CLT: $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

when X_i are i.i.d.

w/ $E(X) = \mu$, $Var(X) = \sigma^2$

If X_i are normally distr.

then CLT is exact for any sample size n.

If non-normal but not too skewed $n \geq 30$ ok.

Also by CLT get that

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

---

ex) $X_i \sim Pois(\lambda = 3)$ w/ $n = 40$

estimate $P(\overline{X} > 3.1)$.

by CLT: $\overline{X} \sim N\left(3, \frac{3}{40}\right)$ so:

$$P(\bar{X} > 3.1) = 1 - \text{pnorm}\left(3.1, \sqrt{\tfrac{3}{40}}\right)$$

$$\approx 0.357$$

ex, con't: Can also calc. w/ sum $\sum X_i$

$$P(\bar{X} > 3.1) = P(n \cdot \bar{X} > n \cdot 3.1)$$

$$= P(\sum X_i > 40 * 3.1)$$

$$= P(\sum X_i > 124)$$

$$\approx 1 - \text{pnorm}(124, 120, \sqrt{120})$$

$$= 0.357$$

Note that $\sum X_i \sim \text{Pois}(\lambda = 3 \cdot 40)$

(see: scaling pois. param.)

so $P(\sum X_i > 124)$

$$= 1 - \text{ppois}(124, 120)$$

$$= \underline{0.336}$$, so the normal approx. is reasonable.

(Student's) t-distribution:

$$X_i \sim N(\mu, \sigma^2)$$

then $$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim T_{n-1}$$

is t-distributed w/ n−1 "degrees of freedom."

---

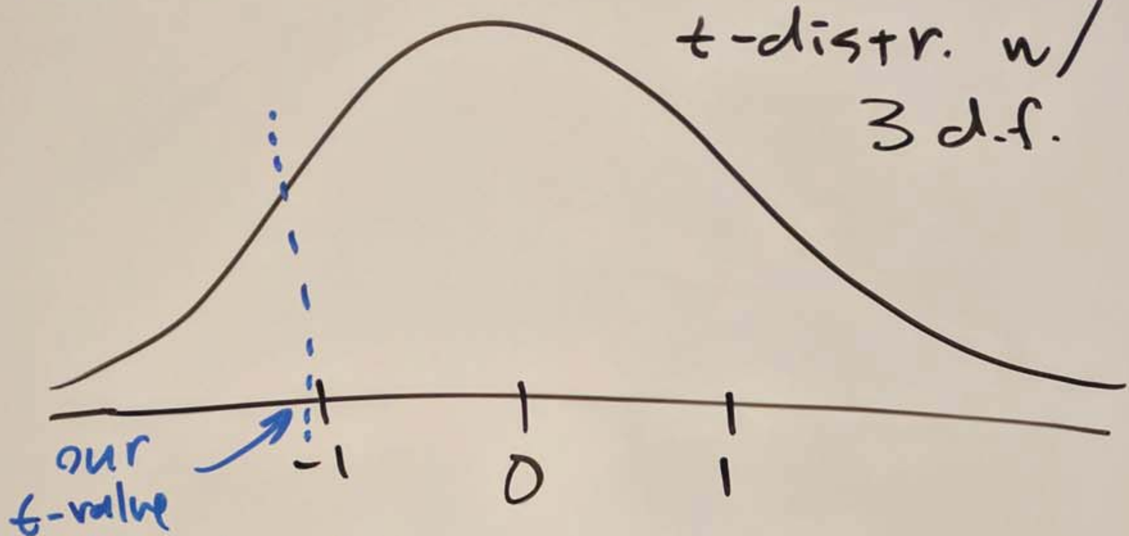
ex: Strength of rope is normally distr. Assume we don't know σ^2 but know $\mu = 120$ lb.

If we have a sample of 4 pieces of rope w/ $\bar{X} = 117.5$ and $s = 5$, we can calc. a t-statistic:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{117.5 - 120}{5/\sqrt{4}}$$

$$= -1$$

ex cont: $t = -1$



Recall 68-95-99.7 rule.

t-distr. is "wider" than Std. normal, so our ± 1 interval captures less data.

$$P(-1 < T_3 < 1) = pt(1,3) - pt(-1,3)$$
$$= 0.609$$

Our t-statistic is typical.

What if $\bar{x} = 80$, $s = 30$, $n = 4$

$\bar{x} = 80$, $s = 5$, $n = 4$

$$t = \frac{80-120}{30/\sqrt{4}} = \frac{-40}{15} = -2.67$$ ← still fairly typical

$$t = \frac{80-120}{5/\sqrt{4}} = \frac{-40}{5/2} = \frac{-80}{5} = -16$$ ← unusually large!!

Confidence Intervals:

$$X_i \sim N(\mu, \sigma^2)$$

std. normal:

$$\Rightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

t-distr.

$$\frac{\overline{X} - \mu}{s/\sqrt{n}} \sim T_{n-1}$$

<u>IDEA</u>: if we don't know μ, we can use this to predict a range of estimates for μ

i.e. μ is likely in interval:

$$(\text{est. } \mu) \pm (\text{error})$$

$(1-\alpha)100\%$ C.I.

$1-\alpha$ = confidence level
usually $\alpha=0.05$ to give 95% C.I.

ex: Let's say we have our rope sample w/ $\bar{x}=90$, $s=5$, $n=4$. Predict μ w/ a 95% C.I.

sol'n: we know $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim T_{n-1}$

so: $P\left(-t^* < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t^*\right) = 95\%$

w/ $t^* = t_{1-\alpha/2,\, n-1} = qt(1-\frac{\alpha}{2}, n-1)$
quantile

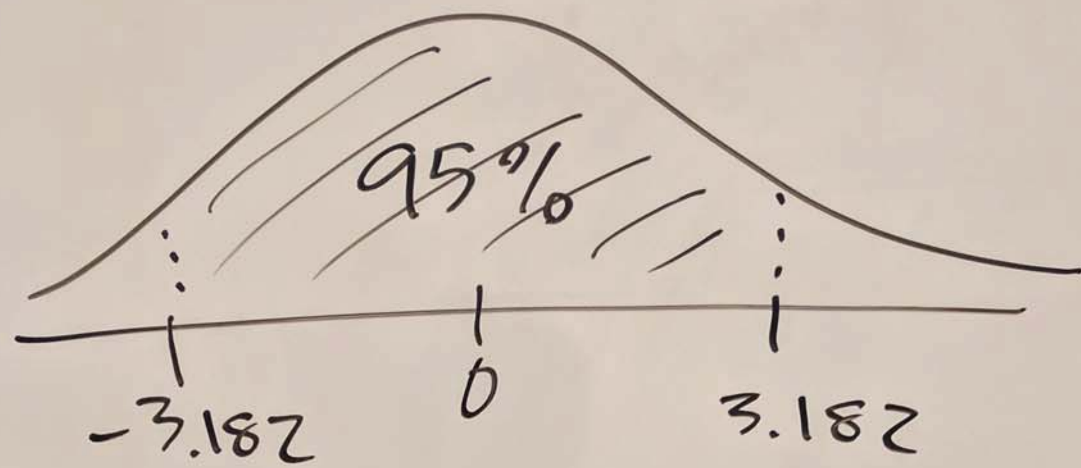
rearrange:

$$-t^*\frac{s}{\sqrt{n}} < \bar{x}-\mu < t^*\cdot\frac{s}{\sqrt{n}}$$

$$\boxed{\bar{x}-t^*\frac{s}{\sqrt{n}} < \mu < \bar{x}+t^*\frac{s}{\sqrt{n}}}$$ our C.I.

$t_{0.975,3} = qt(0.975, 3)$

$= 3.182$



Recall for std. normal ± 1.96 std. devs. gives 95%.

We can see that t-distr. is <u>wider</u>.

So our 95% CI for μ is:

$$\bar{X} \pm t^* \frac{s}{\sqrt{n}} \longrightarrow 90 \pm 3.182 \cdot \frac{5}{\sqrt{4}}$$

$$= 90 \pm 7.955$$

$$= \boxed{(82.045, 97.955)}$$

<u>Consider this</u>: If we wanted the mean strength of the rope to be $\mu = 120$, then our sample suggests $\mu \neq 120$ b/c it's not in our interval.

* This is really what <u>STATISTICS</u> is about...

Estimating μ w/ C.I:

General rules:

σ - known, data normal

$$\bar{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$z_{1-\alpha/2} = \text{qnorm}(1-\alpha/2)$$

σ - unknown, but $n \geq 30$

$$\bar{X} \pm z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

(we are plugging in s for σ)

σ - unknown, $n < 30$

$$\bar{X} \pm t_{1-\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

$$t_{1-\frac{\alpha}{2}, n-1} = qt(1-\tfrac{\alpha}{2}, n-1)$$

** All assume $X_i \sim N(\mu, \sigma^2)$