Hypothesis Testing II: One Sample *t* Tests

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

OCTOBER 11, 2016

What if you don’t know variance?

- *t*-test allows you to test hypothesis if you don’t know variance
- Sometimes called “small sample test”: same as *z* test with enough observations
- William Gossett: check that yeast content matched Guiness’s standard (but couldn’t publish)
- I.e., checking whether yeast content equal to μ_0

t-test statistic

- Need to estimate variance

$$s^2 = \sum_i \frac{(x_i - \bar{x})^2}{N-1} \tag{1}$$

- $n-1$ removes bias (expected value is less than truth)
- Test statistic looks similar

$$T \equiv \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{N}}} \tag{2}$$

Degrees of Freedom

- Like χ^2, t-distribution parameterized by degrees of freedom
- $\nu = N - 1$ degress of freedom

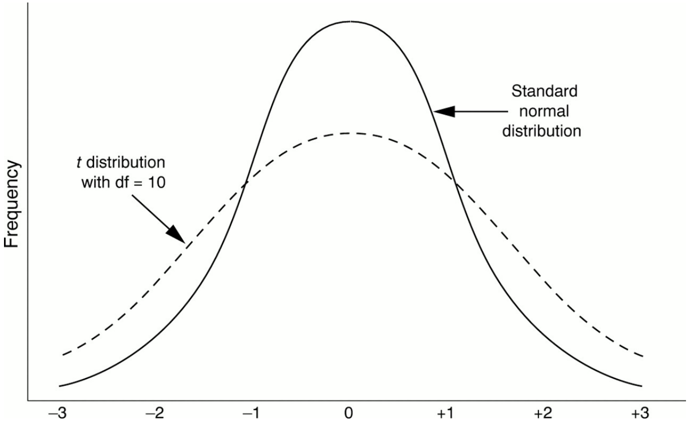
PDF

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{3}$$

CDF

$$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \tag{4}$$

Shape of t-distribution



Example

- Suppose observe $\{0, 1, 2, 3, 4, 5\}$
- Test whether $\mu \neq 1$

Example

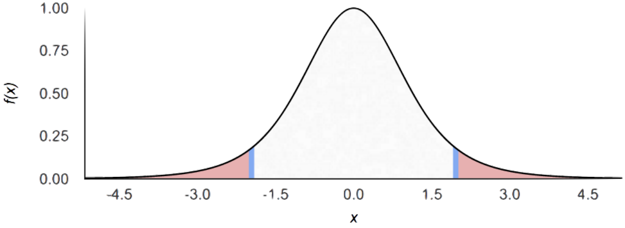
- Suppose observe $\{0, 1, 2, 3, 4, 5\}$
- Test whether $\mu \neq 1$
- $\bar{x} = 2.5$, $s^2 = 3.5$

Example

- Suppose observe $\{0,1,2,3,4,5\}$
- Test whether $\mu \neq 1$
- $\bar{x} = 2.5$, $s^2 = 3.5$
- $T = \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{N}}} = \frac{2.5-1.0}{\sqrt{\frac{3.5}{6}}} = 1.9640$

Example

- Suppose observe $\{0, 1, 2, 3, 4, 5\}$
- Test whether $\mu \neq 1$
- $\bar{x} = 2.5$, $s^2 = 3.5$
- $T = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} = \frac{2.5 - 1.0}{\sqrt{\frac{3.5}{6}}} = 1.9640$
- Double area under the at two tailed CDF



$\mu = E(X) = 0 \qquad \sigma = SD(X) = 1.291 \qquad \sigma^2 = Var(X) = 1.667$