Case of Two Independent Normal populations:

Consider two populations with parameters: μ_1, σ_1^2 and μ_2, σ_2^2

The sampling distributions of the means are: $\overline{X}_1$ and $\overline{X}_2$

We are interested in comparing these two populations based on the analysis of the differences between their sampling distributions: $\mathbf{\overline{X}_1 - \overline{X}_2}$

In general, given random variables $X_1, X_2, ..., X_p$ and constants $c_1, c_2, .., c_p$, then:

If $Y = c_1 X_1 \pm c_2X_2 \pm .... \pm c_2X_p$ is a linear combination of $X_1, X_2, ..., X_p$ then:

$E(Y) = c_1 E(X_1) \pm c_2 E(X_2) \pm .... \pm c_2 E(Xp)$ and

$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \cdots + c_p^2 V(Xp) + 2\sum_{i<j}\sum c_i c_j Cov(X_i X_j)$

the last term is equal to zero for independent variables.

Therefore:

- Expected Value:

$$E[(\overline{X}_1 - \overline{X}_2)] = \mu_{(\overline{X}_1-\overline{X}_2)} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$$

- Variance:

$$\sigma^2{}_{(\overline{X}_1-\overline{X}_2)} = V[(\overline{X}_1 - \overline{X}_2)]$$

$$= V(\overline{X}_1) + V(\overline{X}_2)$$
$$= \sigma^2{}_{\overline{X}_1} + \sigma^2{}_{\overline{X}_2}$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_{21}}$$

- Standard Normal Sampling Distribution of the Mean:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If the parent populations are normal, then Z is exactly normal; otherwise Z is approximately normal if the **central limits** theory can be applied.

Example 7.3

An existing assembly line produces components with mean effective life of 5000 hours with a standard deviation of 40 hours.

$\mu_1 = 5000\ hours \quad ; \quad \sigma_1 = 40\ hours$

An improvement is supposedly introduced so that the mean life is increased to 5050 hours and the precision is improved so that the standard deviation is 3o hours.

$\mu_2 = 5050\ hours \quad ; \quad \sigma_2 = 30\ hours$

We want to test whether this difference is true (has high probability of being confirmed). Therefore, two samples are withdrawn from the two production plans, the first of size $n_1 = 16$ components and the second of size $n_2 = 25$ components.

Solution:

The claimed difference is 50 hours longer lives. As a test for improvement find the probability of the sampling distribution of the difference of the means is more than 25 hours.

i.e. find $P[(\bar{X}_2 - \bar{X}_1) > 25]$

$$\bar{X}_1 = \mu_1 = 5000\ hours \quad \sigma_{\bar{X}_1} = \frac{\sigma_1}{\sqrt{16}} = \frac{40}{4} = 10$$

$$\bar{X}_2 = \mu_2 = 5050\ hours \quad \sigma_{\bar{X}_2} = \frac{\sigma_2}{\sqrt{25}} = \frac{30}{5} = 6$$

$(\bar{X}_2 - \bar{X}_1)$ is normal with mean $=(\mu_2 - \mu_1) = (5050 - 5000) = 50$

And variance given by:

$$\sigma^2_{(\bar{X}_2 - \bar{X}_1)} = \sigma^2_{\bar{X}_2} + \sigma^2_{\bar{X}_1}$$

$$= 6^2 + 10^2 = 136$$

$$Z = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} = \frac{25 - 50}{\sqrt{136}} = -2.14$$

$$P[(\bar{X}_2 - \bar{X}_1) > 25] = P[Z > -2.14) = 0.9838$$

Standard Error: (Standard Deviation of Sampling Distribution)

Introduction

The standard error of a statistic (or an estimator) is actually the standard deviation of the sampling distribution of that statistic (or $\sqrt{variance\ of\ estimator}$). Standard errors reflect how much sampling fluctuation a statistic will show. Increasing the sample size, the Standard Error decreases.

The size of standard error is affected by two values:

1. The Standard Deviation of the population which affects the standard error. The larger the population's standard deviation (σ), the larger is the standard error i.e. $\frac{\sigma}{\sqrt{n}}$
2. The standard error is affected by the number of observations in a sample. A large sample will result in a small standard error of an estimate.

Standard Error of Mean:

The standard error for the mean or standard deviation of the sampling distribution of the mean, measures the deviation / variation in the sampling distribution of the sample mean ($\overline{X}$), denoted by ($\sigma_{\bar{x}}$) and calculated as a function of the standard deviation of the population and the respective size of the sample. For a normal population with SD σ

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

And this is the standard error in the point estimator of the normal population mean μ. When the standard deviation (σ) of the population is unknown, an estimate of the standard error is obtained using the standard deviation of the sample as an estimator (S) of (σ). In this case, an estimate of the standard error of the mean is:

$$\widehat{\sigma_{\bar{x}}} = \frac{s}{\sqrt{n}}$$

Example 7-5:
10 measurements of thermal conductivity (K) of a metal at 100°F and 550 Watts are given in Btu/hr-ft-°F as follows:

41.6 41.48 42.34 41.45 41.86 42.18 41.72 42.26 41.81 42.04

The point estimates of the mean and standard deviation are:

$$\bar{x} = 41.924 \qquad and\ s = 0.284$$

$$\widehat{\sigma_{\bar{x}}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

This means that the standard error is about $\frac{0.0898}{41.924} \cong 0.2$ % of the sample mean implying that the resulting point estimate of μ is precise.

$2\,\widehat{\sigma_{\bar{x}}} = 0.1796 \qquad \rightarrow \qquad 41.744 \le \mu \le 42.104$ with high probability.

Standard Error for Difference between Means:

Standard error for difference between two independent quantities is:

$$\sigma_{(\bar{x}_1 + \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where σ_1 and σ_2 are the respective variances of the two independent populations to be compared and $n_1 + n_2$ are the respective sizes of the two samples drawn from their respective populations.

For unknown variances: $\hat{\sigma}_{(\bar{x}_1 + \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Methods of Point Estimation:

Method of moments:

Equate population moments, which are defined in terms of expected values, to the corresponding sample moments. The population moments are functions of unknown parameters (to be estimated). Then these equations are solved to yield estimates of the unknown parameters.

Moments:

Given random variables X_1, X_2,...., X_p taken from distribution f(x) (discrete or continuous). Then:

k^{th} population distribution moment is $E(X^k)$, k=1,2,....

k^{th} sample moment is $\frac{1}{n}\sum_{i=1}^{n} X_i^k$, k=1,2,.... and n is sample size.

Ex:

1^{st} population moment $E(X) = \mu$

1^{st} sample moment is $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

$\therefore \hat{\mu} = \bar{x}$ sample mean is the moment estimator of the population mean

EX 7-7: Normal Distribution moment estimators

Let $x_1, x_2, \ldots x_n$ be random sample from normal population with μ and σ^2.

Population moments:

1st $E(X) = \mu$ 2nd $E(X^2) = \mu^2 + \sigma^2$

Sample moments.

1st $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ 2nd $\frac{1}{n}\sum_{i=1}^{n} x_i^2$

Equate:

$\mu = \bar{x}$ and $\mu^2 + \sigma^2 = \frac{1}{n}\sum x_i^2$

solving:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2$$

$$= \frac{\sum x_i^2 - n\left(\frac{1}{n}\sum x_i\right)^2}{n}$$

$$= \frac{\sum x_i^2 - n\bar{x}^2}{n}$$

$$= \frac{\sum x_i^2 - \sum \bar{x}^2}{n}$$

$$\therefore \hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

EX (exam question)

X is a random variable with PDF given by

$$f(x) = \frac{1}{\theta} \qquad 0 < x < \theta$$

a) Find E(x)

$$E(X) = \int_0^\theta x \cdot \frac{1}{\theta} \cdot dx = \frac{x^2}{2} \cdot \frac{1}{\theta} \Big|_0^\theta$$

$$= \left(\frac{\theta^2}{2} \cdot \frac{1}{\theta}\right) - (0)$$

$$= \frac{1}{2}\theta$$

b) Find Method of moments estimator for θ

$$E(X) = M_1 \quad \text{(population 1st moment)}$$

$$= \frac{1}{2}\theta$$

$$\therefore \hat{\theta} = 2M_1$$

$$M_1 = \frac{\sum x_i}{n} \quad \text{(sample 1st moment)}$$

equate:

$$\hat{\theta} = 2\left(\frac{\sum x_i}{n}\right)$$

c) Find Method of moments estimate for θ based on

x : 1   0.5   1.4   2   0.25

From above

$$\hat{\theta} = 2\left(\frac{1 + 0.5 + 1.4 + 2 + 0.25}{5}\right)$$

$$= 2\left(\frac{5.15}{5}\right)$$

$$= \underline{\underline{2.06}}$$