College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Hypothesis Testing II: z tests, t tests, and confidence intervals

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

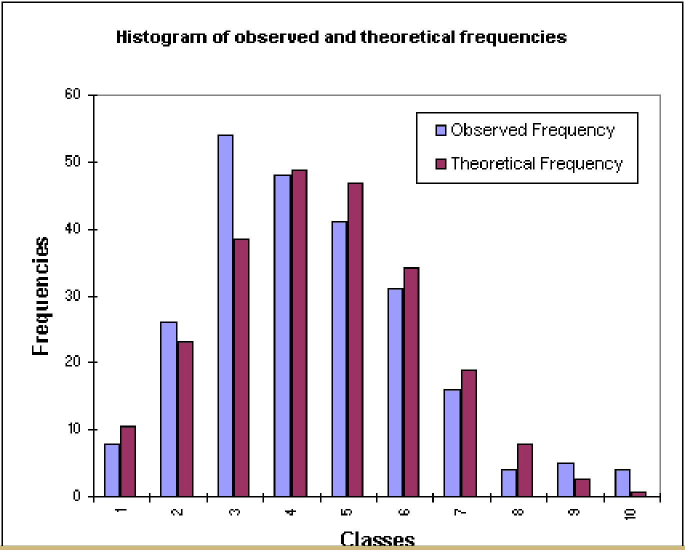
Same Story

- Form a hypothesis
- Get some data
- Compute test statistic
- Reject hypothesis if p-value low enough

Important difference

- Assumes data come from normal distribution
- Can cause problems if diverges too much normal
- (How can you tell?) χ^2 goodness of fit

Goodness of Fit



Hypothesis Testing II: z tests

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

z-test

- Suppose we have one observation from normal distribution with mean μ and variance σ^2
- Given an observation x we can compute the Z score as

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

- H_0: Our observation came from the normal distribution with $\mu = \mu_0$
  - Assume same known variance σ

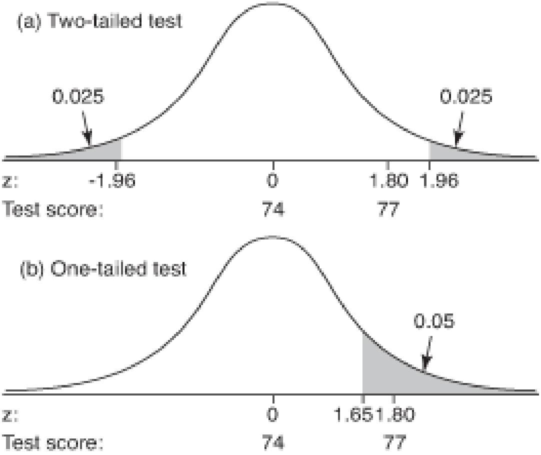
z-test

- Suppose we have one observation from normal distribution with mean μ and variance σ^2
- Given an observation x we can compute the Z score as

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

- H_0: Our observation came from the normal distribution with $\mu = \mu_0$
    - Assume same known variance σ
    - But we need to be more specific!

Two-tailed vs. one-tailed tests



- Two tail: Alternative $\mu \neq \mu_0$
- One tail: Alternative $\mu > \mu_0$

Multiple observations

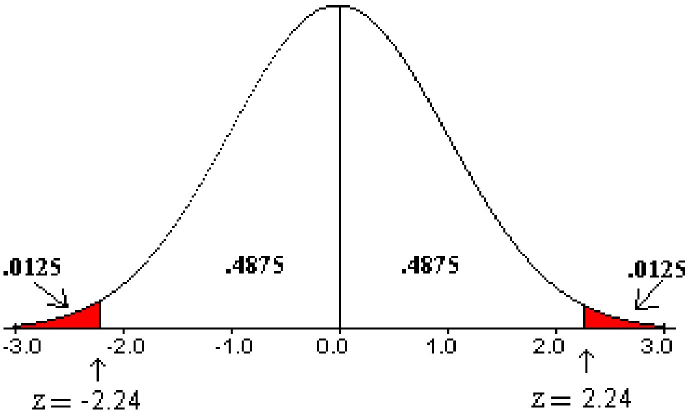
If you observe $x_1 \dots x_N$ from distribution with mean μ, test whether $\mu \neq \mu_0$

- Compute test statistic

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{N}} \tag{2}$$

- If H_0 were true, $\bar{x}$ would be normal distribution with μ_0 and variance $\frac{\sigma^2}{N}$
- Now we can decide when to reject based on normal CDF

When to reject (two-tailed)



College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Hypothesis Testing II: Confidence Intervals

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

Conveying Uncertainty

- Suppose you make a measurement (e.g., a poll)
- Often useful to show all μ_0 that could be excluded given the measurement
- Fixed α
- Confidence interval

z-distribution Confidence

- Assume known variance σ
- You observe $\{x_1 \dots x_N\}$
- Obtain mean $\bar{x}$
- Recall test statistic

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \tag{1}$$

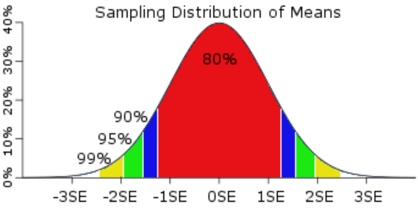
z-distribution Confidence

- Assume known variance σ
- You observe $\{x_1 \dots x_N\}$
- Obtain mean $\bar{x}$
- Recall test statistic

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \tag{1}$$

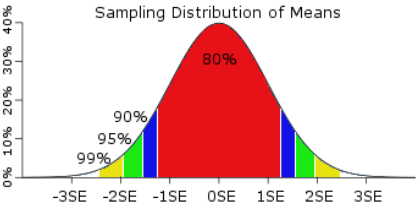
- For what μ would we not reject the null?
- Note that $\bar{x}$ and μ are symmetric

From the Distribution



- Set $\alpha = 0.05$
- Solve for $\mu = \text{NorCDF}(0, 0.025)$

From the Distribution



- Set $\alpha = 0.05$
- Solve for $\mu = \text{NorCDF}(0, 0.025)$
- $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$
- Some people just use 2.0
- More samples $\longrightarrow$ tighter bounds

Hypothesis Testing II: One Sample t Tests

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

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

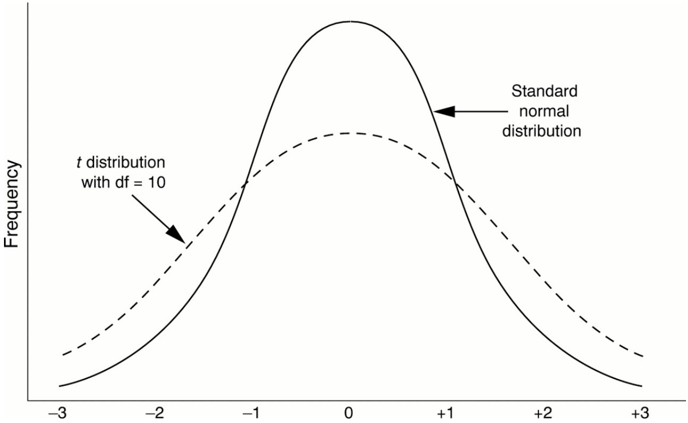
PDF

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

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

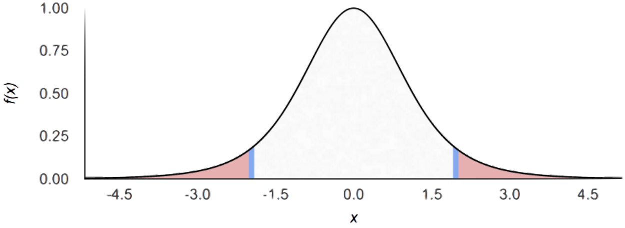
- Suppose observe $\{0, 1, 2, 3, 4, 5\}$
- Test whether $\mu \neq 1$
- $\bar{x} = 2.5$, $s^2 = 3.5$

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



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

Hypothesis Testing II: Two Sample *t* Tests

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

Comparing Two Samples

- Thus far, we've tested whether data is consistent with mean μ
- What if we want to test whether two samples are from the same distribution

Comparing Two Samples

- Thus far, we've tested whether data is consistent with mean μ
- What if we want to test whether two samples are from the same distribution
- Two-Sample t-test

Two-Sample (unpooled)

- Two samples $X_1 = \{x_{1,1}, x_{1,2} \dots x_{1,N_1}\}$ and $X_2 = \{x_{2,1}, x_{2,2} \dots x_{2,N_2}\}$
- Doesn't assume that variance is the same for both samples (unpooled)
- Compute mean and sample variance for sample 1 $(\bar{x}_1, s_1^2)$ and sample 2 $(\bar{x}_2, s_2^2)$

Test Statistic

- T-statistic

$$T = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{1}$$

- Plug into t-distrubtion with

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{2}$$

- Intuition: Difference between $\bar{x}_1$ and $\bar{x}_2$ has variance that's an interpolation between the two samples

Test Statistic

- T-statistic

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{1}$$

- Plug into t-distrubtion with

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{2}$$

- Intuition: Difference between $\bar{x}_1$ and $\bar{x}_2$ has variance that's an interpolation between the two samples
- Two-tailed vs. one-tailed distinction still applies

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

(4)

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

$$= \frac{\left(\frac{1}{4} + \frac{2}{8}\right)^2}{\frac{1}{3}\left(\frac{1}{4}\right)^2 + \frac{1}{7}\left(\frac{2}{8}\right)^2} \tag{4}$$

(5)

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

$$= \frac{\left(\frac{1}{4} + \frac{2}{8}\right)^2}{\frac{1}{3}\left(\frac{1}{4}\right)^2 + \frac{1}{7}\left(\frac{2}{8}\right)^2} \tag{4}$$

$$= \frac{\frac{1}{4}}{\left(\frac{1}{4}\right)^2\left[\frac{1}{3} + \frac{1}{7}\right]} = \frac{4}{\frac{10}{21}} = \frac{42}{5} \tag{5}$$

Hypothesis Testing II

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

APRIL 24, 2017

Details

- Important to pre-register hypothesis
- Check assumptions of tests: distribution, randomness, response
- Many other tests that are possible

Problems with Statistical Tests' Philosophy

- Failing to reject the null does not prove the null
- Cannot exclude data you don't like
- Confusing statistical significance with practical significance ($\mu \neq 0$ could mean $\mu = 0.000001$)

Problems with Statistical Tests' Philosophy

- Failing to reject the null does not prove the null
- Cannot exclude data you don't like
- Confusing statistical significance with practical significance ($\mu \neq 0$ could mean $\mu = 0.000001$)

Often better:

- Present plot with confidence intervals, let people decide for themselves
- Create comprehensive model with explainable parameters, compute intervals for parameters