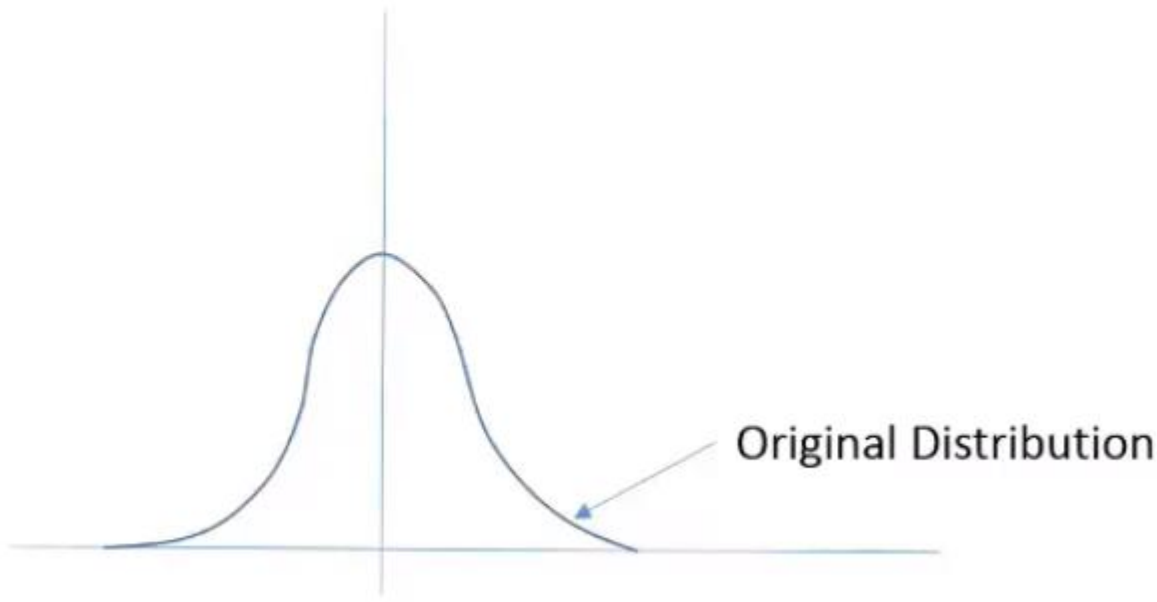
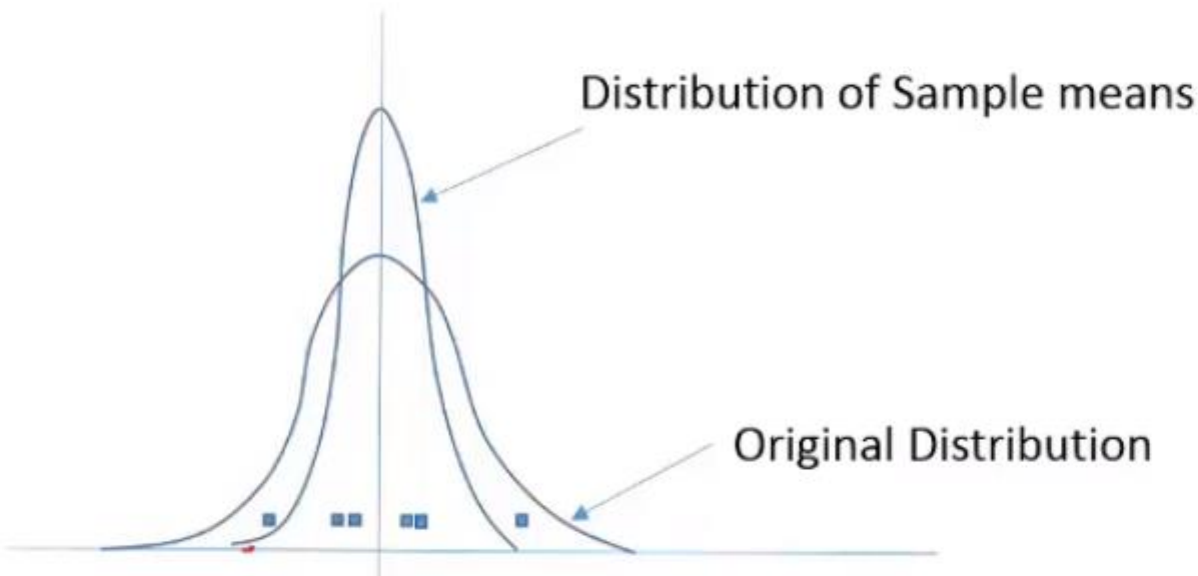
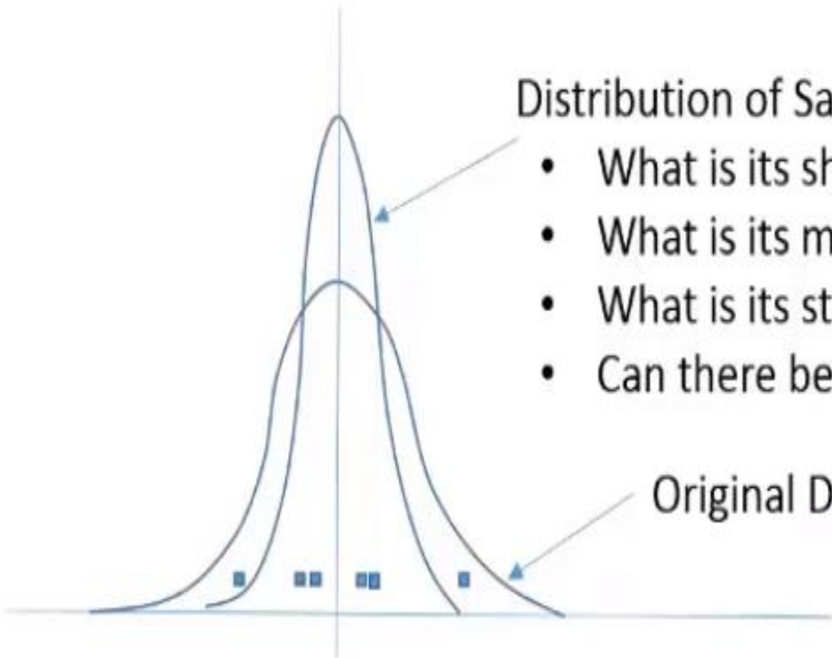
Central Limit Theorem

Sampling Distribution



Normal Distribution N(μ,σ)


Distribution of Sample means
Original Distribution



Distribution of Sample means

- What is its shape?
- What is its mean?
- What is its standard deviation?
- Can there be a distribution for sample standard deviations?

Original Distribution

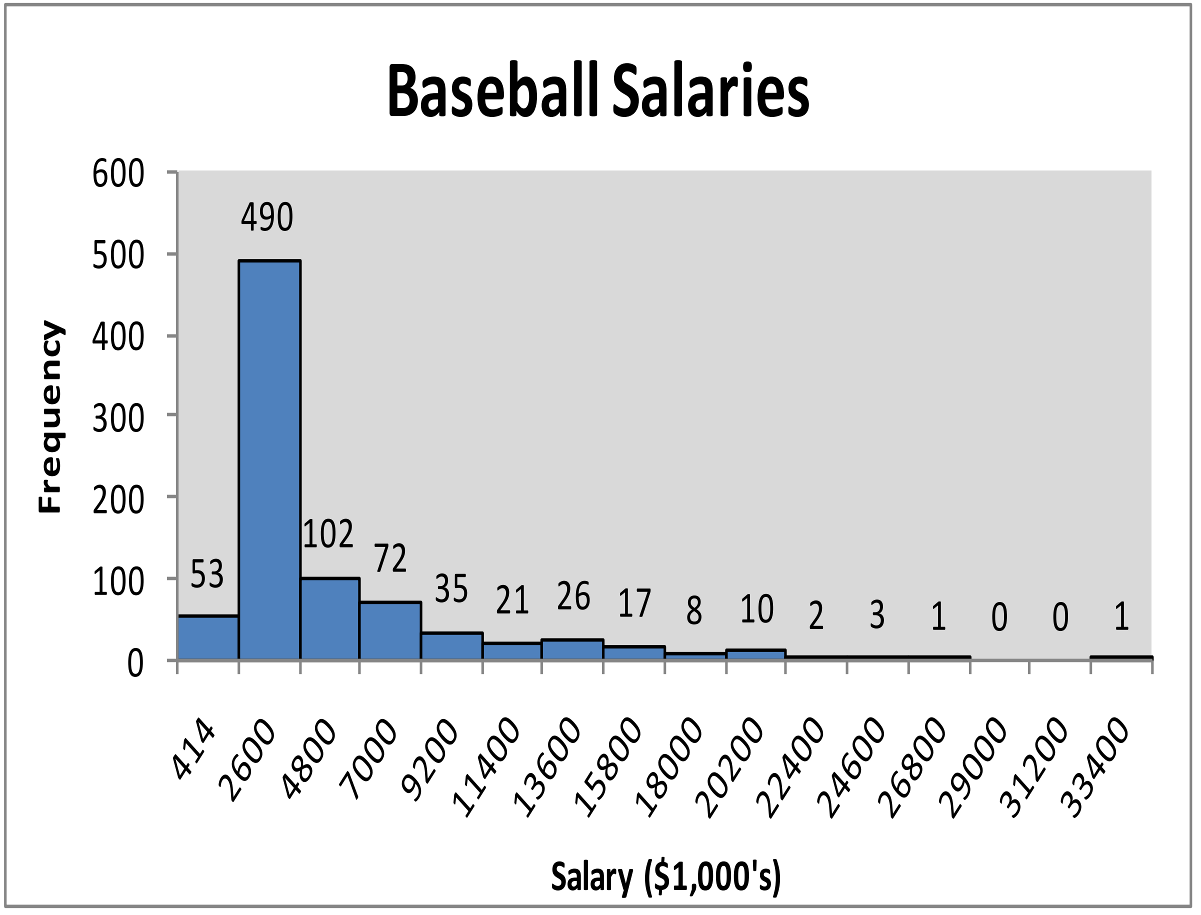
Normal Populations

- **Important Fact:**

  If the population is normally distributed, then the sampling distribution of $\bar{x}$ is normally distributed for any sample size n.

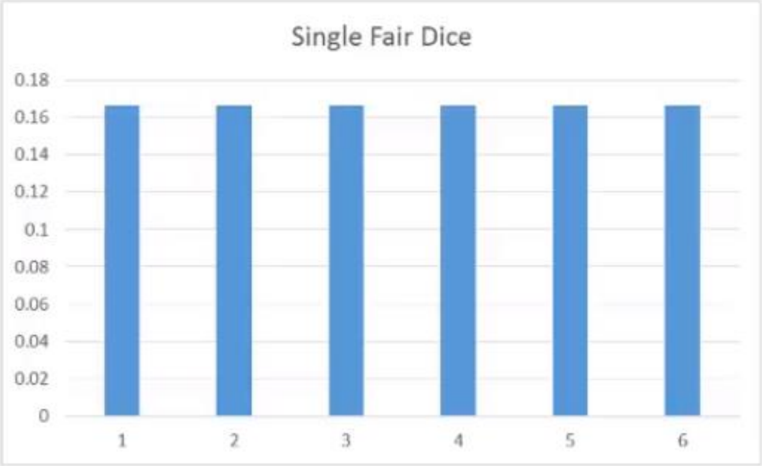
Non-normal Populations

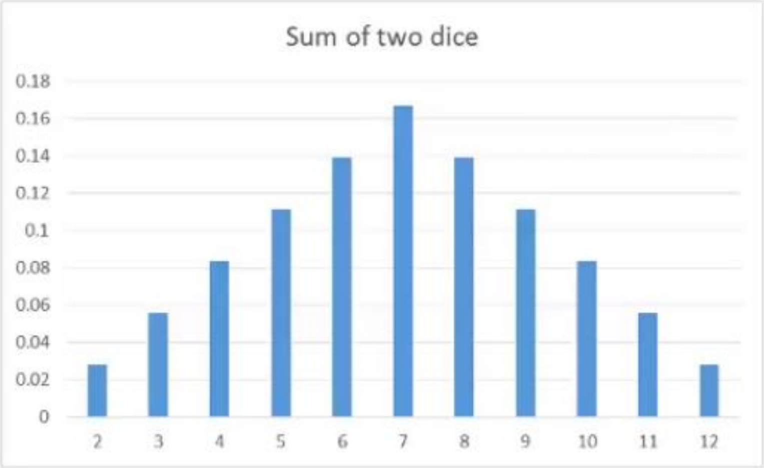
- What can we say about the shape of the sampling distribution of $\overline{x}$ when the population from which the sample is selected is not normal?

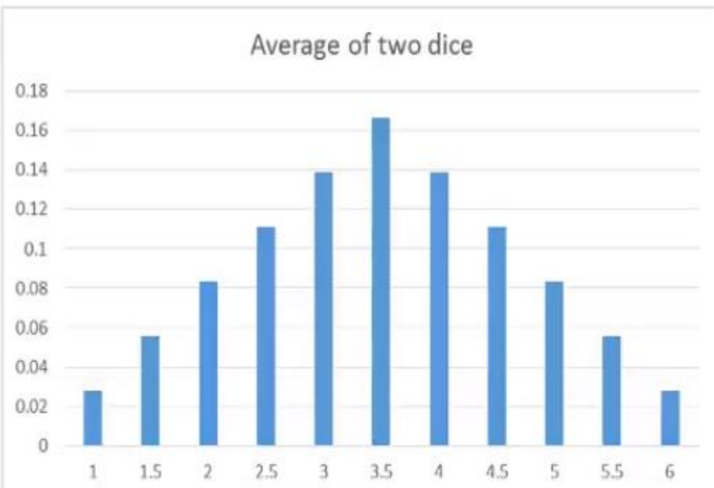


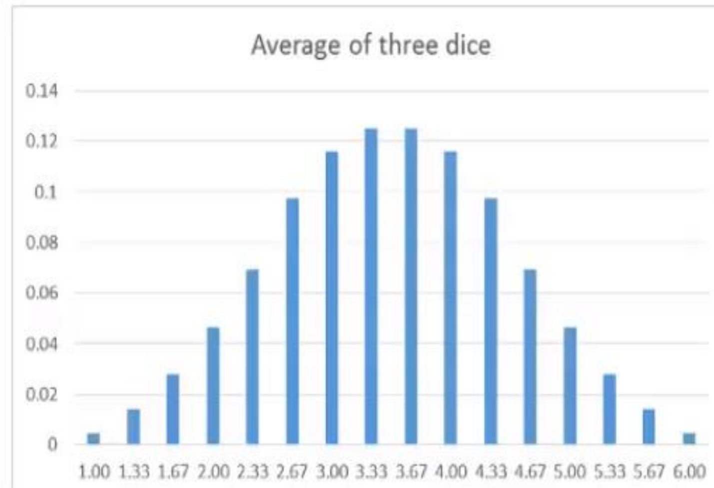
Central Limit Theorem

- The aggregation of a sufficiently large number of independent random variables results in a random variable which will be approximately normal.
- Example






Average of two dice
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6


Average of three dice
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1.00 1.33 1.67 2.00 2.33 2.67 3.00 3.33 3.67 4.00 4.33 4.67 5.00 5.33 5.67 6.00

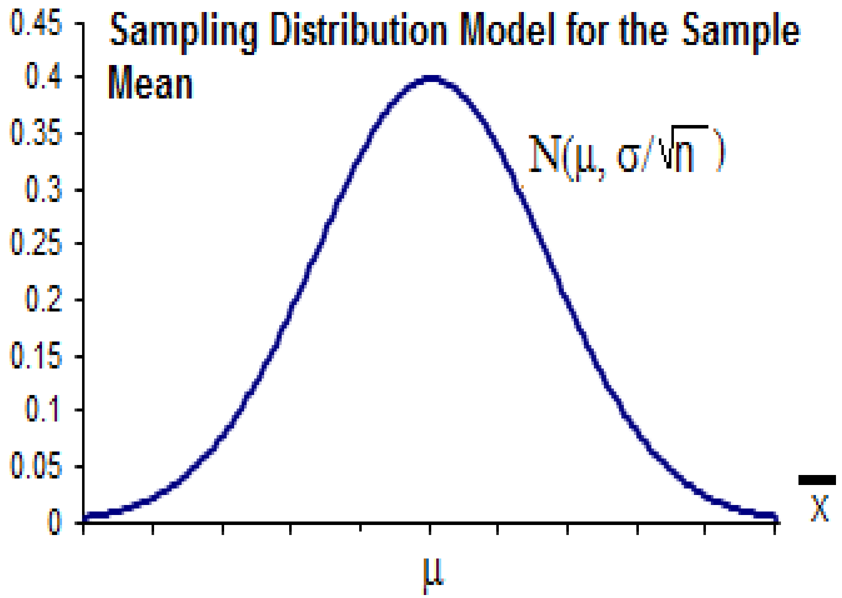
The Central Limit Theorem (for the sample mean $\bar{x}$)

- If a random sample of n observations is selected from a population (**any** population), then when n is sufficiently large, <u>the sampling distribution of $\bar{x}$ will be approximately normal.</u>

(The larger the sample size, the better will be the normal approximation to the sampling distribution of $\bar{x}$.)

The Importance of the Central Limit Theorem

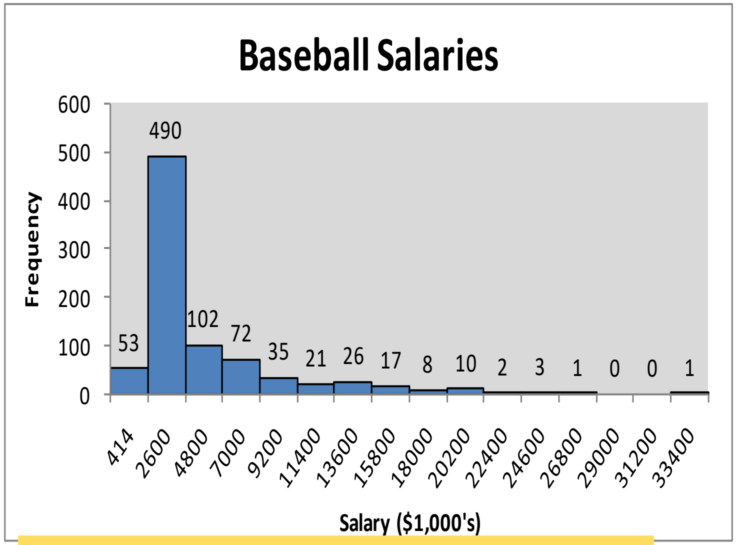
- When we select simple random samples of size n, the sample means $\bar{x}$ will vary from sample to sample. We can model the distribution of these sample means with a probability model that is …



$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

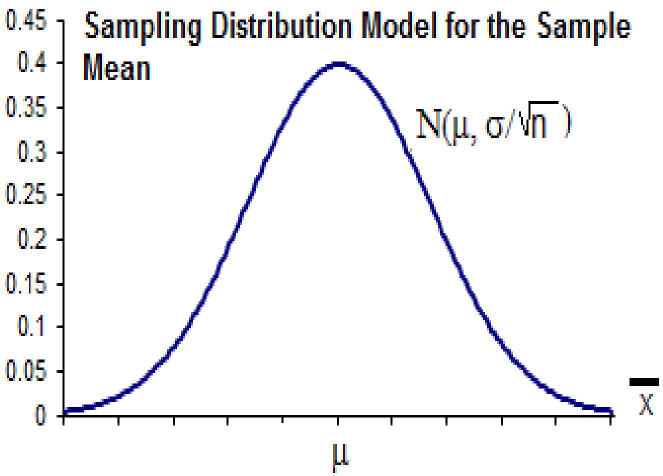
How Large Should n Be?

- For the purpose of applying the Central Limit Theorem, we will consider a sample size to be large when n > 30.



← ← ← *Even if the population from which the sample is selected looks like this …*

… the Central Limit Theorem tells us that a good model for the sampling distribution of the sample mean $\bar{x}$ is … → → →



Example

- The probability distribution of incomes of account executives has mean $20,000 and standard deviation $5,000. 64 account executives are randomly selected. What is the probability that the sample mean exceeds $20,500?

Example

answer E(X) = $20,000

SD(X) = $5,000

$$E(\bar{X}) = \$20,000$$

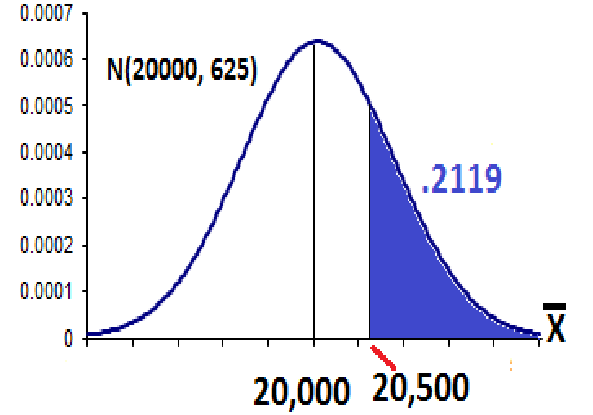
$$SD(\bar{X}) = \frac{SD(x)}{\sqrt{n}} = \frac{5,000}{\sqrt{64}} = 625$$

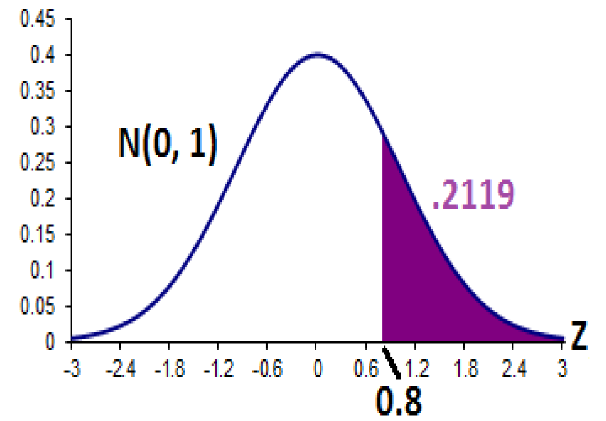
By CLT, $\bar{X} \sim N(20,000, 625)$

$$P(\bar{X} > 20,500) =$$

$$P\left(\frac{\bar{X} - 20,000}{625} > \frac{20,500 - 20,000}{625}\right) =$$

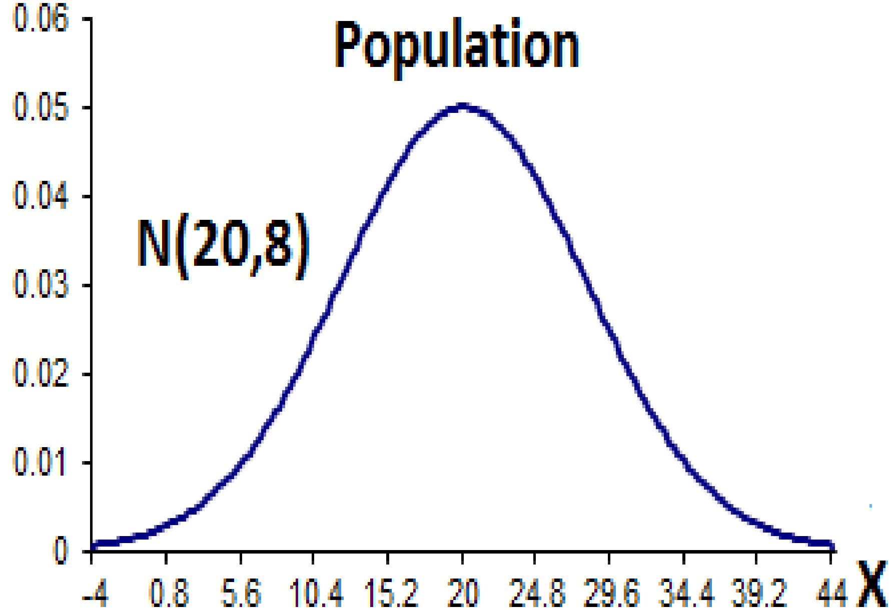
$$P(z > 0.8) = 1 - 0.7881 = 0.2119$$





Example

A sample of size n=16 is drawn from a normally distributed population with E(X)=20 and SD(X)=8. a) What is the probability that sample mean will be more than 24? b)What is the probability that sample mean lies between 16 and 24



Example

$$X \sim N(20, 8);\ \overline{X} \sim N(20, \tfrac{8}{\sqrt{16}})$$

$$a)\ P(\overline{X} \geq 24) = P(\tfrac{\overline{X}-20}{2} \geq \tfrac{24-20}{2})$$

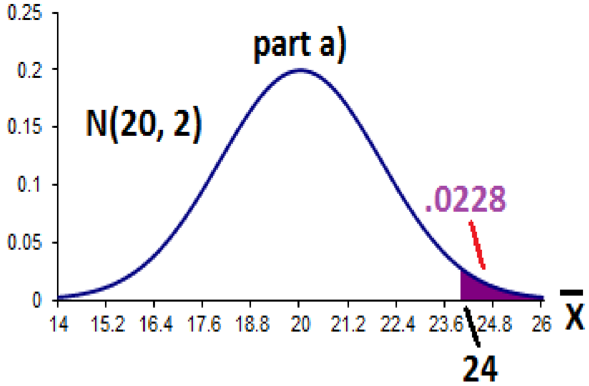
$$= P(z \geq 2) = 1 - 0.9772 = 0.0228$$

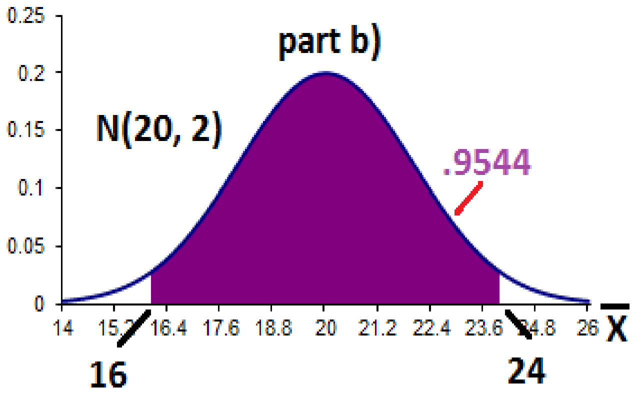
$$b)\ P(16 \leq \overline{X} \leq 24)$$

$$= P\left(\tfrac{16-20}{2} \leq z \leq \tfrac{24-20}{2}\right)$$

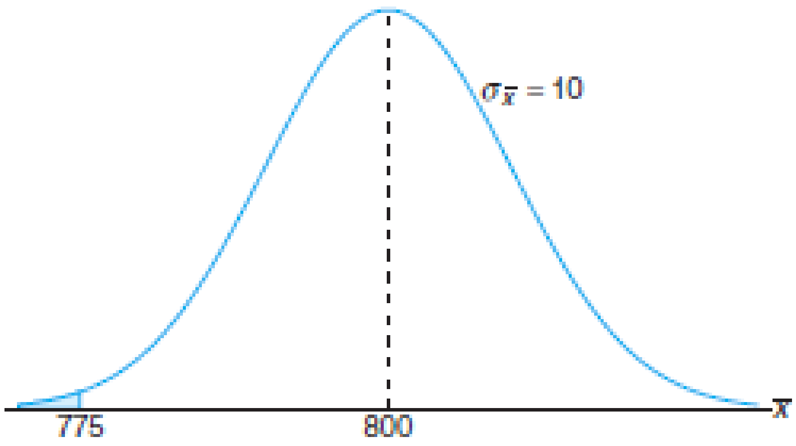
$$= P(-2 \leq z \leq 2)$$

$$= 0.9772 - 0.0228 = 0.9544$$





- An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.



Distribution of sample mean

The sampling distribution of sample mean will be approximately normal, with $\mu\_X = 800$ and $\sigma\_X = 40/\sqrt{16} = 10$.

$z = (775 - 800)/10 = -2.5$, and therefore $P(X < 775) = P(Z < -2.5) = 0.0062$.

Inferences on the Population Mean

- Important application of central limit theorem is to determine reasonable values of the population mean.

- **Automobile Parts**: An important manufacturing process produces cylindrical component parts for the automotive industry. It is important that the process produce parts having a mean diameter of 5.0 millimeters. The engineer involved conjectures that the population mean is 5.0 millimeters. An experiment is conducted in which 100 parts produced by the process are selected randomly and the diameter measured on each. It is known that the population standard deviation is $\sigma = 0.1$ millimeter. The experiment indicates a sample average diameter of $\bar{x} = 5.027$ millimeters. Does this sample information appear to support or refute the engineer's conjecture?

$$P(|\bar{X}-5| \geq 0.027) = P(\bar{X}-5 \geq 0.027) + P(\bar{X}-5 \leq -0.027)$$
$$= 2P\left(\frac{\bar{X}-5}{0.1/\sqrt{100}} \geq 2.7\right).$$

Here we are simply standardizing $\bar{X}$ according to the Central Limit Theorem. If the conjecture $\mu = 5.0$ is true, $\frac{\bar{X}-5}{0.1/\sqrt{100}}$ should follow $N(0,1)$. Thus,

$$2P\left(\frac{\bar{X}-5}{0.1/\sqrt{100}} \geq 2.7\right) = 2P(Z \geq 2.7) = 2(0.0035) = 0.007.$$

Therefore, one would experience by chance that an $\bar{x}$ would be 0.027 millimeter from the mean in only 7 in 1000 experiments. As a result, this experiment certainly does not give supporting evidence to the conjecture that μ =5.0. In fact, it strongly refutes the conjecture!

- Traveling between two campuses of a university in a city via shuttle bus takes, on average, 28 minutes with a standard deviation of 5 minutes. In a given week, a bus transported passengers 40 times. What is the probability that the average transport time was more than 30 minutes? Assume the mean time is measured to the nearest minute.

- In this case, $\mu = 28$ and $\sigma = 3$. We need to calculate the probability $P(\bar{X} > 30)$ with $n = 40$. Since the time is measured on a continuous scale to the nearest minute, an $\bar{x}$ greater than 30 is equivalent to $\bar{x} \geq 30.5$. Hence,

$$P(\bar{X} > 30) = P\left(\frac{\bar{X} - 28}{5/\sqrt{40}} \geq \frac{30.5 - 28}{5/\sqrt{40}}\right) = P(Z \geq 3.16) = 0.0008.$$

Sampling Distribution of the Difference between Two Means

- Suppose that we have two populations, the first with mean μ_1 and variance $\sigma^2{}_1$, and the second with mean μ_2 and variance $\sigma^2{}_2$. Let the statistic $\overline{X}1$ represent the mean of a random sample of size $n1$ selected from the first population, and the statistic $\overline{X}2$ represent the mean of a random sample of size $n2$ selected from the second population, independent of the sample from the first population.

- What can we say about the sampling distribution of the difference $\overline{X}1 - \overline{X}2$ for repeated samples of size $n1$ and $n2$?

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

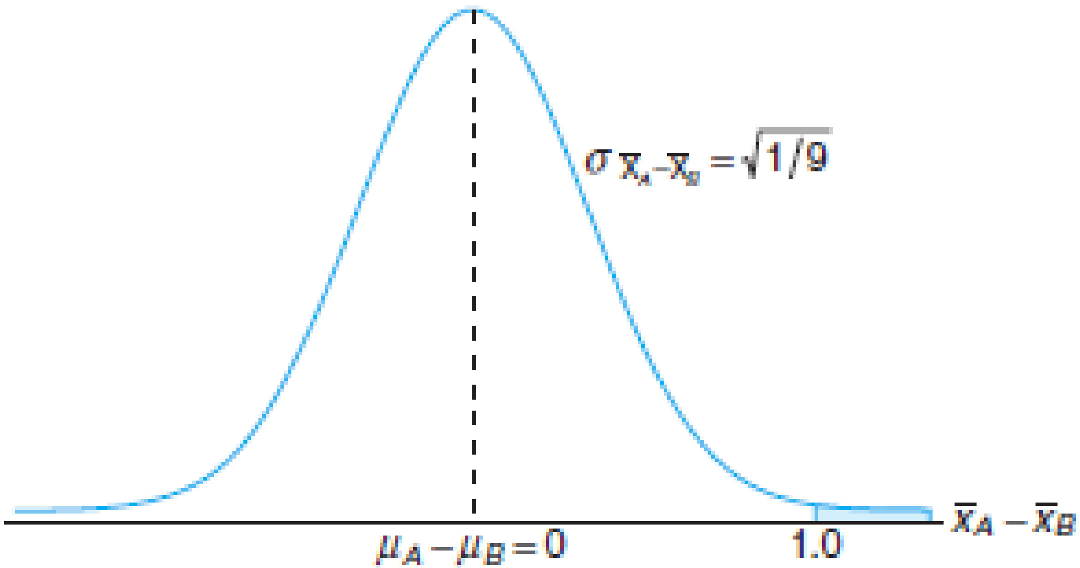
Hence,

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

is approximately a standard normal variable.

- **Paint Drying Time**: Two independent experiments are run in which two different types of paint are compared. Eighteen specimens are painted using type *A*, and the drying time, in hours, is recorded for each. The same is done with type *B*. The population standard deviations are both known to be 1.0. Assuming that the mean drying time is equal for the two types of paint, find $P(\bar{X}_A - \bar{X}_B > 1.0)$, where $\bar{X}_A$ and $\bar{X}_B$ are average drying times for samples of size $n_A = n_B = 18$.

- From the sampling distribution of $\overline{X}_A - \overline{X}_B$, we know that the distribution is approximately normal with mean $\mu_{XA-XB} = \mu_A - \mu_B = 0$ and $\sigma^2_{XA-XB} = 1/9$



$$z = \frac{1 - (\mu_A - \mu_B)}{\sqrt{1/9}} = \frac{1 - 0}{\sqrt{1/9}} = 3.0;$$

so

$$P(Z > 3.0) = 1 - P(Z < 3.0) = 1 - 0.9987 = 0.0013.$$

- The television picture tubes of manufacturer *A* have a mean lifetime of 6.5 years and a standard deviation of 0.9 year, while those of manufacturer *B* have a mean lifetime of 6.0 years and a standard deviation of 0.8 year. What is the probability that a random sample of 36 tubes from manufacturer *A* will have a mean lifetime that is at least 1 year more than the mean lifetime of a sample of 49 tubes from manufacturer *B*?

- $\mu_{\overline{X}1-\overline{X}2} = 6.5 - 6.0 = 0.5$ and $\sigma_{\overline{X}1-\overline{X}2} = 0.189$
- $z = (1.0 - 0.5)/0.189 = 2.65,$
- $P(\overline{X}1 - \overline{X}2 \geq 1.0) = P(Z > 2.65) = 0.004$