Lecture 22: t-tests

Issue: Wald tests with small n

The Wald test for a population mean μ relies on

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{s} \approx N(0, 1)$$

- $Z_n \xrightarrow{d} N(0, 1)$ as $n \to \infty$
- But for small n, Z_n is not normal, even if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

What is the exact distribution of $\frac{\sqrt{n}(\overline{X}_n - \mu)}{s}$?

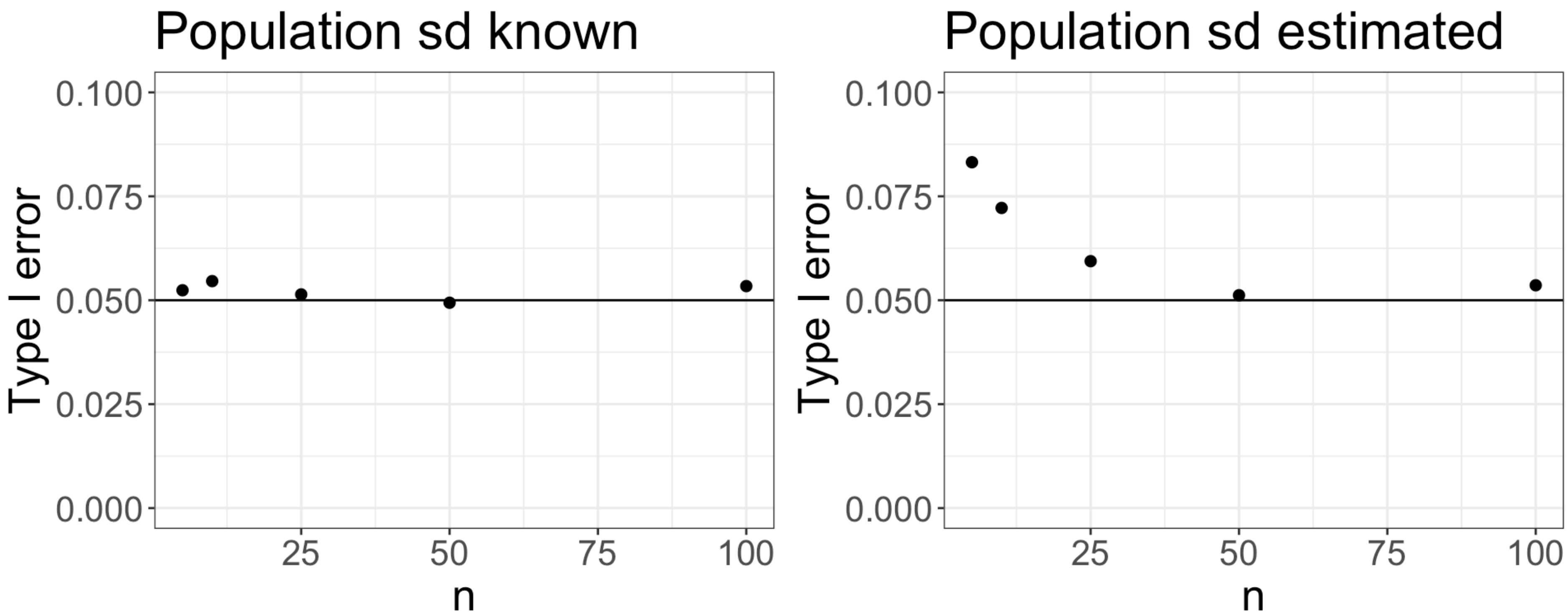
t-tests

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

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

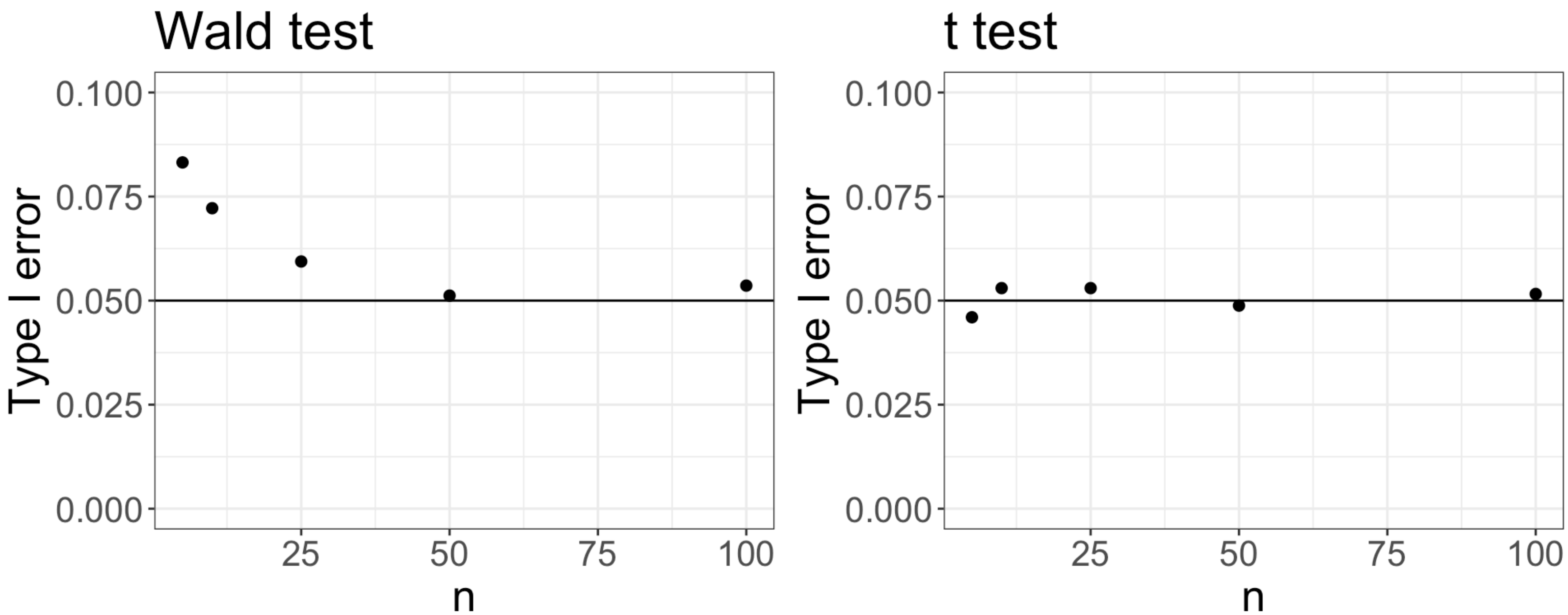
Class activity

Type I error rate with Normal distribution:



Class activity

Wald test vs. t-test:



Philosophical question

- **Position 1:** We should always use a Wald test to test hypotheses about a population mean
- **Position 2:** We should always use a t-test to test hypotheses about a population mean

With which position do you agree?

t distribution

CLT: $X_1, \ldots, X_n$ iid w/ mean μ & var. σ^2

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

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

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

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

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$\frac{s}{\sigma} = \sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}} = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2 (n-1)}}$$

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

Apply: $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{s} = \dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \cdot \dfrac{\sigma}{s}$

- $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$ (b/c data $\overset{iid}{\sim} N(\mu, \sigma^2)$, this is exact $\forall n$)
- $(n-1)\dfrac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, $\bar{X} \perp\!\!\!\perp s^2$ (independent)

if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then $\bar{X}_n \underset{(exact)}{\sim} N(\mu, \frac{\sigma^2}{n})$

more generally: X, Y independent $N(\mu_X, \sigma_X^2)$ $N(\mu_Y, \sigma_Y^2)$

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

---

what we want to show:

(Remember: $V_1 \sim \chi^2_{d_1} \perp V_2 \sim \chi^2_{d_2}$ $\Rightarrow V_1 + V_2 \sim \chi^2_{d_1+d_2}$)

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \frac{(n-1)s^2}{\sigma^2} \perp \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

(lose a degree of freedom by approximating μ)

we know: $\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$ (sum of n indep. squared $N(0,1)$)

Now,

$$\underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2}_{\chi^2_n} = \underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2}_{\chi^2_{n-1}\ ?} + \underbrace{n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2}_{\chi^2_1\ ?}$$

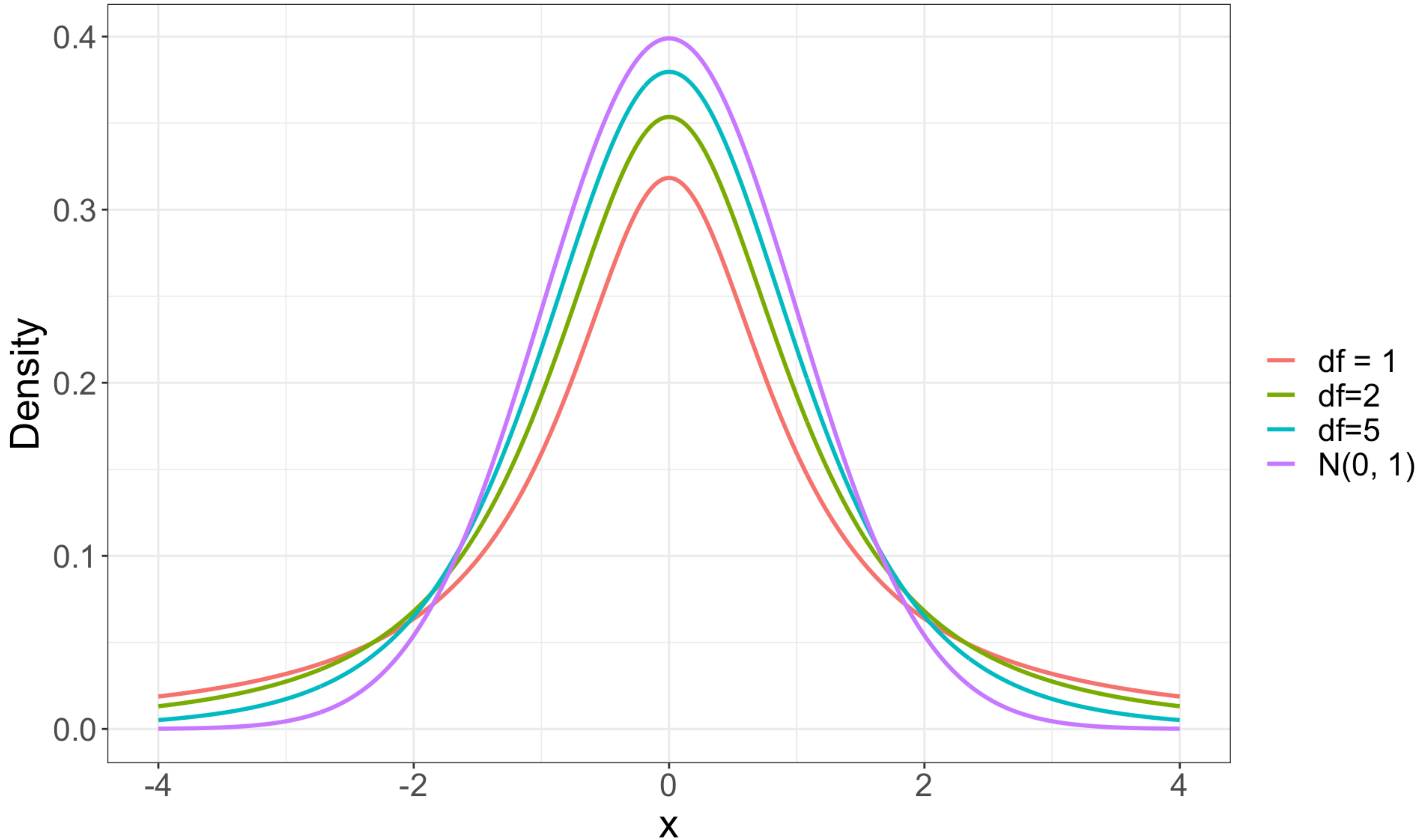
$$\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2 = \sum_i \left(\frac{x_i-\bar{x}+\bar{x}-\mu}{\sigma}\right)^2$$

$$= \sum_i \left[\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 + \left(\frac{\bar{x}-\mu}{\sigma}\right)^2 + 2\left(\frac{x_i-\bar{x}}{\sigma}\right)\left(\frac{\bar{x}-\mu}{\sigma}\right)\right]$$

$$= \sum_i \left(\frac{x_i-\bar{x}}{\sigma}\right)^2 + \underbrace{\sum_i \left(\frac{\bar{x}-\mu}{\sigma}\right)^2}_{n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2} + \cancel{2\left(\frac{\bar{x}-\mu}{\sigma}\right)\underbrace{\sum_i \left(\frac{x_i-\bar{x}}{\sigma}\right)}_{0}}$$

$$\sum_i (x_i-\bar{x}) = 0$$

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

Let $Z_i = \frac{X_i - \mu}{\sigma} \overset{iid}{\sim} N(0,1)$ $\quad Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$

$$\sum_i Z_i^2 = Z^T Z$$

$$\Rightarrow Z^T Z = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2}_{Z^T A_1 Z} + \underbrace{n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2}_{Z^T A_2 Z}$$

want to find A_1, A_2 st $Z^T A_1 Z$, $Z^T A_2 Z$

$$n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 = n(\bar{Z})^2 = \frac{1}{n}\left(\sum_j Z_j\right)^2$$

$$= \frac{1}{n}[Z_1 \cdots Z_n] \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \\ \vdots & & \ddots & \\ 1 & & & 1 \end{bmatrix}}_{J_n = n \times n \text{ matrix of all 1s}} \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

$$= Z^T \left(\frac{1}{n} J_n\right) Z$$

$A_2 = \frac{1}{n} J_n$ $\quad$ $\operatorname{rank}(A_2) = 1$

$$\sum_i \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = \sum_i (Z_i - \bar{Z})^2 = \sum_i \left(Z_i - \frac{1}{n}\sum_j Z_j\right)^2$$

$$= Z^T Z - Z^T\left(\frac{1}{n}J_n\right)Z$$

$$= Z^T\left(I_n - \frac{1}{n}J_n\right)Z$$

$$Z^T Z = Z^T\left(I_n - \frac{1}{n}J_n\right)Z + Z^T\left(\frac{1}{n}J_n\right)Z$$

$$Z^T Z = \underbrace{\sum_i \left(\frac{x_i - \bar{x}}{\sigma}\right)^2} + Z^T\left(\frac{1}{n}J_n\right)Z$$

$$= Z^T Z - Z^T\left(\frac{1}{n}J_n\right)Z$$

$$= Z^T\left(Z - \frac{1}{n}J_n Z\right)$$

$$= Z^T\left(I - \frac{1}{n}J_n\right)Z$$

Cochran's theorem:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\perp \quad n\left(\frac{\bar{x}-\mu}{\sigma}\right) \sim \chi^2_1$$

$A_1 = I - \frac{1}{n}J_n$

$\text{rank}(A_1) = n-1$

$A_2 = \frac{1}{n}J_n$

$\text{rank}(A_2) = 1$

$$\sum_i (z_i - \bar{z})^2$$

$$\sum_i z_i^2 - 2\sum_i \cancel{(z_i - \bar{z})} + \sum_i \boxed{\bar{z}}^2$$

$$= \underbrace{\sum_i z_i^2}_{z^T z} + \underbrace{n(\bar{z})^2}_{z^T\left(\frac{1}{n} J_n\right) z}$$