# İST 292 STATISTICS

## Sections: 05-06

## For Department of Computer Engineering

***LESSON 3 NORMAL DISTRIBUTION AND CENTRAL LIMIT THEOREM***

Dr. Ayten Yiğiter and Dr. Esra Polat Lecture Notes

# 3. NORMAL DISTRIBUTION

- **Many statistical phenomen are modelled by normal distribution. For example, human charecteristics such as height, weight, strength; the speed of any particule in gas, errors in measument of quantities. It has bell-shaped symetric curve and the probability is interpreted as "area under the curve".**

**Characteristics of the Normal Distribution**

- **Symmetric, bell shaped**
- **X random variable is defined as** $-\infty < x < \infty$
- **Two parameters, μ and σ. Note that the normal distribution is actually a family of distributions, since μ and σ determine the shape of the distribution.**
- **The rule for a normal density function is**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

- **The notation N(μ, σ²) means normally distributed with mean μ and variance σ². If we say X~N(μ, σ²) we mean that X is distributed N(μ, σ²).**

## Why is the normal distribution useful?

- **Many things actually are normally distributed, or very close to it. For example, height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution.**

- **The normal distribution is easy to work with mathematically. In many practical cases, the methods developed using normal theory work quite well even when the distribution is not normal.**

- **There is a very strong connection between the size of a sample n and the extent to which a sampling distribution approaches the normal form. Many sampling distributions based on large n can be approximated by the normal distribution even though the population distribution itself is definitely not normal.**

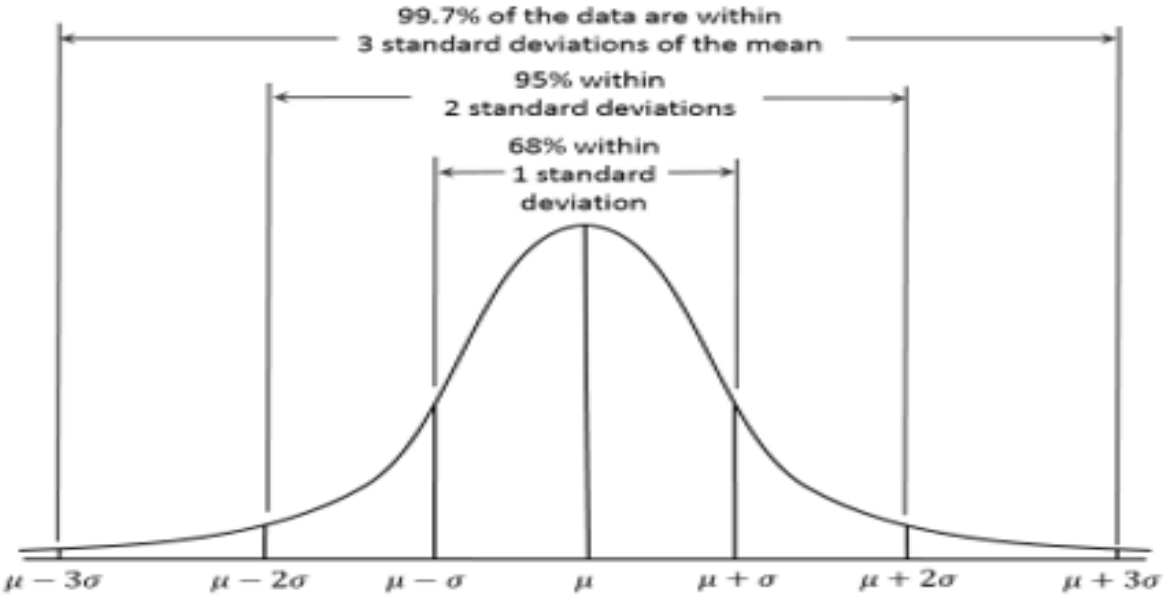



**Figure 1. Empirical rule of normal distribution.**

- About 68% of all cases fall within one standard deviation of the mean, that is
  - $P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6827$
- About 95% of cases lie within 2 standard deviations of the mean, that is
  - $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545$
- About 99.7% of cases lie within 3 standard deviations of the mean, that is

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$

## Standart Normal Distribution

- **Standart normal distribution is a special case of the normal distribution.**

  If X has a normal disribution showed as: $X \sim N(\mu, \sigma^2)$, Z will have a standard normal distribution showed as: $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

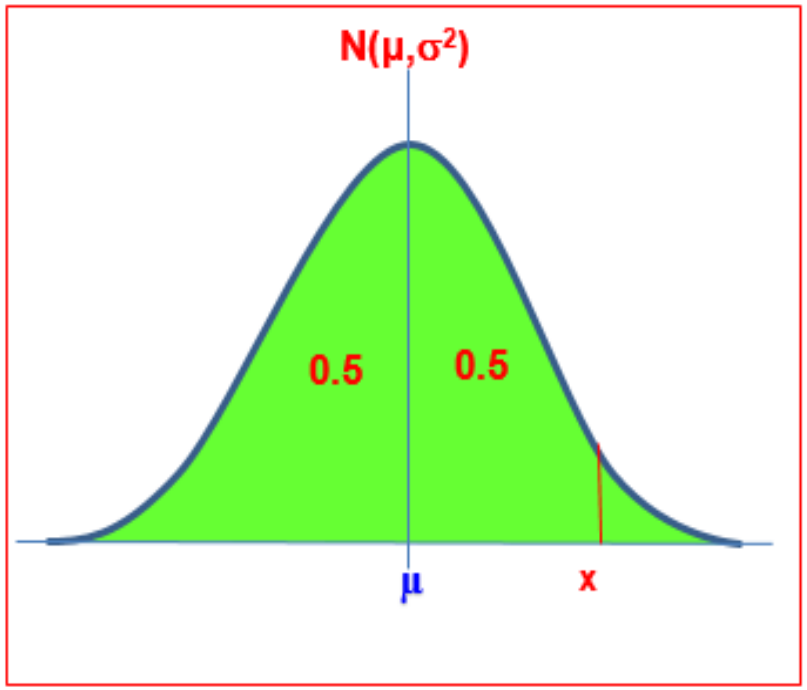
- If random variable Z has a standart normal distribution then:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty$$

$$E(Z) = 0 \text{ and } V(Z) = 1$$

- Every normal random variable X can be transformed into a z score via the following equation: $Z = \frac{X - \mu}{\sigma}$ and we call that **standardized random variable.**

**Normal Distribution Graph**

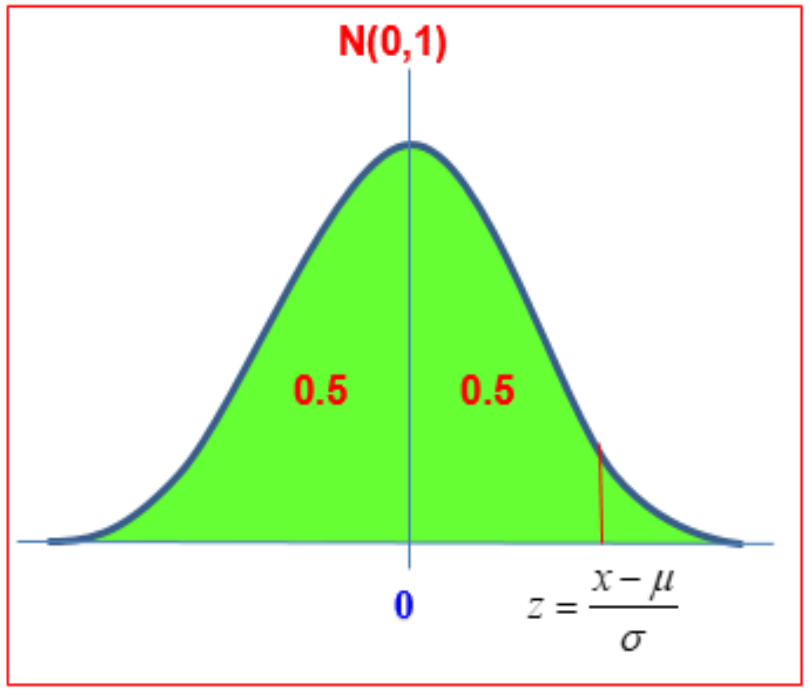



$X \sim N(\mu, \sigma^2)$

**Because of symmtery:**

$P(X > \mu) = P(X < \mu) = 0.5$

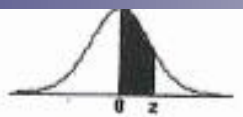
**Standart Normal Distribution Graph**




$Z \sim N(0,1)$

**Because of symmtery:**

$P(Z > 0) = P(Z < 0) = 0.5$

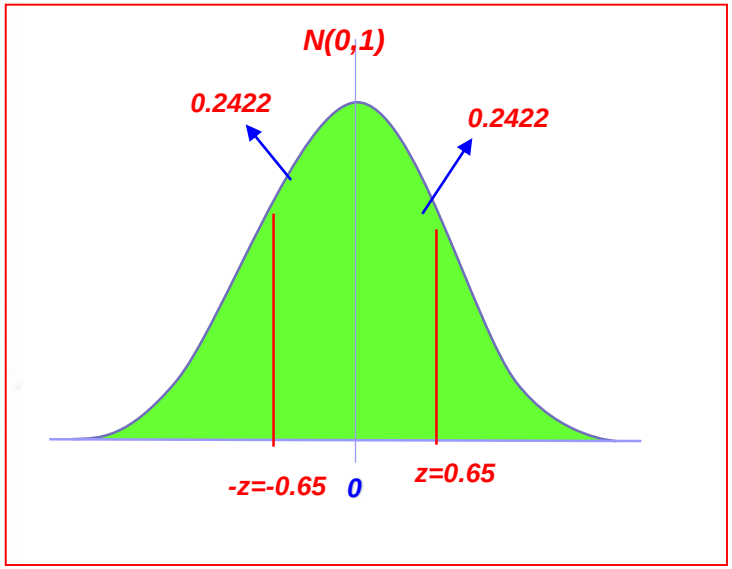



$$P(0 \le Z \le z) = P(-z \le Z \le 0) = \int_0^z f(t)dt$$

| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3436 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.4940 | 0.4941 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.4955 | 0.4956 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.4966 | 0.4967 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.4975 | 0.4976 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4979 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.4982 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.4987 | 0.4987 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |
| 3.1 | 0.4990 | 0.4991 | 0.4991 | 0.4991 | 0.4992 | 0.4992 | 0.4992 | 0.4992 | 0.4993 | 0.4993 |
| 3.2 | 0.4993 | 0.4993 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4995 | 0.4995 | 0.4995 |
| 3.3 | 0.4995 | 0.4995 | 0.4995 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4997 |
| 3.4 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4998 |
| 3.5 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 |

**We will use this Standart Normal Distribution (SND) Table for finding probabilities!!**

**First column in the table gives z values means; the values of random variable Z having standart normal distribution. First row show the decimal part of z values. This table gives probabilities for P(0<Z<z)!! All the values are inside show these probabilities. For example, P(0<Z<0.65)=0.2422 and because of SND is symmetric this means also P(-0.65<Z<0)= 0.2422**



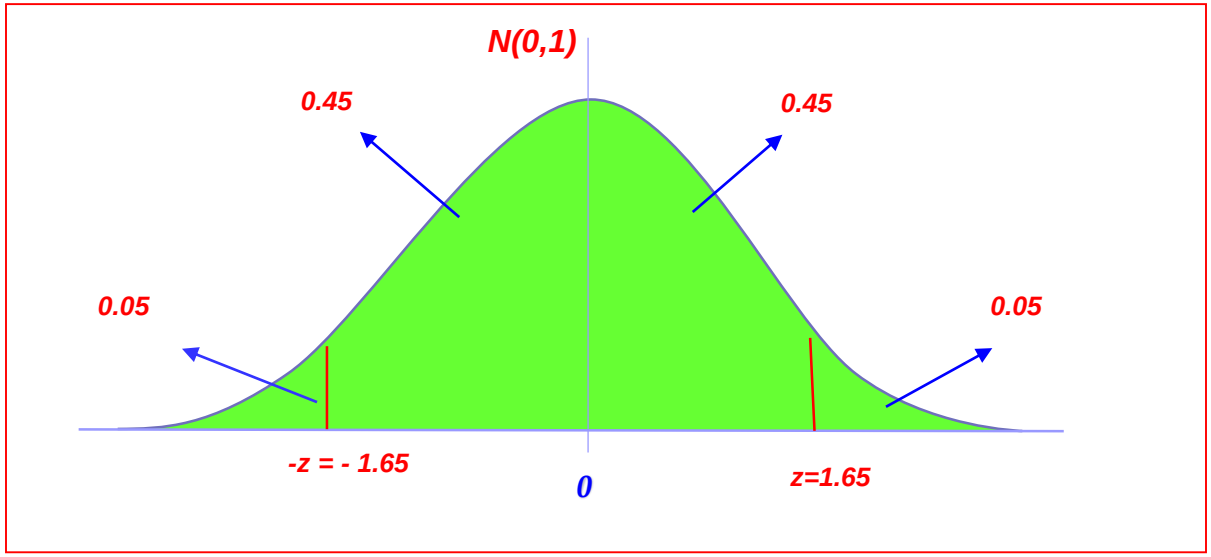
**Example:** Find $P(Z \leq a)$ for a = 1.65, -1.65, 1.0, -1.0

**To solve:** For positive values of a, look up and report the value for $\Phi(a)$ given in SND Table. For negative values of a, look up the value for $\Phi(-a)$ (i.e. $\Phi$ (absolute value of a)) and report $1 - \Phi(-a)$. **Note: $\Phi(a)$ ise cumulative distribution function for SND such as $\Phi(a)=P(Z \leq a)$**

**Be Careful:** From Standart Normal Distrubition Table $P(0<Z<1.65) \cong 0.45$

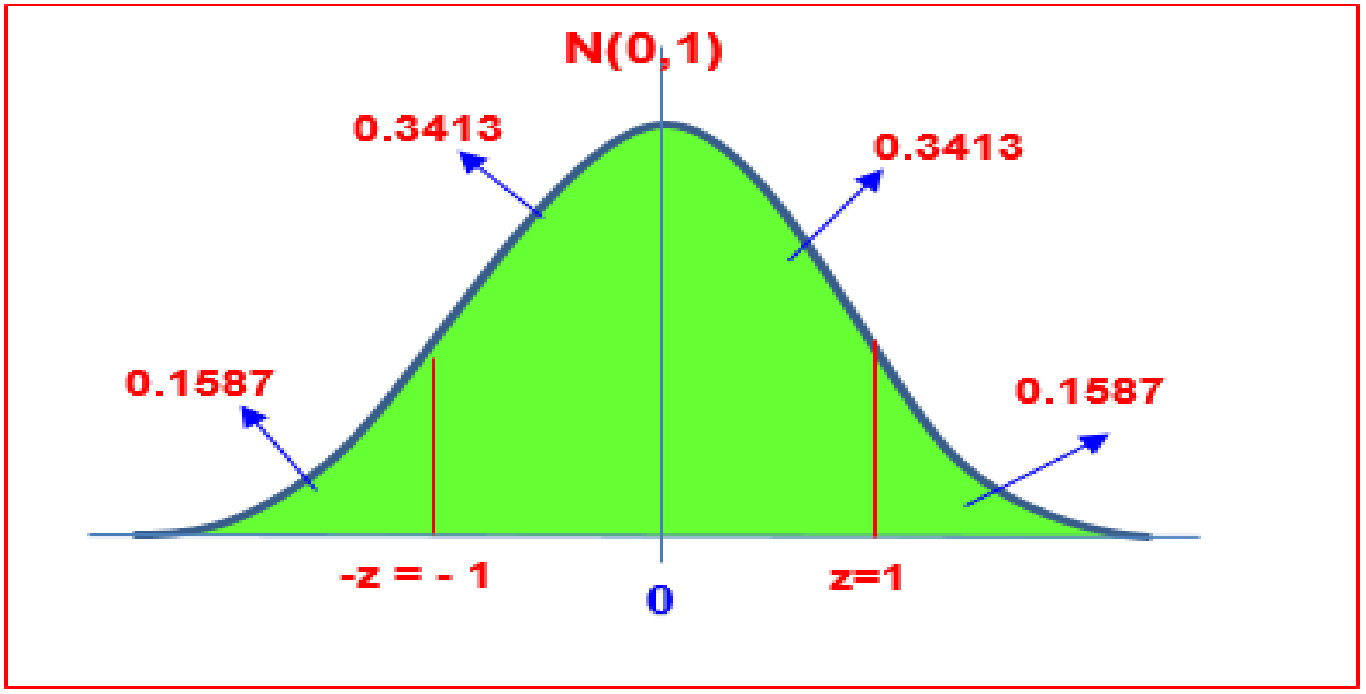
$P(Z \leq 1.65) = \Phi(1.65) = 0.5 + P(0 \leq Z \leq 1.65) = 0.5 + 0.45 = 0.95$

$P(Z \leq -1.65) = \Phi(-1.65) = 1 - \Phi(1.65) = 0.05$ (Easy Way: 0.5-0.45=0.05)



$P(Z \leq 1.0) = 0.5 + P(0 \leq Z \leq 1.0) = 0.5 + 0.3413 = 0.8413$

$P(Z \leq -1.0) = 0.5 - P(-1.0 \leq Z \leq 0) = 0.5 - 0.3413 = 0.1587$

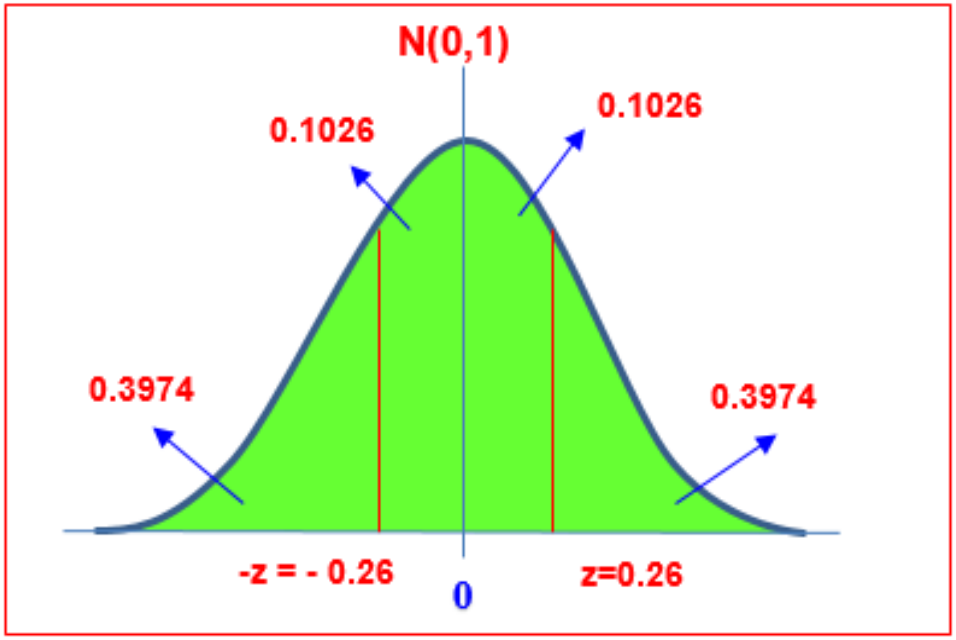


**Example:** Find a for $P(Z \leq a) = 0.6026, 0.9750, 0.3446$

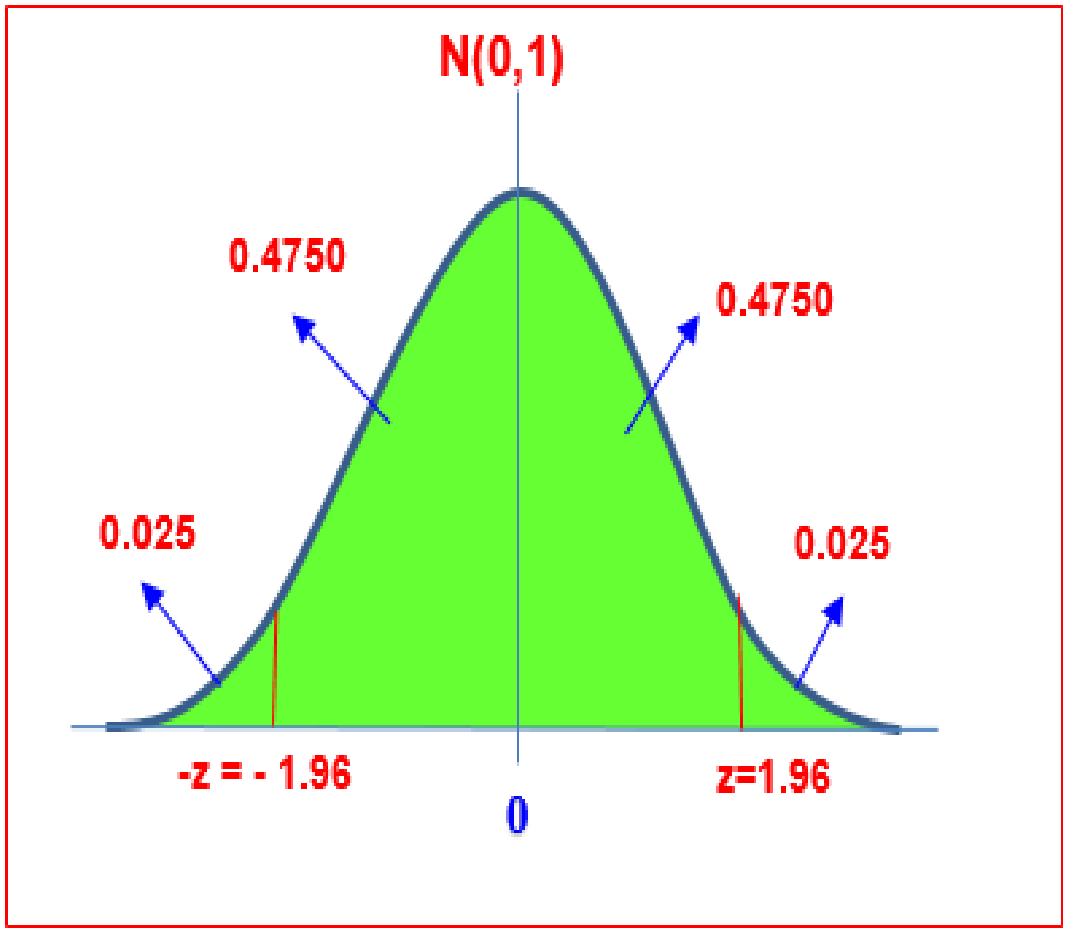
**To solve:**

- For $p \geq 0.5$, find the probability value in SND Table, and report the corresponding value for Z.
- For $p < 0.5$, compute $1 - p$, find the corresponding Z value, and report the negative of that value, i.e. -Z.

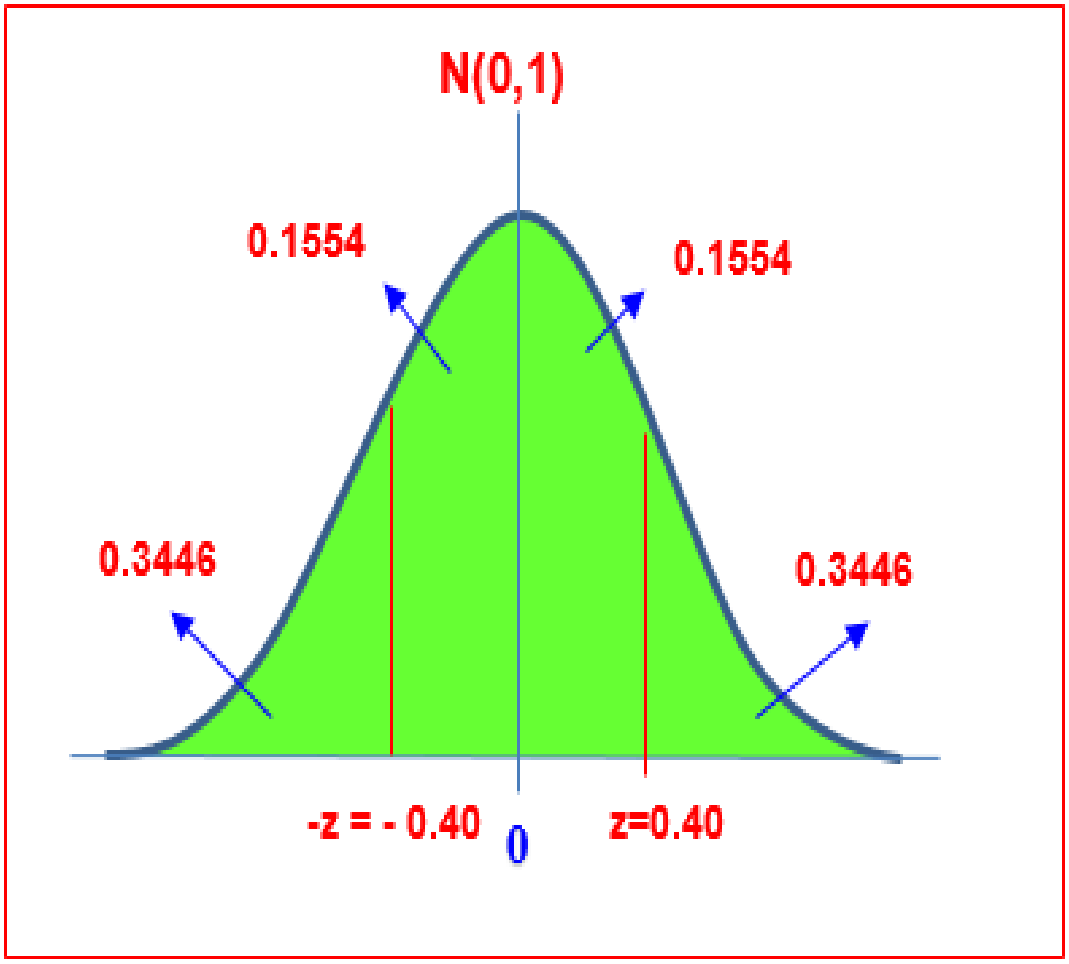
$P(Z \leq 0.26) = 0.5+0.1026=0.6026$ (SND Table $P(0<Z<0.26)=0.1026$) so a=0.26



$P(Z \leq 1.96) = 0.5+0.4750=0.9750$ (SND Table $P(0<Z<1.96)=0.4750$) so $a=1.96$



$P(Z \leq -0.40) = 0.5 - 0.1554 = 0.3446$ (SND Table $P(0<Z<0.40) = 0.1554$) so $a = -0.40$



# EXAMPLES

1. **The top 5% of applicants (as measured by GRE scores) will receive scholarships.**

**If GRE ~ N(500, $100^2$), how high does your GRE score have to be to qualify for a scholarship?**

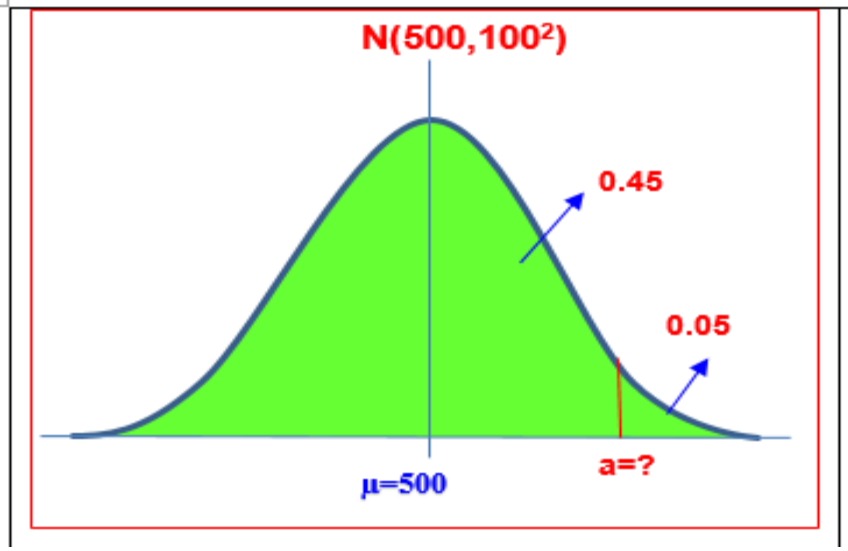
**Solution:** X ~ N(500, $100^2$) and $P(X > a) = 0.05$

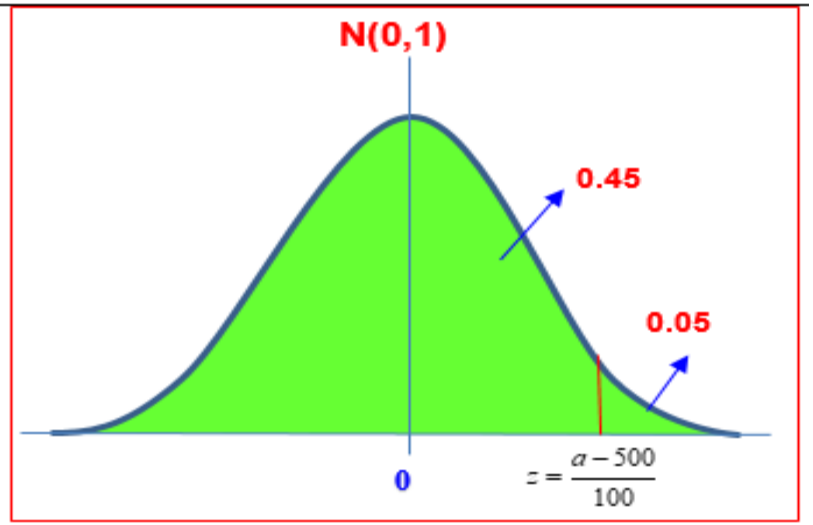
$$P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(\frac{X-500}{100} > \frac{a-500}{100}\right) = 0.05$$

$$P\left(Z > \frac{a-500}{100}\right) = 1 - P\left(Z \le \frac{a-500}{100}\right) = 1 - \Phi\left(\frac{a-500}{100}\right) = 0.05$$

$$\frac{a-500}{100} = 1.65 \Rightarrow a = 665$$

**If your GRE score ≥ 665 will receive scholarship.**



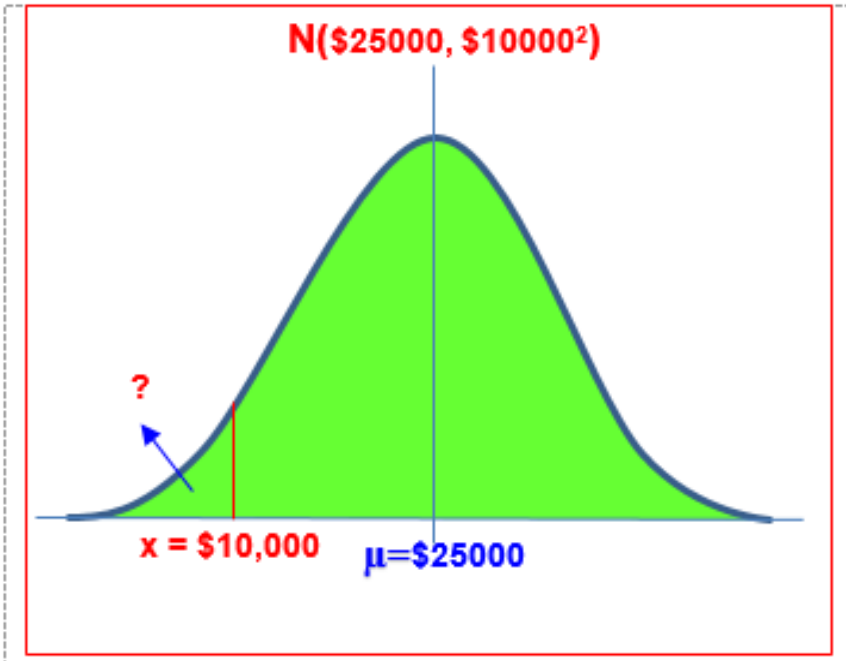



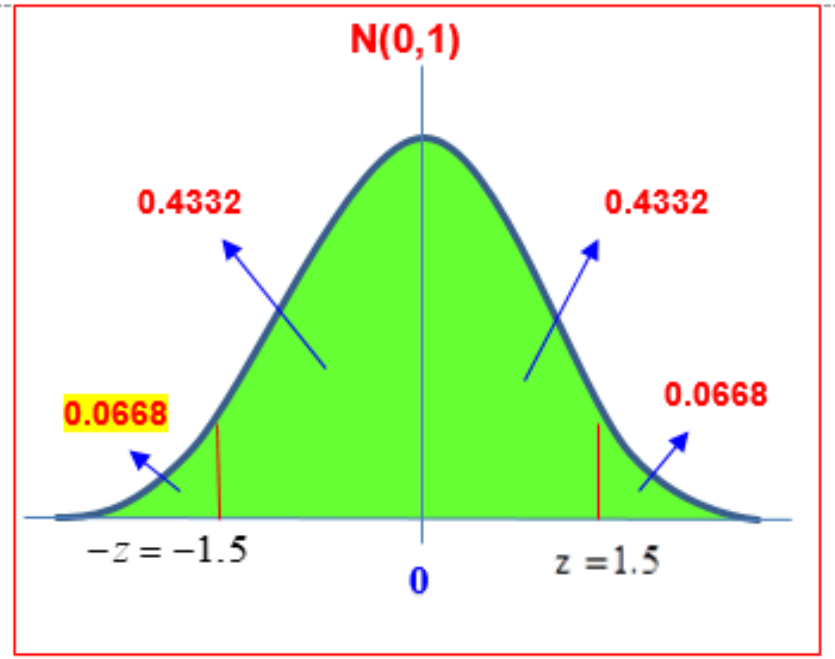
2. Family income ~ N($25000, $10000$^2$). If the poverty level is $10,000, what percentage of the population lives in poverty?

Solution: Let X = Family income. We want to find P(X ≤ $10,000). This is too hard to compute directly, so let Z = (X - $25,000)/$10,000. If x = $10,000, then $z = \frac{\$10,000 - \$25,000}{\$10,000} = -1.5$

So, P(X ≤ $10,000) = P(Z ≤ -1.5) = Φ(-1.5) = 1-Φ(1.5) = 1 -0.9332 =0.0668

(Easy Way: 0.5-0.4332=0.0668), hence, a little under 7% of the population lives in poverty.



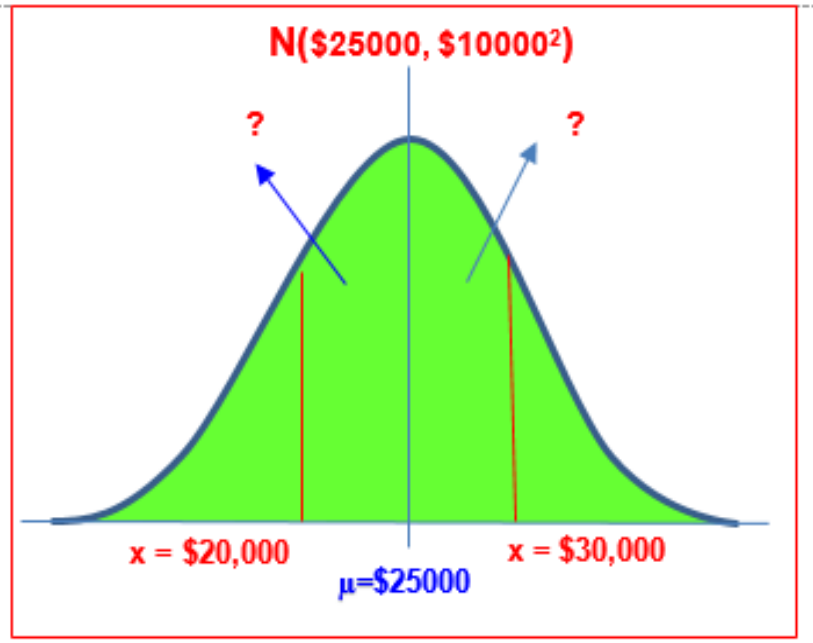



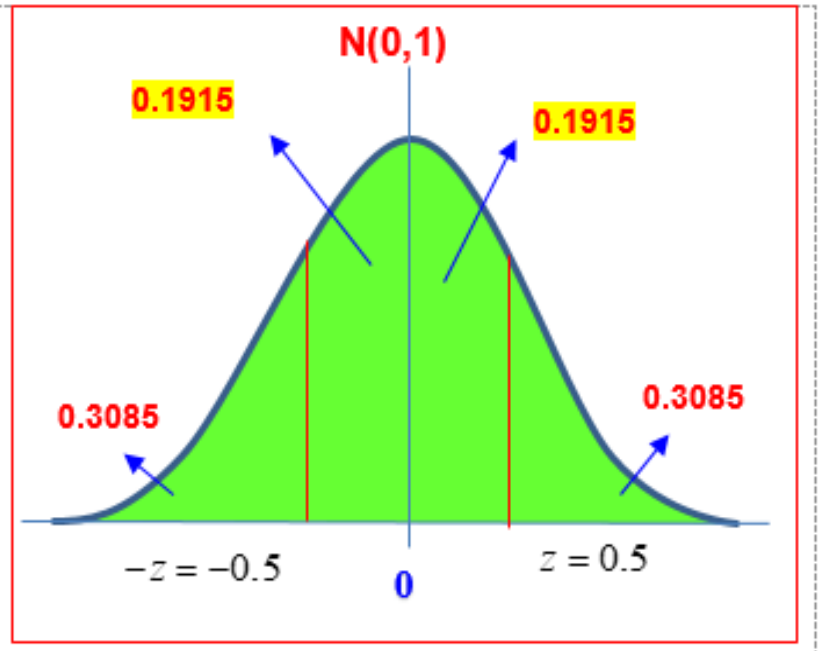
3. A new tax law is expected to benefit "middle income" families, those with incomes between $20,000 and $30,000. If Family income ~ N($25000, $10000$^2$), what percentage of the population will benefit from the law?

**Solution:** Let X = Family income. We want to find P($20,000 X ≤ $30,000)=? Z = (X - $25,000)/$10,000. If x = $20,000, $z = \frac{\$20,000 - \$25,000}{\$10,000} = -0.5$ and if x = $30,000, $z = \frac{\$30,000 - \$25,000}{\$10,000} = 0.5$. Hence,

P($20,000 ≤ X ≤ $30,000) = P(-0.5 ≤ Z ≤0.5) = 2×0.1915=0.383

Thus, about 38% of the taxpayers will benefit from the new law.



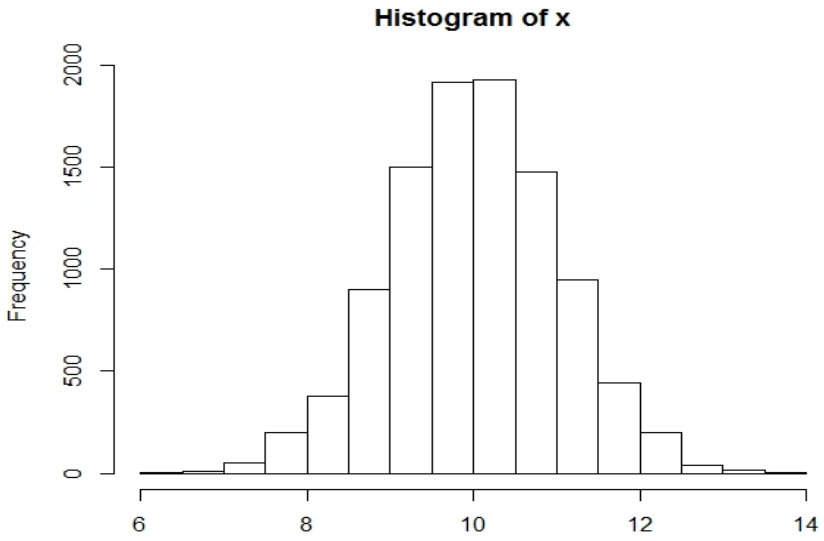
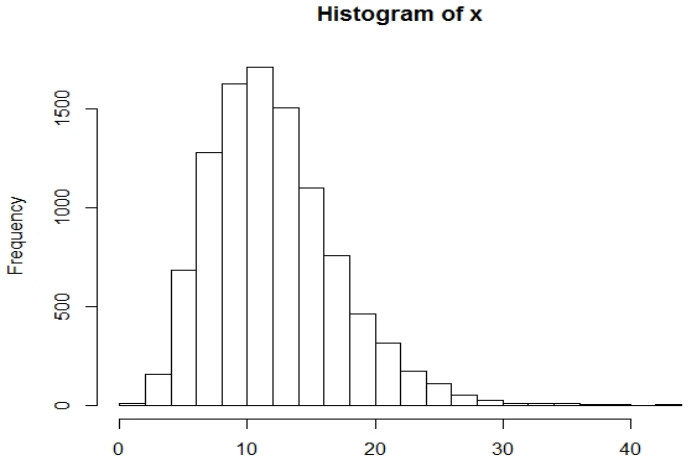



Figure 2 is the histogram of a normal data set. The histograms of normal data often reach their peaks at the sample median and then decrease on both sides of this point in a bell-shaped symmetric fashion.

**Figure 2. Histogram of a normal data set.**



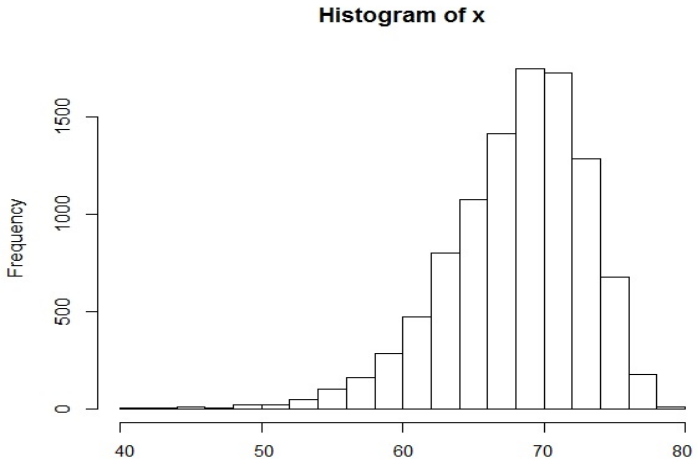




**Figure3. (a) Histogram of a data set skewed to the right.** **(b) Histogram of a data set skewed to the left.**

## Distribution of Sample Mean $\bar{X}$

Suppose a population has normal distrbution with μ and $\sigma^2$ parameters. Suppose a sample of n independent measurements is selected from this population:

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

Sample mean (sample statistic) $\frac{\sum_{i=1}^{n} X_i}{n}$ is a random variable hence it has probability distribution and so $E(\bar{X})$ and $V(\bar{X})$ are found as follow:

$$E(\bar{X}) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left[\sum_{i=1}^{n} E(X_i)\right] = \frac{n\mu}{n} = \mu$$

, (it is not required that $X_1, X_2, \ldots, X_n$ are independent)

$$V(\bar{X}) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} V(X_i)\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

, if $X_1, X_2, \ldots, X_n$ are independent.

# CENTRAL LIMIT THEOREM

**Central Limit Theorem:** For large sample size, the mean $\bar{X}$ of the a sample from a population with mean $\mu$ and standard deviation $\sigma$ possesses a sampling distribution that is ***approximately normal-regardless of the probability distribution of the sampled population.*** For the large the sample size, the approximation will be better.

From the theorem, the distibution of $\bar{X}$, $n \to \infty \quad \bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ where the term of ∞ is coresponded to $n \geq 30$.

Using the result of theorem, the standardized $\bar{X}$ has a standard normal distibution:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathrm{N}(0,1)$$

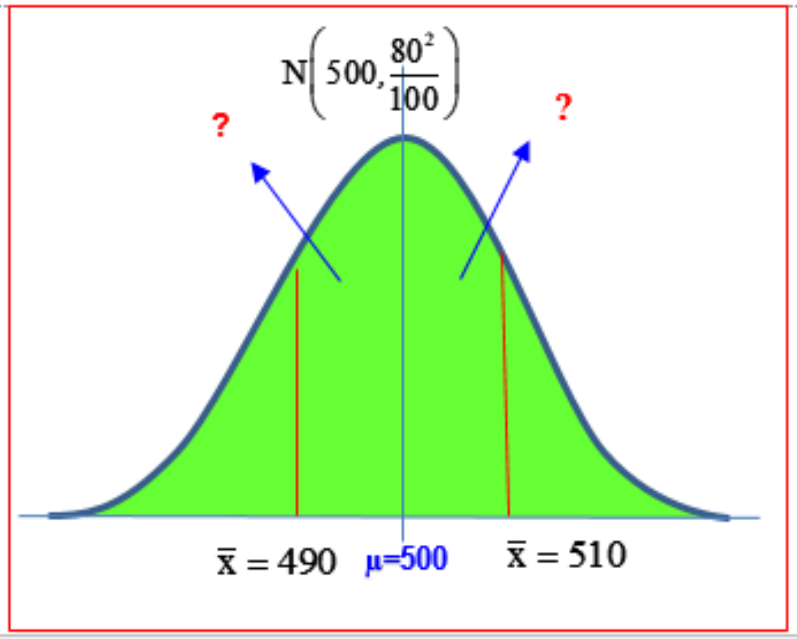
In addition, the distribution of $\sum_{i=1}^{n} X_i$ : $n \to \infty$ $\sum_{i=1}^{n} X \sim \mathrm{N}\left(n\mu, n\sigma^2\right)$

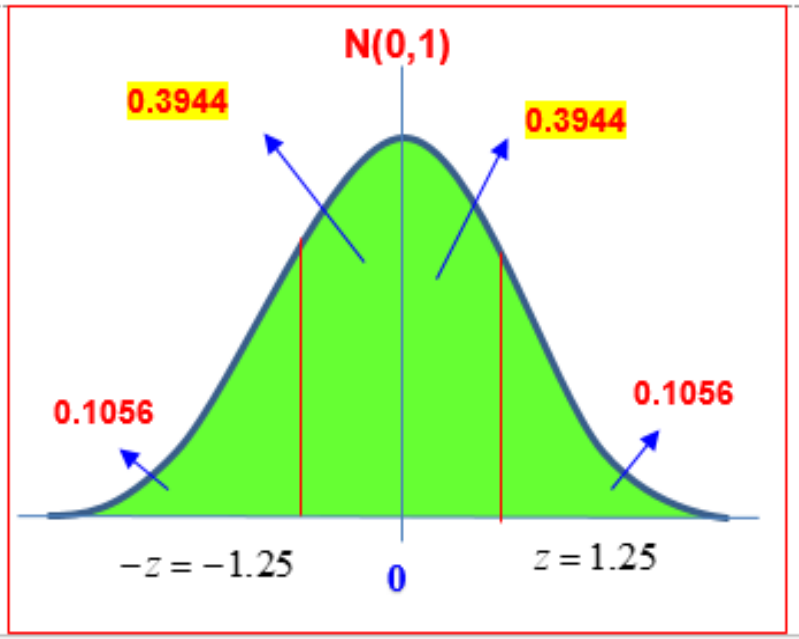
Then the standardized $\sum_{i=1}^{n} X_i$ has a standard normal distibution: $Z = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \sim \mathrm{N}(0,1)$

**Example:** The life of electronic devices has a distribution with mean 500 hours and standard deviation 80. Suppose a system used 100 devices. **Find probability if the expected life of system is at least 490 and at most 510.**

**Solution:** If $X \sim N(500, 80^2)$ **(since expected life system means average life of system)** then from CLT since n=100> 30 you can write $\bar{X} \sim N\left(500, \frac{80^2}{100}\right)$ and so that $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$ you can use SND table to find the probability.

$$P(490 < \bar{X} < 510) = P\left(\frac{490-500}{80/10} < Z < \frac{510-500}{80/10}\right) = P(-1.25 < Z < 1.25) = 2 \times 0.3944 = 0.7888$$



**Example:** From past experience, it is known that the number of tickets purchased by a student standing in line at the ticket window for the football match of UCLA against USC follows a distribution that has mean $\mu = 2.4$ and standard deviation $\sigma = 2$. Suppose that few hours before the start of one of these matches there are 100 eager students standing in line to purchase tickets. If only 250 tickets remain, what is the probability that all 100 students will be able to purchase the tickets they desire?

**Solution:** We are given that $\mu = 2.4$; $\sigma = 2$; $n = 100$. $X \sim N(2.4, 2^2)$. There are 250 tickets available, so the 100 students will be able to purchase the tickets they want if all together ask for less than 250 tickets.

$\sum_{i=1}^{100} X_i$ shows the total number of tickets asked by 100 students. We want to find

$$P\left(\sum_{i=1}^{100} X_i \le 250\right) = ?$$

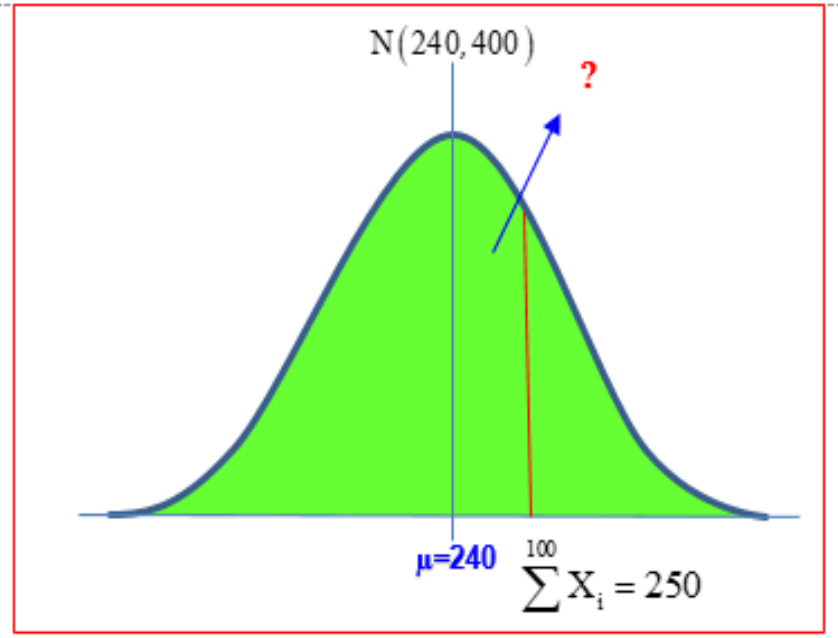
Since $n = 100 > 30$ from CLT $n \to \infty$ $\sum_{i=1}^{n} X \sim N(n\mu, n\sigma^2)$ then $Z = \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1)$

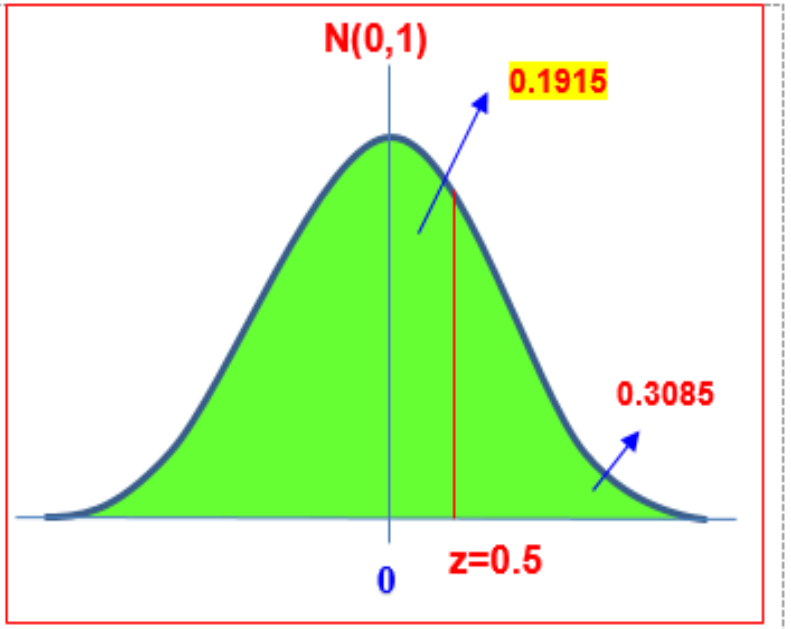
**The probability for that is**

$$E\left(\sum_{i=1}^{100} X_i\right) = 100 \times 2.4 \qquad V\left(\sum_{i=1}^{100} X_i\right) = 100 \times 2^2$$

$$P(Total\ number\ of\ tickets : \sum_{i=1}^{100} X_i \le 250) = P\left(\frac{\sum_{i=1}^{100} X_i - E\left(\sum_{i=1}^{100} X_i\right)}{\sqrt{V\left(\sum_{i=1}^{100} X_i\right)}} \le \frac{250-100\times 2.4}{2\sqrt{100}}\right)$$

$$= P\left(Z \le \frac{250-100\times 2.4}{2\sqrt{100}}\right) = P(Z \le 0.5) = 0.5 + 0.1915 = 0.6915$$



## EXAMPLES

1. A large freight elevator (nakliye/yük asansörü) can transport a maximum of 9800 pounds. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean $\mu$=205 pounds and standard deviation $\sigma$=15 pounds. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

**Solution:**

**X: Weight of each box**

$$X \sim \text{with } \mu = 205,\ \sigma^2 = 15^2 = 225$$

**When we load all boxes (49) onto the elevator:**

$$\text{From CLT } \sum_{i=1}^{49} X_i \sim N(49\times 205,\ 49\times 225)$$

$$E\left(\sum_{i=1}^{49} X_i\right) = n\mu = 49\times 205 = 10045 \quad \text{and} \quad V\left(\sum_{i=1}^{49} X_i\right) = n\sigma^2 = 49\times 225 = 11025$$

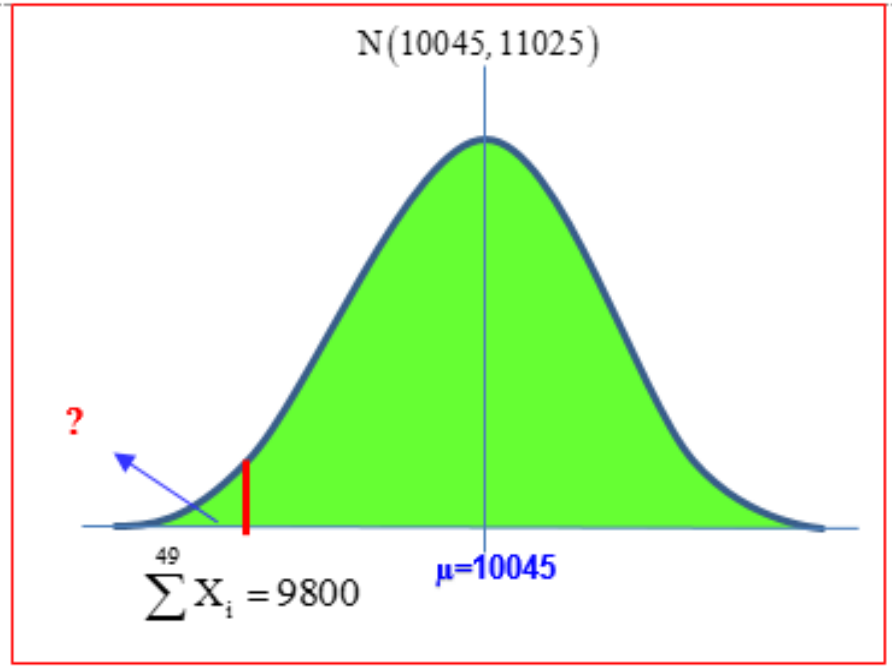
$$n \to \infty \quad \sum_{i=1}^{49} X_i \sim N(10045, 11025)$$

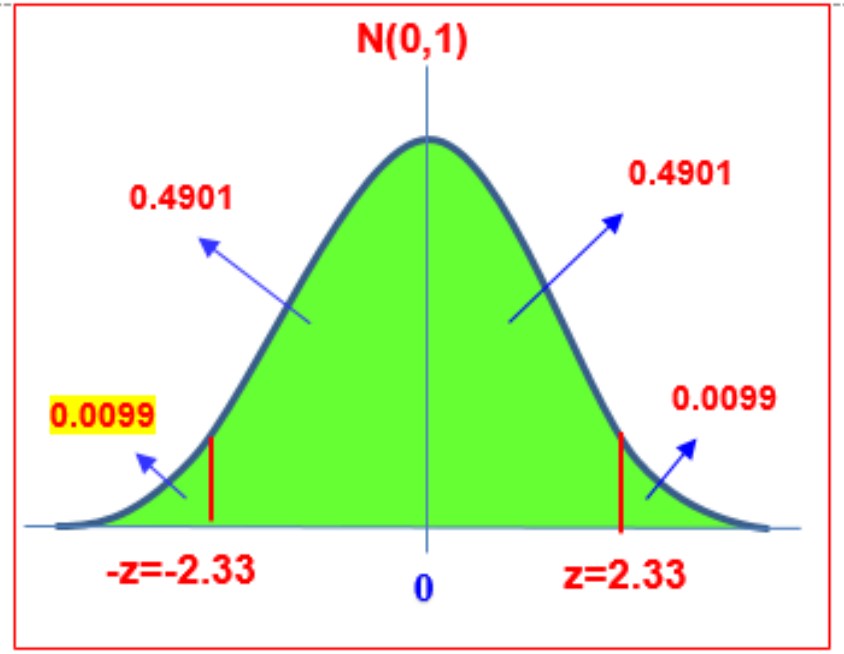
$$P\left(\sum_{i=1}^{49} X_i < 9800\right) = P\left(\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} < \frac{9800-10045}{\sqrt{11025}}\right)$$

$\downarrow$

total weight of 49 boxes

$$= P(Z < -2.33) = 0.5 - 0.4901 = 0.0099$$



2. The amount of mineral water consumed by a person per day on the job is normally distributed with mean 19 ounces and standard deviation 5 ounces. A company supplies its employees with 2000 ounces of mineral water daily. The company has 100 employees.

a) Find the probability that the mineral water supplied by the company will not satisfy the water demanded by its employees.

**Solution:**

$$X \sim N(19, 25) \quad \Rightarrow \text{CLT} \quad n \to \infty \quad \sum X_i \sim N(1900, 2500)$$

$$\downarrow \qquad\qquad (n = 100 \geq 30)$$

amount of water

consumed by a person per day

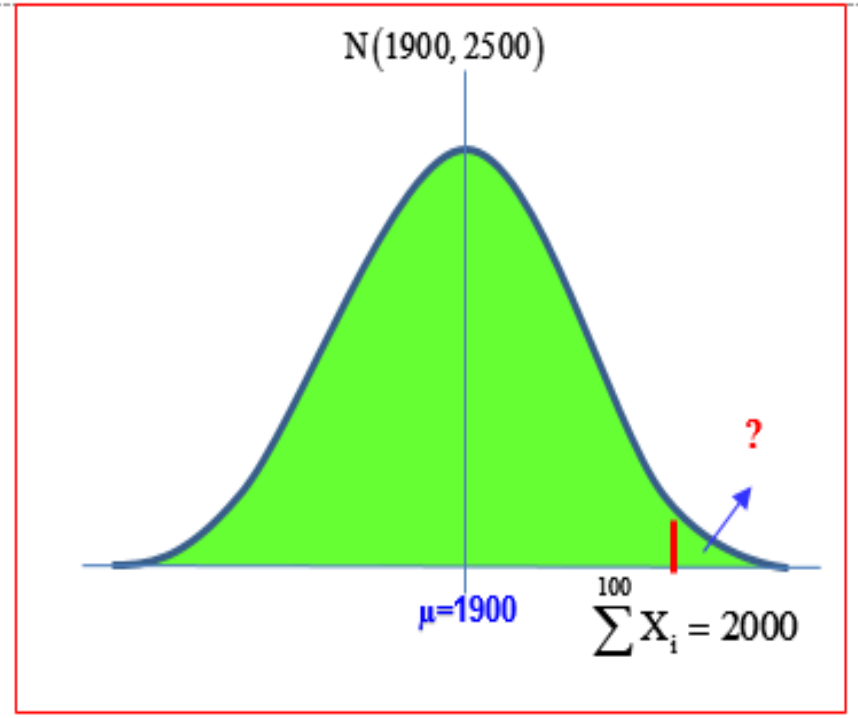
If the total amount of water demand by 100 employees per day exceeds 2000 ounces the company will not satisfy the water demand of employees. So we will find the

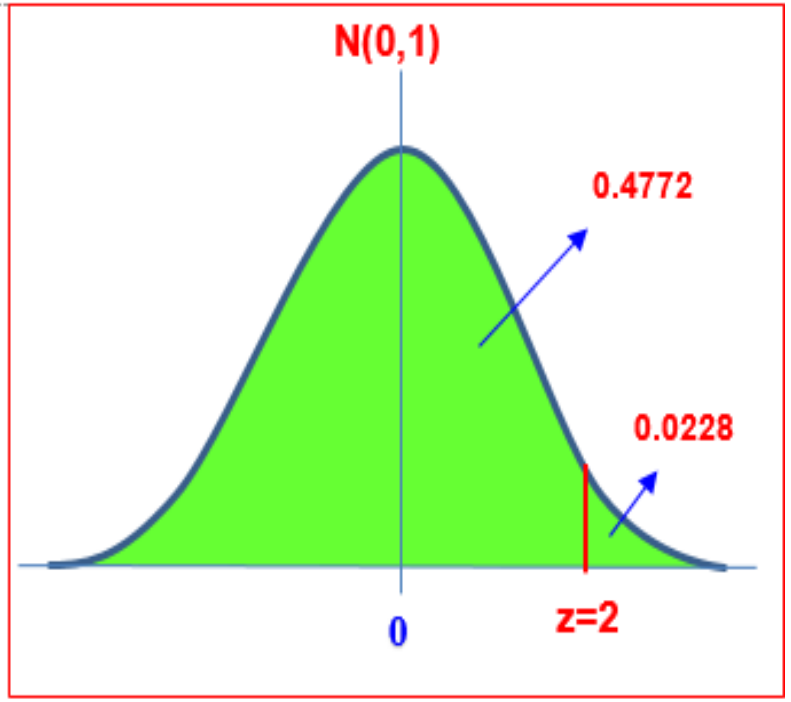
$$P\left(\sum_{i=1}^{100} X_i > 2000\right) = ?$$

$$P\left(\sum X_i > 2000\right) = P\left(Z > \frac{2000 - \overbrace{1900}^{\mu}}{\sqrt{\underbrace{2500}_{\sigma^2}}}\right) = P(Z > 2) = 0.5 - 0.4772 = 0.0228$$

↓

corresponds to the lack of water supplement for all employees.



**b)** Find the probability that in the next 4 days the company will not satisfy the water demanded by its employees on at least 1 of these 4 days. Assume that the amount of mineral water consumed by the employees of the company is independent from day to day.

**Solution:**

**X: The number of day of 4 days in which the company would face with the problem lack of water for their employees.**

We will find $P(X \geq 1) = ?$

$$X \sim \text{Binomial}(n, p)$$

$$\text{Binomial Distribution}: p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n$$

$$E(X) = np \quad \text{and} \quad V(X) = np(1-p)$$

$p = 0.0228$ is the probability that the mineral water supplied by the company will not satisfy the water demanded by its employees per day.

$$P(X \geq 1) = \sum_{x=1}^{4} \binom{4}{x} 0.0228^x \, 0.9772^{4-x}$$

c) Find the probability that during the next year (365 days) the company will not satisfy the water demanded by its employees on more than 15 days.

X: The number of days during the next year (365 days) the company will not satisfy the water demanded by its employees

$$X \sim \text{Binomial}(n = 365, p = 0.0228) \qquad E(X) = np \qquad V(X) = np(1-p)$$

**Exact solution:** $$P(X > 15) = \sum_{x=16}^{365} \binom{365}{x} 0.0228^{x}\, 0.9772^{365-x}$$

## Approximate solution from Central Limit Theorem

$P(X > 15) = ?$ . **If X has a binomial distribution then** $Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$ **(as result of central limit theorem,** $n \to \infty \quad X \sim N(np, np(1-p))$**) then n=365, p=0.0228 and**

$$N(8.322, 8.1322)$$

$$P\left(\frac{X - np}{\sqrt{npq}} > \frac{15 - 365 \times 0.0228}{\sqrt{365 \times 0.0228 \times 0.9772}}\right) = P\left(Z > \frac{6.678}{\sqrt{8.1322}}\right)$$

$$P(Z > 2.34) = 0.5 - 0.4904 = 0.0096$$