STAT 515 fa 2023 Lec 12 slides

Confidence interval for the mean when variance unknown

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Recall: If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then

$$\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

is a $(1 - \alpha) \times 100\%$ CI for μ.

But what if we don't know σ?

Using $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ is okay if n is large, but not if n is small. . .

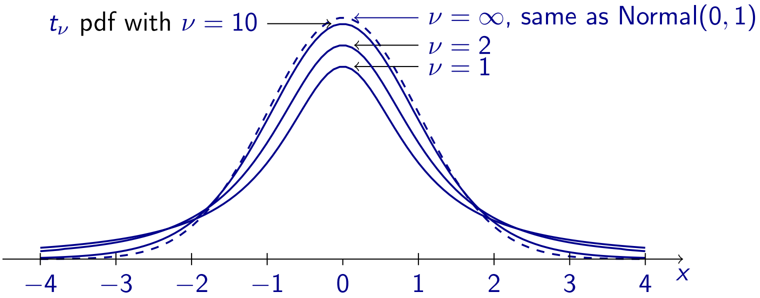
The t-distributions

The probability distribution with pdf given by

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\nu\pi\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \text{where } \Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du,$$

for $\nu > 0$ is called the *t-distribution* with *degrees of freedom ν*.

We write $T \sim t_\nu$ if a rv T has this distribution.



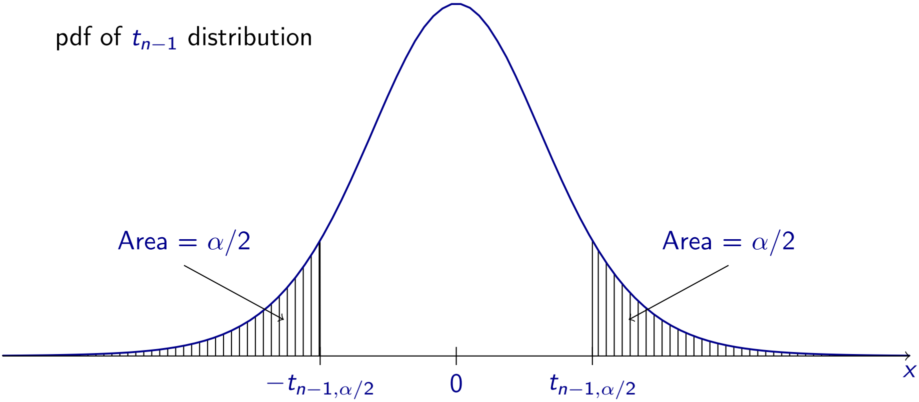
Sampling distribution of studentized mean

If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Exercise: Show above using this: If $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$ are ind. then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$



Can use function `qt()` or a t-table to look up the values, e.g.

$$t_{19,0.025} = \texttt{qt(.975,19)} = 2.093024$$
$$t_{19,0.005} = \texttt{qt(.995,19)} = 2.860935.$$

Confidence interval for mean of a Normal population with σ unknown

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$. Then a $(1-\alpha) \times 100\%$ CI for μ is

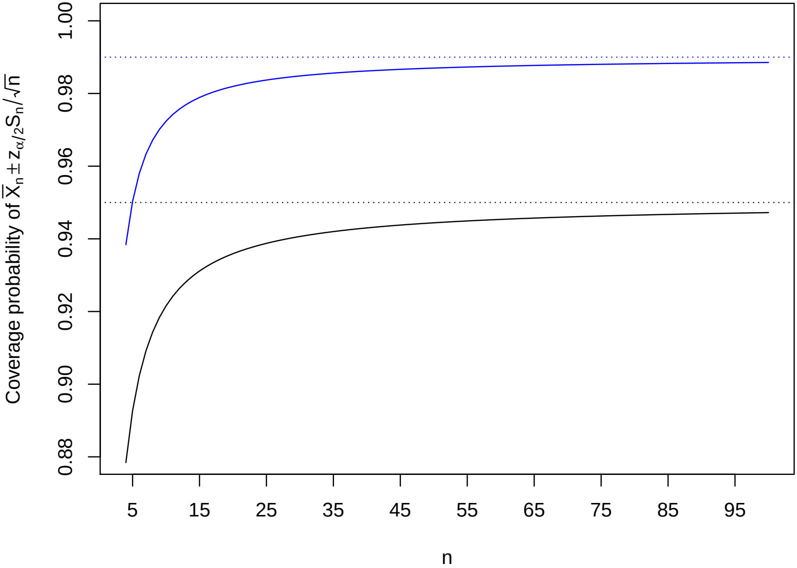
$$\bar{X}_n \pm t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}}.$$

Show where this CI comes from.

Exercise: These are the commute times (sec) to class of a sample of students.

1832	1440	1620	1362	577	934	928	998	1062	900
1380	913	654	878	172	773	1171	1574	900	900

1. Make a Q-Q plot to check Normality of the population.
2. Construct a 95% confidence interval for the mean commute time of all students.
3. Construct a 99% confidence interval for the mean commute time of all students.
4. What if the intervals $\bar{X}_n \pm z_{\alpha/2} \cdot S_n/\sqrt{n}$ are used? How are they different?


Coverage probability of $\overline{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$
0.88
0.90
0.92
0.94
0.96
0.98
1.00
5
15
25
35
45
55
65
75
85
95
n

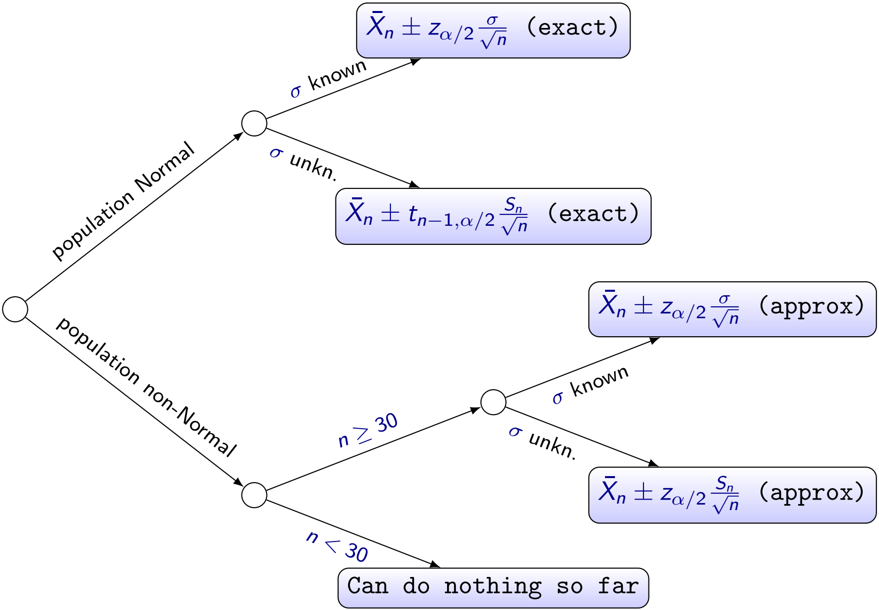
CI for mean of non-Normal population with σ unknown

Let $X_1, \dots, X_n$ be a rs from a pop. with mean μ, and with $\mu_4 < \infty$, then

$$\bar{X}_n \pm z_{\alpha/2} \cdot \frac{S_n}{\sqrt{n}}$$

is an approximate $(1-\alpha) \times 100\%$ CI for μ when n is large (≥ 30, say).

In the above $\mu_4 = \mathbb{E}|X_1|^4$. This limits the heavy-tailedness of the population.



$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (exact)
σ known
population Normal
σ unkn.
$\bar{X}_n \pm t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}$ (exact)
$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (approx)
population non-Normal
σ known
$n \geq 30$
σ unkn.
$\bar{X}_n \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}}$ (approx)
$n < 30$
Can do nothing so far