Sec “7”

Central Limit Theorem

Sampling Distribution of the

- Sample Mean.
- Difference between two sample means.

	Population	Sample
	Parameter	**Statistic**
Mean	μ	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	σ^2	$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$
Standard deviation	σ	$S = \sqrt{S^2}$

n: Number of items in the sample.

Central Limit Theorem

- The central limit theorem in statistics states that, given sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of that variable's distribution in the population. This fact holds especially true for sample sizes over 30.
- Therefore, as a sample size increases, the sample mean, and standard deviation will be closer in value to the population mean and standard deviation

Sampling Distribution of Sample Mean

- Given a **population** with a mean of **μ** and a standard deviation of **σ**, the sampling distribution of the sample mean has a mean of **μ** and a standard deviation $\frac{\sigma}{\sqrt{n}}$.

- We know that $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

- **The mean of $\bar{X}$:**

i.e., $E(ax) = aE(x)$

μ

$$\mu_{\bar{X}} = E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$

?

$$= \frac{1}{n}\left(\sum_{i=1}^{n} \mu\right) = \frac{1}{n}(n\mu) = \mu$$

$E(x_1 + x_2 + \cdots + x_n) = E(x_1) + \ldots + E(x_n)$

$$\therefore \mu_{\bar{X}} = \mu$$

- $\sigma_{\overline{X}}^2 = Var(\overline{X})$

$$= Var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right)$$

$Var(ax) = a^2 Var(x)$

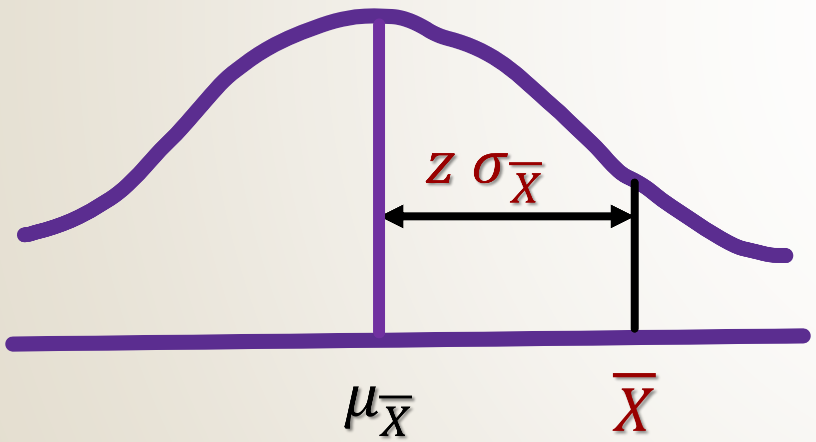
$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}\left(\sum_{i=1}^{n} \sigma^2\right) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$Var(x_1 + \cdots + x_n) = Var(x_1) + \cdots + Var(x_n)$

$$\therefore \boldsymbol{\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}} \longrightarrow \boldsymbol{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

Central Limit Theorem

- As n increase, the sampling distribution of sample means approaches a normal distribution with $\mu_{\overline{X}} = \mu$, and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.



Thus,

$$\overline{X} = \mu_{\overline{X}} + z\,\sigma_{\overline{X}},$$

And the standard score

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Exercise

(1) Given a normal population with a mean of 63 and a standard deviation of 12, find the probability that a random sample of size 16 has a mean greater than 61.5?

Solution

$\mu = 63, \qquad \sigma = 12, \qquad n = 16$

The probability that a random sample has a mean greater than 61.5 is given by

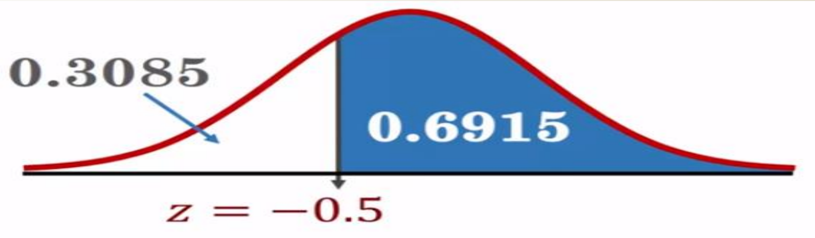
$$p(\overline{X} > 61.5) = p\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{61.5 - 63}{\frac{12}{\sqrt{16}}}\right)$$

$$= p\left(z > \frac{-1.5}{3}\right) = p(z > -0.5)$$

$$= 1 - p(z < -0.5) = 1 - 0.3085$$

$$= 0.6915$$

z	0.0
-0.5	$\boxed{0.3085}$



(2) An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size 25 are drawn randomly from the population

(a) Find the probability that the sample mean is between 85 and 92.

(b) Find the average value that is 2 standard deviations above the mean of the averages.

Solution

$\mu = 90, \qquad \sigma = 15, \qquad n = 25$

(a) $p(85 < \overline{X} < 92) = p\left(\frac{85-90}{\frac{15}{\sqrt{25}}} < z < \frac{92-90}{\frac{15}{\sqrt{25}}}\right)$

$= p\left(\frac{-5}{3} < z < \frac{2}{3}\right) = p(-1.6667 < z < 0.6667)$

$= p(-1.67 < z < 0.67)$

$= p(z < 0.67) - p(z < -1.67)$

$= 0.7486 - 0.0475 = 0.7011.$

(b) $z = +2$, and

$$\overline{X} = \mu_{\overline{X}} + z\,\sigma_{\overline{X}} = \mu + z\frac{\sigma}{\sqrt{n}} = 90 + (2)\left(\frac{15}{\sqrt{25}}\right) = 96.$$

(3) The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a mean of 2 hours and a standard deviation of 0.5 hours. A sample of size n = 50 is drawn randomly from the population. Find the probability that the sample mean is between 1.8 hours and 2.3 hours.

Sampling Distribution of Difference Between Two Means

- The **Sampling Distribution of the Difference between Two Means** shows the distribution of means of two samples drawn from the two independent populations, such that the difference between the population means can possibly be evaluated by the difference between the sample means
- If two populations follow each normal distributions, $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ (or both of them follow any distribution with these means and SD), and each samples are big enough in size n_1 and n_2, then the **sampling distribution of difference between means** follows a normal distribution with mean $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$ and standard deviation $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

- $\mu_{\overline{X}_1-\overline{X}_2} = E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$

We know that $\mu_{\overline{X}} = \mu$, therefore $\mu_{\overline{X}_1} = \mu_1$ and $\mu_{\overline{X}_2} = \mu_2$

$$\boxed{\therefore \mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2}$$

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

$= Var(\overline{X}_1) + Var(\overline{X}_2) = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}$

We know that $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$, therefore $\sigma^2_{\overline{X}_1} = \frac{\sigma_1^2}{n_1}$ and $\sigma^2_{\overline{X}_2} = \frac{\sigma_2^2}{n_2}$

$$\boxed{\therefore \sigma^2_{\overline{X}_1-\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \longrightarrow \boxed{\therefore \sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- The standard normal distribution of the difference between two sample means comes by the expression:

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

Exercise:

(4) Population 1 has a mean of 20 and a variance of 100. Population 2 has a mean of 15 and a variance of 64. You sample 20 scores from Pop 1 and 16 scores from Pop 2. What is the mean and the standard deviation of the sampling distribution of the difference between means (Pop 1 - Pop 2)?

Solution

Population 1

$$\mu_1 = 20$$
$$\sigma_1^2 = 100$$
$$n_1 = 20$$

Population 2

$$\mu_2 = 15$$
$$\sigma_2^2 = 64$$
$$n_2 = 16$$

- $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2 = 20 - 15 = 5$
- $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{100}{20} + \frac{64}{16}} = 3$

(5) The mean height of 15-year-old boys (in cm) is 175 and the variance is 64. For girls, the mean is 165 and the variance is 64. If eight boys and eight girls were sampled, what is the probability that the mean height of the sample of girls would be higher than the mean height of the sample of boys?

Solution

Population 1: *Girls*	**Population 2: *Boys***
$\mu_1 = 165\ cm$	$\mu_2 = 175\ cm$
$\sigma_1^2 = 64\ cm$	$\sigma_2^2 = 64\ cm$
$n_1 = 8$	$n_2 = 8$

$$\mathrm{P}(\overline{X}_1 > \overline{X}_2) = \mathrm{P}(\overline{X}_1 - \overline{X}_2 > 0) = \mathrm{P}\left(\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{0 - (165 - 175)}{\sqrt{\frac{64}{8} + \frac{64}{8}}}\right)$$

$$= \mathrm{p}\left(z > \frac{10}{4}\right) = \mathrm{p}(z > 2.5) = 1 - \mathrm{p}(z < 2.5) = 1 - 0.9938 = 0.0062$$

(6) an electrical appliance made by Company A has an average life of 2500 hours, with a SD of 500 hours. Another Company B, makes this appliance with an average life of 2300 hours, and a SD of 800 hours. We take 300 Company A devices and 200 Company B devices. Calculate the probability that the average life of the sample from Company A isn't 100 hours more than the average life of the sample from Company B.

Company A

$\mu_1 = 2500\ Hrs$

$\sigma_1 = 500\ Hrs$

$n_1 = 300$

Solution

Company B

$\mu_2 = 2300\ Hrs$

$\sigma_2 = 800\ Hrs$

$n_2 = 200$

$$P\left(\overline{X}_1 - \overline{X}_2 \leq 100\right) = P\left(\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \frac{100 - (2500 - 2300)}{\sqrt{\frac{(500)^2}{300} + \frac{(800)^2}{200}}}\right)$$

$$= p\left(z \leq \frac{100 - 200}{63.509}\right) = p\left(z \leq \frac{-100}{63.509}\right) = p(z \leq -1.57) = 0.0582$$

		Sampling distribution		
Parameter	**statistic**	**Center(mean)**	**Spread(standard deviation"error")**	**Z-Score**
• $\mu =$ population mean, • $\sigma =$ Population standard deviation	$\overline{X} =$ sample mean	$\mu_{\overline{X}} = \mu$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$	$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
• $\mu_1, \mu_2 =$ the mean of population1,2 • $\sigma_1, \sigma_2 =$ the standard deviation of population 1,2.	$\overline{X}_1 - \overline{X}_2$	$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$	$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$