Set 21: Central Limit Theorem

Stat 260: July 5, 2024

Standardization Theorem: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with a normal distribution with mean μ and standard deviation σ. Then,

- The sample mean $\overline{X}$ has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

  $E[\bar{x}] = E[x_i] = \mu \quad V(\bar{x}) = \frac{V(x_i)}{n} = \frac{\sigma^2}{n} \quad SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

- The random variable $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has the standard normal distribution:

  $Z \sim \text{Normal}(\mu=0, \sigma=1)$

Terminology: $\dfrac{\sigma}{\sqrt{n}}$ is called the ***standard error (s.e.)*** of the sample mean $\overline{X}$.

Central Limit Theorem (CLT): Let $X_1, X_2, \ldots, X_n$ be an random sample of size n, where n is large ($n \geq 30$) from **any** distribution, with mean μ and standard deviation σ. Then,

- The sample mean $\overline{X}$ has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. $\rightarrow E[\bar{x}] = \mu$
- The random variable $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has the standard normal distribution: $SD[\bar{x}] = \frac{\sigma}{\sqrt{n}}$

Example 1: Suppose that the average lifespan of a certain population of cougar (*Puma concolor*) is 10 years, with a standard deviation of 1 year. Suppose that you take up a research assistant position and the observe the lifespans of 50 wild Vancouver Island cougars. What is the probability that the mean lifespan of the cougars in your study is at least 10.2 years?

Let X_i be the lifespan of Cougar i, for $i = 1, 2, \ldots, 50$

$X_i \sim$ unknown $\quad \mu = 10, \sigma = 1$

By CLT since $n = 50 \geq 30$, $\bar{X}$ (mean of $X_1, X_2, \ldots, X_{50}$) is normally distributed

$\bar{X} \sim \text{Normal}\left(\mu_{\bar{x}} = 10, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

$P(\bar{X} \geq 10.2) = 1 - P(\bar{X} \leq 10.2) = 1 - P\left(\frac{\bar{X} - \overset{\mu_{\bar{x}}}{\mu}}{\sigma/\sqrt{n}} \leq \frac{10.2 - 10}{1/\sqrt{50}}\right)$

$= 1 - P(Z \leq 1.41) = 1 - 0.9207 = 0.0793$

Extra Example: A certain publisher knows that 10% of all the books it prints contain some variety of printing error (torn pages, smeared ink, off-centre pages, etc); assume that printing errors in books are independent of one another. Suppose that the publisher sells a new cookbook in cases of 20 to book distributors. If a certain distributor purchases 10 cases of the cookbook, what is the probability that the average number cookbooks with printing errors in each case is greater than 3?

Answer: 0.0091

Readings: Swartz 5.6 [EPS 4.4]
Practice problems: EPS : 4.17, 4.21, 4.25.