Lecture 12

Quantitative Political Science

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/10/10

Slides Updated: 2023-10-09

Agenda

1. Small Sample Significance Tests
2. Degrees of Freedom
3. Review of Inference with One Variable

Small Sample Significance Tests

- What happens with n is small?
- Start as before, $Y_1, Y_2, \ldots, Y_n$ are a random sample drawn from Normal population with $\bar{Y}$ and S_U^2 as before
- Want construct CI for μ when $VAR(Y_i) = \sigma^2$ is unknown and n is small
- Since n is small, we can't rely on the CLT to assume that $\bar{Y} \sim \mathcal{N}(\mu, \sigma^2)$

Small Samples

- Assume: $Y \sim \mathcal{N}(\mu, \sigma^2)$
- Theorem: Linear combination of independent, Normally-distributed RVs is itself Normally distributed
  - See Chapter 6 for proof if interested (Fang!)
- "Linear combination": sum of products of RVs and scalars: $\sum_i^J a_i Y_i$
- $\bar{Y}$ is one such linear combination where a_is are $\frac{1}{n}$
- Thus, if we assume $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $\bar{Y}$ is too

Small Samples

- Given that each Y_i is itself a random variable, we can standardize each just like before

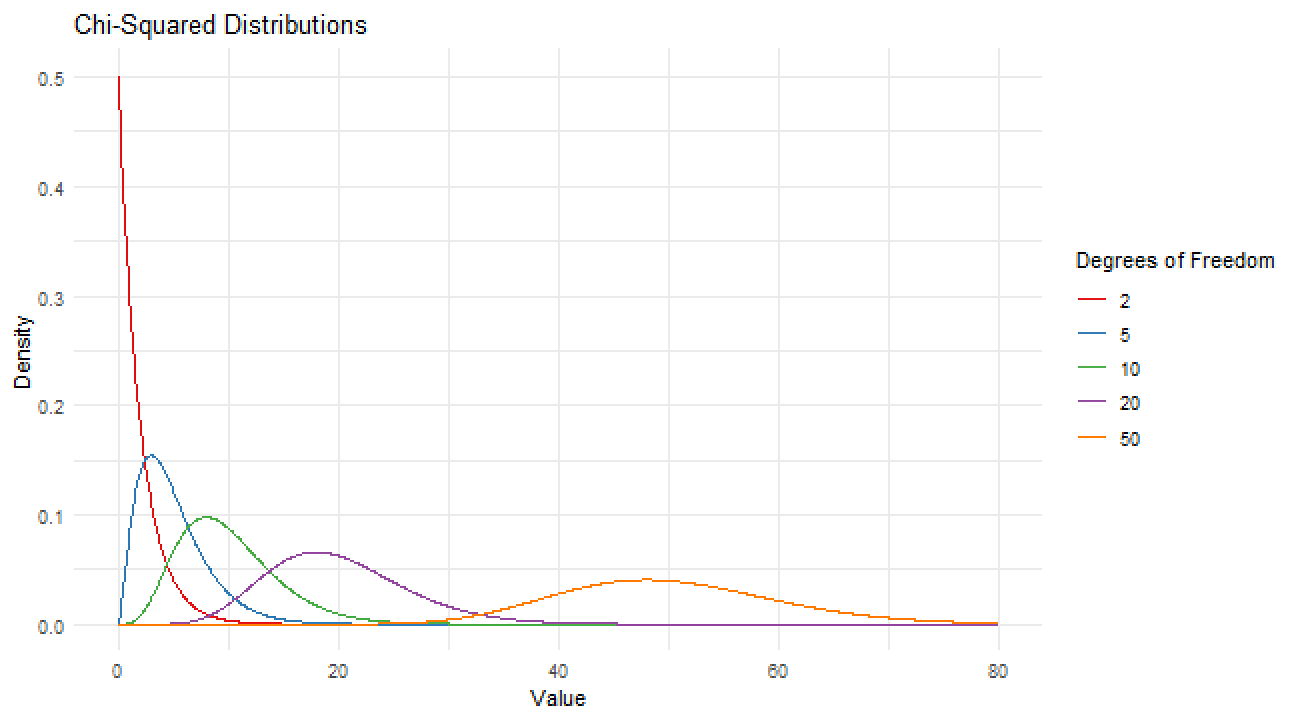
$$Z_i = \frac{Y_i - \mu}{\sigma}$$

- Consider the **sum of squares** of this quantity

$$\sum Z_i^2 = \sum_i \left(\frac{Y_i - \mu}{\sigma} \right)^2$$

- This sum of squares takes on a **chi-squared (χ^2) distribution with n degrees of freedom**
- NB: *any* sum of squares of a normally distributed RV will take on this distribution

Chi-Squared Distributions



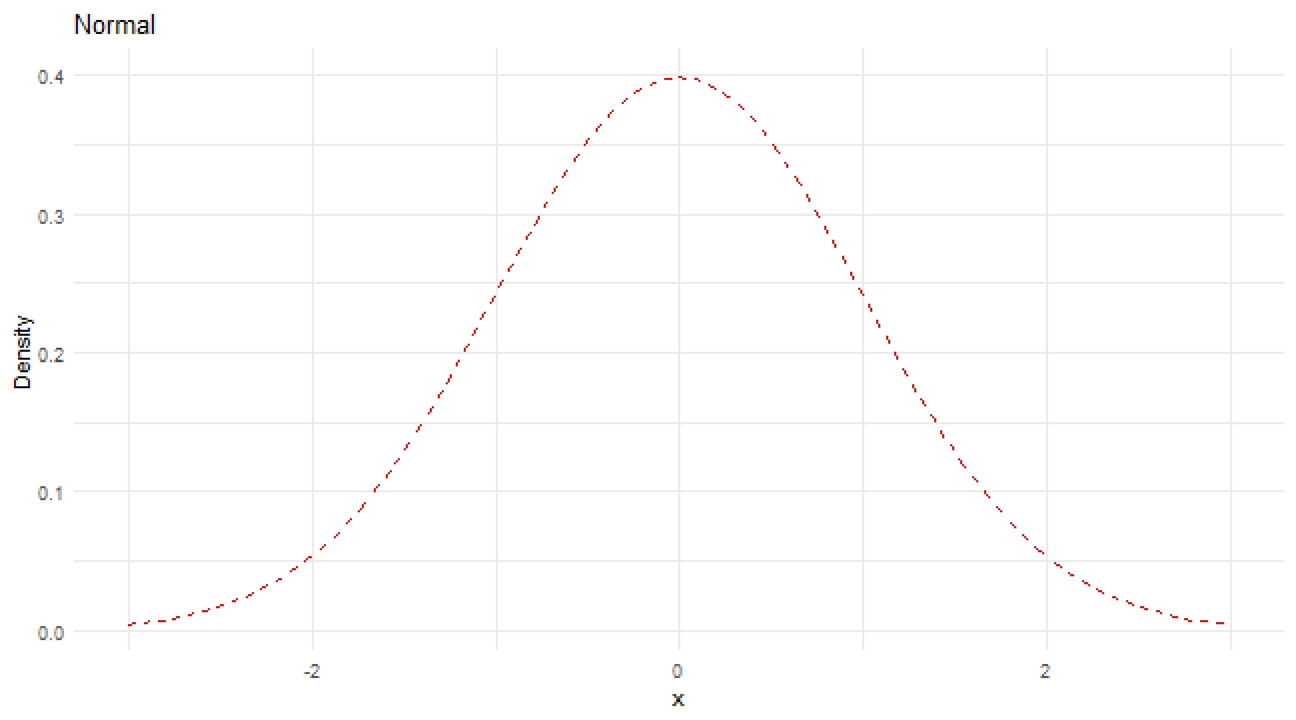
Degrees of Freedom

- "Degrees of Freedom": number of independent pieces of information on which the statistic is based
  - Pieces of information you have (n) minus the number you need to generate the statistic
  - I.e., to calculate $\bar{Y}$ from $n = 3$ observations, we calculate $\bar{Y} = \frac{Y_1+Y_2+Y_3}{3}$
  - Contrast with the sample standard deviation (unbiased)

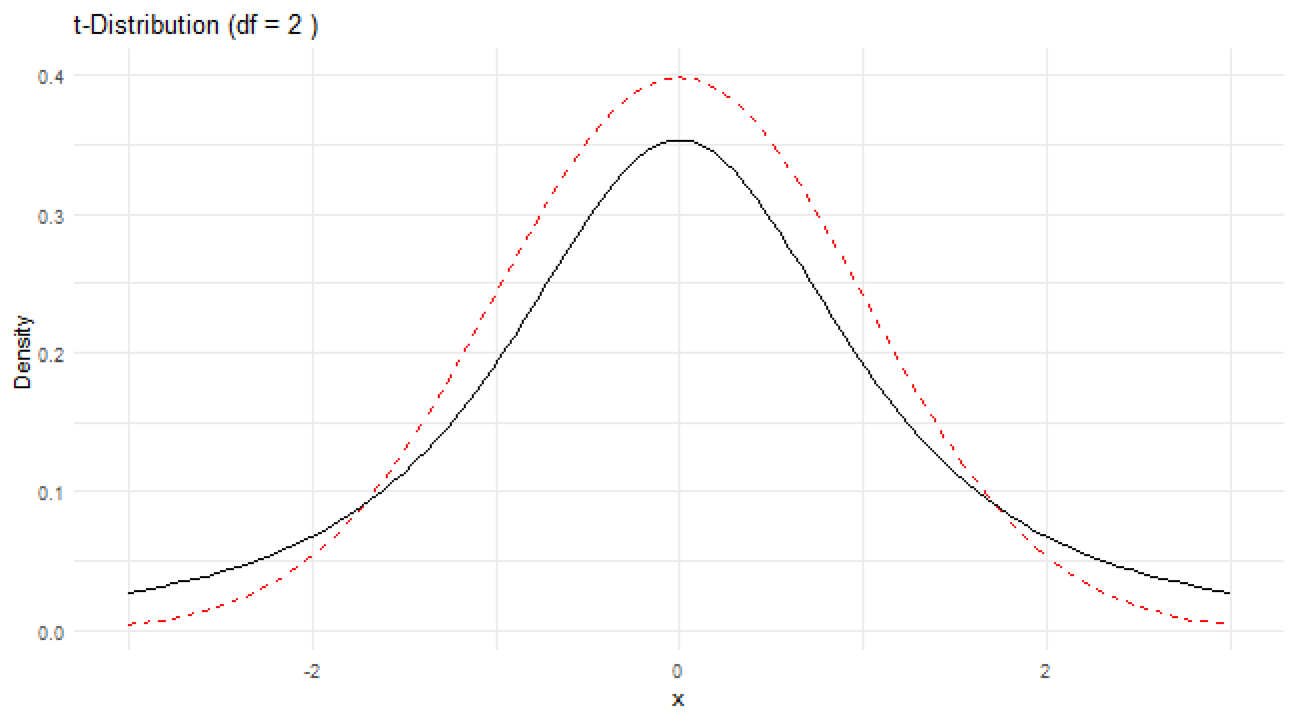
Student's t-distribution

- Thus far, we have been focused on a Normal distribution
  - CLT gave us a pass to rely on the normal for inference
- But in practice, we actually rely on something called the **Student's t-distribution**
- Defined as $T = \frac{Z}{\sqrt{W/\nu}}$
  - Standard normal RV Z over square root of χ^2 RV W divided by its degrees of freedom ν
- Similar to the Normal:
  - Gnarly formula: $f(y) = \frac{\Gamma\left[\frac{\nu+1}{2}\right]}{\Gamma\left[\frac{\nu}{2}\right]} * \frac{1}{\sqrt{\nu\pi}\left(1+\frac{y^2}{\nu}\right)^{\frac{\nu+1}{2}}}$
  - Symmetric around 0
- But..."fatter tails"

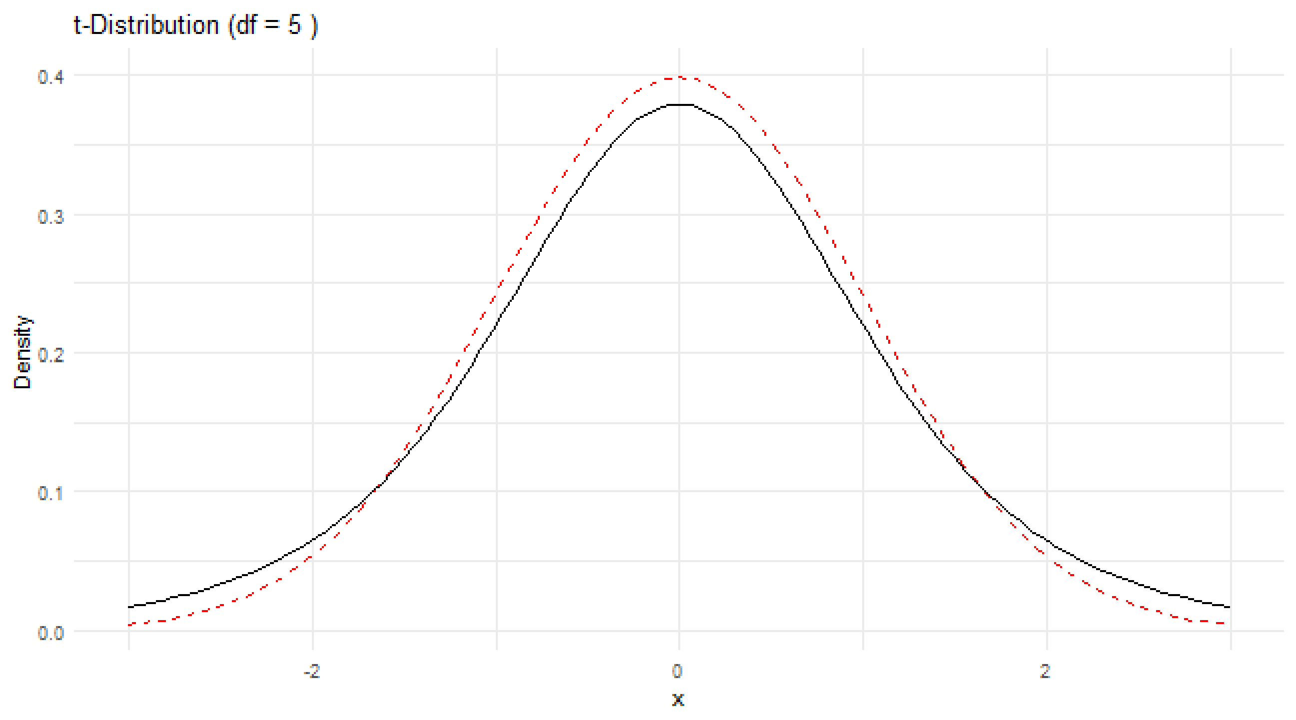
Fatter Tails



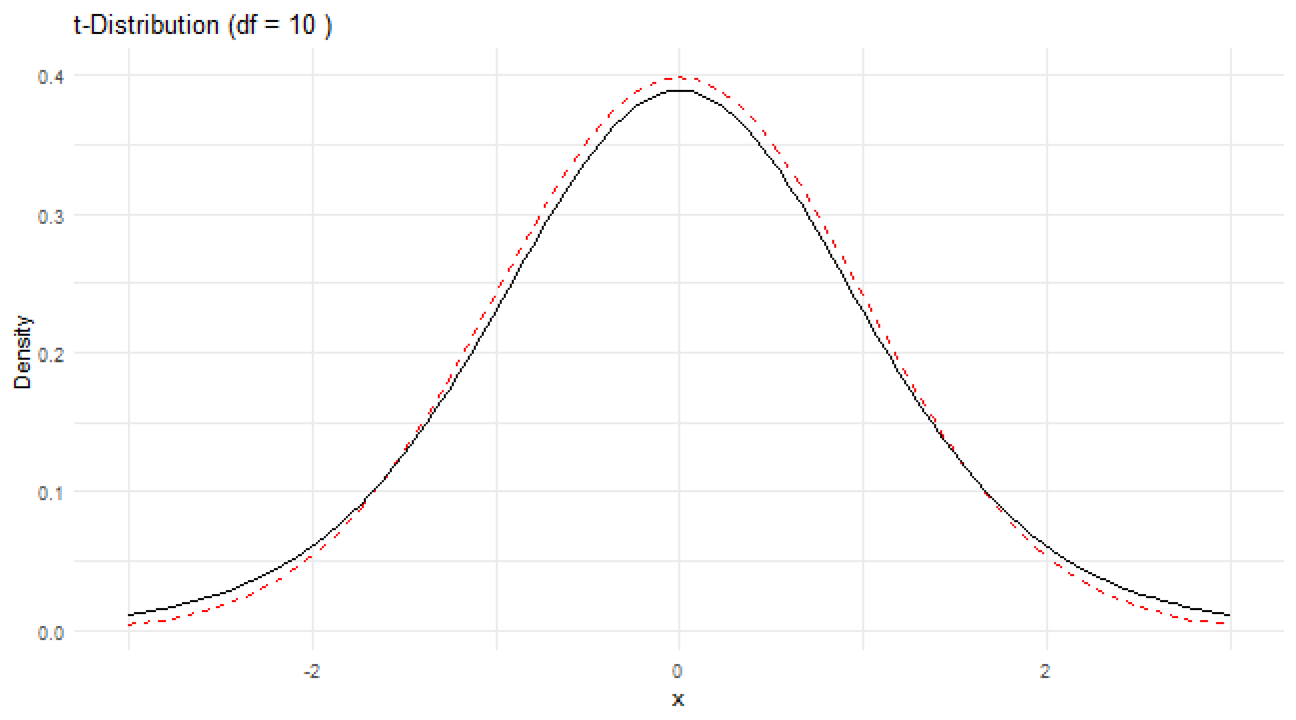
Fatter Tails



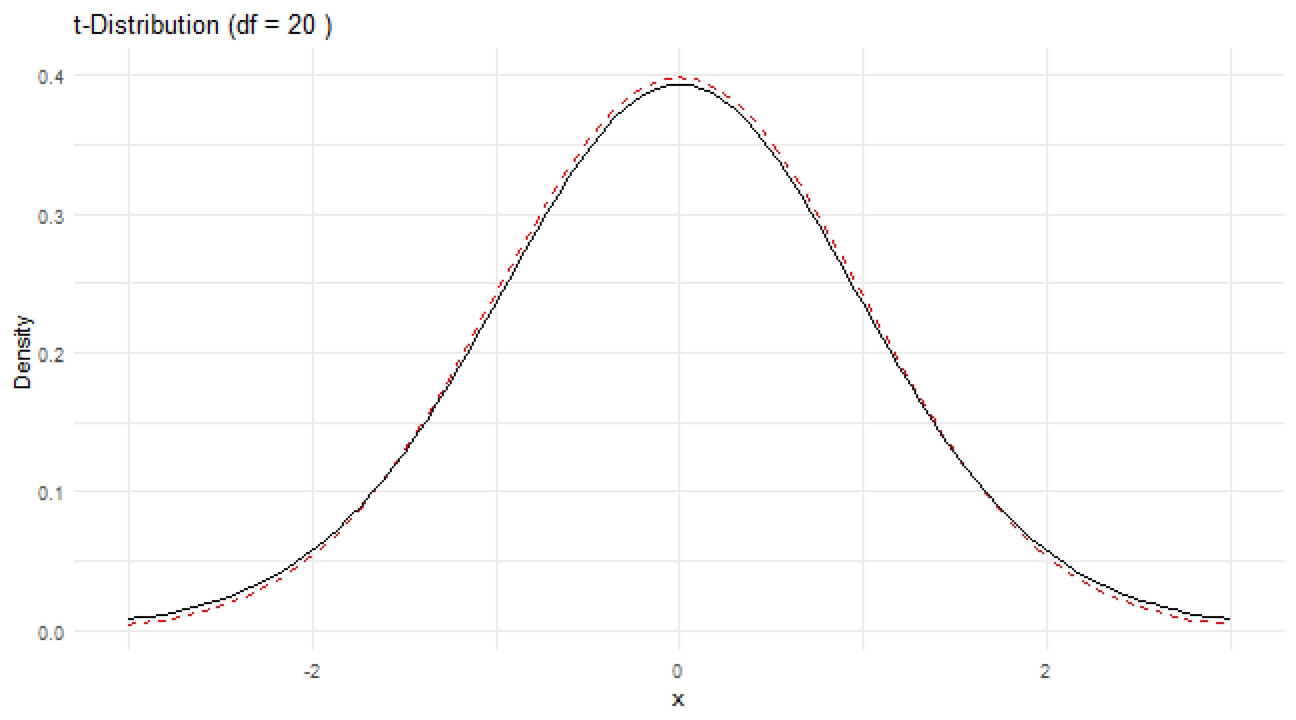
Fatter Tails



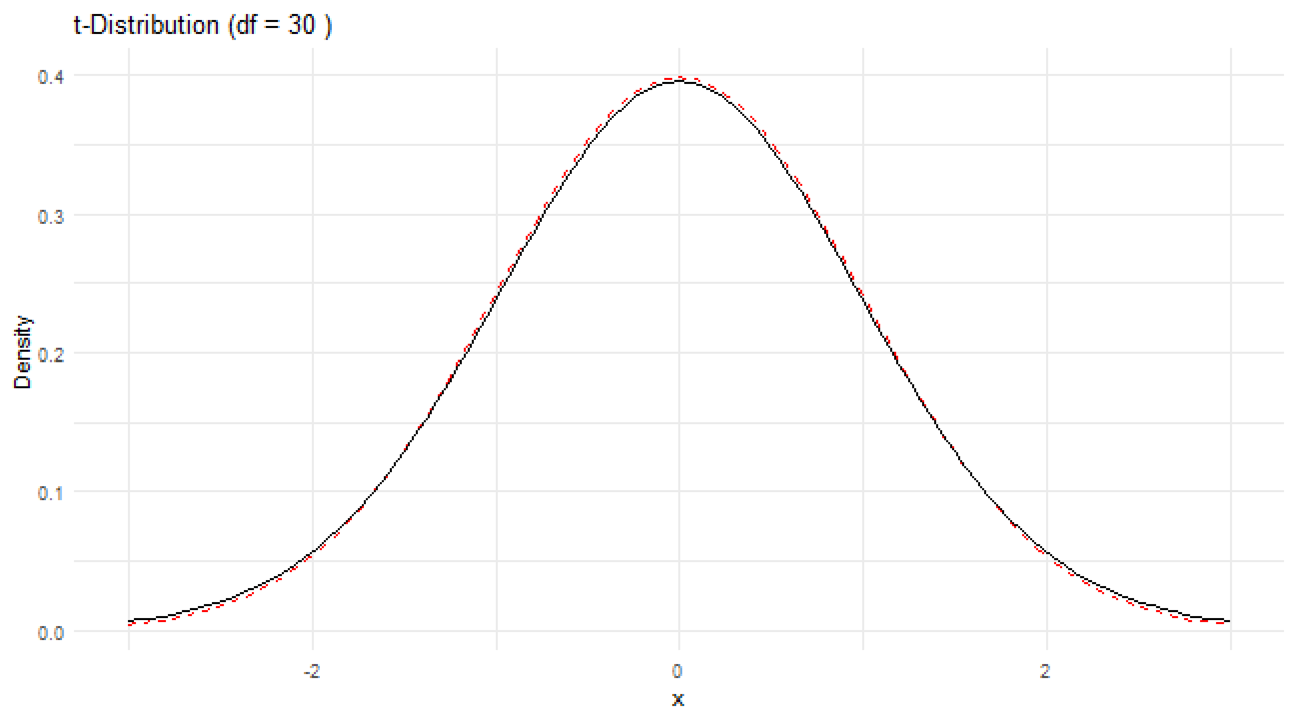
Fatter Tails



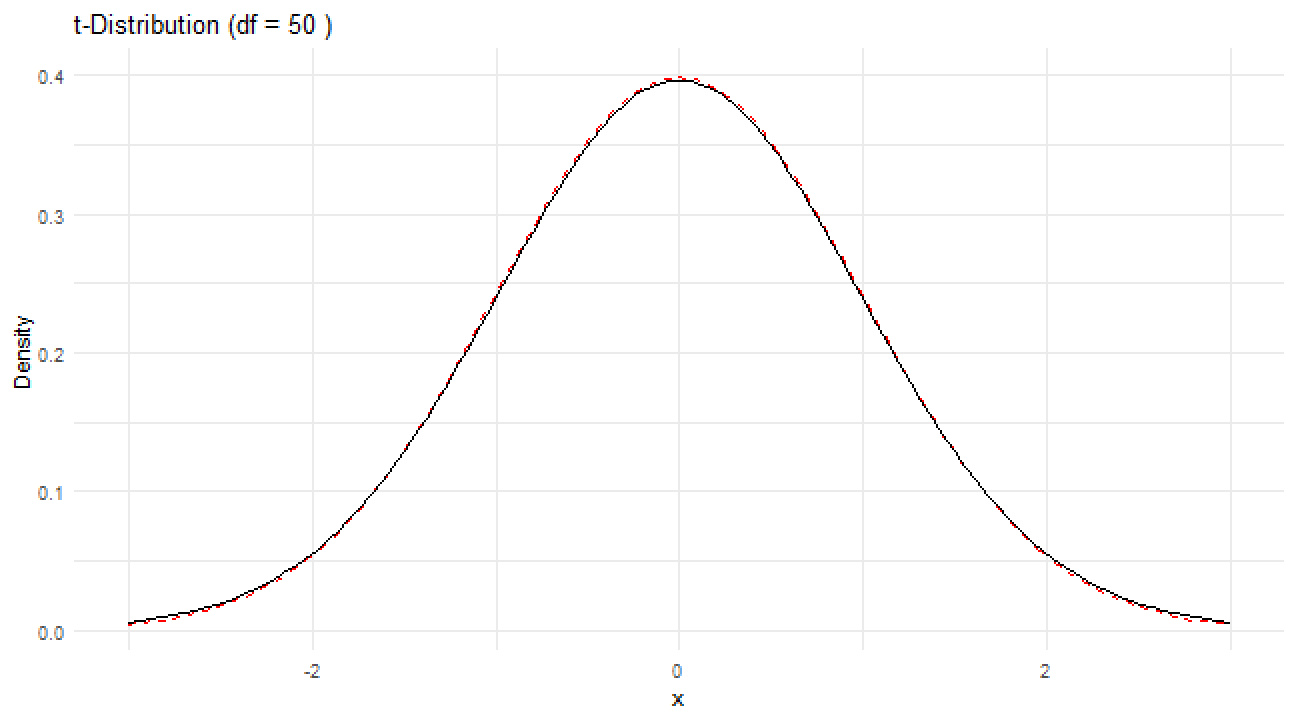
Fatter Tails



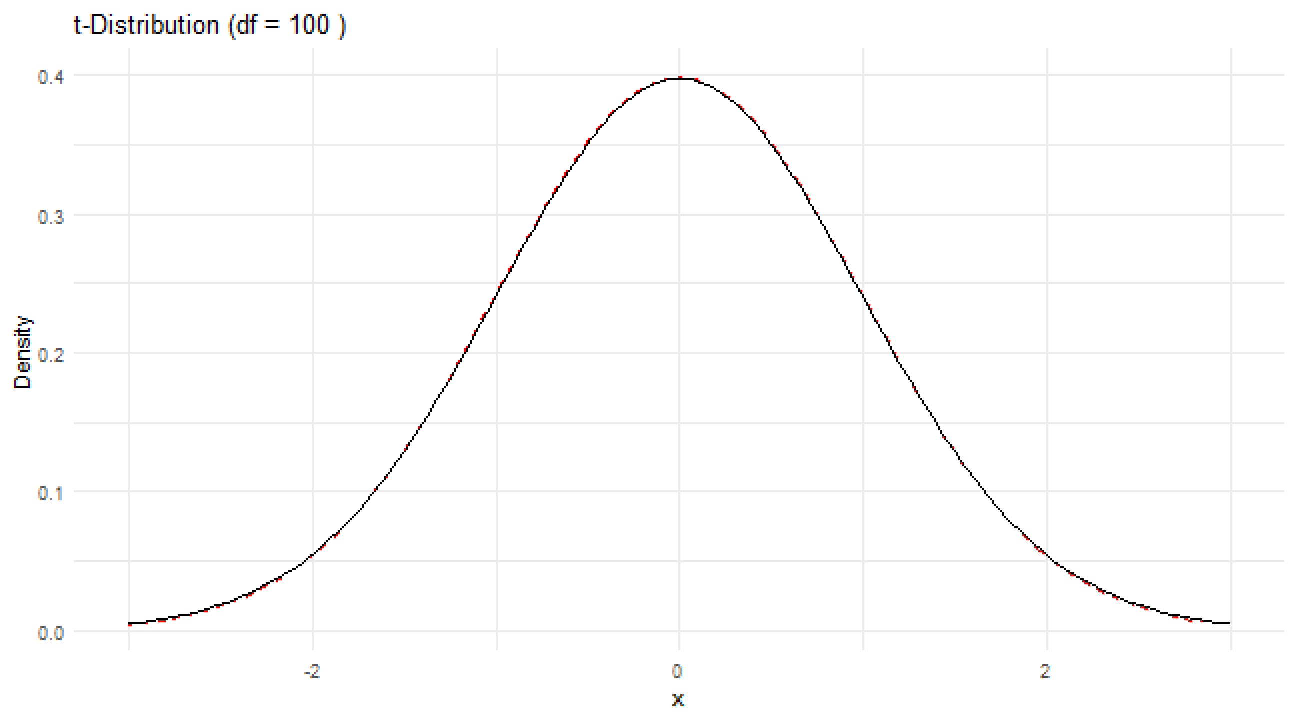
Fatter Tails



Fatter Tails



Fatter Tails



Putting it together

- So we have a distribution with known quantities for the sum of squared standardized RVs
- However, as always, we don't know σ
- Recall that we proposed $S_u^2 = \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$
- Turns out that $\frac{(n-1)S_u^2}{\sigma^2} \sim \chi^2_{n-1}$ (call this χ^2-RV W)

$$\begin{aligned} W &= \frac{(n-1)S_u^2}{\sigma^2} \\ &= \frac{(n-1)\frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2}{\sigma^2} \\ &= \frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2 \end{aligned}$$

- Since $\sum_i \left(\frac{Y_i - \mu}{\sigma}\right)^2 \sim \chi^2_{n-1}$, $W \sim \chi^2_{n-1}$

Putting it together

- Remember, we've been using $\frac{\bar{Y}-\mu}{S_u/\sqrt{n}}$ to calculate our test statistics
- But look at this more closely...this is equivalent to $\frac{Z}{\sqrt{W/\nu}}$!

$$
\begin{aligned}
\frac{\bar{Y}-\mu}{S_u/\sqrt{n}} &= \sqrt{n}\frac{\bar{Y}-\mu}{S_u} \\
&= \frac{\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma}}{\frac{S_U}{\sigma}} \\
&= \frac{\frac{(\bar{Y}-\mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_U^2}{\sigma^2}/(n-1)}} \\
&= \frac{Z}{\sqrt{W/\nu}}
\end{aligned}
$$

Small Sample Inference

- These findings allow us to construct $100(1-\alpha)$% CIs around estimates of μ drawn from **small samples** of a Normally distributed population
- For example:

$$
\begin{aligned}
P(-t_{\alpha/2,\nu} \leq T \leq t_{\alpha/2,\nu}) &= 1-\alpha \\
P\left(-t_{\alpha/2,\nu} \leq \frac{\bar{Y}-\mu}{S_U/\sqrt{n}} \leq t_{\alpha/2,\nu}\right) &= 1-\alpha \\
P\left(-t_{\alpha/2,\nu}\frac{S_U}{\sqrt{n}} - \bar{Y} \leq -\mu \leq t_{\alpha/2,\nu}\frac{S_U}{\sqrt{n}} + \bar{Y}\right) &= 1-\alpha \\
P\left(t_{\alpha/2,\nu}\frac{S_U}{\sqrt{n}} + \bar{Y} \geq \mu \geq t_{\alpha/2,\nu}\frac{S_U}{\sqrt{n}} - \bar{Y}\right) &= 1-\alpha
\end{aligned}
$$

Small Sample Inference

- Note that the CI here is VERY similar to that we used earlier
  - Instead of $\bar{Y} \pm z_{\alpha/2} \frac{S_U}{\sqrt{n}}$ we use $\bar{Y} \pm t_{\alpha/2,\nu} \frac{S_U}{\sqrt{n}}$
- Also, hypothesis testing is almost identical
  - H_0: $\mu = \mu_0$
  - H_A: $\mu \neq \mu_0$ (two-tailed) or $\mu > \mu_0$ or $\mu < \mu_0$ (one-tailed)
  - Test stat: $T = \frac{\bar{Y} - \mu}{S_U/\sqrt{n}}$
  - RR: $|t| > t_{\alpha/2,\nu}$

Difference in Means

- $\mu_1 - \mu_2$ style tests are very similar as well
    - Recall in the large-n case, the test statistic was $Z = \frac{\bar{y}_1 - \bar{y}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
- Here, we make an additional assumption that $\sigma_1^2 = \sigma_2^2$. We do this for two reasons:
    1. With small n, it is hard to get good estimates of σ^2 (remember it takes lots of n for the consistency of S_U^2 to kick in)
    2. It's a minor sin: we are already assuming Y_1 and Y_2 are Normal!
- We can then calculate the "s-squared pooled" as $s_p^2 = \frac{(n_1 - 1)S_{U1}^2 + (n_2 - 1)S_{U2}^2}{n_1 + n_2 - 2}$
- And the test statistic then becomes $T = \frac{\bar{y}_1 - \bar{y}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Conclusion

- To conduct small n inference, we have been piling up the assumptions
- What if the underlying random variables are not Normal?
  - Statisticians test sensitivity by sampling from known nonnormal distributions
  - Moderate departures from normality have little effect on the probability distribution of the test statistic
  - (But beware!)
- Note that the t is indistinguishable from the Normal at high DF, a t-test is indistinguishable from a z-test at most sample sizes we typically deal with
  - Thus, we are always claiming to be running t-tests when in fact we might as well be running z-tests

End of first half

- Recap of assumptions:

1. **Random sample:** yields *independent* and *identically distributed* observations
   - Assures us that $\bar{Y}$ is unbiased estimator of μ
2. **Large enough n:** buys us three assumptions!
   - **CLT:** with large enough n, sampling distribution for $\bar{Y}$ is Normal
   - **Consistency:** with large enough n, variance of estimator goes to zero
   - **Slutzky's Theorem:** Ratio of function that converges to standard Normal over a function that converges to 1 will itself converge to standard Normal
3. If n is small, then need $Y \sim \mathcal{N}(\mu, \sigma^2)$