

Lecture 27

# Central Limit Theorem

Sections 7.3 & 7.5

# Point Estimation

- Suppose we are given the following dataset:

**Salary & Training Status for EAI Managers**

<b>Annual Salary</b>	<b>Training Program</b>	<b>Annual Salary</b>	<b>Training Program</b>
\$49,094.30	Yes	\$55,924.00	Yes
\$53,263.90	Yes	\$49,092.30	Yes
\$49,643.50	Yes	\$57,268.40	No
\$49,894.90	Yes	\$55,688.80	Yes
\$47,621.60	No	\$51,564.70	No

# Point Estimation

- Suppose we would like to know the average annual salary and standard deviation for EAI managers, but we do not have the salaries for ALL managers. Therefore, we would need to estimate that average. The same would apply if we wanted to find what proportion of managers completed the training program.
- Remember we discussed early in class that we could estimate  $\mu$  with  $\bar{x}$

# Point Estimation

- Sample Mean:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{519,056.40}{10} = \mathbf{\$51,905.64}$$

- Sample Standard Deviation:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \mathbf{\$75.94}$$

- Sample Proportion:

$$\bar{p} = \frac{x}{n} = \frac{3}{10} = \mathbf{0.3}$$

# Point Estimation

- **Sample Statistic** – A value used to estimate a population parameter.
  - $\mu \rightarrow \bar{x}$ ,  $\sigma/\sigma^2 \rightarrow s/s^2$ ,  $\pi \rightarrow \bar{p}$
- **Point Estimator** – The symbol used to denote the estimation of a population parameter.
  - $\bar{x}$ ,  $s$ ,  $\bar{p}$
- **Point Estimate** – The numerical value obtained for the point estimator.
  - \$51,905.64    \$75.94    0.3

# Sampling Distribution of $\bar{x}$

- The sampling distribution of  $\bar{x}$  is the probability distribution of all possible values of the sample mean  $\bar{x}$ .
- Example:
  - Suppose that the average age of all Purdue undergraduates is 20 years old. If I were to use this STAT 225 class as my sample, I might find that our average age is 21 but another class average is 19, etc.

# Central Limit Theorem

- This phenomena is called the Central Limit Theorem.
- It was first stated by Abraham De Moivre and generalized by Pierre Laplace; therefore, it is sometimes called the De Moivre – Laplace Theorem.
- [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html)

# Central Limit Theorem

- If a measurement is repeated many times, then the **sum or average** of the measurements is **approximately normally** distributed.
  - Don't let this get mixed up with the Normal approximation to the Binomial distribution.

# Central Limit Theorem

- **Theorem:** Let  $X_1, \dots, X_n$  be independent samples with  $E(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2$  for  $i = 1, \dots, n$ . Then for large  $n$  ( $n > 30$ ), the sum and the average of the  $X_i$ 's has approximately a Normal distribution.

Average: 
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \text{Normal} \left( \mu, \frac{\sigma^2}{n} \right)$$

Sum: 
$$\sum_{i=1}^n X_i \sim \text{Normal} (n\mu, n\sigma^2)$$

# Central Limit Theorem

- The larger  $n$  is, the more closely the PDF of the sum or average will resemble a Normal PDF.
- If the  $X_i \sim N(\mu, \sigma^2)$  to start with, then you do not need to meet the  $n > 30$  criteria for the sum and average to be normally distributed.

# CLT Example 1

- One has 100 light bulbs whose lifetimes are independent exponentials with mean 5 hours. If the bulbs are used one at a time, with a failed bulb being replaced immediately by a new one, what is the probability that there is still a working bulb after 525 hours?

# CLT Example 1

$$X_i \sim \text{Exp}\left(\frac{1}{5}\right) \quad P\left(\sum_{i=1}^{100} X_i > 525\right)$$

$$S = \sum_{i=1}^{100} X_i \sim N(\mu = 100(5), \sigma^2 = 100(25))$$

$$S \sim N(\mu = 500, \sigma^2 = 2500)$$

$$Z = \frac{525 - 500}{\sqrt{2500}} = 0.5$$

$$P(S > 525) = P(Z > 0.5) = 1 - P(Z < 0.5) = \mathbf{0.3085}$$

# CLT Example 2

- How often do you have to roll a fair 6-sided die so that the probability that the average of the scores being between 3.2 and 3.8 is at least 0.9?

$$P(3.2 < \bar{X} < 3.8) = 0.9$$

- **What is  $n$ ?**

# CLT Example 2

$$P(z_1 < Z < z_2) = 0.9 \Rightarrow z_1 = -1.645, z_2 = 1.645$$

$$A = \bar{X} \sim N\left(\mu = 3.5, \sigma^2 = \frac{2.92}{n}\right)$$

$$-1.645 = \frac{3.2 - 3.5}{\sqrt{\frac{2.92}{n}}} \Rightarrow n \approx 88$$