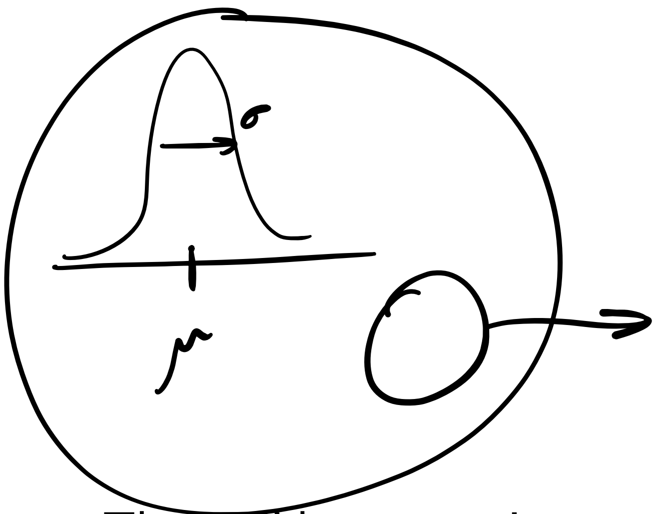
STAT 515 Lec 12 slides

Confidence interval for the mean when variance unknown

Pop: μ, σ^2

Karl Gregory

University of South Carolina



$X_1, \ldots, X_n$ $\quad \bar{X}_n$, S_n^2

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

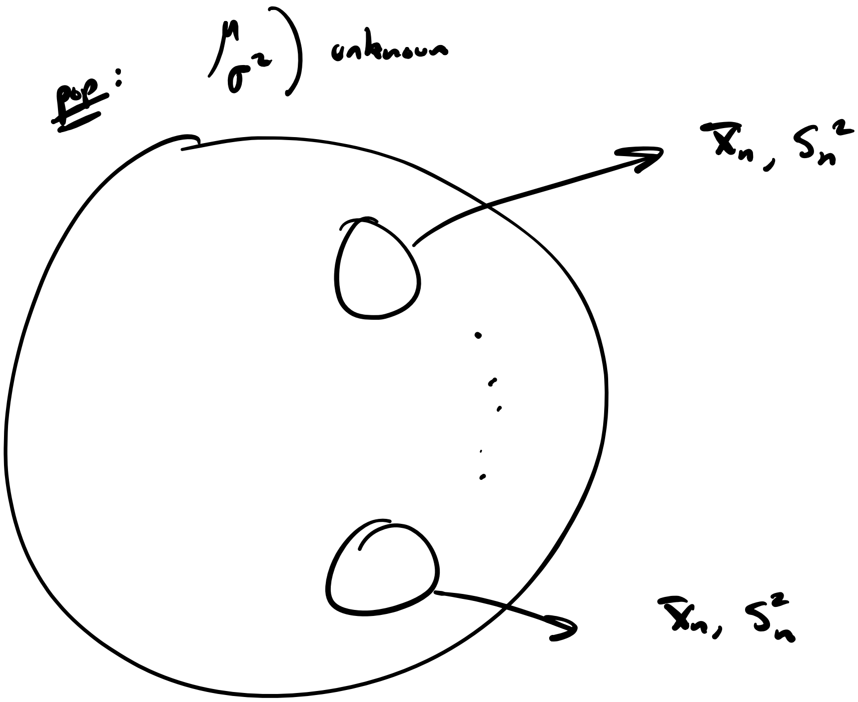
replace unknown σ with S_n??

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

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...



previously:

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

now, if σ unknown

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \not\sim N(0,1)$$

$$\sim t_{n-1}$$

"t dist. with degrees of freedom n-1."

Sampling distribution of "studentized" mean

If $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then $\dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$.

In the above, t_{n-1} represents the t-distribution with $n-1$ degrees of freedom...

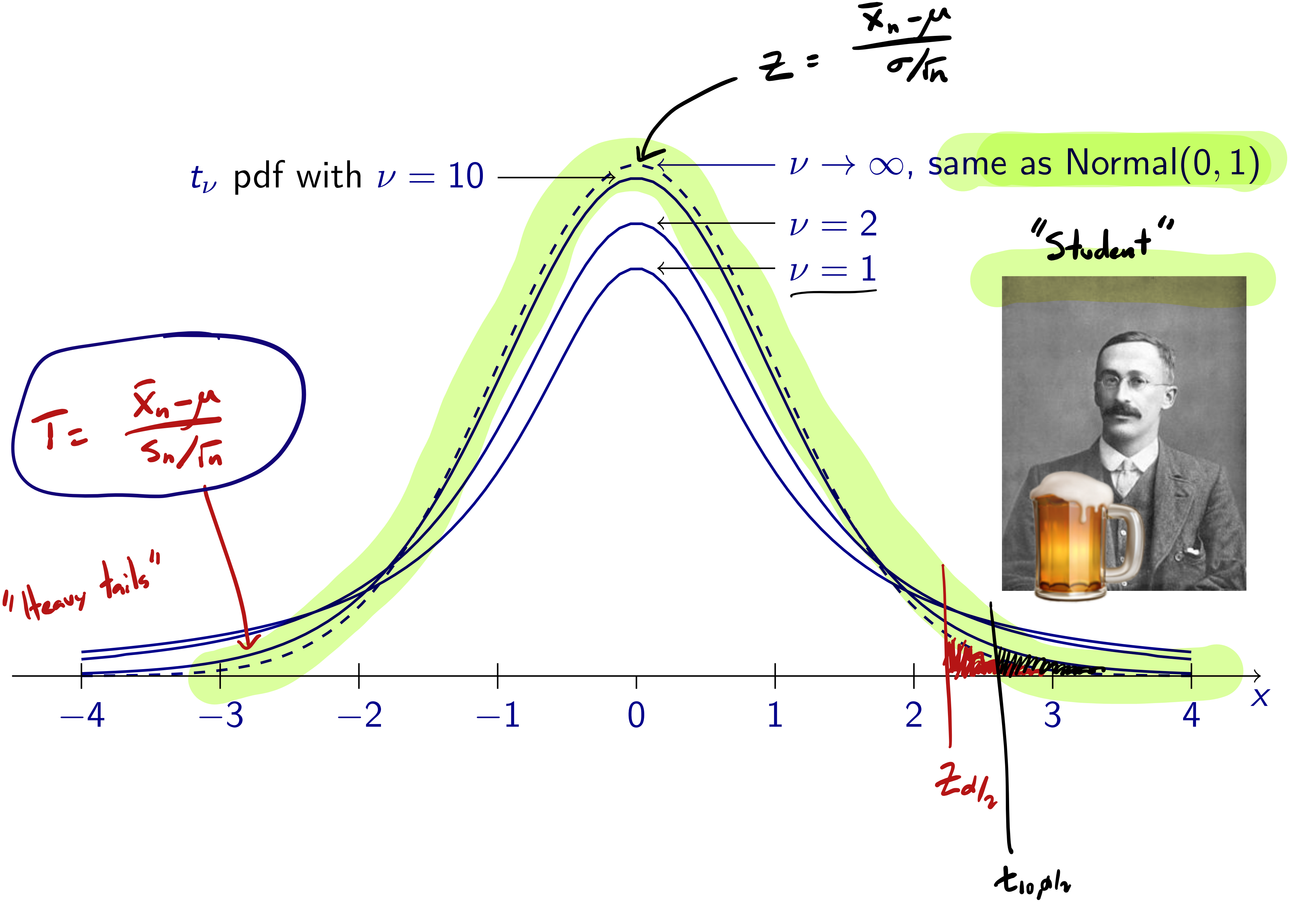
The t-distributions

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*.

For a random variable T with this distribution we write $T \sim t_\nu$.



$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$
t_ν pdf with $\nu = 10$
$\nu \to \infty$, same as Normal(0, 1)
$\nu = 2$
$\nu = 1$
"Student"
$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$
"Heavy tails"
−4
−3
−2
−1
0
1
2
3
4
x
$z_{\alpha/2}$
$t_{10,\alpha/2}$

How to build a t-distributed random variable.

Take $Z \sim N(0,1)$ and a $W \sim \chi^2_k$, independent

Then

$$T = \frac{Z}{\sqrt{W/k}} \sim t_k .$$

Why do we have $T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$?

Because

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \frac{\sigma}{S_n}$$

$$= \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S_n^2}{\sigma^2}}}$$

$$= \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2} \Big/ (n-1)}} \sim t_{n-1}$$

$\sim N(0,1)$

χ^2_{n-1}

How to build a t_ν random variable

If $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$ are independent rvs, then

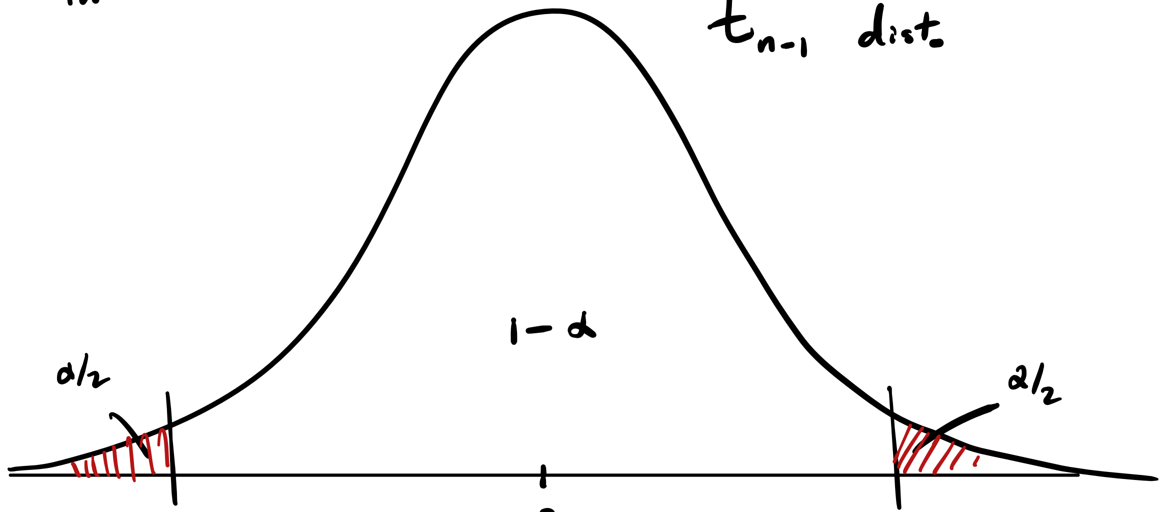
$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$

Discuss: From the above we can show that $\dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$.

σ known:

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

PDF of t_{n-1} dist.



$t_{n-1,1-\alpha/2}$

$=$

$-t_{n-1,\alpha/2}$

$t_{n-1,\alpha/2}$

$$\underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\text{"standardized"}}$$

Now we can write

$$P\left(-t_{n-1,\alpha/2} \le \underbrace{\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}}_{\text{"studentized"}} \le t_{n-1,\alpha/2} \right) = 1-\alpha$$

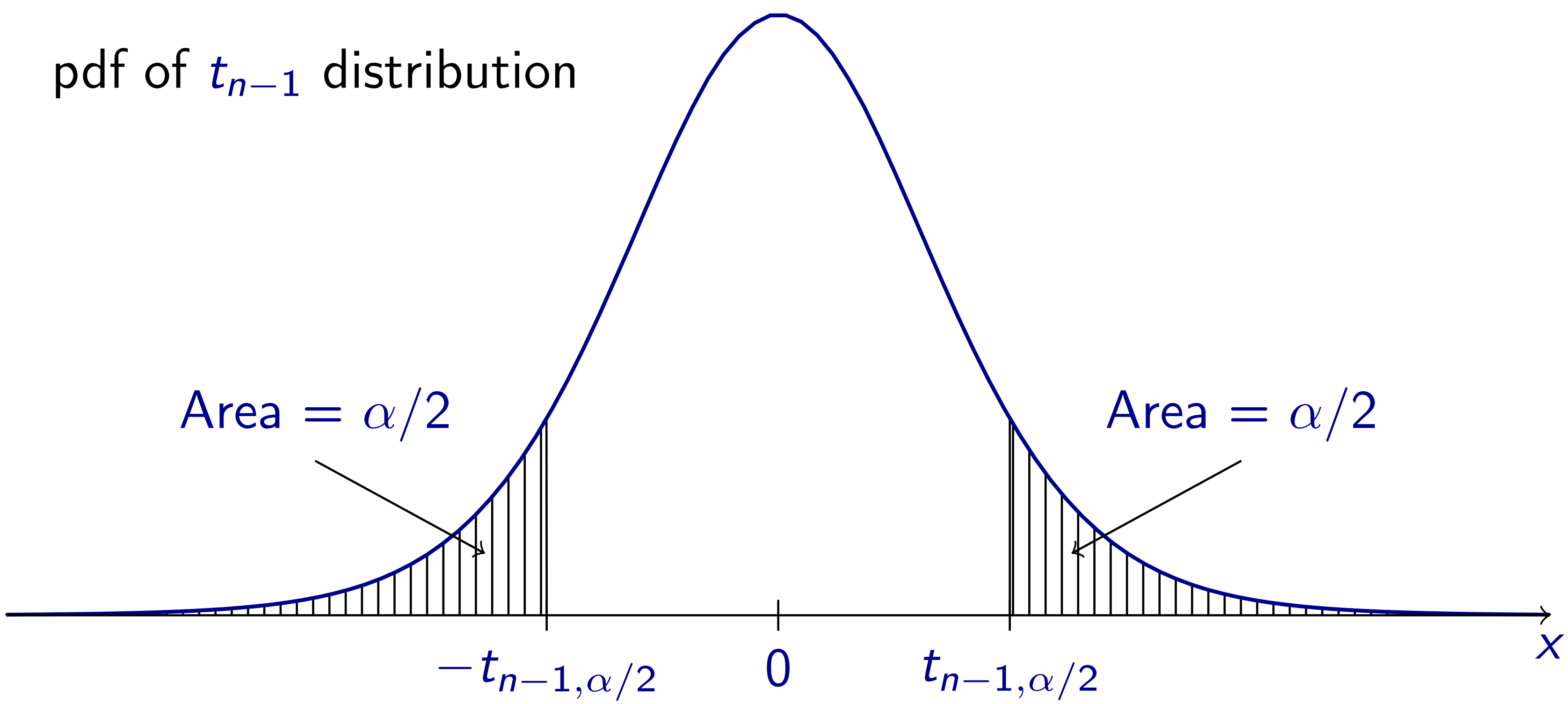
<=>

$$P\left(\bar{X}_n - t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}} \le \mu \le \bar{X}_n + t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}} \right) = 1-\alpha$$

Giving the $(1-\alpha)*100\%$ C.I. for μ with endpoints

$$\bar{X}_n \pm \underbrace{t_{n-1,\alpha/2}}_{\text{larger than } z_{\alpha/2}.} \frac{S_n}{\sqrt{n}}$$

Upper quantile notation for the t distributions:



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

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

σ known:

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

unknown

Confidence interval for mean of a Normal population with σ unknown

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

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

Discuss: How to obtain the above formula.

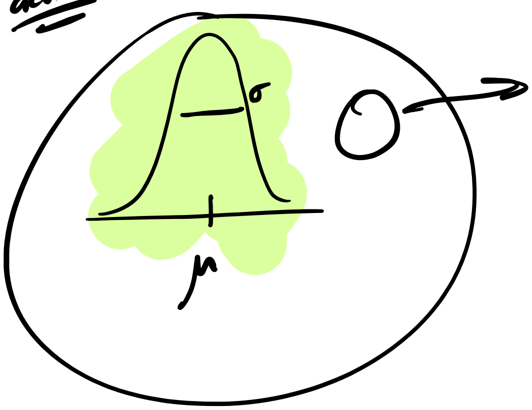
Exercise: These are the diameters of a sample of live oak acorns.

0.455	0.375	0.428	0.390	0.388	0.420	0.430	0.420	0.350
0.440	0.450	0.430	0.500	0.480	0.460	0.429	0.427	0.424
0.512	0.428	0.426	0.430	0.400	0.460	0.440	0.400	0.450
0.410	0.398	0.420	0.453	0.449	0.428			

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?

acorns

μ, σ^2 unknown



$\bar{x}_n, s_n^2$

$\boxed{\bar{x}_n = 0.4303}$

$s_n = 0.03299$

$n = 33$

$\alpha = 0.05$

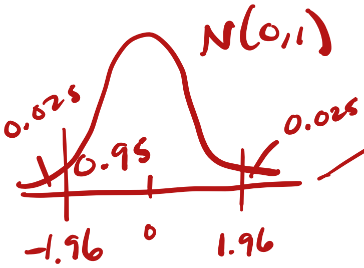
$$\bar{x}_n \pm t_{n-1,\alpha/2} \frac{s_n}{\sqrt{n}}$$

(2) 95%

$\boxed{t_{30,0.05} \approx t_{33-1,\, 0.025}.}$

$$0.4303 \pm \boxed{2.0423} \frac{0.03299}{\sqrt{33}} = [0.419,\ 0.442]$$

$t_{33-1,0.025}$



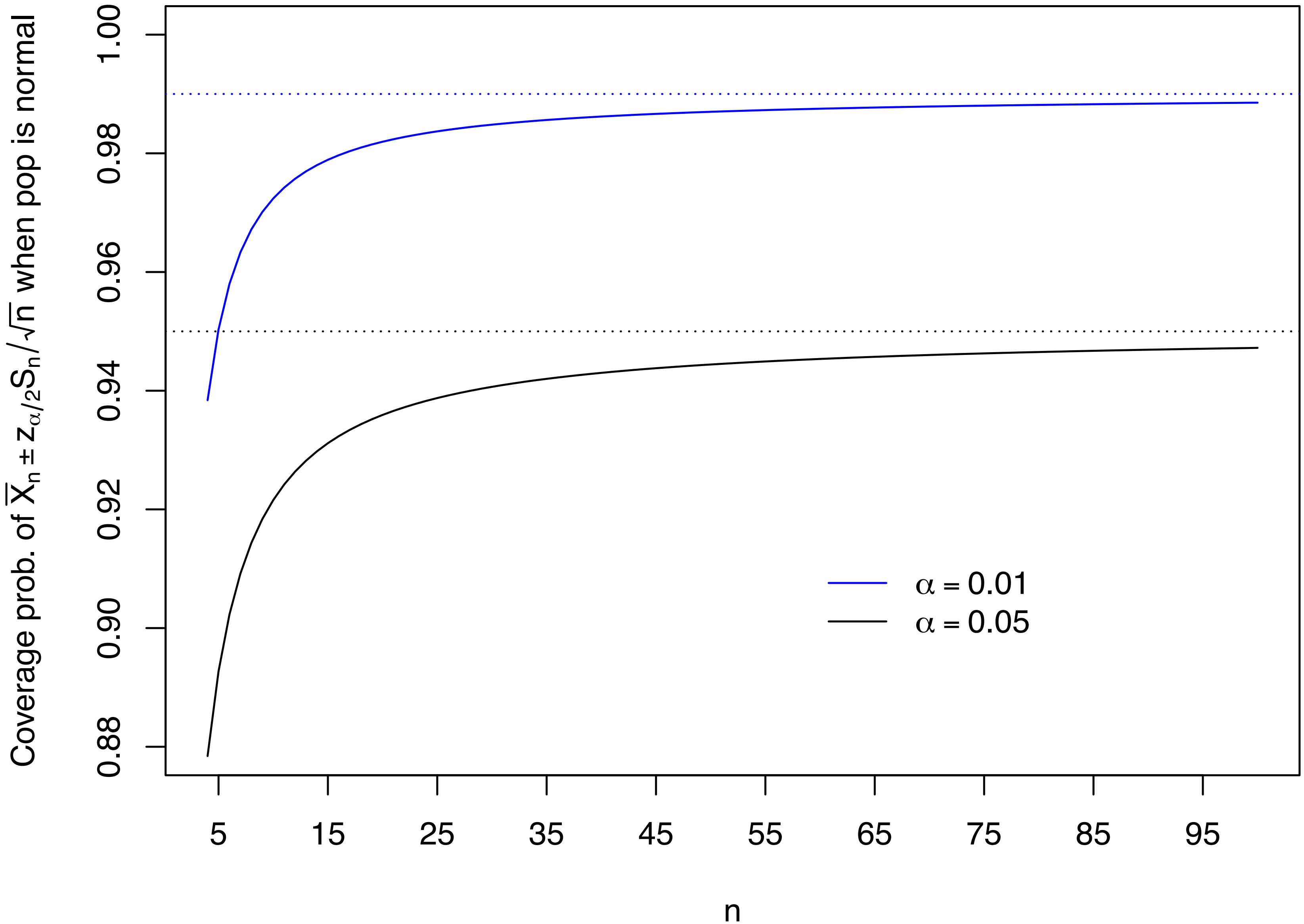
$$\bar{x}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(3) $\alpha = 0.01$

$$0.4303 \pm \boxed{2.75} \frac{0.03299}{\sqrt{33}} = [0.415,\ 0.446]$$

$t_{30,0.005} \approx t_{33-1,0.005}$

```r
X <- c(0.455,0.375,0.428,0.390,0.388,0.420,0.430,0.420,0.350,0.440,
       0.450,0.430,0.500,0.480,0.460,0.429,0.427,0.424,0.512,0.428,
       0.426,0.430,0.400,0.460,0.440,0.400,0.450,0.410,0.398,0.420,
       0.453,0.449,0.428)
qqnorm(X)
Xbar <- mean(X)
Sn <- sd(X)
n <- length(X)
alpha <- 0.05
tval <- qt(1-alpha/2,n-1)
lo <- Xbar - tval * Sn / sqrt(n)
up <- Xbar + tval * Sn / sqrt(n)
```


Coverage prob. of $\overline{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ when pop is normal
1.00
0.98
0.96
0.94
0.92
0.90
0.88
5
15
25
35
45
55
65
75
85
95
n
$\alpha = 0.01$
$\alpha = 0.05$

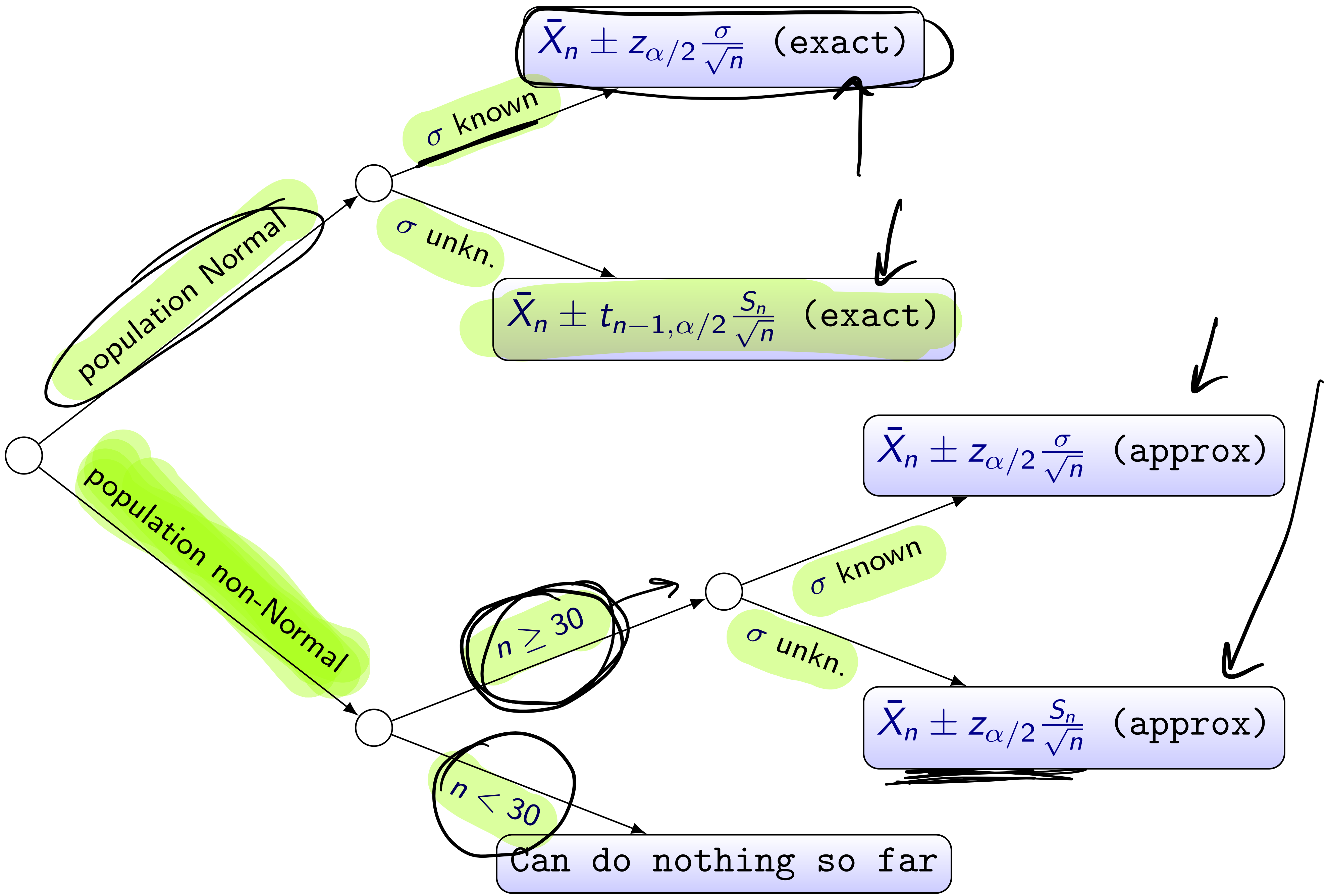
CI for mean of non-Normal population with σ unknown

Let $X_1, \dots, X_n$ be a rs from a pop. with mean μ and variance $\boxed{\sigma^2 < \infty}$ then

random sample

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

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


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