The Central Limit Theorem

Sybil Shaver

May 18, 2017

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean
- $\bar{X} := \frac{\sum_{i=1}^{n} X_i}{n}$

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean
- $\bar{X} := \frac{\sum_{i=1}^{n} X_i}{n}$
- and the sample variance is

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean
- $\bar{X} := \frac{\sum_{i=1}^{n} X_i}{n}$
- and the sample variance is
- $S^2 := \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean
- $\bar{X} := \frac{\sum_{i=1}^{n} X_i}{n}$
- and the sample variance is
- $S^2 := \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$
- These are RVs whose values depend on a random sample: they are examples of **statistics**.

background: Sampling Distribution of the Mean

- For a random sample $X_1, X_2, ...X_n$ define the sample mean
- $\bar{X} := \frac{\sum_{i=1}^{n} X_i}{n}$
- and the sample variance is
- $S^2 := \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
- These are RVs whose values depend on a random sample: they are examples of **statistics**.
- The following three theorems refer to the distribution of the sample mean for a given sample size n.

First theorem

These theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n, then the distribution of the sample mean $\bar{X}$ has mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Note: The standard deviation of the sample means, $\sigma_{\bar{X}}$, is also called the **standard error of the mean**.

Second theorem

These theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

- Theorem

  Central Limit Theorem
  If $X_1, X_2, \ldots, X_n$ is a random sample of size n, then as $n \to \infty$, the distribution of the sample mean $\bar{X}$ approaches a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Second theorem

These theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

- Theorem
  Central Limit Theorem
  If $X_1, X_2, \ldots, X_n$ is a random sample of size n, then as $n \to \infty$, the distribution of the sample mean $\bar{X}$ approaches a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.
  - Another way to say the same thing: The statistic

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

    has a distribution which approaches the standard normal distribution as $n \to \infty$.

Third theorem

These theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

- Theorem

  Samples taken from a normally distributed population
  If X has a normal distribution and $X_1, X_2, \ldots, X_n$ is a random sample of size n, then the sample mean $\bar{X}$ has a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Third theorem

These theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

- Theorem

  Samples taken from a normally distributed population
  If X has a normal distribution and $X_1, X_2, \ldots, X_n$ is a random sample of size n, then the sample mean $\bar{X}$ has a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

  - Another way to say the same thing: If X has a normal distribution, then the statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

has a standard normal distribution.

Return to Example 1 to illustrate what these say:

Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 1 tells us that for samples of size 25 taken from this population, the sample means will have mean $\mu_{\bar{X}} = 8.0$ and standard deviation $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{25}} = 0.06$.

Return to Example 1 to illustrate what these say:

Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 1 tells us that for samples of size 25 taken from this population, the sample means will have mean $\mu_{\bar{X}} = 8.0$ and standard deviation $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{25}} = 0.06$.
- We can think of this as saying that the average sample mean is the same as the average weight of the coffee in the cans, and its standard deviation is small (only one-fifth of the standard deviation of the weights).

Return to Example 1 to illustrate what these say:

Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 1 tells us that for samples of size 25 taken from this population, the sample means will have mean $\mu_{\bar{X}} = 8.0$ and standard deviation $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{25}} = 0.06$.
- We can think of this as saying that the average sample mean is the same as the average weight of the coffee in the cans, and its standard deviation is small (only one-fifth of the standard deviation of the weights).
- Also, if we would increase the sample size to 100, the standard deviation of the sample means (standard error of the mean) would decrease to $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{100}} = 0.03$ So by taking larger samples we can make the standard error of the mean as small as desired.

Return to Example 1 to illustrate what these say (continued):

Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 2 tells us that the sample means also have a distribution which is close to a normal distribution for large enough n. In practice, as a rule of thumb, we use this result when $n \geq 30$ (but this can be relaxed, especially if the distribution of X is not too far from normal).

Return to Example 1 to illustrate what these say (continued):

Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 3 tells us that since the distribution of the weights is normal, then the distribution of the sample means is also normal regardless of n - Theorem 2 is not needed in this case!