VANDERBILT
School *of* Engineering

CE 3300-01 – RISK, RELIABILITY, AND RESILIENCE ENGINEERING

Confidence Intervals

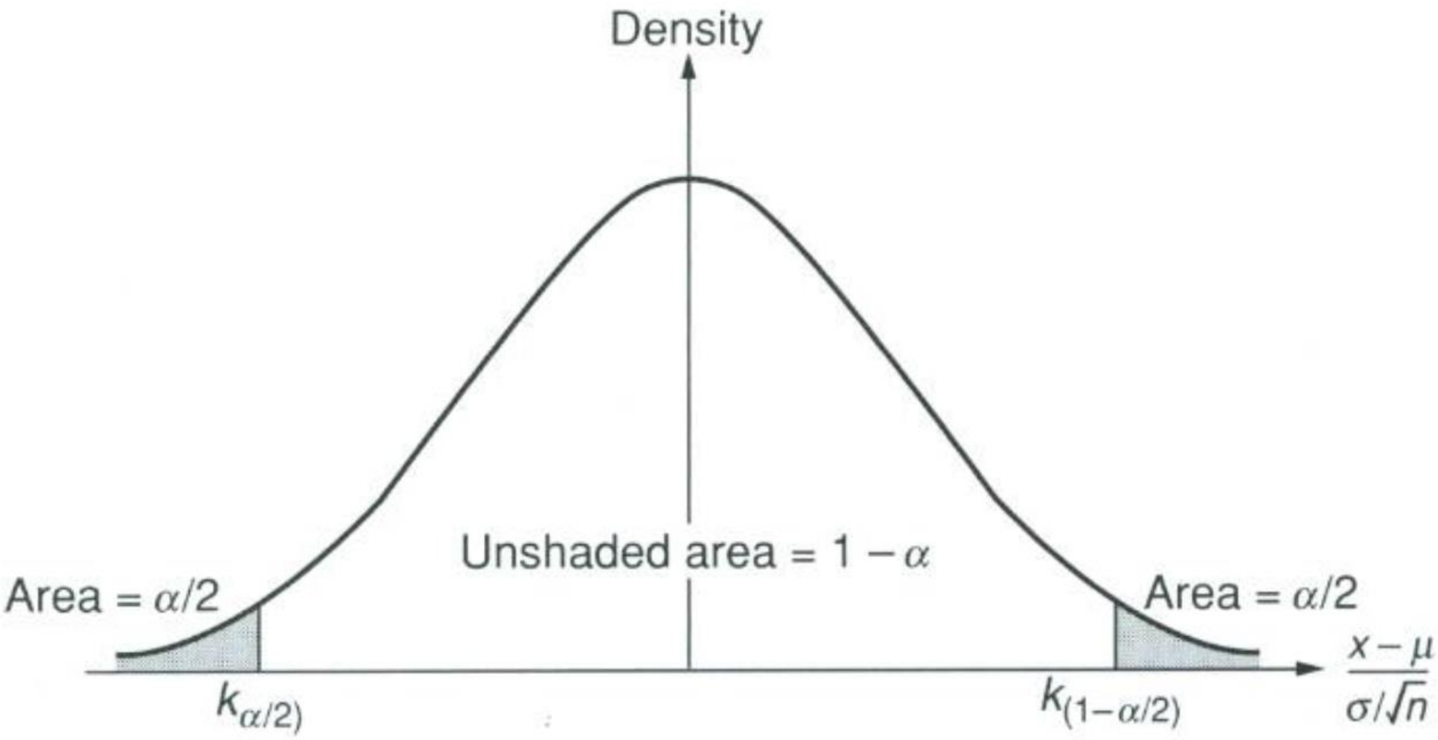
Book Reference: Chapters 6 – 6.2.1, 6.4.1, 6.4.2

TA: Mohamad Kazma

March 25th-26th, 2025

Instructor: Dr. Hiba Baroud

Today: Confidence Interval



Confidence Interval – Learning Objectives

- **Construct** confidence intervals
- **Interpret** results
- **Make inferences** about population parameters

Recall: How to estimate a population parameter?

Given a statistical model of a population, a **point estimate** is a value used to estimate a model parameter.

1. Select a random and representative sample.
2. Collect information from members of the sample.
3. Calculate the value of the sample statistic of interest.
4. Assign values to the population parameter.

How to “assign” values to the population parameter?

Point Estimate	Interval	Test
$\bar{x} \to \mu$ $x \sim \text{Bin}(n, p)$ $\bar{x} = np = \mu$	Point Estimate $\pm$ margin of error [,] Confidence Interval	$\mu > 70$? $\mu \neq 0$?

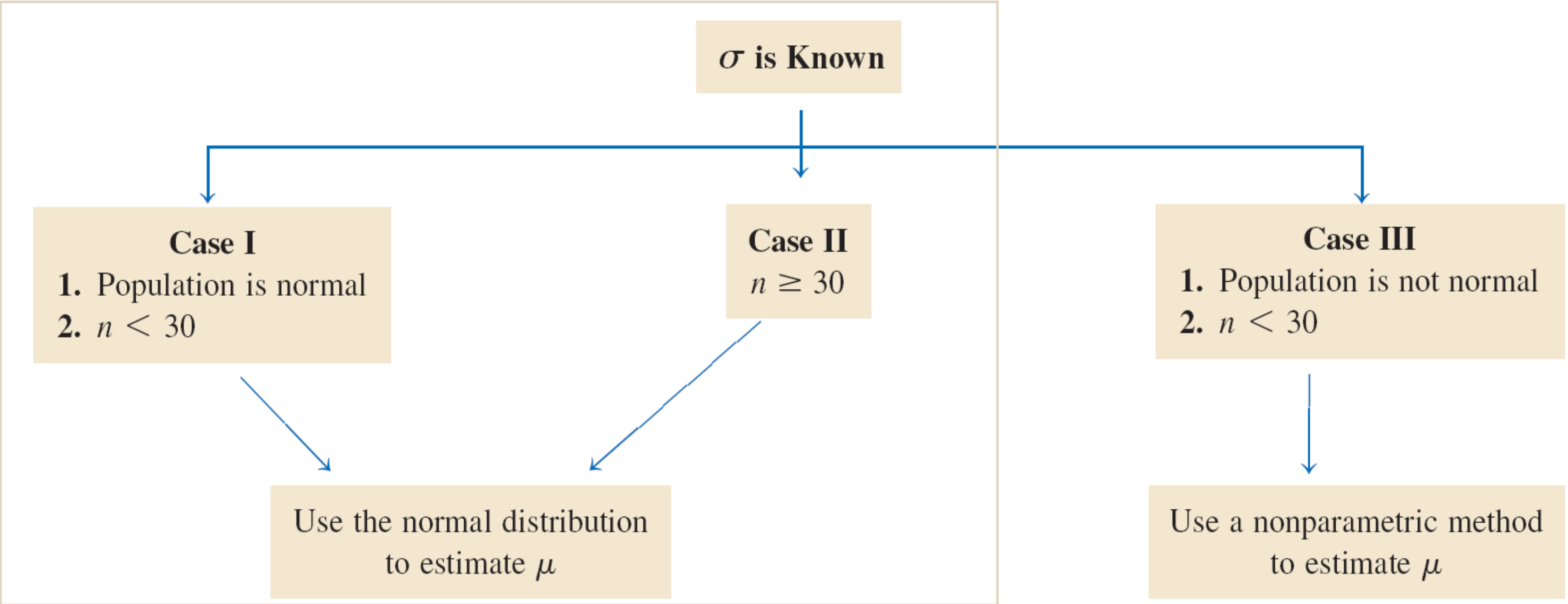
What is a Confidence Interval?

A **confidence interval** gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data with a specified level of confidence (e.g., 90%, 95%).

It quantifies the uncertainty in the estimate, via a confidence level α,
α is usually small
$\alpha = 0.05$ implies a 95% confidence interval
$\alpha = 0.10$ implies a 90% confidence interval

Confidence interval for the mean (μ)

When the population variance is known



$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Recall CLT:

$X \rightarrow$ Normal

$\bar{x} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

x is not normal make sure n is large

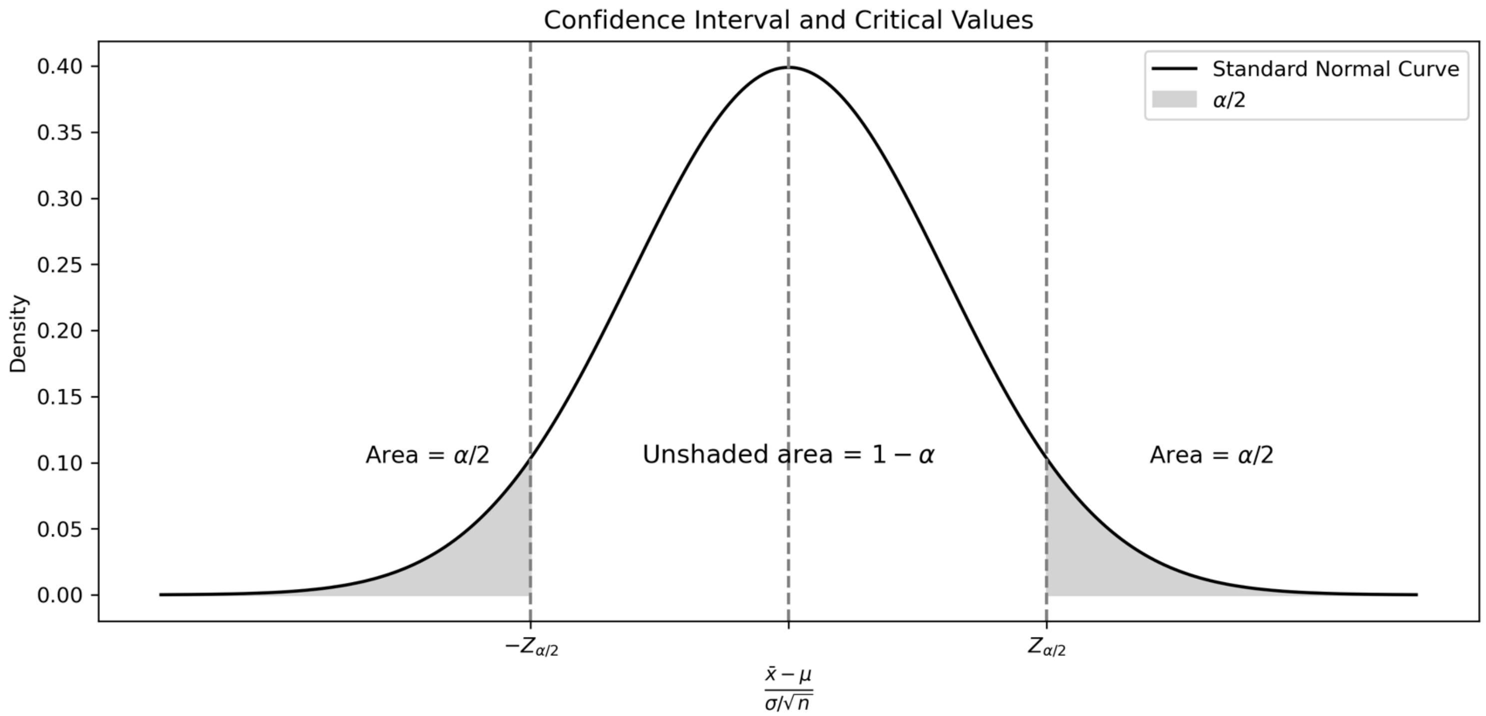
What is a Confidence Interval?

point estimate ± margin of error

Confidence level: $(1-\alpha)\%$
Significance level: $\alpha\%$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = (1-\alpha)\%$$



How to construct the confidence interval (μ)?

$\overline{\mathrm{X}} \pm Z_{\frac{\alpha}{2}} \sigma_{\overline{\mathrm{X}}}$ where $\sigma_{\overline{\mathrm{X}}} = \frac{\sigma}{\sqrt{n}}$

In other words:
The $(1-\alpha)\%$ confidence interval of μ is $\left(\overline{\mathrm{X}} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{\mathrm{X}} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$

Example 1 – Body Temperature

You read in a medical journal that the distribution of human body temperature has a standard deviation of 0.7°F.
You asked 42 friends to measure their temperature and found their mean temperature to be 98.2°F.

- What is the 90% confidence interval for the mean body temperature of the population?

$$\sigma = 0.7 \rightarrow \sigma \text{ is known}$$

$$\bar{x} = 98.2 \text{ and } n = 42$$

Stanford MEDICINE | News Center

FRONT PAGE ALL NEWS TOPICS MULTIMEDIA

Email Share 5.4K Tweet

Human body temperature has decreased in United States, study finds

Stanford researchers have determined that average human body temperature in the United States has decreased since the 1800s.

JAN 7 2020

Since the 19th century, the average human body temperature in the United States has dropped, according to researchers at the Stanford University School of Medicine.

"Our temperature's not what people think it is," said Julie Parsonnet, MD, professor of medicine and of health research and policy. "What everybody grew up learning, which is that our normal temperature is 98.6, is wrong."

That standard of 98.6 degrees Fahrenheit was made famous by German physician Carl Reinhold August Wunderlich, who published the figure in a book in 1868. Modern studies, however, have called that number into question, suggesting that it's too high. A recent study, for example, found the average temperature of 25,000 British patients to be 97.9 F.

Modern studies have called the "normal" human temperature of 37 degrees Celsius (or 98.6 degrees Fahrenheit) into question, suggesting that it's too high.
Dinga/Shutterstock.com

Example 1 – Body Temperature

You read in a medical journal that the distribution of human body temperature has a standard deviation of 0.7°F.
You asked 42 friends to measure their temperature and found their mean temperature to be 98.2°F.

- What is the 90% confidence interval for the mean body temperature of the population?

$$\sigma = 0.7 \rightarrow \sigma \text{ is known}$$

$$\bar{x} = 98.2 \text{ and } n = 42$$

$$P(z < z_{0.05}) = 0.95$$

$$\rightarrow z_{0.05} = 1.645$$

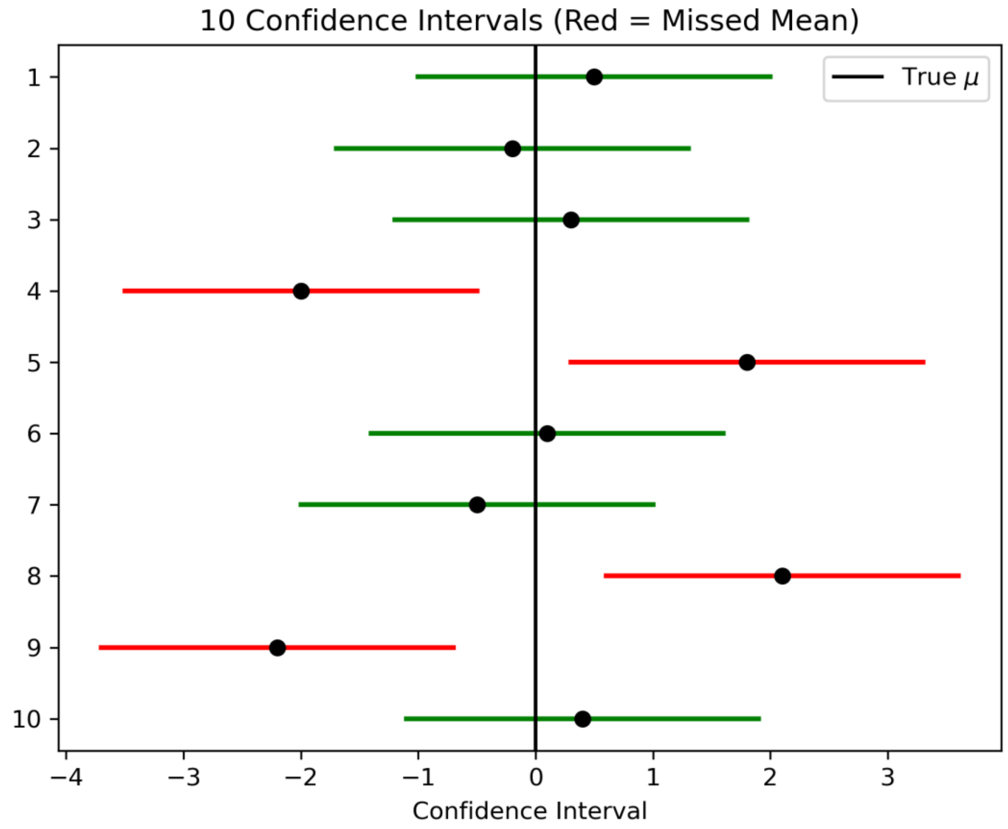
$$98.2 \pm z_{0.05} \times \frac{0.7}{\sqrt{42}}$$

$$CI = (98.022, 98.378)$$

How to Interpret Confidence Intervals?

- The interpretation is a <u>probabilistic statement about the interval</u> NOT the population parameter.
- The population mean <u>(μ) is a constant. There's no probability</u> associated with its value.

What does 90% confidence level mean?



Remember that we only collect <u>**one sample**</u> and construct <u>**one interval**</u>

Example 2 – T/F

$\bar{x}$=98.20 °F, CI=[98.022, 98.378]. Indicate whether the statement is true or false.

- The average human body temperature is 98.20 °F. False
- It's probably true that the average human body temperature is 98.20 °F. False
- We don't know exactly what the average human body temperature is but we know that it's between 98.0220 °F and 98.3780 °F. False
- We don't know exactly what the average human body temperature is but the interval from 98.0220 °F to 98.3780 °F probably contains the true average. True

Additional Material

Sample size

For a desired margin of error and a desired confidence level, what is a good sample size? (p. 267)

Sample Size

On some occasions, we may have

1. a desired margin of error
2. a desired confidence level for that margin of error

And we want to find out the sample size we should collect.

If E = margin of error, then

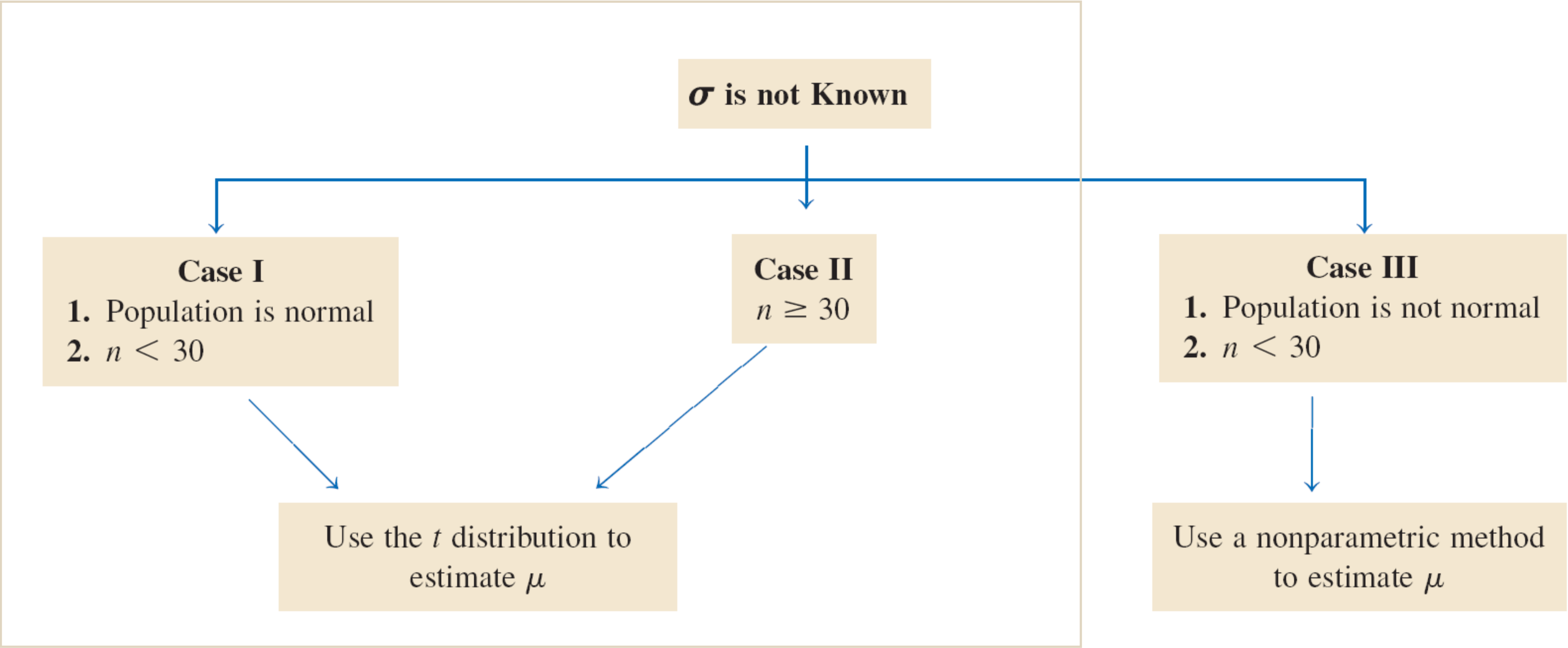
$$\bar{x} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{E} \quad \rightarrow \quad E = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

Additional Material

Confidence interval for the mean when σ is unknown
p. 263-264

Confidence interval for the mean (μ) - Continued

When the population variance is **unknown**



Confidence interval for the mean (μ) when σ is unknown

When the population variance is **unknown**

In most practical situations, the population variance may not be known. In such cases, we must also estimate the sample variance s^2, besides the sample mean $\bar{x}$, from a sample of n.

$$\bar{x} \pm t_{df,\alpha/2} s_{\bar{x}}$$

where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

$s =$ sample standard deviation

William Sealy Gossett

It was assumed that a good estimate of the population standard deviation was the sample standard deviation divided by the root of the sample size, $s/\sqrt{n}$.

A guy named William S. Gossett noticed something wrong with the standard errors (that is, $s/\sqrt{n}$) when sample sizes were small

t-distribution

Student's t distribution depends on a parameter called DEGREES OF FREEDOM

Degrees of freedom, denoted, $df = n - 1$

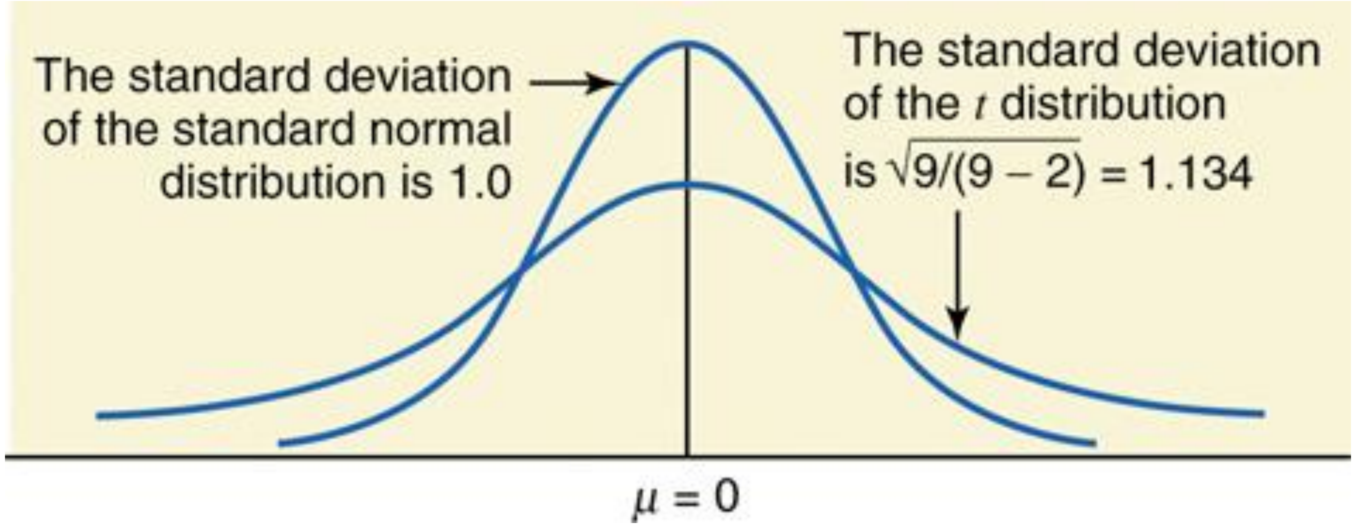
The value of t for a given value of degrees of freedom is denoted t_df

It's a much more practical distribution for drawing inference about means as, regardless of sample size, we don't have to know the population standard deviation and we can use the sample standard deviation

t-distribution

Student's *t* distribution is a specific bell-shaped distribution with a lower height and a wider spread than the standard normal distribution

As sample size becomes larger, the *t* distribution approaches the standard normal distribution.



t-distribution

Table A.3 in the textbook (p. 392)

1. $df = n - 1 = 30 - 1 = 29$

90% CI, $n = 30$

$$t_{29,\, 0.095} = 1.6991$$

2. $df = n - 1 = 25 - 1 = 24$

95% CI, $n = 25$

$$t_{24,\, 0.0975} = 2.0639$$

TABLE A.3 Critical Values of t-Distribution at Confidence Level $(1-\alpha) = p$

d.o.f.	$p = 0.900$	$p = 0.950$	$p = 0.975$	$p = 0.990$	$p = 0.995$	$p = 0.999$
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008
9	1.3803	1.8331	2.2622	2.8214	3.2498	4.2968
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	1.3634	1.7959	2.2001	2.7181	3.1058	4.0247
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852
31	1.3095	1.6955	2.0395	2.4528	2.7440	3.3749
32	1.3086	1.6939	2.0369	2.4487	2.7385	3.3653
33	1.3077	1.6924	2.0345	2.4448	2.7333	3.3563
34	1.3070	1.6909	2.0322	2.4411	2.7284	3.3479
35	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400
36	1.3055	1.6883	2.0281	2.4345	2.7195	3.3326
37	1.3049	1.6871	2.0262	2.4314	2.7154	3.3256
38	1.3042	1.6806	2.0244	2.4286	2.7116	3.3190
39	1.3036	1.6849	2.0227	2.4258	2.7079	3.3128
40	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069
45	1.3006	1.6794	2.0141	2.4121	2.6896	3.2815
50	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614
55	1.2971	1.6703	2.0040	2.3961	2.6682	3.2451
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317
70	1.2938	1.6794	1.9944	2.3808	2.6479	3.2108
80	1.2922	1.6759	1.9901	2.3739	2.6387	3.1953
90	1.2910	1.6750	1.9867	2.3685	2.6316	3.1833
∞	1.2824	1.6449	1.9600	2.3264	2.5759	3.0903

Summary: Confidence Intervals

- Population variance is known

$\rightarrow \overline{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\overline{X}}$

- Population variance is not known

$\rightarrow \overline{X} \pm t_{df,\frac{\alpha}{2}} S_{\overline{X}}$

- Interpretation of CI

$\rightarrow$ Probabilistic statement about the interval, not the population parameter

Example 3 – Body Temperature

In order to estimate the mean cholesterol level for all adult men in a town, a sample of 25 men was taken. The mean cholesterol level for this sample is 186 with a standard deviation of 12.

Assuming that the cholesterol level for all adult men is approximately normally distributed. Construct a 95% confidence interval for the population mean.

- Sample size, $n = 25$
- Sample mean, $\bar{x} = 186$
- Sample standard deviation, $s = 12$
- Confidence level $= 95\%$
- Population is approximately normally distributed

Example 3 – Body Temperature

- Sample size, $n = 25$
- Sample mean, $\bar{x} = 186$
- Sample standard deviation, $s = 12$
- Confidence level $= 95\%$
- Population is approximately normally distributed

•Degrees of freedom
$df = n - 1 = 24$
•From a t-distribution table
$t_{0.975,24} \approx 2.064$

$$\rightarrow s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{25}} = \frac{12}{5} = 2.4$$

$$\rightarrow \bar{x} \pm t_{df,\alpha/2} s_{\bar{x}} \rightarrow 186 \pm 2.0639 \times 2.4$$

$$\rightarrow \quad [181.05,\ 190.95]$$

Additional Material

- **Sample size**

For a desired margin of error and a desired confidence level, what is a good sample size? (p. 267)

- Confidence interval for the mean **when σ is unknown**

p. 263-264

- Confidence interval for the **proportion**

Chapter 6 – 6.4.2

Confidence interval for the proportion (p)

The confidence interval for the proportion p (the occurrence probability)

- CLT for proportions
  - $\hat{p} \sim Normal\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$
  - Large sample size: $n\hat{p} > 4\,, n(1-\hat{p}) > 4$

  - Confidence interval for the proportion
    - $\hat{p} \pm Z_{\frac{\alpha}{2}}\, S_{\hat{p}}$
    - $\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$