Lecture 17

Inference for One Population Mean

STAT 8010 Statistical Methods I
September 30, 2019

Whitney Huang
Clemson University

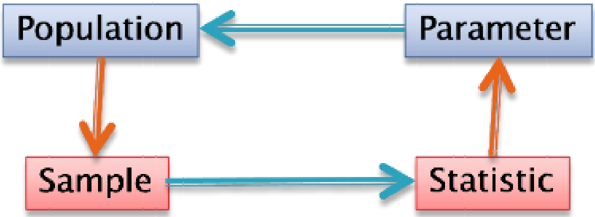
Statistical Inference

For the rest of the semester, we will focus on conducting statistical inferences for the following tasks:

- Estimating one population mean
- Comparing two population means
- Comparing more than two population means
- Estimating population proportions
- Estimating relationship between response variable and predictors

Statistical Science: Use Sample to Learn About the Population

- We use parameters to describe the population
  Example: mean (μ_X); variacen (σ_X^2)



- We use statistics of a sample (given that the sampling was done properly) to infer the population
  Example: sample mean $(\bar{X})$; sample variance (s_X^2)

Estimating One Population Mean

Goal: To estimate the population mean using a (representative) sample:

- The sample mean, $\bar{X}_n = \frac{\sum_i^n X_i}{n}$, is a reasonable point estimate of the population mean μ_X
- Need to quantify the level of uncertainty of the point estimate $\Rightarrow$ Interval estimation
- Need to figure out the sampling distribution of $\bar{X}_n$ in order to construct interval estimates $\Rightarrow$ Central Limit Theorem (CLT)

Central Limit Theorem (CLT)

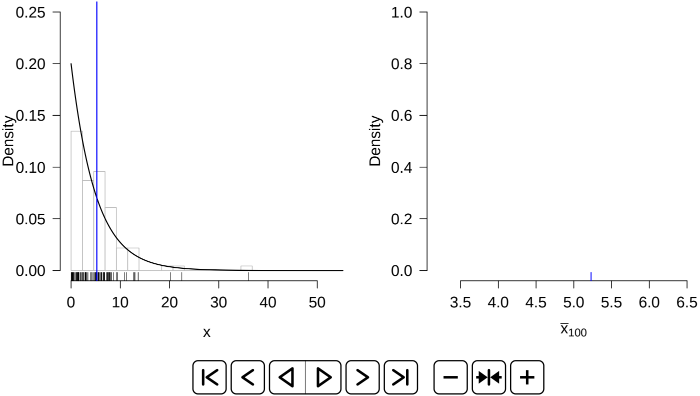
CLT

The **sampling distribution** of $\bar{X}_n$ will become approximately **normally distributed** as the **sample size (n) becomes "large"**, **regardless of the shape of the population distribution!**

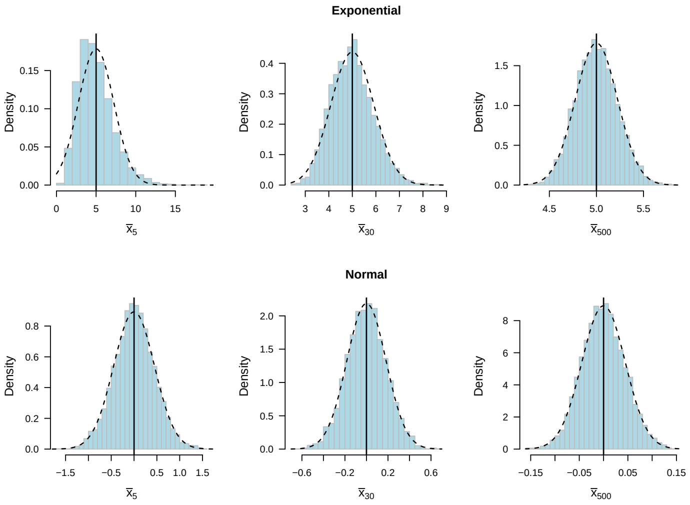
Let $X_1, X_2, \cdots, X_n$ be a random sample from a population X with $\mu = \mathbb{E}[X]$ and $\sigma^2 = \text{Var}[X]$. Then $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{d} \text{N}(\mu, \frac{\sigma^2}{n})$ as $n \to \infty$.

CLT In Action

1. Generate 100 (n) random numbers from an Exponential distribution (population distribution)
2. Compute the sample mean of these 100 random numbers
3. Repeat this process 120 times



CLT: Sample Size (n) and the Normal Approximation



Why CLT is important?

- CLT tells us the **distribution** of our estimator

$$\bar{X}_n \approx \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

  - The distribution of $\bar{X}_n$ is center around the true mean μ
  - The variance of $\bar{X}_n$ is decrease with n
  - With normality approximation of the sampling distribution of $\bar{X}_n$, we can perform interval estimation about μ

- Applications: Confidence Interval, Hypothesis testing

Confidence Intervals (CIs) for μ

CLEMSON UNIVERSITY

- Let's assume we know the population σ^2 (will relax this assumption later on)
- $(1-\alpha) \times 100\%$ CI for μ:

$$\left[\bar{X}_n - z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}\right],$$

  where $z_{(1-\frac{\alpha}{2})}$ is the $1-\frac{\alpha}{2}$ percentile of $Z \sim \mathrm{N}(0,1)$

- $\frac{\sigma}{\sqrt{n}}$ is the standard error of $\bar{X}_n$, that is, the standard deviation of its sampling distribution

Making Sense of Confidence Intervals

For any $\alpha \in (0, 1)$:

$$\begin{aligned}
&\mathbb{P}\left(\bar{X}_n - z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-z_{(1-\frac{\alpha}{2})} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{(1-\frac{\alpha}{2})}\right) \\
&= \mathbb{P}\left(-z_{(1-\frac{\alpha}{2})} \leq Z \leq z_{(1-\frac{\alpha}{2})}\right) \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha
\end{aligned}$$

Making Sense of Confidence Intervals Cont'd