Confidence Intervals for the Mean when \( \sigma \) is unknown

If the sample size is small, the distribution as before must be at least approximately normal. If the sample size is large (generally 30 or bigger), the population need not be normal according to the central limit theorem.

To compensate for not knowing the population standard deviation, the \( t \)-distribution will be used

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

and where \( n \) is the sample size, \( s \) is the sample standard deviation, and \( d.f. = n - 1 \).

The \( t \)-distribution is bell shaped, but has thicker tails than the standard normal distribution. However, as \( n \) increases, the \( t \)-distribution approaches the standard normal distribution.

A table of critical values for \( t \)-distribution’s can be found in the back cover of the text, or in Table 6, Appendix II.

Confidence intervals for \( \mu \) when \( \sigma \) is not known and the population is approximately normal or the sample size is large enough are

\[
\bar{x} - t_{c} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{c} \frac{s}{\sqrt{n}}
\]

where \( t_{c} \) is found on the table with \( d.f. = n - 1 \) (or closest \( d.f. \) smaller than \( n - 1 \)).

Example 1. (Section 8.2 # 17) Price to earnings ratios of stock from a random selection of 51 large company stocks are given (see your text). Using the formulas

\[
\bar{x} = \frac{\sum x}{n} \quad \text{and} \quad s = \sqrt{\frac{\sum x^2 - (\sum x)^2}{n - 1}}
\]

it was computed that \( \bar{x} = 25.2 \) and \( s \approx 15.5 \).

(a) Find a 90% confidence interval for the P/E population mean for all large U.S. companies.

(b) Repeat (a) with a 99% confidence level.

(c) How do companies with P/E’s of 12, 72 and 24 compare to the population mean?

(d) Is it necessary to assume that \( x \) is approximately normal for this case?

Answer. (a) For this case \( \bar{x} = 25.2, s \approx 15.5, \) and \( n = 51 \). Therefore, \( d.f. = 50 \), and using the table we find that \( t_{.9} = 1.676 \). The endpoints of the confidence interval are

\[
25.2 \pm 1.676 \frac{15.5}{\sqrt{51}}
\]

and so the interval is 21.56 to 28.84.

(b) The only thing that changes is \( t_{c} = 2.678 \) in this case, so the interval is 19.39 to 31.01.

(c) 24 is probably close to the mean, but neither 72 nor 12 are in the 99% confidence interval, so we are very certain that those P/E ratios are higher and lower than the mean respectively.
(d) No, the central limit theorem says that the sampling distribution is approximately normal for sufficiently large \( n \).

**Example 2.** (Section 8.2 #11) Total calcium. A total calcium level below 6mg/dl is related to severe muscle spasms. Recently, the patients total calcium tests gave the following readings (in mg/dl).

\[
\begin{align*}
9.3 & \quad 8.8 & \quad 10.1 & \quad 8.9 & \quad 9.4 & \quad 9.8 & \quad 10.0 & \quad 9.9 & \quad 11.2 & \quad 12.1
\end{align*}
\]

(a) Verify that \( \bar{x} = 9.95 \) and \( s \approx 1.02 \). What formulas did you use?

(b) Find a 99.9\% confidence interval for the population mean of total calcium in this patients blood.

(c) Do you think the patient still has a calcium deficiency?

(d) What properties should the distribution possess for the confidence interval in (b) to be valid?

**Answer.** (a) Use the formulas

\[
\bar{x} = \frac{\sum x}{n} \quad \text{and} \quad s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}
\]

as you did in chapter 3.

(b) For this, \( n = 10 \), \( \bar{x} = 9.95 \), \( s \approx 1.02 \), \( d.f. = 9 \), and so \( t_{.999} = 4.781 \) and the confidence interval has endpoints

\[
9.95 \pm 4.781 \frac{1.02}{\sqrt{10}}
\]

which means the 99.9\% interval for the population mean is from 8.41 to 11.49.

(c) No, the lowest point of the confidence interval is 8.41, so we are very certain (99.9\%) that the mean is above 8.41 (and below 11.49).

(d) Because the sample size is small (less than 30), the original distribution must be normal, or at least approximately normal.