R} = X_{\max} - X_{\min}}$$

5.2.2 Sample Variance. Standard Deviation. Range

Definition 6 (Sample variance, standard deviation, and range)

1. Sample variance:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

   The computed value of S^2 for a given sample is denoted by s^2.

2. Sample standard deviation:

$$\boxed{S = \sqrt{S^2}}$$

3. Let $X_{\max}$ denote the largest of the X_i values and $X_{\min}$ the smallest. Sample range:

$$\boxed{\mathbf{R} = X_{\max} - X_{\min}}$$

5.2.2 Sample Variance. Standard Deviation. Range

Example 4

A comparison of coffee prices at 4 randomly selected grocery stores in San Diego showed increases from the previous month of 12, 15, 17, and 20 cents for a 1-pound bag. Find the variance of this random sample of price increases.

5.2.2 Sample Variance. Standard Deviation. Range

Solution

Calculating the sample mean, we get

$$\overline{x} = \frac{12 + 15 + 17 + 20}{4} = 16 \quad \text{cents.}$$

Therefore,

$$\begin{aligned} s^2 &= \frac{1}{3}\sum_{i=1}^{4}(x_i - 16)^2 \\ &= \frac{(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2}{3} \\ &= \frac{34}{3}. \end{aligned}$$

5.2.2 Sample Variance. Standard Deviation. Range

Theorem 1

If S^2 is the variance of a random sample of size n, we may write

$$\boxed{S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]}$$

5.2.2 Sample Variance. Standard Deviation. Range

Example 5

Find the variance of the data 3, 4, 5, 6, 6, and 7, representing the number of trout caught by a random sample of 6 fishermen on June 19, 1996, at Lake Muskoka.

5.2.2 Sample Variance. Standard Deviation. Range

Solution

We find that $\sum_{i=1}^{6} x_i^2 = 171$, $\sum_{i=1}^{6} x_i = 311$, and $n = 6$. Hence,

$$s^2 = \frac{1}{(6)(5)}[(6)(171) - (31)^2] = \frac{13}{6}.$$

Thus, the sample standard deviation $s = \sqrt{13/6} = 1.47$ and the sample range is $7 - 3 = 4$.

5.2.2 Sample Variance. Standard Deviation. Range

Note

Sample	Population
n: number of measurements in the sample	N: number of measurements in the population
$\overline{x}$: sample mean	μ: population mean
s^2: sample variance	σ^2: population variance
s: sample standard deviation	σ: population standard deviation

5.2.2 Sample Variance. Standard Deviation. Range

Calculating Mean and Standard Deviation Using 570 VN Plus

The following commands show how to calculate the mean and standard deviation by using the STAT mode on a CASIO FX 570 VN PLUS (similar for other CASIO models).

Steps

1. Enter the data
   - **MODE** $\to$ **3** $\to$ **AC**
   - **SHIFT** $\to$ **MODE** $\to \Downarrow \to$ **4(STAT)** $\to$ **1(ON)**
   - **SHIFT** $\to$ **1** $\to$ **1(TYPE)** $\to$ **1(1-VAR)**
     To finish entering the data press the AC button.
2. Calculating numerical summaries
   - The sample mean $\overline{x}$: **SHIFT** $\to$ **1** $\to$ **4(VAR)** $\to$ **2**
   - The sample standard deviation s: **SHIFT** $\to$ **1** $\to$ **4** $\to$ **4**

5.2.2 Sample Variance. Standard Deviation. Range

Calculating Mean and Standard Deviation Using 570 VN Plus

The following commands show how to calculate the mean and standard deviation by using the STAT mode on a CASIO FX 570 VN PLUS (similar for other CASIO models).

Steps

1. Enter the data
   - **MODE** $\rightarrow$ **3** $\rightarrow$ **AC**
   - **SHIFT** $\rightarrow$ **MODE** $\rightarrow \Downarrow \rightarrow$ **4(STAT)** $\rightarrow$ **1(ON)**
   - **SHIFT** $\rightarrow$ **1** $\rightarrow$ **1(TYPE)** $\rightarrow$ **1(1-VAR)**

     To finish entering the data press the AC button.
2. Calculating numerical summaries
   - The sample mean $\overline{x}$: **SHIFT** $\rightarrow$ **1** $\rightarrow$ **4(VAR)** $\rightarrow$ **2**
   - The sample standard deviation s: **SHIFT** $\rightarrow$ **1** $\rightarrow$ **4** $\rightarrow$ **4**

5.2.2 Sample Variance. Standard Deviation. Range

Example 6

Find the mean and standard deviation of the following distribution

Variable (x_i)	20	21	22	23	24
Frequency (n_i)	5	8	11	10	6

Solution

22.1; 1.2567.

5.2.2 Sample Variance. Standard Deviation. Range

Example 6

Find the mean and standard deviation of the following distribution

Variable (x_i)	20	21	22	23	24
Frequency (n_i)	5	8	11	10	6

Solution

22.1; 1.2567.

5.3 Sampling Distributions

Definition 7

The probability distribution of a statistic is called a sampling distribution.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Introduction

The first important sampling distribution to be considered is that of the mean $\overline{X}$. Suppose that a random sample of n observations is taken from a normal population with mean μ and variance σ^2. Each observation X_i, $i = 1, 2, \ldots, n$, of the random sample will then have the same normal distribution as the population being sampled.

Sampling Distributions of Means

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

has a normal distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Theorem 2 (Central Limit Theorem)

If $\overline{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

as $n \to \infty$, is the standard normal distribution $\mathcal{N}(0, 1)$.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Note

1. The normal approximation for $\overline{X}$ will generally be good if $n \geq 30$, provided the population distribution is not terribly skewed.
2. If $n < 30$, the approximation is good only if the population is not too different from a normal distribution and, as stated above, if the population is known to be normal, the sampling distribution of $\overline{X}$ will follow a normal distribution exactly, no matter how small the size of the samples.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Note

1. The normal approximation for $\overline{X}$ will generally be good if $n \geq 30$, provided the population distribution is not terribly skewed.
2. If $n < 30$, the approximation is good only if the population is not too different from a normal distribution and, as stated above, if the population is known to be normal, the sampling distribution of $\overline{X}$ will follow a normal distribution exactly, no matter how small the size of the samples.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Note

Figure 1 illustrates how the theorem works. It shows how the distribution of $\overline{X}$ becomes closer to normal as n grows larger, beginning with the clearly nonsymmetric distribution of an individual observation ($n = 1$).

5.3.1 Sampling Distributions of Means and Central Limit Theorem

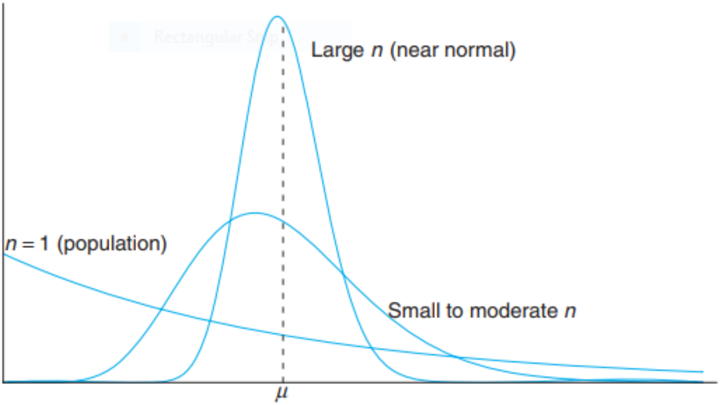


Figure 1: Illustration of the Central Limit Theorem (distribution of $\overline{X}$ for $n = 1$, moderate n, and large n).

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Example 7

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with a mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

5.3.1 Sampling Distributions of Means and Central Limit Theorem

Solution

The sampling distribution of $\overline{X}$ will be approximately normal, with $\mu_{\overline{X}} = 800$ and $\sigma_{\overline{X}} = 40/\sqrt{16} = 10$. The desired probability is given by the area of the shaded region in Figure 2. Corresponding to $\overline{x} = 775$, we find that $z = \frac{775-800}{10} = -2.5$ and therefore

$$P[\overline{X} < 775] = P[Z < -2.5] = 0.0062.$$

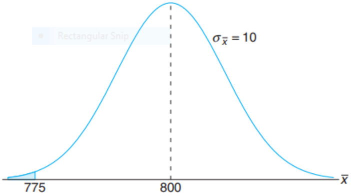


Figure 2: Area for Example 7

5.3.1 Sampling Distributions of Means and Central Limit Theorem

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62556	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68739
0,5	69146	69447	69847	70194	70544	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	**97500**	97558	97615	97670

5.3.1 Sampling Distributions of Means and Central Limit Theorem

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

x	0	1	2	3	4	5	6	7	8	9
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	99461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99285	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99261	99477	99492	99506	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99763
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3,0	0,99865	3,1	99903	3,2	99931	3,3	99952	3,4	99966	
3,5	99977	3,6	99984	3,7	99989	3,8	99993	3,9	99995	
4,0	999968									
4,5	999997									
5,0	99999997									

5.3.2 Sampling Distribution of the Difference between Two Means

Theorem 3

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of means, $\overline{X}_1 - \overline{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma^2_{\overline{X}_1-\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Hence,

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

is approximately a standard normal variable $\mathcal{N}(0, 1)$.

5.3.2 Sampling Distribution of the Difference between Two Means

Example 8 (Paint Drying Time)

Two independent experiments are run in which two different types of paint are compared. Eighteen specimens are painted using type A, and the drying time, in hours, is recorded for each. The same is done with type B. The population standard deviations are both known to be 1.0. Assuming that the mean drying time is equal for the two types of paint, find $P(\overline{X}_A - \overline{X}_B > 1.0)$, where $\overline{X}_A$ and $\overline{X}_B$ are average drying times for samples of size $n_A = n_B = 18$.

5.3.2 Sampling Distribution of the Difference between Two Means

Solution

From the sampling distribution of $\overline{X}_A - \overline{X}_B$, we know that the distribution is approximately normal with mean

$$\mu_{\overline{X}_A - \overline{X}_B} = \mu_A - \mu_B = 0$$

and variance

$$\sigma^2_{\overline{X}_A - \overline{X}_B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}.$$

5.3.2 Sampling Distribution of the Difference between Two Means

Solution (continuous)

The desired probability is given by the shaded region in Figure 3. Corresponding to the value $\overline{X}_A - \overline{X}_B = 1.0$, we have

$$z = \frac{1.0 - (\mu_A - \mu_B)}{\sqrt{1/9}} = \frac{1}{\sqrt{1/9}} = 3.0,$$

so

$$P(Z > 3.0) = 1 - P(Z < 3.0) = 1 - 0.9987 = 0.0013.$$

5.3.2 Sampling Distribution of the Difference between Two Means

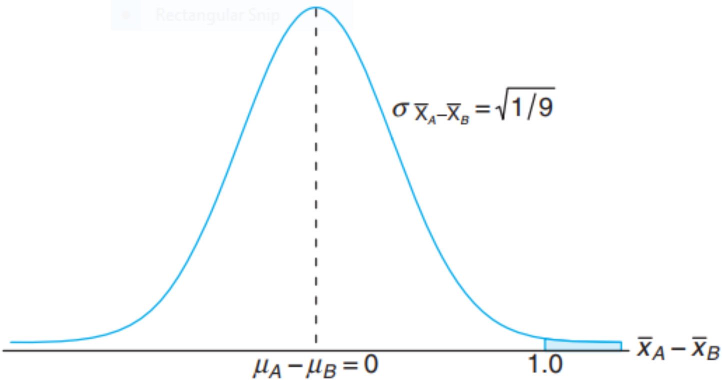


Figure 3: Area for Case Study in Example 8

5.3.3 Sampling Distribution of S^2

Theorem 4

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2}$$

has a chi-squared distribution with $\nu = n - 1$ degrees freedom.

Note

The values of the random variable χ^2 are calculated from each sample by the formula

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}.$$

5.3.3 Sampling Distribution of S^2

Note

It is customary to let χ^2_α represent the χ^2 value above which we find an area of α. This is illustrated by the shaded region in Figure 4.

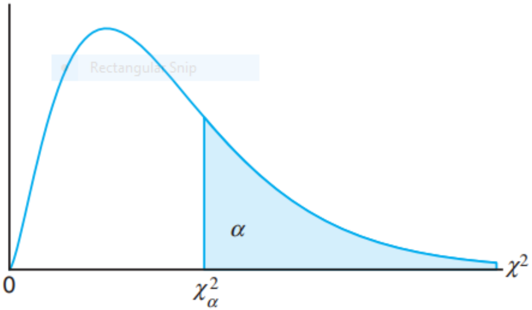


Figure 4: The chi-squared distribution

5.3.4 t-Distribution

Theorem 5

Let Z be a standard normal random variable and V a chi-squared random variable with ν degrees of freedom. If Z and V are independent, then the distribution of the random variable T, where

$$T = \frac{Z}{\sqrt{V/(n-1)}}$$

is given by the density function

$$h(t) = \frac{\Gamma(v+1)/2}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}, \quad -\infty < t < +\infty.$$

This is known as the t-distribution with ν degrees of freedom.

5.3.4 t-Distribution

Corollary 1

Let $X_1, X_2, \ldots, X_n$ be independent random variables that are all normal with mean μ and standard deviation σ. Let

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Then the random variable $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ has a t-distribution with $\nu = n-1$ degrees of freedom.

5.3.4 t-Distribution

Note

The probability distribution of T was first published in 1908 in a paper written by W.S. Gosset. At the time, Gosset was employed by an Irish brewery that prohibited publication of research by members of its staff. To circumvent this restriction, he published his work secretly under the name "Student." Consequently, the distribution of T is usually called the Student t-distribution or simply the t-distribution.

5.3.4 t-Distribution

What does the t-distribution look like?

1. The distribution of T is similar to the distribution of Z in that they both are symmetric about a mean of zero.
2. Both distributions are bell shaped, but the t-distribution is more variable, owing to the fact that the t-values depend on the fluctuations of two quantities, $\overline{X}$ and S^2, whereas the Z-values depend only on the changes in $\overline{X}$ from sample to sample.
3. The distribution of T differs from that of Z in that the variance of T depends on the sample size n and is always greater than 1.
4. Only when the sample size $n \to \infty$ will the two distributions become the same.

5.3.4 t-Distribution

What does the t-distribution look like?

1. The distribution of T is similar to the distribution of Z in that they both are symmetric about a mean of zero.
2. Both distributions are bell shaped, but the t-distribution is more variable, owing to the fact that the t-values depend on the fluctuations of two quantities, $\overline{X}$ and S^2, whereas the Z-values depend only on the changes in $\overline{X}$ from sample to sample.
3. The distribution of T differs from that of Z in that the variance of T depends on the sample size n and is always greater than 1.
4. Only when the sample size $n \to \infty$ will the two distributions become the same.

5.3.4 t-Distribution

What does the t-distribution look like?

1. The distribution of T is similar to the distribution of Z in that they both are symmetric about a mean of zero.
2. Both distributions are bell shaped, but the t-distribution is more variable, owing to the fact that the t-values depend on the fluctuations of two quantities, $\overline{X}$ and S^2, whereas the Z-values depend only on the changes in $\overline{X}$ from sample to sample.
3. The distribution of T differs from that of Z in that the variance of T depends on the sample size n and is always greater than 1.
4. Only when the sample size $n \to \infty$ will the two distributions become the same.

5.3.4 t-Distribution

What does the t-distribution look like?

1. The distribution of T is similar to the distribution of Z in that they both are symmetric about a mean of zero.
2. Both distributions are bell shaped, but the t-distribution is more variable, owing to the fact that the t-values depend on the fluctuations of two quantities, $\overline{X}$ and S^2, whereas the Z-values depend only on the changes in $\overline{X}$ from sample to sample.
3. The distribution of T differs from that of Z in that the variance of T depends on the sample size n and is always greater than 1.
4. Only when the sample size $n \to \infty$ will the two distributions become the same.

5.3.4 t-Distribution

What does the t-distribution look like?

In Figure 5, we show the relationship between a standard normal distribution ($\nu = \infty$) and t-distributions with 2 and 5 degrees of freedom. The percentage points of the t-distribution are given in Table 1.

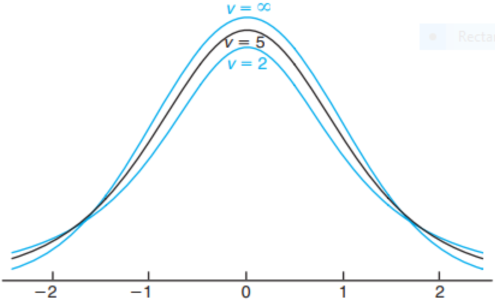
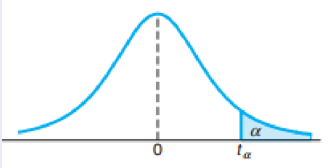


Figure 5: The t-distribution curves for $\nu = 2, 5$, and ∞

5.3.4 t-Distribution

Table 1

Critical Values of the t-Distribution.



5.3.4 t-Distribution

d.f. \ α	$0,10$	$0,05$	$0,025$	$0,01$	$0,005$	$0,002$	0.0005
1	3,078	6,314	12,706	31,821	63,526	318,309	363,6
2	1,886	2,920	4,303	6,965	9,925	22,327	31,600
3	1,638	2,353	3,128	4,541	5,841	10,215	12,922
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,705	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,606	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850

5.3.4 t-Distribution

d.f. \ α	$0,10$	$0,05$	$0,025$	$0,01$	$0,005$	$0,002$	0.0005
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,796	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
$+\infty$	1,282	1,645	1,960	2,326	2,576	3,090	3,291

5.3.4 t-Distribution

Note

1. It is customary to let t_α represent the t-value above which we find an area equal to α.
2. The t-value with 10 degrees of freedom leaving an area of 0.025 to the right is $t = 2.228$.
3. Since the t-distribution is symmetric about a mean of zero, we have $t_{1-\alpha} = -t_\alpha$; that is, the t-value leaving an area of $1-\alpha$ to the right and therefore an area of α to the left is equal to the negative t-value that leaves an area of α in the right tail of the distribution (see Figure 6. That is, $t_{0.95} = -t_{0.05}$, $t_{0.99} = -t_{0.01}$, and so forth.

5.3.4 t-Distribution

Note

1. It is customary to let t_α represent the t-value above which we find an area equal to α.
2. The t-value with 10 degrees of freedom leaving an area of 0.025 to the right is $t = 2.228$.
3. Since the t-distribution is symmetric about a mean of zero, we have $t_{1-\alpha} = -t_\alpha$; that is, the t-value leaving an area of $1-\alpha$ to the right and therefore an area of α to the left is equal to the negative t-value that leaves an area of α in the right tail of the distribution (see Figure 6. That is, $t_{0.95} = -t_{0.05}$, $t_{0.99} = -t_{0.01}$, and so forth.

5.3.4 t-Distribution

Note

1. It is customary to let t_α represent the t-value above which we find an area equal to α.
2. The t-value with 10 degrees of freedom leaving an area of 0.025 to the right is $t = 2.228$.
3. Since the t-distribution is symmetric about a mean of zero, we have $t_{1-\alpha} = -t_\alpha$; that is, the t-value leaving an area of $1 - \alpha$ to the right and therefore an area of α to the left is equal to the negative t-value that leaves an area of α in the right tail of the distribution (see Figure 6. That is, $t_{0.95} = -t_{0.05}$, $t_{0.99} = -t_{0.01}$, and so forth.

5.3.4 t-Distribution

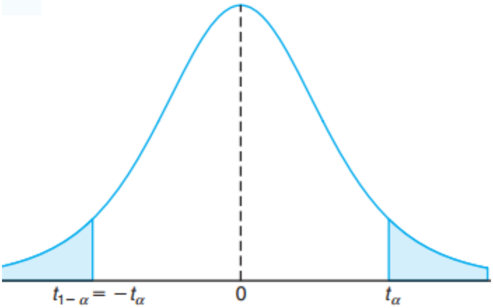


Figure 6: Symmetry property (about 0) of the t-distribution

5.3.4 t-Distribution

Example 9

The t-value with $\nu = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is

$$t_{0.975} = -t_{0.025} = -2.145$$

Example 10

Find $P[-t_{0.025} < T < t_{0.05}]$.

Solution

Since $t_{0.05}$ leaves an area of 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of $1 - 0.05 - 0.025 = 0.925$ between $-t_{0.025}$ and $t_{0.05}$. Hence

$$P[-t_{0.025} < T < t_{0.05}] = 0.925.$$

5.3.4 t-Distribution

Example 9

The t-value with $\nu = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is

$$t_{0.975} = -t_{0.025} = -2.145$$

Example 10

Find $P[-t_{0.025} < T < t_{0.05}]$.

Solution

Since $t_{0.05}$ leaves an area of 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of $1 - 0.05 - 0.025 = 0.925$ between $-t_{0.025}$ and $t_{0.05}$. Hence

$$P[-t_{0.025} < T < t_{0.05}] = 0.925.$$

5.3.4 t-Distribution

Example 9

The t-value with $\nu = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is

$$t_{0.975} = -t_{0.025} = -2.145$$

Example 10

Find $P[-t_{0.025} < T < t_{0.05}]$.

Solution

Since $t_{0.05}$ leaves an area of 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of $1 - 0.05 - 0.025 = 0.925$ between $-t_{0.025}$ and $t_{0.05}$. Hence

$$P[-t_{0.025} < T < t_{0.05}] = 0.925.$$

5.3.4 t-Distribution

Example 11

Find k such that $P[k < T < -1.761] = 0.045$ for a random sample of size 15 selected from a normal distribution and $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$.

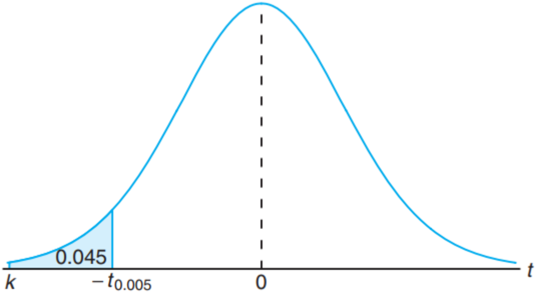


Figure 7: The t-values for Example 11

5.3.4 t-Distribution

Solution

From Table 1 we note that 1.761 corresponds to $t_{0.05}$ when $\nu = 14$. Therefore, $-t_{0.05} = -1.761$. Since k in the original probability statement is to the left of $-t_{0.05} = -1.761$, let $k = -t_\alpha$. Then, from Figure 7, we have $0.045 = 0.05 - \alpha$, or $\alpha = 0.005$. Hence, from Table 1 with $\nu = 14$, $k = -t_{0.005} = -2.977$ and

$$P[-2.977 < T < -1.761] = 0.045.$$

5.3.4 t-Distribution

Example 12

A chemical engineer claims that the population mean yield of a certain batch process is 500 grams per milliliter of raw material. To check this claim he samples 25 batches each month. If the computed t-value falls between $-t_{0.05}$ and $t_{0.05}$, he is satisfied with this claim. What conclusion should he draw from a sample that has a mean $\overline{x} = 518$ grams per milliliter and a sample standard deviation $s = 40$ grams? Assume the distribution of yields to be approximately normal.

5.3.4 t-Distribution

Solution

From Table 1 we find that $t_{0.05} = 1.711$ for 24 degrees of freedom. Therefore, the engineer can be satisfied with his claim if a sample of 25 batches yields a t-value between -1.711 and 1.711. If $\mu = 500$, then

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2.25,$$

a value well above 1.711. The probability of obtaining a t-value, with $\nu = 24$, equal to or greater than 2.25 is approximately 0.02. If $\mu > 500$, the value of t computed from the sample is more reasonable. Hence, the engineer is likely to conclude that the process produces a better product than he thought.

5.3.4 t-Distribution

What is the t-distribution used for?

1. The reader should note that use of the t-distribution for the statistic

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

requires that $X_1, X_2, \ldots, X_n$ be normal.

2. The use of the t-distribution and the sample size consideration do not relate to the Central Limit Theorem. The use of the standard normal distribution rather than T for $n \geq 30$ merely implies that S is a sufficiently good estimator of σ in this case.

5.3.4 t-Distribution

What is the t-distribution used for?

1. The reader should note that use of the t-distribution for the statistic
$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$
requires that $X_1, X_2, \ldots, X_n$ be normal.
2. The use of the t-distribution and the sample size consideration do not relate to the Central Limit Theorem. The use of the standard normal distribution rather than T for $n \geq 30$ merely implies that S is a sufficiently good estimator of σ in this case.

Problems

Problem 5.1

The lengths of time, in minutes, that 10 patients waited in a doctor's office before receiving treatment were recorded as follows: 5, 11, 9, 5, 10, 15, 6, 10, 5, and 10. Treating the data as a random sample, find (a) the mean; (b) the median; (c) the mode.

Problem 5.2

The numbers of incorrect answers on a true-false competency test for a random sample of 15 students were recorded as follows: 2, 1, 3, 0, 1, 3, 6, 0, 3, 3, 5, 2, 1, 4, and 2. Find (a) the mean; (b) the median; (c) the mode.

Problems

Problem 5.3

The grade-point averages of 20 college seniors selected at random from a graduating class are as follows: 3.2, 1.9, 2.7, 2.4, 2.8, 2.9, 3.8, 3.0, 2.5, 3.3, 1.8, 2.5, 3.7, 2.8, 2.0, 3.2, 2.3, 2.1, 2.5, 1.9. Calculate the standard deviation.

Problem 5.4

(a) Find $t_{0.025}$ when $\nu = 14$. (b) Find $-t_{0.10}$ when $\nu = 10$. (c) Find $t_{0.995}$ when $\nu = 7$.

Problem 5.5

(a) Find $P(T < 2.365)$ when $\nu = 7$. (b) Find $P(T > 1.318)$ when $\nu = 24$. (c) Find $P(-1.356 < T < 2.179)$ when $\nu = 12$. (d) Find $P(T > -2.567)$ when $\nu = 17$.

Problems

Problem 5.6

Given a random sample of size 24 from a normal distribution, find k such that (a) $P(-2.069 < T < k) = 0.965$; (b) $P(k < T < 2.807) = 0.095$; (c) $P(-k < T < k) = 0.90$.

Problem 5.7

A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed t-value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of $\overline{x} = 27.5$ hours and a standard deviation of $s = 5$ hours? Assume the distribution of battery lives to be approximately normal.