t-tests

Recap: t distribution

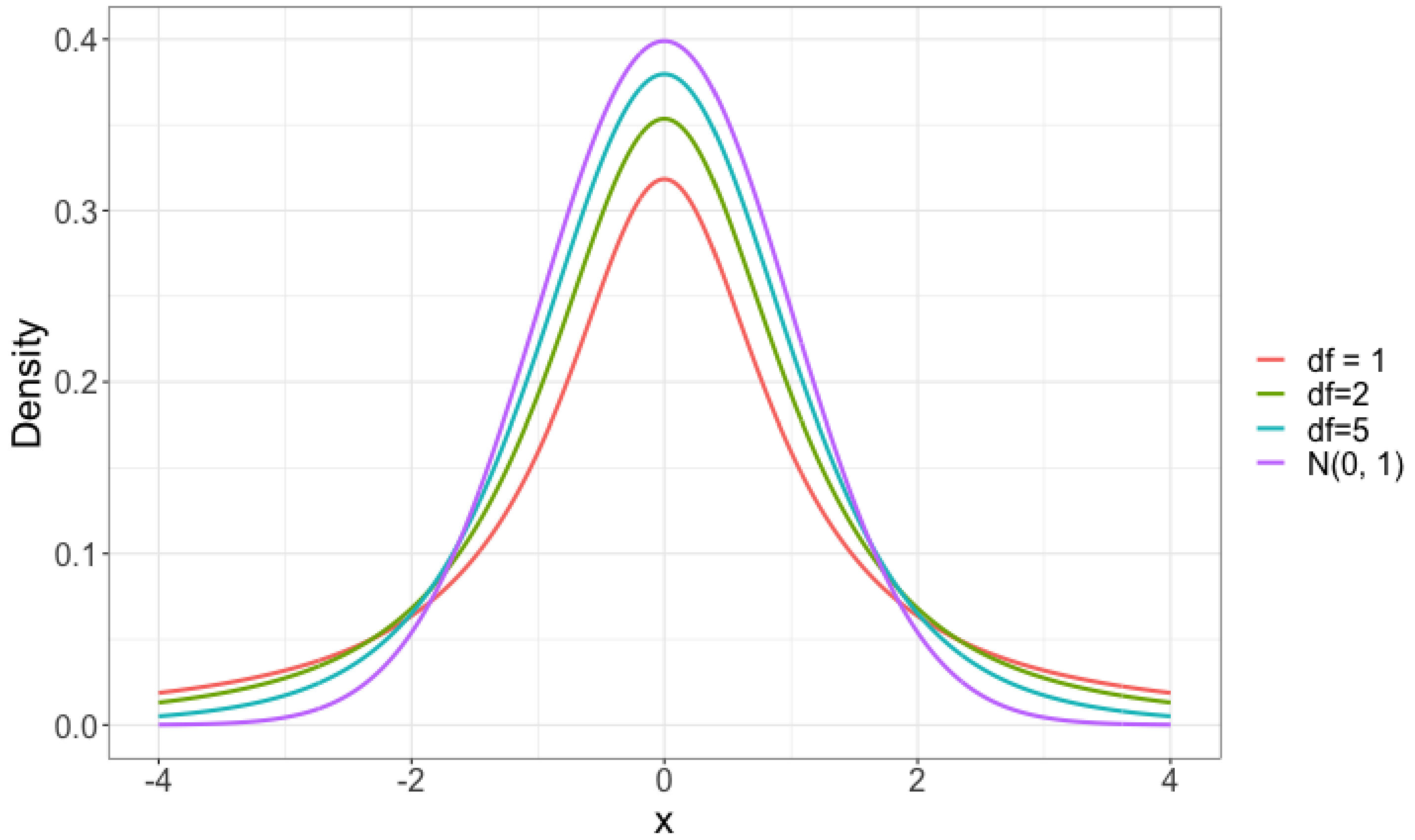
If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{s} \sim t_{n-1}$$

Definition: Let $Z \sim N(0,1)$ and $V \sim \chi^2_d$ be independent. Then

$$T = \frac{Z}{\sqrt{V/d}} \sim t_d$$

t-distribution



Cochran's theorem

Let $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1)$, and let $Z = [Z_1, \ldots, Z_n]^T$. Let $A_1, \ldots, A_k \in \mathbb{R}^{n \times n}$ be symmetric matrices such that $Z^T Z = \sum_{i=1}^{k} Z^T A_i Z$, and let $r_i = rank(A_i)$. Then the following are equivalent:

- $r_1 + \cdots + r_k = n$
- The $Z^T A_i Z$ are independent
- Each $Z^T A_i Z \sim \chi^2_{r_i}$

Application to t-tests

Global F tests for linear regression

Test for a population mean

Suppose $Y_1, \dots, Y_n \overset{iid}{\sim} Bernoulli(p)$. We want to test

$$H_0 : p = p_0 \qquad H_A : p \neq p_0$$

Wald test:

> Why is a t-test not appropriate?

Test for logistic regression

$$Y_i \sim Bernoulli(p_i) \qquad \log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i$$

We want to test

$$H_0 : C\beta = \gamma_0 \qquad H_A : C\beta \neq \gamma_0$$

> Why is a t-test not appropriate?

Philosophical question

- If $X_1, \ldots, X_n$ are iid from a population with mean μ and variance σ^2, then $\frac{\sqrt{n}(\overline{X}_n - \mu)}{s} \xrightarrow{d} N(0, 1)$
- If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\frac{\sqrt{n}(\overline{X}_n - \mu)}{s} \sim t_{n-1}$
- **Position 1:** For any reasonable sample size, the test statistic is approximately normal. And we never really have data from a normal distribution, so the t distribution is an approximation anyway. So always use the normal distribution
- **Position 2:** We always have a finite sample size, so our test statistic is never truly normal. And the t distribution is more conservative than the normal (heavier tails). So always use the t distribution

> With which position do you agree?