



Hypothesis Testing II: Confidence Intervals

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

OCTOBER 11, 2016

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

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \quad (1)$$

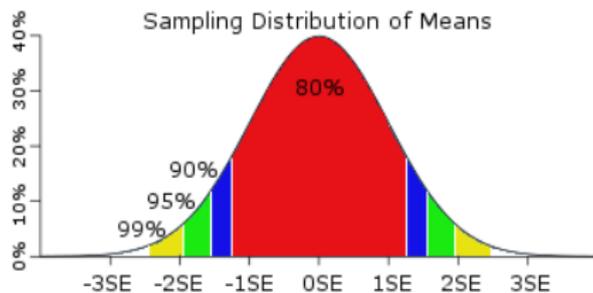
z-distribution Confidence

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

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \quad (1)$$

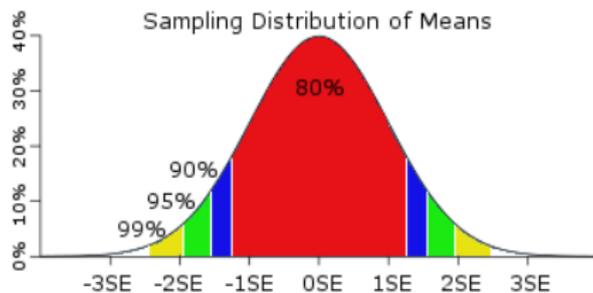
- For what μ would we not reject the null?
- Note that \bar{x} and μ are symmetric

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 \rightarrow tighter bounds