Hypothesis Testing II: z tests

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

OCTOBER 6, 2016

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

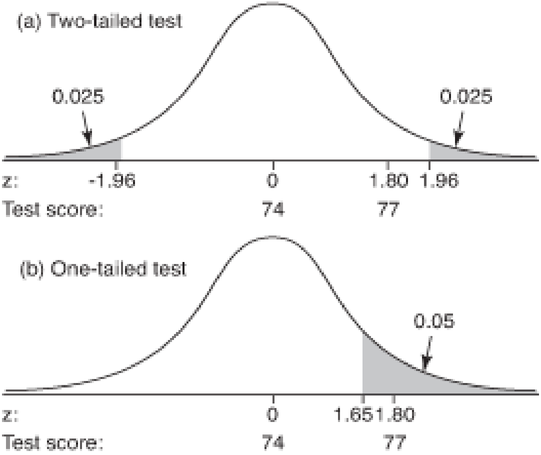
z-test

- 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

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

- 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)