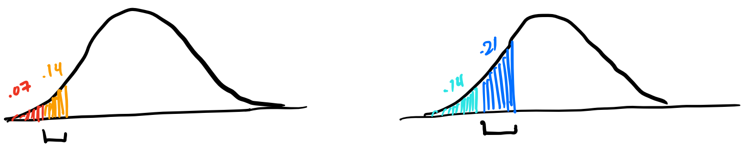
1. Decide which is bigger and explain why. Only use R to confirm your answer.

a. $t_{0.07,n} - t_{0.14,n}$ or $t_{0.14,n} - t_{0.21,n}$

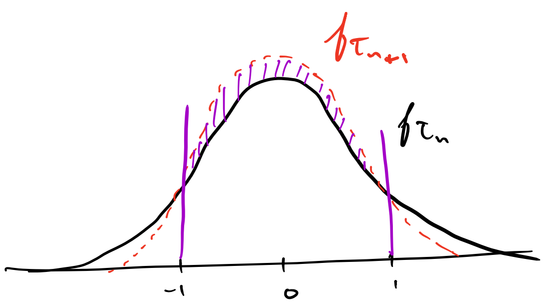


The second is be larger. α indicates an area under the distribution. If n is fixed and we compare the distances (or difference) between two small t-values on the left side, we return a small value. When two larger distances are compared we return a larger t-val.

b. $P(|T_n| < 1)$ or $P(|T_{n+1}| < 1)$

$P(-1 < T_n < 1)$

$P(-1 < T_{n+1} < 1)$



As the degree of freedom increases, the area under the curve (when restricted to $-1 \to 1$) grows in size.

$$P(|T_{n+1}| < 1) > P(|T_n| < 1)$$

2. Let $X_1, \ldots, X_{16}$ be a random sample from a normal distribution with mean 0. For what k is the below inequality true? Explain your reasoning.

$$P\left(\left|\frac{4\overline{X}}{S}\right| > k\right) = 0.08$$

$$\frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{\bar{x} - 0}{S/\sqrt{16}} = \frac{\bar{x}}{S/4} = \frac{4\bar{x}}{S}$$

> pt(k, 15, lower = F) = .08

→ >qt(.04, 15, lower = F)

1.878

$$P\left(\left|\frac{4\bar{x}}{S}\right| > 1.478\right) = .08$$

$k = 1.478$

3. Each day, Zelda starts the morning with a cup of coffee from her Keurig machine. Lately, she has begun to think the machine is malfunctioning because the amount dispensed is different than the advertised average amount, 12 oz. To explore this, she picks 7 random days from the next month and actually weighs her coffee using a calibrated kitchen scale. She believes that the coffee dispensing amounts are normally distributed, and her seven data points give $\overline{x} = 12.9$ and $s = 0.7$ oz.

$n = 7$

$\bar{X} = 12.9$

$S = .7$

$\alpha = .06$

a. Conduct a hypothesis test for Zelda's predicament using $\alpha = 0.06$.

Test $\quad H_0: \mu = 12\text{oz} \qquad H_1: \mu \neq 12\text{oz}$

Critical Region: If $t_{up} > t_{\alpha, n-1}$ reject H_0 otherwise maintain.

$\boxed{\text{P-Val: If } \frac{P(t < t_{n-1})}{2} = P/2 < \alpha/2}$

Zelda's data give t-score: $\frac{\bar{X} - \mu_0}{S\sqrt{n}} = \frac{12.9 - 12}{.7\sqrt{7}} = 3.402$

> qt(.03, 6, lower = F)
2.313

$\rightarrow t_{\alpha/2, n-1} = 2.313$ and $3.402 > 2.313$

So we reject H_0

Her Keurig is not producing the correct amount of coffee.

b. Find a confidence interval for the true average coffee dispensing amount that has duality with the hypothesis test from part a.

H_0 keep ~~iff~~ $\quad -t_{\alpha/2} < \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < t_{\alpha/2}$

$\hookrightarrow \quad \bar{X} - t_{\alpha/2, n-1}(S/\sqrt{n}) < \mu_0 < \bar{X} + t_{\alpha/2, n-1}(S/\sqrt{n}) \iff (1-\alpha)100\%$ CI

$\alpha = .06 \rightarrow$ We want a 94% CI

$t_{\alpha/2} = t_{.03}$ > qt(.03, 6, lower = F)
2.313

$\bar{X} + t_{\alpha/2, n-1}(S/\sqrt{n}) = 12.9 + 2.313(.7/\sqrt{7}) = 13.131$

$\bar{X} - t_{\alpha/2, n-1}(S/\sqrt{n}) = 12.669$

94% CI for μ is $(12.669, 13.131)$ $\quad \mu = 12 \notin$ CI

We reject H_0.

4. Earlier in the class, we explored home-field advantage (HFA) using win percentages. Now we revisit the question using the "margin of victory", which is amenable to means. Looking at data from 317 college (American) football games involving top-25-ranked teams, researchers found an average margin of victory (home team score − away team score) of $\bar{y} = 4.57$ with $s = 18.29$. Do these data support the notion of HFA?

$n = 317$

$\bar{y} = 4.57$

$S = 18.29$

a. Conduct an appropriate hypothesis test using $\alpha = 0.03$. Then, create a one-sided CI that has duality with this hypothesis test.

Test $\quad \underbrace{H_0 : \mu = 0}_{} \qquad \underbrace{H_1 : \mu > 0}_{}$

avg. margin of victory = 0 $\Rightarrow$ No HFA

avg. MOV is positive HFA may exist

Decision Rule: if $t_{score} > t_{\alpha, n-1}$ reject H_0

$$t_{score} = \frac{\bar{y} - \mu}{S/\sqrt{n}} = \frac{4.57 - 0}{18.29/\sqrt{317}} = 4.449$$

$\alpha = .03$

>qt(.03, 316, lower=F)

$1.888 = t_{\alpha, n-1}$

$\Rightarrow 4.449 > 1.888$

We reject H_0

$\alpha = .03 \longrightarrow$ We want a 97% CI

>qt(.03, 316, lower=F)

$1.888 = t_{\alpha, n-1}$

WANT LOWER BOUND FOR μ

$$\bar{y} - t_{\alpha, n-1}(S/\sqrt{n}) = 4.57 - 1.888(18.29/\sqrt{317})$$

$$= 2.631$$

reasonable values for $\mu \in (2.631, \infty)$

b. After publishing your findings, a rival academic argues that you have failed to establish the normality of the population being considered (margins of error for college football games involving top-25-ranked teams). Respond to this criticism.

$n = 317 > 30$ so the distribution $t_{n-1} \approx \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$

It is roughly normal

5. Show that the pdf of T_n converges to the pdf of $N(0,1)$ as $n \to \infty$. When doing this problem you may use Stirling's Formula: $n! \approx \sqrt{2\pi n}\cdot n^n e^{-n}$ and a helpful fact from Calculus I: $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$. Also, it is fine if you assume $\Gamma(r) = (r-1)!$, even when r is not an integer.

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{(n+1)/2}} \quad \text{for } -\infty < t < \infty$$

WANT TO CONVERGE TO

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\Gamma\left(\frac{n+1}{2}\right) = \left(\frac{n+1}{2}-1\right)! = \left(\frac{n+1-2}{2}\right)! = \left(\frac{n-1}{2}\right)!$$

$$\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2}-1\right)! = \left(\frac{n-2}{2}\right)!$$

$$\lim_{n\to\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{n+1/2}}$$

$$= \underbrace{\lim_{n\to\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}}\cdot \lim_{n\to\infty}\frac{1}{\left(1+\frac{t^2}{n}\right)^{n+1/2}}$$

$$\lim_{n\to\infty}\frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}$$

$$= \lim_{n\to\infty}\frac{1}{\left(\left(1+\frac{t^2}{n}\right)^{n+1}\right)^{1/2}}$$

$$= \lim_{n\to\infty}\left[\frac{1}{\left(1+\frac{t^2}{n}\right)^n}\right]^{1/2}\cdot \underbrace{\lim_{n\to\infty}\left[\frac{1}{\left(1+\frac{t^2}{n}\right)^1}\right]^{1/2}}_{1}$$

$$= \frac{1}{e^{t^2/2}\cdot 1}$$

$$= e^{-t^2/2} \checkmark$$

$$\lim_{n\to\infty}\frac{\left(\frac{n-1}{2}\right)!}{\sqrt{n\pi}\left(\frac{n-2}{2}\right)!} = \lim_{n\to\infty}\frac{\left(\frac{n}{2}-\frac{1}{2}\right)!}{\sqrt{n\pi}\left(\frac{n}{2}-1\right)!}$$

$$n! = \sqrt{2\pi n}\cdot n^n e^{-n}$$

$$= \lim_{n\to\infty}\frac{\sqrt{2\pi\left(\frac{n}{2}-\frac{1}{2}\right)}\cdot\left(\frac{n}{2}-\frac{1}{2}\right)^{\frac{n}{2}-\frac{1}{2}} e^{-\left(\frac{n}{2}-\frac{1}{2}\right)}}{\sqrt{n\pi}\sqrt{2\pi\left(\frac{n}{2}-1\right)}\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1} e^{-\left(\frac{n}{2}-1\right)}}$$

$$= \lim_{n\to\infty}\frac{\cancel{\sqrt{2\pi}}\left(\frac{n}{2}-\frac{1}{2}\right)^{1/2}\left(\frac{n}{2}-\frac{1}{2}\right)^{n/2-1/2}\cancel{e^{-n/2}}\,e^{1/2}}{\cancel{\sqrt{2\pi}}\sqrt{n\pi}\left(\frac{n}{2}-1\right)^{1/2}\left(\frac{n}{2}-1\right)^{n/2-1}\cancel{e^{-n/2}}\,e^{1}} = \lim_{n\to\infty}\frac{\left(\frac{n}{2}-\frac{1}{2}\right)^{1/2}\left(\frac{n}{2}-\frac{1}{2}\right)^{n/2-1/2}e^{1/2}}{\sqrt{n\pi}\left(\frac{n}{2}-1\right)^{n/2-1/2}e}$$

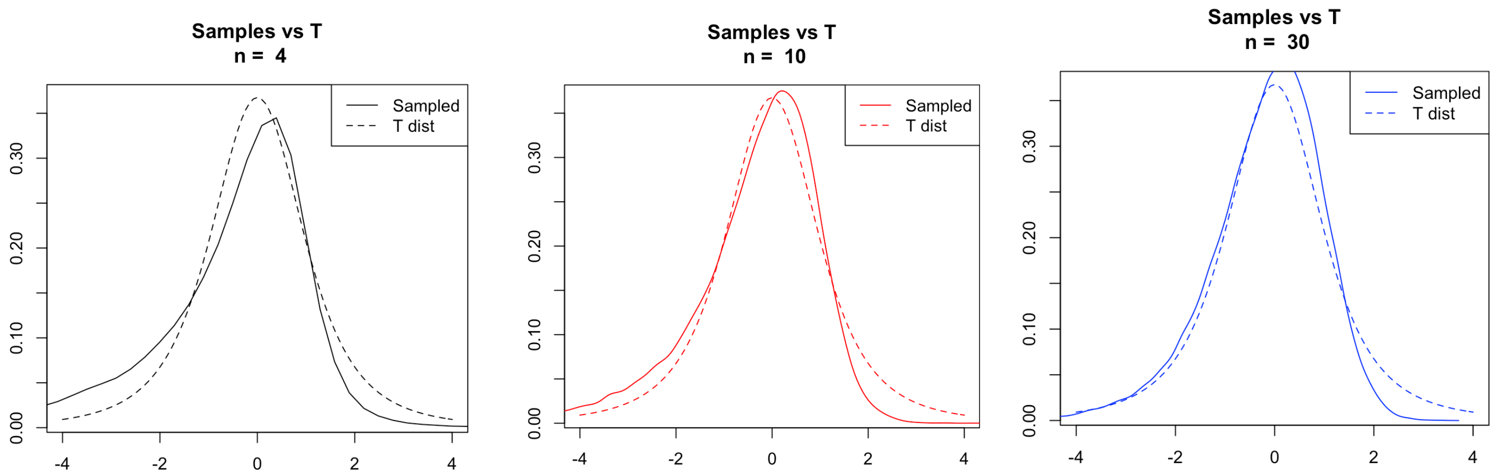
$$= \lim_{n\to\infty}\frac{\frac{1}{2}^{1/2}(n-1)^{1/2}\cancel{\frac{1}{2}^{(n/2-1/2)}}(n-1)^{n/2-1/2}e^{1/2}}{\sqrt{n\pi}\cancel{\frac{1}{2}^{(n/2-1/2)}}(n-2)^{n/2-1/2}e} = \lim_{n\to\infty}\frac{(n-1)^{n/2}e^{1/2}\sqrt{1/2}}{\sqrt{n\pi}(n-2)^{n/2-1/2}e}$$

$$= \lim \frac{(n-1)^{n/2}e^{1/2}(n-2)^{1/2}}{\sqrt{n\pi}(n-2)^{n/2}e} \rightsquigarrow \underbrace{\lim\frac{(n-1)^{n/2}}{(n-2)^{n/2}}}_{=\sqrt{e}}\cdot\underbrace{\lim\frac{\sqrt{1/2}\,(n-2)^{1/2}}{\sqrt{n}\sqrt{\pi}\,e^{1/2}}}_{\frac{1}{\sqrt{2}\sqrt{\pi}\sqrt{e}}} = \frac{1}{\sqrt{2\pi}}$$

$$\text{So } \lim_{n\to\infty} f_{T_n} = \frac{1}{\sqrt{2\pi}}e^{-t^2/2} \sim N(0,1$$

6. In class, I claimed that if a population distribution has moderate or severe skew, then as long as the sample size was about 30 or more, we could count on $T_{n-1} \approx \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$. Our goal here is to empirically show this. To begin, let $n = 4$ and draw the T_{n-1} density on the interval $(-4, 4)$ using a dashed line (use the lty parameter in the plot function to make a dashed line). Now, let $X \sim Exp(\lambda = 3)$, draw a sample of size n, and compute $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$. Repeat this process to get a total of 50000 t-scores. Plot the density of these atop T_{n-1} using a solid line. Both plots should be in the same color. Next, repeat the above process for $n = 10$, 30, and 60 (use a new plot and new color for each n value). Include your code and a sketch of the four plots.

Note: This example will also show that $n \approx 30$ is not some absolute rule. You might be quite bothered by the difference in your two graphs when $n = 30$. In general, the larger the skew in the population, the larger n should be to overwhelm its effect. There is no perfect guideline here: data analysis and statistics involve *human, imperfect* decision making. Sorry to break your quantitative heart.




```r
color = 'green'

n = 60
mu = 1/3
tVals = c()
for (i in 1:50000)
{

  sampleDraw <- rexp(n, 3)
  xBar <- mean(sampleDraw)
  s = sd(sampleDraw)
  tScore <- (xBar - mu)/(s/sqrt(n))
  tVals[i] <- tScore
}

plot(curve(dt(x, df=3), from=-4, to=4), type = "l", lty = 2,
     col= color,xlab="",ylab="")

sampleDistribution <- density(tVals)
lines(sampleDistribution, col = color)
subTitle = paste("Samples vs T\n","n = ", n)
title(main = subTitle)

legend("topright", legend = c("Sampled", "T dist"), lwd = 1,
       lty = c(1, 20), col = c(color, color))
```