Lecture 18: t-tests

Ciaran Evans

Previously: Wald tests for a population mean

Suppose $X_1, ..., X_n$ are an iid sample from a population with mean μ and variance σ^2. We wish to test

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$$

- If σ^2 is known:

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

- If σ^2 is unknown:

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

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

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

t-tests

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

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

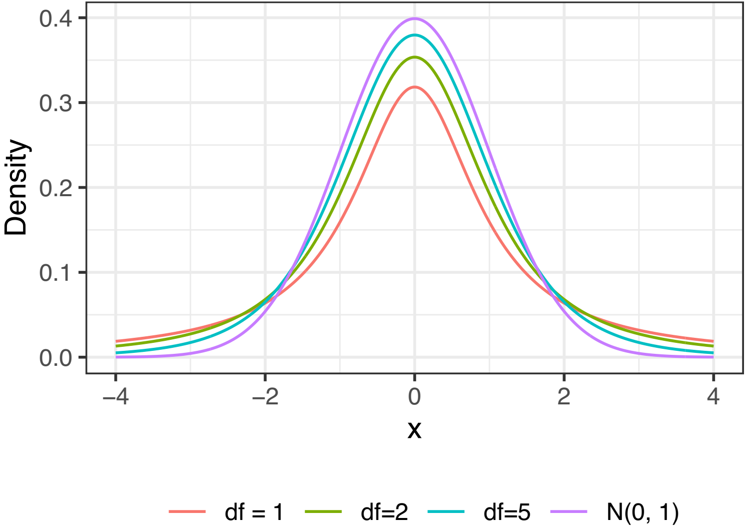
$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{s} \sim t_{n-1}$$

t distribution
with n-1 df

t-distribution

Formal: $X_1, X_2, X_3, \ldots$ $X_i \sim t_i$

$X_n \xrightarrow{d} N(0,1)$



As df $\uparrow$, $t \to N(0,1)$

t distribution

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

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

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$$

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

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

$$\frac{\sqrt{n}(\bar{X} - \mu)}{s} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{s/\sigma} = \frac{Z}{s/\sigma} = \frac{Z}{\sqrt{((n-1)s^2/\sigma^2)/(n-1)}}$$

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

$\Rightarrow$ WTS: $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ and $\dfrac{(n-1)s^2}{\sigma^2} \perp\!\!\!\perp \dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$

(independent)

What we want to show

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

$$(n-1)\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_i (X_i-\overline{X})^2 = \sum_i \left(\frac{X_i-\overline{X}}{\sigma}\right)^2$$

know: $\frac{X_i-\mu}{\sigma} \overset{iid}{\sim} N(0,1) \Rightarrow \sum_i \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n$

know: if $V_1 \sim \chi^2_{d_1}$, $V_2 \sim \chi^2_{d_2}$ independent, then $V_1+V_2 \sim \chi^2_{d_1+d_2}$

idea: $$\underbrace{\sum_i \left(\frac{X_i-\mu}{\sigma}\right)^2}_{\chi^2_n} = \underbrace{\sum_i \left(\frac{X_i-\overline{X}}{\sigma}\right)^2}_{\text{hope is } \chi^2_{n-1}} + \underbrace{\text{something}}_{\chi^2_1}$$

Decomposing the sum of squares

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

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 + n\left(\frac{\overline{X}-\mu}{\sigma}\right)^2$$

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

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

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

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

Cochran's theorem

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

Let $Z_1, ..., Z_n \overset{iid}{\sim} N(0,1)$, and let $Z = [Z_1, ..., Z_n]^T$. Let $A_1, ..., A_k \in \mathbb{R}^{n \times n}$ 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 $Z_i = \frac{X_i - \mu}{\sigma} \overset{iid}{\sim} N(0,1)$ $\quad Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$

$$\underbrace{\sum_i \left(\frac{X_i - \mu}{\sigma}\right)^2}_{Z^T Z} = \sum_i \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2$$

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

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

$$= \frac{1}{n}[Z_1 \cdots Z_n] \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & 1 \\ 1 & 1 & & \end{bmatrix}}_{J_n} \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

$J_n = n \times n$ matrix of all 1s

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

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

Application to t-tests

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

$$Z^TZ - Z^T(\tfrac{1}{n}J_n)Z$$

$$= Z^T(I - \tfrac{1}{n}J_n)Z$$

$$\Rightarrow \quad A_1 = I - \tfrac{1}{n}J_n \qquad\qquad A_2 = \tfrac{1}{n}J_n$$

Now want to find the ranks:

Fact: if A is idempotent $(A = A^2)$, then

$$\text{rank}(A) = \text{tr}(A)$$

$$\text{tr}(A_2) = \tfrac{1}{n}\,\text{tr}(J_n) = \tfrac{1}{n}\cdot n = 1$$

$$\text{tr}(A_1) = \text{tr}(I - \tfrac{1}{n}J_n) = \text{tr}(I) - \text{tr}(\tfrac{1}{n}J_n)$$

$$= n - 1$$

Application to t-tests

So: $\underbrace{\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2}_{\substack{Z^T Z \\ \sim \chi^2_n}} = \underbrace{\sum_i \left(\frac{x_i-\bar{x}}{\sigma}\right)^2}_{Z^T A_1 Z} + \underbrace{n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2}_{Z^T A_2 Z}$

$\text{rank}(A_1) = n-1 \qquad \text{rank}(A_2) = 1$

By Cochran's theorem:

- $Z^T A_1 Z \perp\!\!\!\perp Z^T A_2 Z \Rightarrow \frac{(n-1)s^2}{\sigma^2} \perp\!\!\!\perp \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right)$
- $Z^T A_1 Z \sim \chi^2_{n-1}$, $\quad n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2 \sim \chi^2_1$

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

$\Rightarrow \quad \frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{\sqrt{n}(\bar{x}-\mu)/\sigma}{\sqrt{((n-1)s^2/\sigma^2)/(n-1)}} \sim t_{n-1} \qquad //$