Hypothesis Testing II: $z$ tests

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul

OCTOBER 11, 2016
Suppose we have one observation from normal distribution with mean $\mu$ and variance $\sigma^2$.

Given an observation $x$ we can compute the $Z$ score as

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

$H_0$: Our observation came from the normal distribution with $\mu = \mu_0$

- Assume same known variance $\sigma$
**z-test**

- Suppose we have one observation from normal distribution with mean $\mu$ and variance $\sigma^2$.
- Given an observation $x$ we can compute the $Z$ score as
  \[
  Z = \frac{x - \mu}{\sigma} \quad (1)
  \]
- $H_0$: Our observation came from the normal distribution with $\mu = \mu_0$
  - Assume same known variance $\sigma$
  - But we need to be more specific!
Two-tailed vs. one-tailed tests

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

![Diagram](image)
Multiple observations

If you observe \( x_1 \ldots x_N \) from distribution with mean \( \mu \), 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 \( \mu_0 \) and variance \( \frac{\sigma^2}{N} \)

- Now we can decide when to reject based on normal CDF
When to reject (two-tailed)