Estimation of Means and Proportions

Small-Sample Estimation of a Population Mean

Dr.Mamatha.H.R
Professor
Department of Computer Science and Engineering
PES University
Bangalore

Course material created using various Internet resources and text book

Introduction

- When the sample size is small, the estimation procedures used for large samples are not appropriate.
- Point estimators remain the same
- There are small sample interval estimators/confidence intervals for
  - ✓ μ, the mean of a normal population
  - ✓ $\mu_1 - \mu_2$, the difference between two normal population means

The Sampling Distribution of the Sample Mean

- When we take a sample from a normal population, the sample mean $\bar{x}$ has a normal distribution for any sample size *n*, and

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

has a standard normal distribution.

- But if σ is unknown, and we must use *s* to estimate it, the resulting statistic **is not normal**.

$$\frac{\bar{x} - \mu}{s / \sqrt{n}} \text{ is not normal!}$$

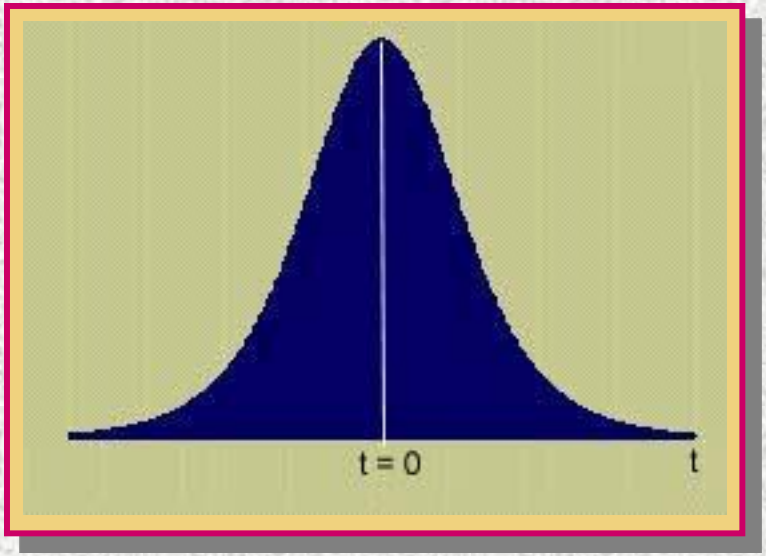
Student's t Distribution

- Fortunately, this statistic does have a sampling distribution that is well known to statisticians, called the **Student's t distribution**, with ***n*-1 degrees of freedom**.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

- We can use this distribution to create estimation procedures for the population mean μ.

Properties of Student's *t*



- **Mound-shaped** and symmetric about 0.
- **More variable than *z***, with "heavier tails"

- Shape depends on the sample size *n* or the **degrees of freedom, *n*-1**.
- As *n* increases the shapes of the *t* and *z* distributions become almost identical.

Summary

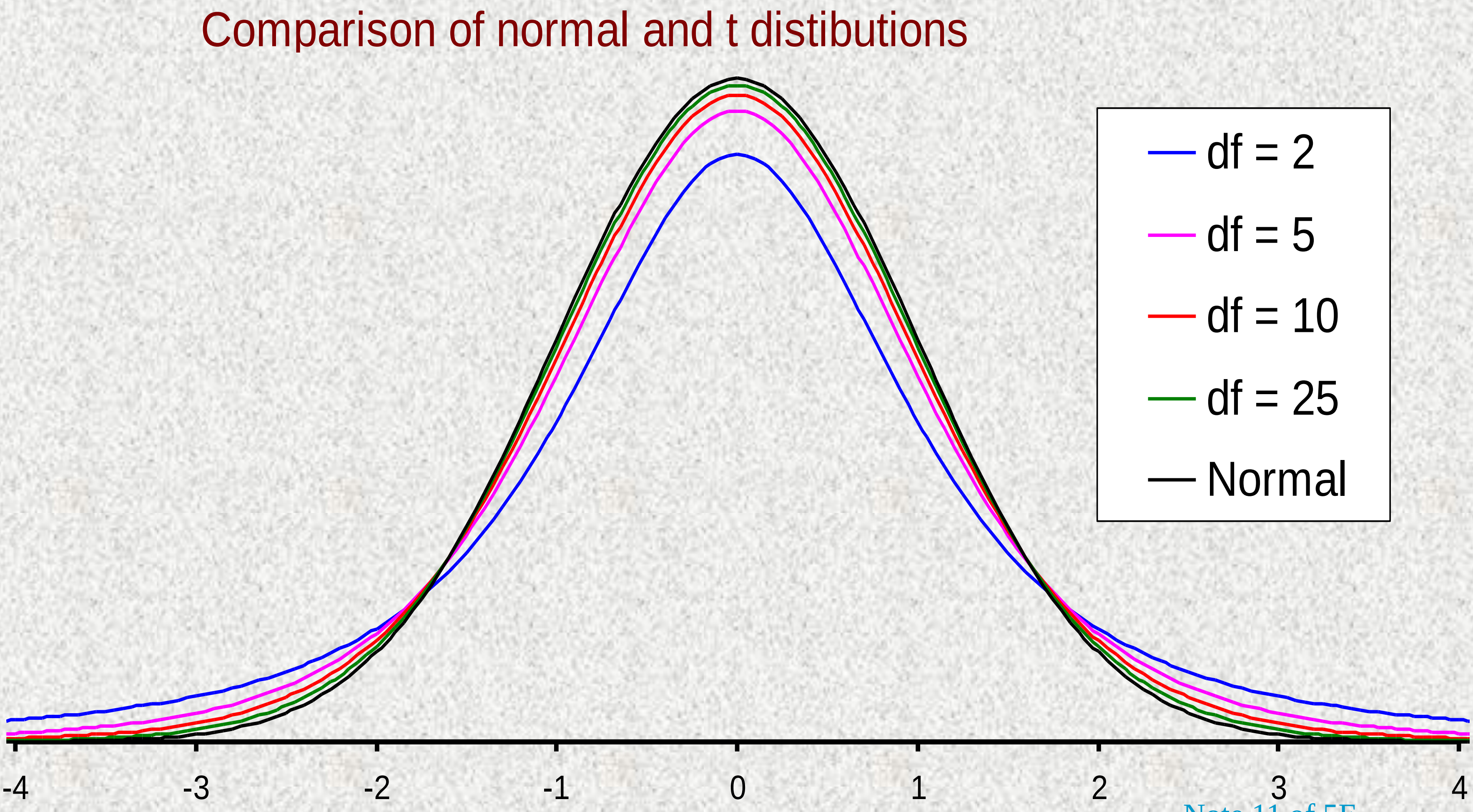
Let $X_1, \ldots, X_n$ be a *small* (e.g., $n < 30$) sample from a *normal* population with mean μ. Then the quantity

$$\frac{\overline{X} - \mu}{s/\sqrt{n}}$$

has a Student's t distribution with $n - 1$ degrees of freedom, denoted t_{n-1}.

When n is large, the distribution of the quantity $(\overline{X} - \mu)/(s/\sqrt{n})$ is very close to normal, so the normal curve can be used, rather than the Student's t.

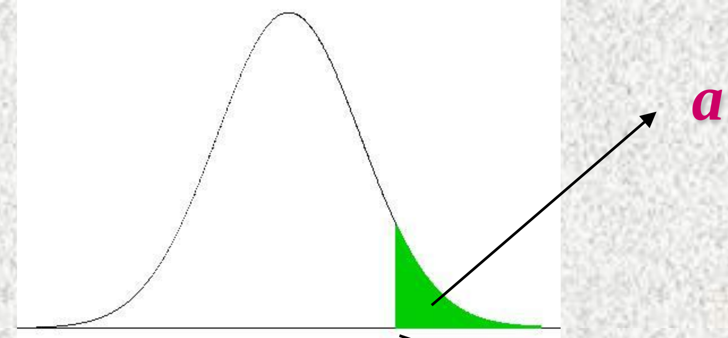
t distribution



RELATIONSHIP TO THE NORMAL CURVE

- As the df increase, the t distribution approaches the standard normal distribution (μ=0.0, σ=1.0).
- The standard normal curve is a special case of the t distribution when df= infinity.
- For practical purposes, the t distribution approaches the standard normal distribution relatively quickly, such that when df=30 the two are almost identical.
- The t-distribution is symmetric and bell-shaped, like the normal distribution, but has heavier tails, meaning that it is more prone to producing values that fall far from its mean.

Using the *t*-Table



- Table 4 gives the values of *t* that cut off certain critical values in the right tail of the *t* distribution.
- Use index ***df*** and the appropriate tail area ***a*** to find t_a, the value of *t* with area ***a*** to its right.

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602

For a random sample of size *n* = 10, find a value of *t* that cuts off .025 in the right tail.

Row = *df* = *n* –1 = 9

Column subscript = *a* = .025

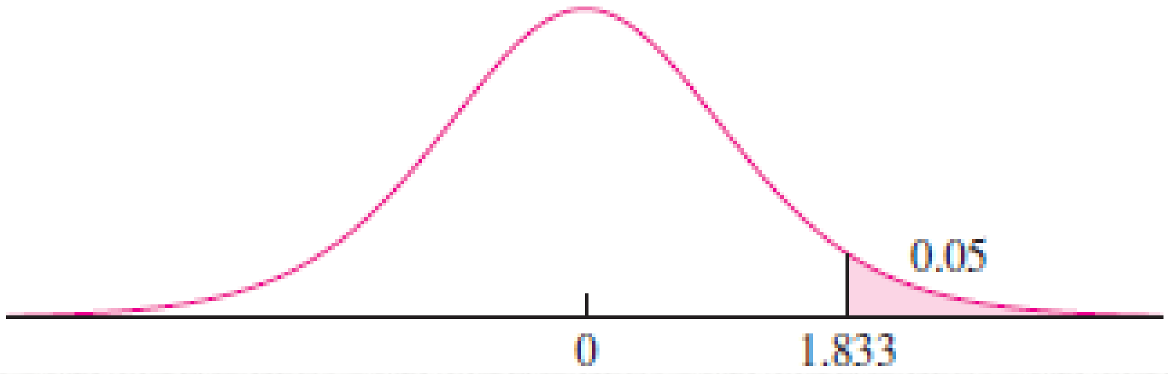
$t_{.025} = 2.262$

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602

A random sample of size 10 is to be drawn from a normal distribution with mean 4. The Student's t statistic $t = (\overline{X} - 4)/(s/\sqrt{10})$ is to be computed. What is the probability that $t > 1.833$?

Solution

This t statistic has $10 - 1 = 9$ degrees of freedom. From the t table, $P(t > 1.833) = 0.05$. See Figure 5.10.

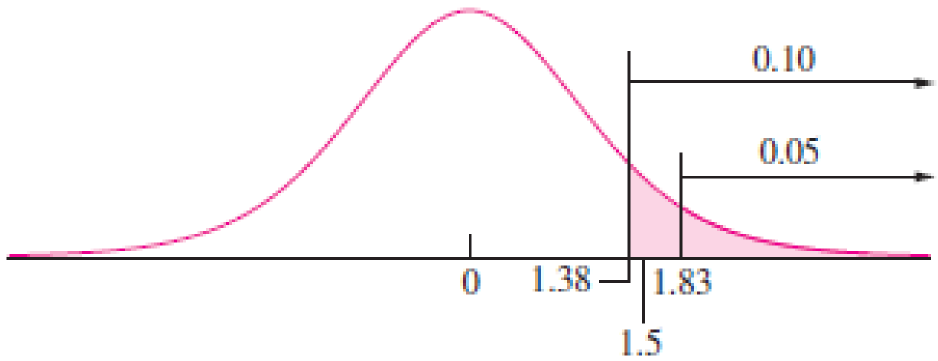


. Find $P(t > 1.5)$.

Solution

Looking across the row corresponding to 9 degrees of freedom, we see that the t table does not list the value 1.5. We find that $P(t > 1.383) = 0.10$ and $P(t > 1.833) = 0.05$. We conclude that $0.05 < P(t > 1.5) < 0.10$. See Figure 5.11. If a more precise result were required, linear interpolation could be used as follows:

$$P(t > 1.5) \approx 0.10 - \frac{1.5 - 1.383}{1.833 - 1.383}(0.10 - 0.05) = 0.0870$$



Find the value for the t_{12} distribution whose upper-tail probability is 0.025.

Solution

Look down the column headed "0.025" to the row corresponding to 12 degrees of freedom. The value for t_{12} is 2.179.

Find the value for the t_{14} distribution whose lower-tail probability is 0.01.

Solution

Look down the column headed "0.01" to the row corresponding to 14 degrees of freedom. The value for t_{14} is 2.624. This value cuts off an area, or probability, of 1% in the upper tail. The value whose lower-tail probability is 1% is −2.624.

Don't Use the Student's t Statistic If the Sample Contains Outliers

For the Student's t statistic to be valid, the sample must come from a population that is approximately normal. Such samples rarely contain outliers.

Therefore, methods involving the Student's t statistic should not be used for samples that contain outliers.

Small Sample Confidence Interval for Population Mean μ

Small - Sample $(1-\alpha)100\%$ confidence interval

of the population mean μ is

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

where $t_{\alpha/2}$ is the value of t that cuts off area $\alpha/2$

in the right tail of a t - distribution with $df = n - 1$.

Assumption: population must be normal

Summary

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1 - \alpha)\%$ confidence interval for μ is

$$\overline{X} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \tag{5.9}$$

Example

Ten randomly selected students were each asked to list how many hours of television they watched per month. The results are

82	66	90	84	75
88	80	94	110	91

Find a 90% confidence interval for the true mean number of hours of television watched per month by students.

Example Continued

Calculating the sample mean and standard deviation we have $n = 10,\ \bar{x} = 86,\ \text{and}\ s = 11.842$
We find the critical t value of 1.833 by looking on the t table in the row corresponding to df = 9, in the column with label $t_{.050}$. The 90% confidence interval for μ is

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 86 \pm (1.833)\frac{11.842}{\sqrt{10}} = 86 \pm 6.86$$

$$(79.14,\ 92.86)$$

How Do I Determine Whether the Student's *t* Distribution Is Appropriate?

- The Student's *t* distribution is appropriate whenever the sample comes from a population that is approximately normal.
- one must decide whether a population is approximately normal by examining the sample.
- Unfortunately, when the sample size is small, departures from normality may be hard to detect.

- A reasonable way to proceed is to construct a boxplot or dot plot of the sample.
- If these plots do not reveal a strong asymmetry or any outliers, then in most cases the Student's *t* distribution will be reliable.
- In principle, one can also determine whether a population is approximately normal by constructing a probability plot.
- With small samples, however, boxplots and dot plots are easier to draw.

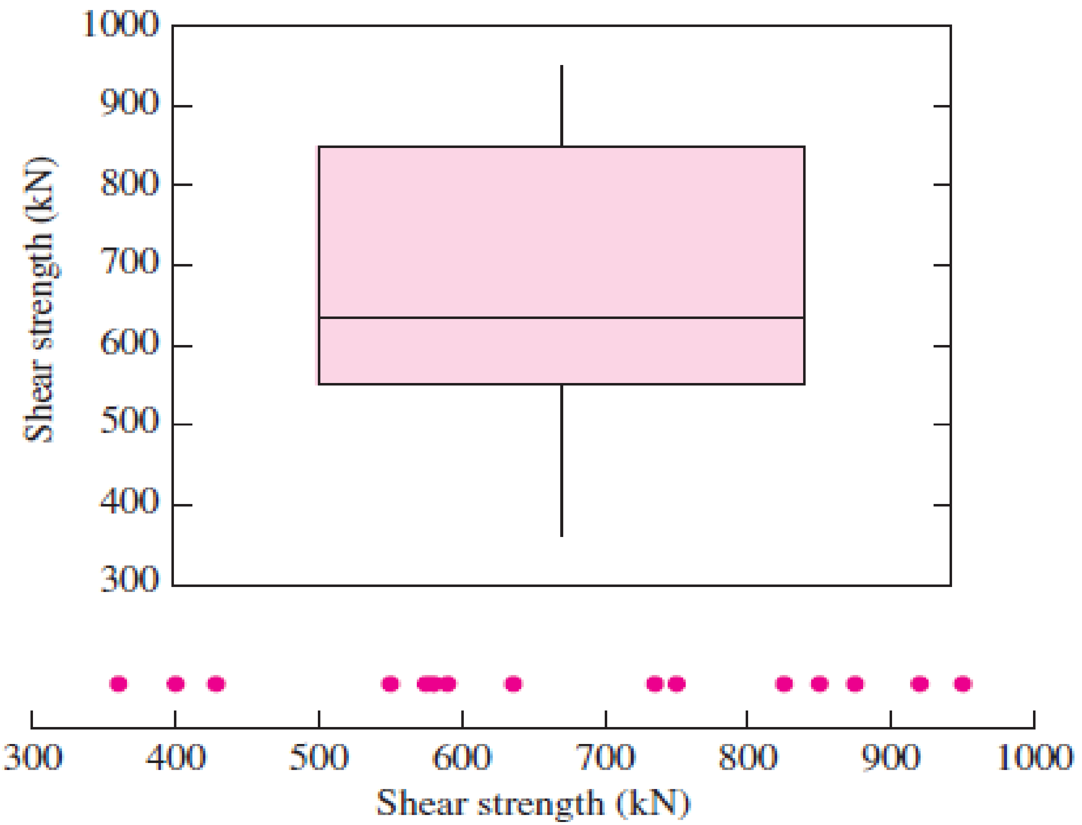
The article "Direct Strut-and-Tie Model for Prestressed Deep Beams" (K. Tan, K. Tong, and C. Tang, *Journal of Structural Engineering*, 2001:1076–1084) presents measurements of the nominal shear strength (in kN) for a sample of 15 prestressed concrete beams. The results are

580	400	428	825	850	875	920	550
575	750	636	360	590	735	950	

Is it appropriate to use the Student's t statistic to construct a 99% confidence interval for the mean shear strength? If so, construct the confidence interval. If not, explain why not.

Solution

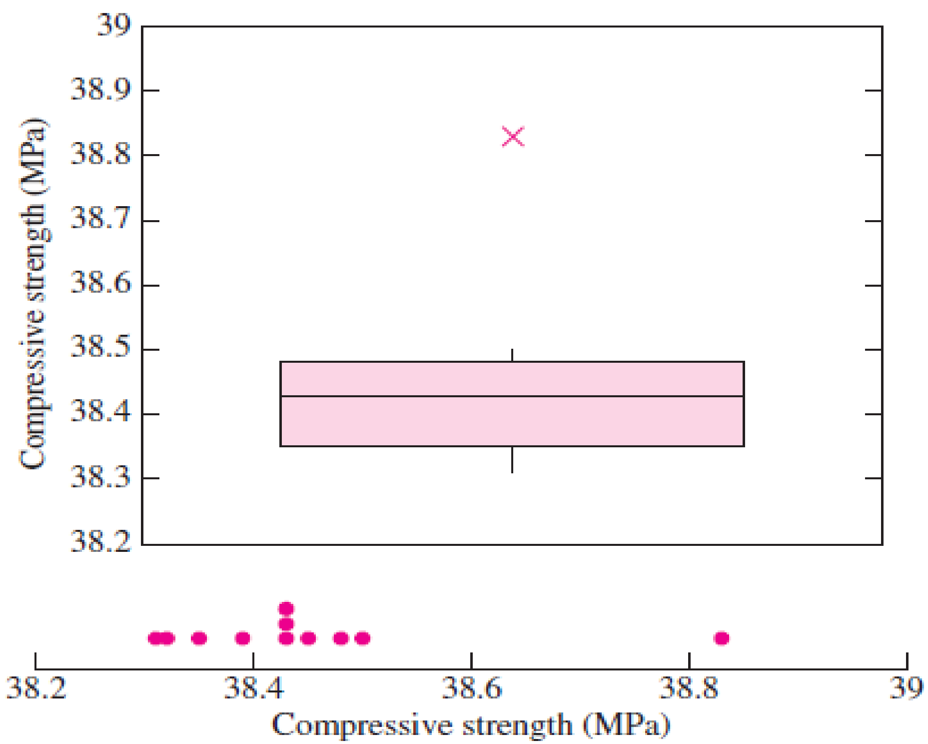
To determine whether the Student's t statistic is appropriate, we will make a boxplot and a dotplot of the sample. These are shown in the following figure.



In the article referred to in Example 5.19, cylindrical compressive strength (in MPa) was measured for 11 beams. The results were

38.43 38.43 38.39 38.83 38.45 38.35 38.43 38.31 38.32 38.48 38.50

Is it appropriate to use the Student's t statistic to construct a 95% confidence interval for the mean cylindrical compressive strength? If so, construct the confidence interval. If not, explain why not.



Note 11 of 5E

There is an outlier in this sample. The Student's t statistic should not be used.

An engineer reads a report that states that a sample of 11 concrete beams had an average compressive strength of 38.45 MPa with standard deviation 0.14 MPa. Should the t curve be used to find a confidence interval for the mean compressive strength?

Solution

No. The problem is that there is no way of knowing whether the measurements cam
from a normal population. For example, if the measurements contained an outlier (a
in Example 5.20), the confidence interval would be invalid.

It is important to remember that when the sample size is small, the population must be approximately normal, whether or not the standard deviation is known.

The Student's t distribution can be used to compute one-sided confidence intervals. The formulas are analogous to those used with large samples.

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1 - \alpha)\%$ upper confidence bound for μ is

$$\overline{X} + t_{n-1,\alpha}\frac{s}{\sqrt{n}} \tag{5.10}$$

and a level $100(1 - \alpha)\%$ lower confidence bound for μ is

$$\overline{X} - t_{n-1,\alpha}\frac{s}{\sqrt{n}} \tag{5.11}$$

Use z, Not t, If σ Is Known

Summary

Let $X_1, \ldots, X_n$ be a random sample (of any size) from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1-\alpha)\%$ confidence interval for μ is

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{5.12}$$

Summary

Let X be a single value sampled from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1-\alpha)\%$ confidence interval for μ is

$$X \pm z_{\alpha/2}\sigma \tag{5.13}$$

Refer to Example 5.19. Assume that on the basis of a very large number of previous measurements of other beams, the population of shear strengths is known to be approximately normal, with standard deviation $\sigma = 180.0$ kN. Find a 99% confidence interval for the mean shear strength.

Solution

We compute $\overline{X} = 668.27$. We do not need to compute s, because we know the population standard deviation σ. Since we want a 99% confidence interval, $\alpha/2 = 0.005$. Because we know σ, we use $z_{\alpha/2} = z_{.005}$, rather than a Student's t value, to compute the confidence interval. From the z table, we obtain $z_{.005} = 2.58$. The confidence interval is $668.27 \pm (2.58)(180.0)/\sqrt{15}$, or $(548.36, 788.18)$.

Key Concepts

I. One Population Mean

For normal population, $(1-\alpha)100\%$ confidence interval of the population mean μ is

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

where $t_{\alpha/2}$ is the value of t that cuts off area $\alpha/2$ in the right tail of a t-distribution with $df = n-1$