Chapter 7. Statistical Intervals Based on a Single Sample

Chapter 7: Statistical Intervals Based on A Single Sample

- 7.1. Basic Properties of Confidence Intervals
- 7.2. Larger-Sample Confidence Intervals for a Population Mean and Proportion
- 7.3 Intervals Based on a Normal Population Distribution

Chapter 7 Introduction

- Introduction

- A point estimation provides no information about the **precision** and **reliability** of estimation.
- For example, using the statistic $\overline{X}$ to calculate a point estimate for the true average breaking strength (g) of paper towels of a certain brand, and suppose that $\overline{X} = 9322.7$. Because of sample **variability**, it is virtually never the case that $\overline{X} = \mu$. The point estimate says nothing about how close it might be to μ.
- An alternative to reporting a single sensible value for the parameter being estimated is to calculate and report an entire interval of plausible values—an interval estimate or confidence interval (CI)

7.1 Basic Properties of Confidence Intervals

7.1 Basic Properties of Confidence Intervals

- Considering a Simple Case

Suppose that the parameter of interest is a **population mean μ** and that

1. The population distribution is normal.
2. The value of the population standard deviation σ is known

- Normality of the population distribution is often a reasonable assumption.
- If the value of μ is unknown, it is implausible that the value of σ would be available.

In later sections, we will develop methods based on less restrictive assumptions.

7.1 Basic Properties of Confidence Intervals

- **Example 7.1**

Industrial engineers who specialize in ergonomics are concerned with designing workspace and devices operated by workers so as to achieve high productivity and comfort. A sample of $n = 31$ trained typists was selected , and the preferred keyboard height was determined for each typist. The resulting sample average preferred height was 80.0 cm. Assuming that preferred height is normally distributed with $\sigma = 2.0$ cm. Please obtain a CI for μ, the true average preferred height for the population of all experienced typists.

Given the confidence level is 95%, find confidence intervals?

Consider a random sample X_1, X_2, … X_n from the normal distribution with mean value μ and standard deviation σ . Then the sample mean is normally distribution with expected value μ and standard deviation $\sigma/\sqrt{n}.$

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

Because the area under the standard normal curve between –1.96 and 1.96 is .95,

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = .95 \qquad \textbf{(7.2)}$$

Now let's manipulate the inequalities inside the parentheses in (7.2) so that they appear in the equivalent form $l < \mu < \mu$, where the endpoints l and u involve X and $\sigma/\sqrt{n}$.

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \qquad \textbf{(7.3)}$$

To interpret (7.3), think of a **random interval** having left endpoint $\overline{X} - 1.96 \cdot \sigma/\sqrt{n}$ and right endpoint $\overline{X} + 1.96 \cdot \sigma/\sqrt{n}$. In interval notation, this becomes

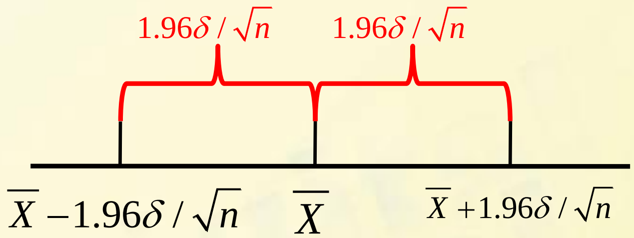
$$\left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) \qquad \textbf{(7.4)}$$

7.1 Basic Properties of Confidence Intervals

- Example 7.1 (Cont')

The CI of 95% is:

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$



Interpreting a CI: It can be paraphrased as "the probability is 0.95 that the random interval includes or covers the true value of μ.

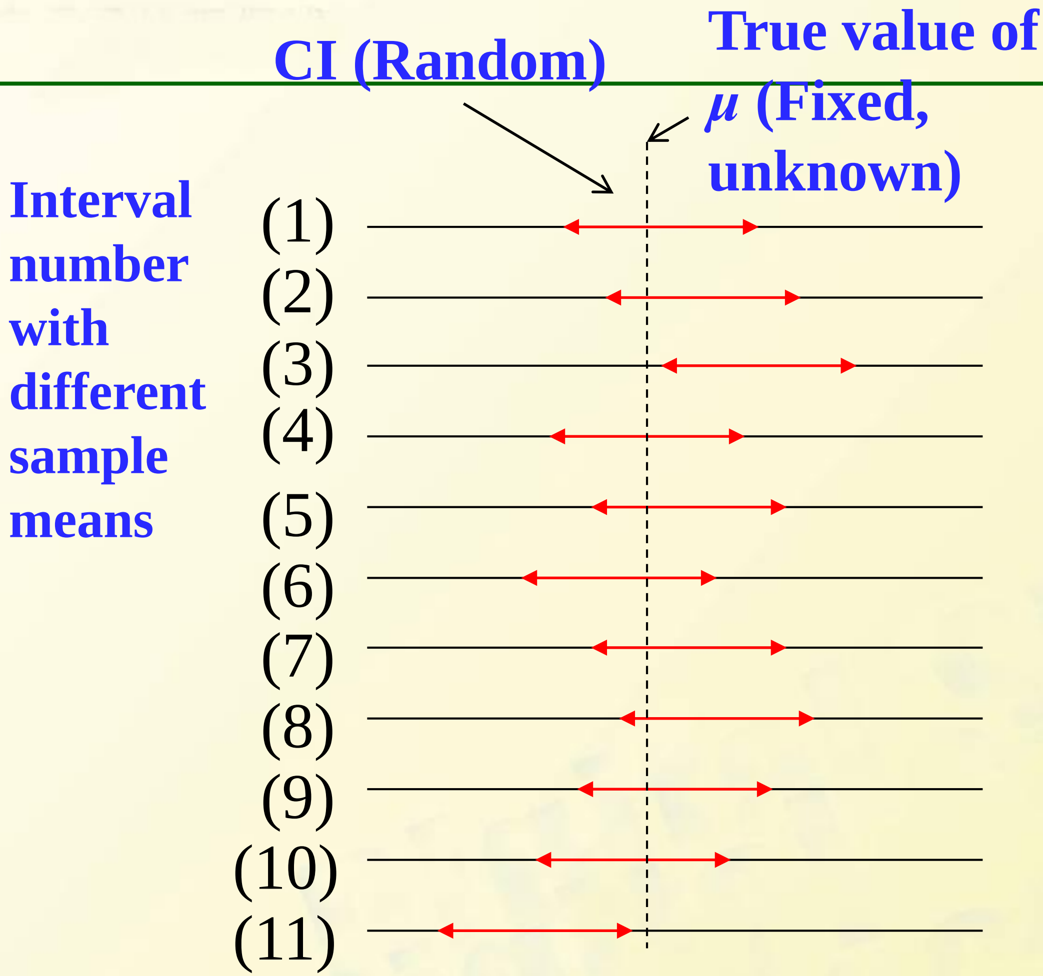


Fig 7.3 Repeated construction of 95% CIs

Notice that of the 11 interval pictured, only intervals 3 and 11 fail to contain ***μ. In the long run, 95% of the intervals will contain μ, only 5% of the intervals will fail to contain μ.***

7.1 Basic Properties of Confidence Intervals

- **Example 7.2 (Ex. 7.1 Cont')**

  The quantities needed for computation of the 95% CI for average preferred height are σ =2, n=31and $\bar{x}=80$.
  The resulting interval is

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 80.0 \pm (1.96)\frac{2.0}{\sqrt{31}} = 80.0 \pm .7 = (79.3, 80.7)$$

That is, we can be highly confident that 79.3 < μ < 80.7. This interval is relatively narrow, indicating that μ has been rather precisely estimated.

7.1 Basic Properties of Confidence Intervals

- **Definition**

If after observing $X_1=x_1$, $X_2=x_2$, … $X_n=x_n$, we compute the observed sample mean $\bar{x}$. The resulting fixed interval is called **a 95% confidence** interval for μ. This CI can be expressed either as

$$\left(\bar{x}-1.96\cdot\frac{\sigma}{\sqrt{n}},\bar{x}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right)$$

is a 95% CI for μ

or as $$\bar{x}-1.96\cdot\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\cdot\frac{\sigma}{\sqrt{n}}$$ with a 95% confidence

Lower Limit

Upper Limit

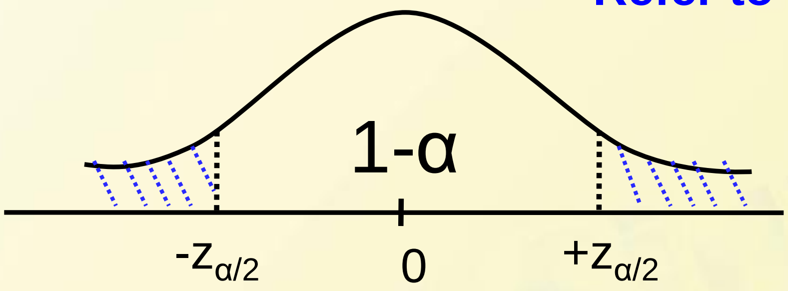
7.1 Basic Properties of Confidence Intervals

▪ Other Levels of Confidence

Why is Symmetry?
Refer to pp. 276 Ex.8

$P(a<z<b) = 1-\alpha$

Refer to pp.156 for the Definition Z_α



A 100(1- α)% confidence interval for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \text{ or, } \quad \bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

For instance, the 99% CI is $\bar{x} \pm 2.58 \cdot \sigma/\sqrt{n}$

7.1 Basic Properties of Confidence Intervals

- ***Example 7.3***

  Let's calculate a confidence interval for true average hole diameter using a confidence level of 90%.

  This requires that $100(1-\alpha) = 90$, from which $\alpha = 0.1$ and $z_{\alpha/2} = z_{0.05} = 1.645$. The desired interval is then

$$5.426 \pm (1.645) \cdot \frac{0.100}{\sqrt{40}} = 5.426 \pm 0.26 = (5.400, 5.452)$$

7.1 Basic Properties of Confidence Intervals

- **Confidence Level, Precision, and Choice of Sample Size**

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Then the width (Precision) of the CI

$$w = 2 \times z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Independent of the sample mean

Higher confidence level (larger $z_{\alpha/2}$) ➔ A wider interval

Reliability ⟺ **Precision**

Larger σ ➔ A wider interval

Smaller n ➔ A wider interval

Given a desired confidence level (α) and interval width (w), then we can determine the necessary sample size n, by

$$n = \left(2 z_{a/2} \cdot \frac{\sigma}{w}\right)^2$$

7.1 Basic Properties of Confidence Intervals

- **Example 7.4**

  Extensive monitoring of a computer time-sharing system has suggested that response time to a particular editing command is normally distributed with standard deviation 25 millisec. A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment.

  Assuming that response times are still normally distributed with $\sigma = 25$, what sample size is necessary to ensure that the resulting 95% CI has a width of no more than 10?

Solution:

The sample size *n* must satisfy

$$10 = 2 \cdot (1.96)\left(25/\sqrt{n}\right)$$

$$\Rightarrow \quad \sqrt{n} = 2 \cdot (1.96)(25)/10 = 9.80$$

$$\Rightarrow \quad n = (9.80)^2 = 96.04$$

Since *n* must be an integer, a sample size of 97 is required.

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

- The CI for μ given in the previous section assumed that the population distribution is normal and that the value of σ is known. We now present a **large-sample** CI whose validity does not require these assumptions.

- Let X_1, X_2, … X_n be a random sample from a population having a mean μ and standard deviation σ (**any population, normal or un-normal**).

  Provided that n is large (Large-Sample), the Central Limit Theorem (CLT) implies that $\overline{X}$ has approximately a normal distribution whatever the nature of the population distribution.

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \quad \Rightarrow P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}) \approx 1 - \alpha$$

Therefore, $\bar{x} \pm z_{\alpha/2} \cdot \sigma / \sqrt{n}$ is a large-sample CI for μ with a confidence level of **approximately** $100(1-\alpha)\%$.

That is , **when n is large, the CI for μ given previously remains valid whatever the population distribution, provided that the qualifier "approximately" is inserted in front of the confidence level.**

When σ is not known, which is generally the case, we may consider the following standardized variable

$$Z = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

$S \approx \sigma$

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

- **Proposition**

If n is sufficiently large (**usually, n>40**), the standardized variable

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has approximately a standard normal distribution, meaning that

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Compared with (7.5) in pp.272

$s \approx \delta$

is a large-sample confidence interval for μ with confidence level approximately 100(1-α)%.

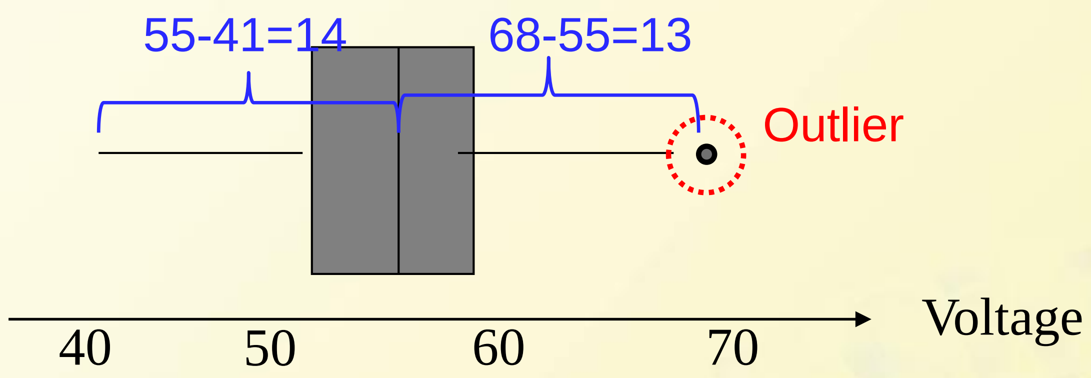
Note: This formula is valid regardless of the shape of the population distribution.

- **Example 7.6**

The alternating-current breakdown voltage of an insulating liquid indicates its dielectric strength. The article "test practices for the AC breakdown voltage testing of insulation liquids," gave the accompanying sample observations on breakdown voltage of a particular circuit under certain conditions.

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55
57	48	63	57	57	55	53	59	53	52	50	55	60	50	56	58

- Example 7.6 (Cont')



Summary quantities include

$$n = 48, \sum x_i = 2626, \sum x_i^2 = 144950 \Rightarrow \bar{x} = 54.7 \text{ and } s = 5.23$$

The 95% confidence interval is then

$$54.7 \pm 1.96 \frac{5.23}{\sqrt{48}} = 54.7 \pm 1.5 = (53.2, 56.2)$$

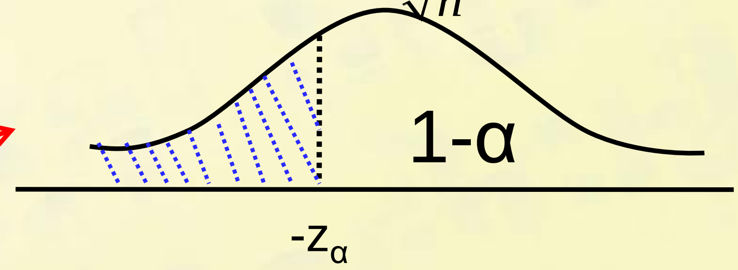
7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

- One-Sided Confidence Intervals **(Confidence Bounds)**

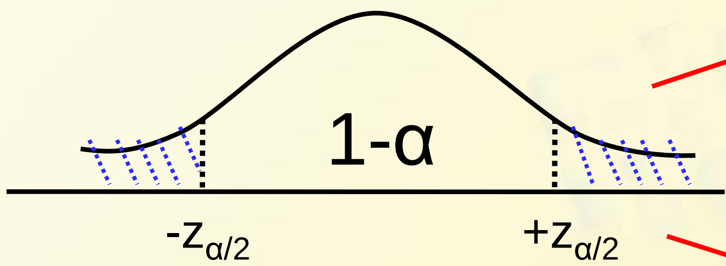
So far, the confidence intervals give both a lower confidence bound and an upper bound for the parameter being estimated.

In some cases, we will want only the upper confidence or the lower one.

$$\mu < \bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}} \ (upper\ \ bound)$$

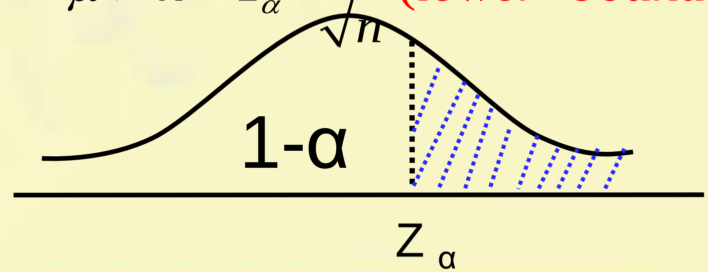


$$\mu \in (\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}})$$



Standard Normal Curve

$$\mu > \bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}} \ (lower\ \ bound)$$



7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

- Proposition

A large-sample upper confidence bound for μ is

$$\mu < \overline{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

and a large-sample lower confidence bound for μ is

$$\mu > \overline{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

Compared the formula (7.8) in pp.277

- **Example 7.10**

A sample of 48 shear strength observations gave a sample mean strength of 17.17 N/mm^2 and a sample standard deviation of 3.28 N/mm^2.

Then A lower confidence bound for true average shear strength μ with confidence level 95% is

$$17.17-(1.645)\frac{(3.28)}{\sqrt{48}}=17.17-0.78=16.39$$

Namely, with a confidence level of 95%, the value of μ lies in the interval (16.39, ∞).

7.3 Intervals Based on a Normal Population Distribution

7.3 Intervals Based on a Normal Population Distribution

- The CI for μ presented in the previous section is valid provided that n is large. The resulting interval can be used whatever the nature of the population distribution (with unknown μ and σ).
- **If n is small, the CLT can not be invoked. In this case we should make a specific assumption**.
- Assumption

  The population of interest is normal, $X_1, X_2, \ldots X_n$ constitutes a random sample from a normal distribution with both μ and σ unknown.

- **Theorem**

When $\overline{X}$ is the mean of a random sample of size n from a normal distribution with mean μ. Then the rv

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

has a probability distribution called a t distribution with n-1 degrees of freedom (df) .

only n-1 of these are "freely determined"

S is based on the **n** deviations $(X_1 - \overline{X}), (X_2 - \overline{X}), \ldots, (X_n - \overline{X})$

Notice that $\sum_{i=1}^{n} (X_i - \overline{X}) = 0$

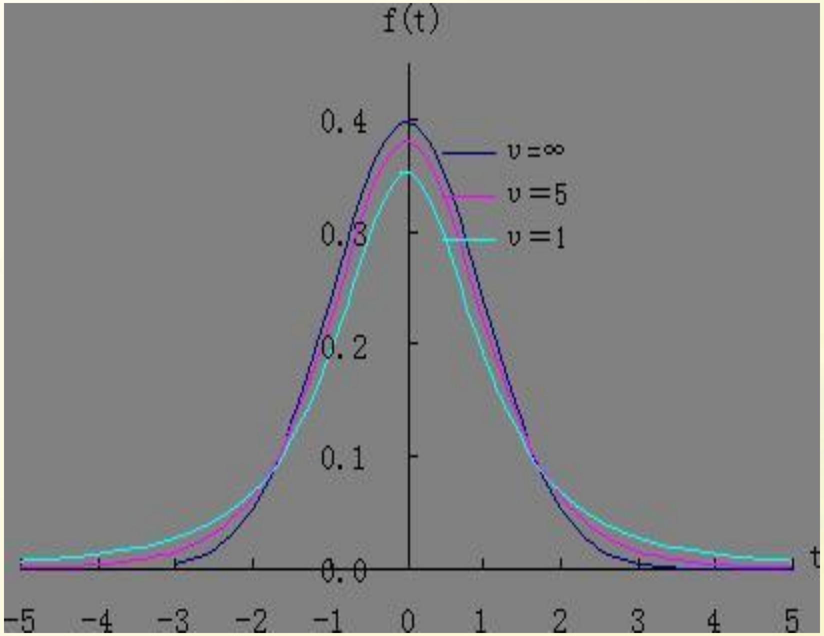
7.3 Intervals Based on a Normal Population Distribution

- Properties of *t* Distributions

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

The only one parameter in T is the number of df: v=n-1

Let t_v be the density function curve for v df



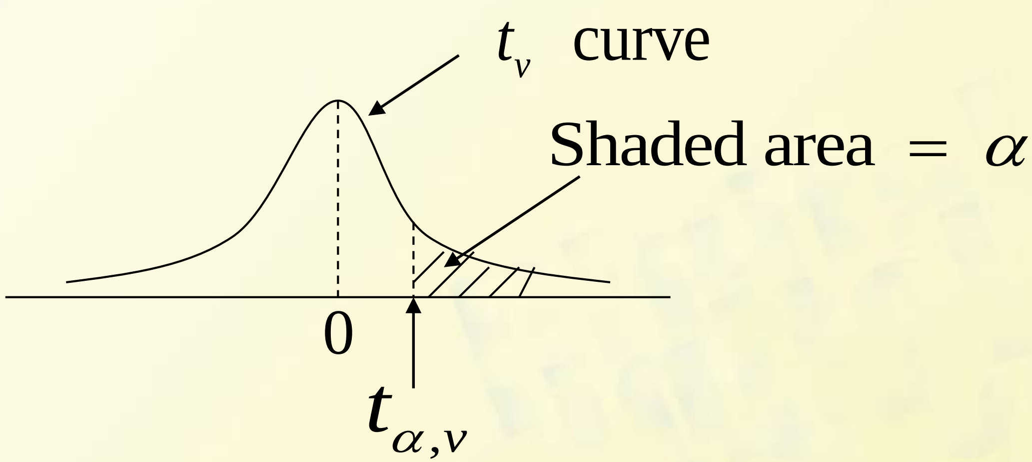
1. Each t_v curve is bell-shaped and centered at 0.
2. Each t_v curve is more spread out than the standard normal curve.
3. As v increases, the spread of the corresponding t_v curve decreases.
4. As v→ ∞, the sequence of t_v curves approaches the standard normal curve $N(0,1)$.

Rule: $v \geq 40 \sim N(0,1)$

7.3 Intervals Based on a Normal Population Distribution

- **Notation**

Let $t_{\alpha,v}$ = the value on the measurement axis for which the area under the t curve with v df to the right of $t_{\alpha,v}$ is α; $t_{\alpha,v}$ is called a ***t* critical value**



Fixed α, v↑, $t_{\alpha,v}$↓

Fixed v, α↑, $t_{\alpha,v}$↓

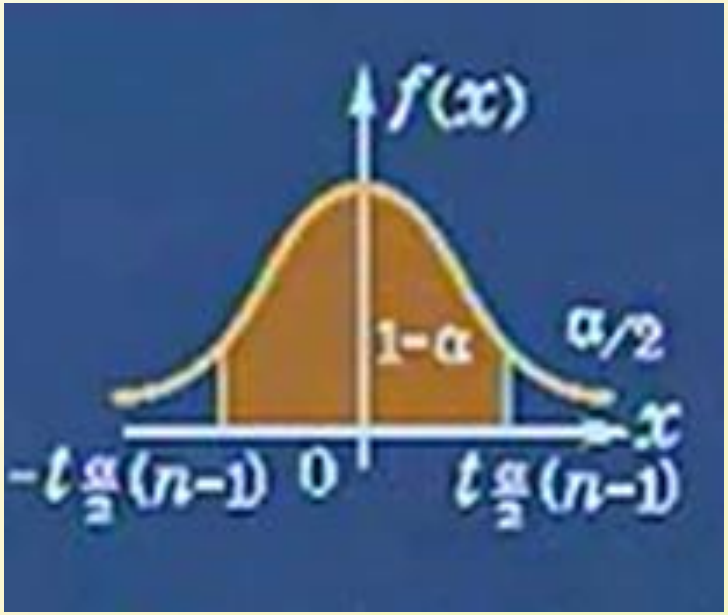
Figure 7.7 A pictorial definition of $t_{\alpha,v}$ **(Refer to: Appendix Table A.5)**

Refer to pp.156 for the similar definition of Z_α

▪ The One-Sample t confidence Interval

The standardized variable T has a t distribution with n-1 df, and the area under the corresponding t density curve between $-t_{\alpha/2,n-1}$ and $t_{\alpha/2,n-1}$ is $1-\alpha$, so

$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1-\alpha$$



Proposition

Let x and s be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a 100(1-α)% confidence interval for μ is

$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right)$$

Or, compactly $\bar{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$

An upper confidence bound with 100(1-α)% confidence level for μ is $\bar{x} + t_{\alpha,n-1} \cdot s/\sqrt{n}$. Replacing + by – gives a lower confidence bound for μ.

- Example 7.11

Consider the following observations

10490	**16620**	**17300**	**15480**	**12970**	**17260**	**13400**	**13900**
13630	**13260**	**14370**	**11700**	**15470**	**17840**	**14070**	**14760**

1. Approximately normal by observing the probability plot.
2. $n = 16$ is small, and the population deviation σ is unknown, so we choose the statistic T with a t distribution of $n - 1 = 15$ df. The resulting 95% CI is

$$\bar{x} \pm t_{.025,15} \cdot \frac{s}{\sqrt{n}} = 14,532.5 \pm (2.131) \frac{2055.67}{\sqrt{16}} = (13,437.3,\ 15,627.7)$$

- Example 7.12

Consider the following sample of fat content (in percentage) of n = 10 randomly selected hot dogs

25.2 21.3 22.8 17.0 29.8 21.0 25.5 16.0 20.9 19.5

Assume that these were selected from a normal population distribution.

Please give a 95% **CI** for the population mean fat content.

$$\bar{x} \pm t_{.025,9} \cdot \frac{s}{\sqrt{n}} = 21.90 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = 21.90 \pm 2.96$$

$$= (18.94, 24.86)$$

Summary of Chapter 7

- General method for deriving CIs (2 properties, p.273)

Case #1: (7.1)

CI for μ of a normal distribution with known σ;

Case #2: (7.2)

Large-sample CIs for μ of General distributions with unknown σ

Case #3: (7.3)

Small-sample CIs for μ of Gaussian distributions with unknown σ

- Both Sided Vs. One-sided CIs (p.283)