Probability Distribution

Md. Mahfuzur Rahman

Senior Lecturer, Statistics

Probability Distribution

Some name of the discrete and continuous distribution.

Discrete Distribution	**Continuous distribution**
Bernoulli distribution	Normal distribution
Binomial distribution	Exponential distribution
Poisson distribution	Gamma distribution
Hypergeommetric distribution	Beta distribution
Geometric distribution	Lognormal distribution
Negative binomial distribution	Weibull distribution
Discrete Uniform distribution	Continuous uniform distribution

Binomial Distribution

Bernoulli trial:

A trial that has only two possible outcomes (often called 'Success' and 'Failure')

Outcome	Success	failure
Probability	p	1-p

Let,

- n independent Bernoulli trials are performed
- Each trial has the same probability of success, p

Binomial Distribution

Let,

X= number of success in n trials

Then, X is a binomial random variable with distribution function (pmf),

$$p(x) = {}^{n}C_{x}\, p^{x}\, (1-p)^{n-x} \qquad ; \; x = 0, 1, 2, \dots, n$$

$$= \frac{n!}{(n-x)!\; x!}\; p^{x}\, (1-p)^{n-x}$$

Here, n!= n(n-1)(n-2) … 1 0!=1 1!=1 2!= 2*1= 2 3!= 3*2*1=6

We write it as, $X \sim binomial\,(n, p)$

Binomial Distribution

Mean of the binomial distribution, $\mu = E(X) = \sum x\, p(x) = np$

Variance of binomial distribution, $\sigma^2 = E(X^2) - \mu^2 = np(1-p) = npq$

Standard deviation of binomial distribution, $\sigma = \sqrt{npq}$

Binomial Distribution

Example 3:

There are 3 multiple choice questions in a MCQ test. Each MCQ consists of four possible choices and only one of them is correct. If an examinee answers those MCQ randomly (without knowing the correct answers)

a. What is the probability that exactly any two of the answers will be correct?
b. What is the probability that at least two of the answers will be correct?
c. What is the probability that at most two of the answers will be correct?
d. What will be the average or expected number of correct answers?
e. Also, find the standard deviation of number of correct answers.

Binomial Distribution

Solution:

Let,

X= number of correct answers selected in 3 MCQs

Here, p = probability of selecting correct answer per question = ¼ = 0.25

$$\therefore X \sim binomial\ (n = 2, p = 0.25)$$

$$p(x) = {}^{3}C_{x}\ (0.25)^{x}\ (1 - 0.25)^{3-x} \qquad ;\ x = 0, 1, 2, 3$$

$$= \frac{3!}{(3 - x)!\ x!}(0.25)^{x}\ (0.75)^{n-x}$$

Binomial Distribution

Solution (contd.)**:**

a. probability that exactly any two of the answers will be correct-

$$P(X = 2) = \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2}$$

$$= \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1 = \frac{3*2*1}{1*(2*1)} * 0.0625 * 0.75 = 0.141$$

b. probability that at least two of the answers will be correct-

$$P(X \geq 2) = P(X = 2) + P(X = 3)$$

$$= \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2} + \frac{3!}{(3-3)!\,3!}(0.25)^3\,(0.75)^{3-3}$$

$$= \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1 + \frac{3!}{0!\,3!}(0.25)^3\,(0.75)^0 = 0.141 + 0.016 = 0.157$$

Binomial Distribution

Solution **(contd.):**

c. probability that at most two of the answers will be correct-

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= \frac{3!}{(3-0)!\,0!}(0.25)^0\,(0.75)^{3-0} + \frac{3!}{(3-1)!\,1!}(0.25)^1\,(0.75)^{3-1}$$

$$+ \frac{3!}{(3-2)!\,2!}(0.25)^2\,(0.75)^{3-2}$$

$$= \frac{3!}{3!\,0!}(0.25)^0\,(0.75)^3 + \frac{3!}{2!\,1!}(0.25)^1\,(0.75)^2 + \frac{3!}{1!\,2!}(0.25)^2\,(0.75)^1$$

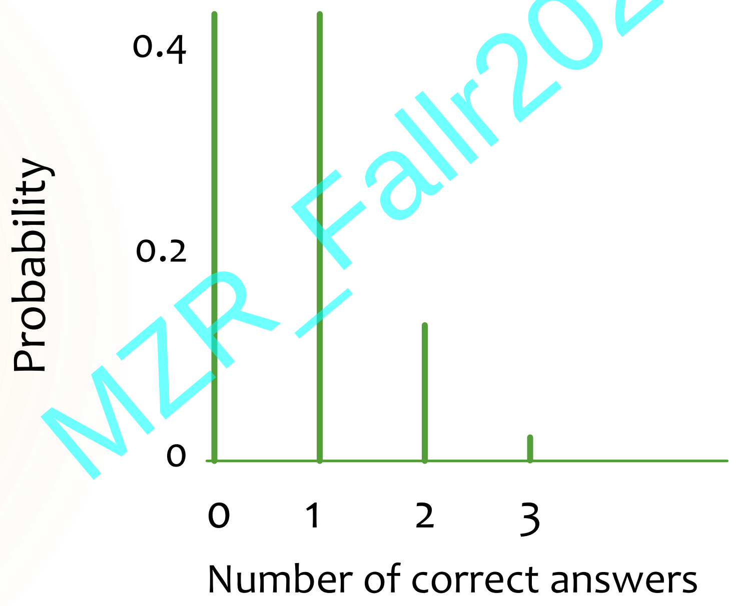
$$= 0.422 + 0.422 + 0.141 = 0.985$$

d. $E(X) = np = 3 * .25 = 0.75$

e. $SD(X) = \sqrt{npq} = \sqrt{3 * 0.25 * 0.75} = \ 0.75$

Binomial Distribution

X	0	1	2	3
P(x)	0.422	0.422	0.141	0.016



Geometric Distribution

- Suppose that independent Bernoulli trials each having probability p of success are performed until a success occurs.

- Let, X= No. of trails required to get the first success.
- Then X~ geometric (p)

Pmf:

P (X=x) = $(1-p)^{x-1}p$

- E(X)=1/ p
- V(X)= 1- p /p^2

Geometric Distribution

- A fair die is thrown until a “6” occurs.
- i) What is the probability that at most 3 tosses will be required?
- Ii) What will be the average no. of tosses required?

Geometric Distribution

- Coming home from work, you always seem to hit every light. You calculate the odds of making it through a light to be 0.2. How many lights can you expect to hit before making it through one? With what std. dev. ? What's the probability of the 3^{rd} light being the first one that is green?

Geometric Distribution

- Coming home from work, you always seem to hit every light. You calculate the odds of making it through a light to be 0.2. How many lights can you expect to hit before making it through one? With what std. dev. ? What's the probability of the 3rd light being the first one that is green?
- Solution:
- Mean= $\mu = \frac{1}{p} = \frac{1}{0.2} = \mathbf{5}\ lights$
- Std. Dev.= $\sigma = \frac{\sqrt{1-p}}{p} = \frac{\sqrt{q}}{p} = \frac{\sqrt{0.8}}{0.2} = \mathbf{4.47}\ lights$
- Probability: $P(X = 3)$ = $(1-p)^{x-1}\ p$ = $(0.8)^2\ (0.2)$ = **0.128**

Poisson Distribution

Let,

X= a random variable usually counts or number of occurrences

Then, X is a Poisson random variable with distribution function (pmf),

$$p(x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} \quad ; \; x = 0, 1, 2, \ldots$$

We write it as, $X \sim Poisson\ (\lambda)$

Poisson Distribution

Mean of the Poisson distribution, $\mu = E(X) = \sum x\, p(x) = \lambda$

Variance of Poisson distribution, $\sigma^2 = E(X^2) - \mu^2 = \lambda$

Standard deviation of Poisson distribution, $\sigma = \sqrt{\lambda}$

* Count data with no upper limit

- Example:
- Some random quantities that can be modeled by Poisson distribution:
- (i) Number of patients in a waiting room in an hour.
- (ii) Number of surgeries performed in a month.
- (iii) Number of car accidents daily in a city.
- (iv) Number of rats in each house in a particular city.

- **Note**
- λ is the average (mean) of the distribution.
- If X = The number of patients seen in the emergency unit in a day, and if

  $X \sim \text{Poisson}(\lambda)$, then:
- The average (mean) of patients seen every day in the emergency unit = λ.
- The average (mean) of patients seen every month in the emergency unit =30λ.
- The average (mean) of patients seen every year in the emergency unit = 365λ.
- The average (mean) of patients seen every hour in the emergency unit = $\lambda/24$.

- Also, notice that:
- (i) If Y = The number of patients seen every month, then:
- Y ~ Poisson (λ^*), where $\lambda^*=30\lambda$
- (ii) W = The number of patients seen every year, then:
- W ~ Poisson (λ^*), where $\lambda^*=365\lambda$
- (iii) V = The number of patients seen every hour, then:
- V ~ Poisson (λ^*), where $\lambda^*= \lambda/24$

Poisson Distribution

Example 4:

The average number of errors on a page of a certain magazine is 0.2. What is the probability that the next page (or a randomly selected page) you read contains

i. 0 (zero) error?

ii. 2 or more errors?

iii. What is the average error per page?

iv. Also, find standard deviation of the number of errors.

Poisson Distribution

Solution:

Let,

X= number of errors in a page

Here, λ = average number of errors per page= 0.2

$$\therefore X \sim Poisson\ (\lambda = 0.2)$$

$$p(x) = \frac{e^{-\lambda}\ \lambda^{x}}{x!} \quad ;\ x = 0, 1, 2, \ldots$$

$$= \frac{e^{-0.2}\ 0.2^{x}}{x!}$$

Poisson Distribution

Solution:

i. $P(X = 0) = \frac{e^{-0.2}\, 0.2^{x}}{x!} = \frac{e^{-0.2}\, 0.2^{0}}{0!} = \frac{e^{-0.2} * 1}{1} = 0.8187$

ii. $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$

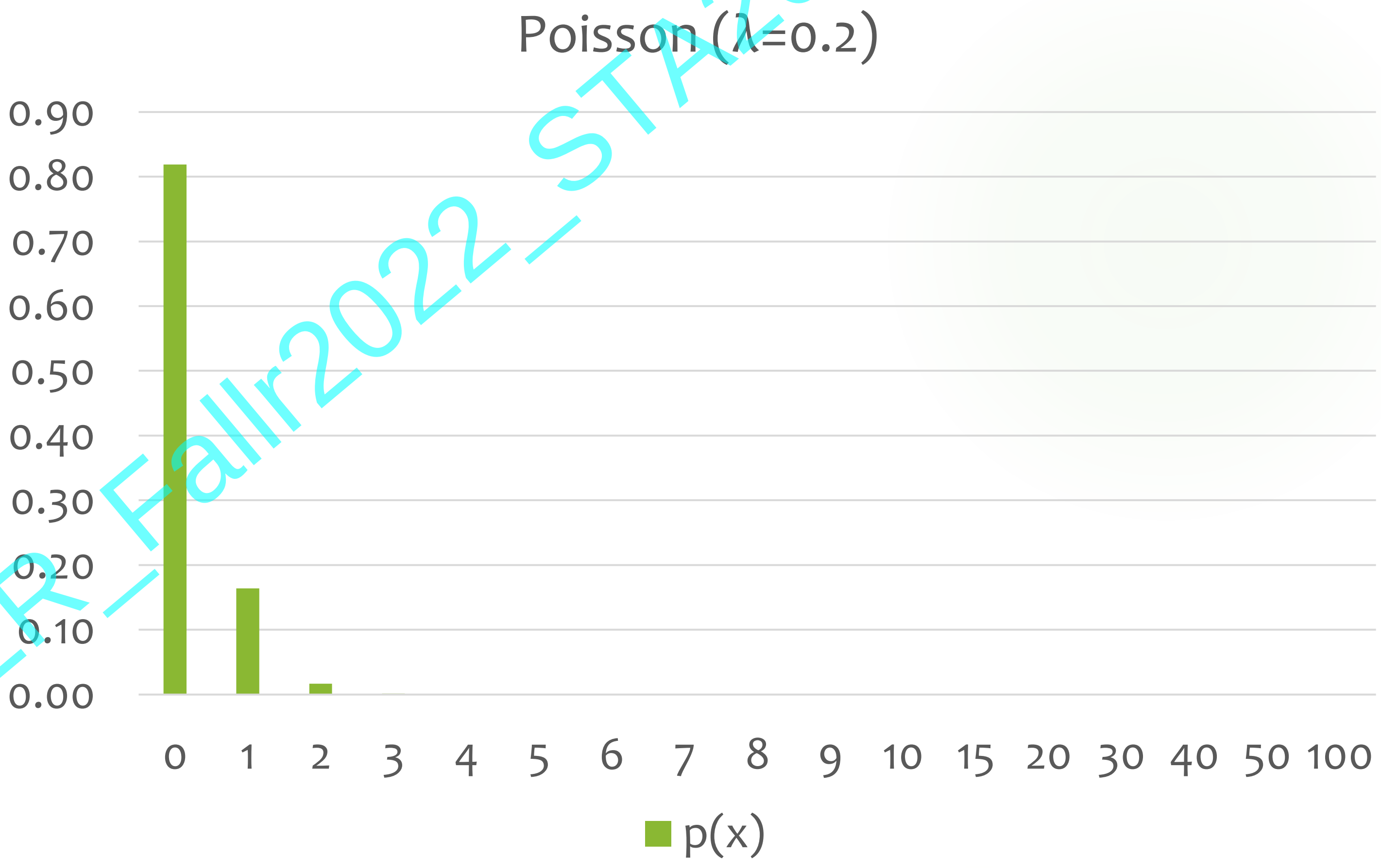
$= \frac{e^{-0.2}\, 0.2^{0}}{0!} + \frac{e^{-0.2}\, 0.2^{1}}{1!} = 1 - [e^{-0.2} + e^{-0.2} * 0.2] = 0.01756$

iii. Average number of errors, $E(X) = \lambda = 0.2$

iv. Standard deviation, $SD(X) = \sqrt{\lambda} = \sqrt{0.2} = 0.45$

Poisson Distribution

x	p(x)
0	0.82
1	0.16
2	0.02
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
10	0.00
15	0.00
20	0.00
30	0.00
40	0.00
50	0.00
100	0.00



Suppose that the number of accidents per day in a city has a Poisson distribution with average 2 accidents.

(1) What is the probability that in a day:

(i) The number of accidents will be 5?

(ii) The number of accidents will be less than 2?

(2) What is the probability that there will be 6 accidents in 2 days?

(3) What is the probability that there will be no accidents in an hour?

- (1) X= number of accidents in a day

  $X \sim$ Poisson (2) (λ=2)

- (ii) P (X=5) = 0.036089 (ii) P (X<2) =0.406005

- (2) Y = number of accidents in 2 days

  Y ~ Poisson(4) (λ^*=4)

  P (Y=6) = 0.1042

W = number of accidents in an hour

W ~ Poisson (0.083) (λ^{**}= λ /24= 2/24= 0.083)

P (W=0) = 0.9204

Normal Distribution

Let,

X is a continuous random variable

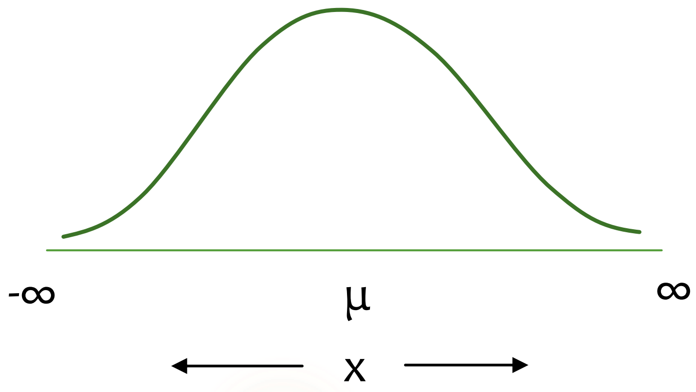
Then, if X has a probability density function (pdf),

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

We write it as, $X \sim N\left(\mu, \sigma^2\right)$

Mean, $E(X) = \mu$

Variance, $V(X) = \sigma^2$



Standard Normal Distribution

Let,

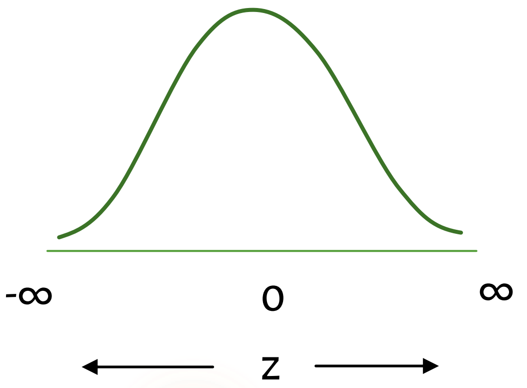
$$Z = \frac{X - \mu}{\sigma}$$

Then, Mean, $E(Z) = 0$

Variance, $V(Z) = 1$

And, if Z has a probability density function (pdf),

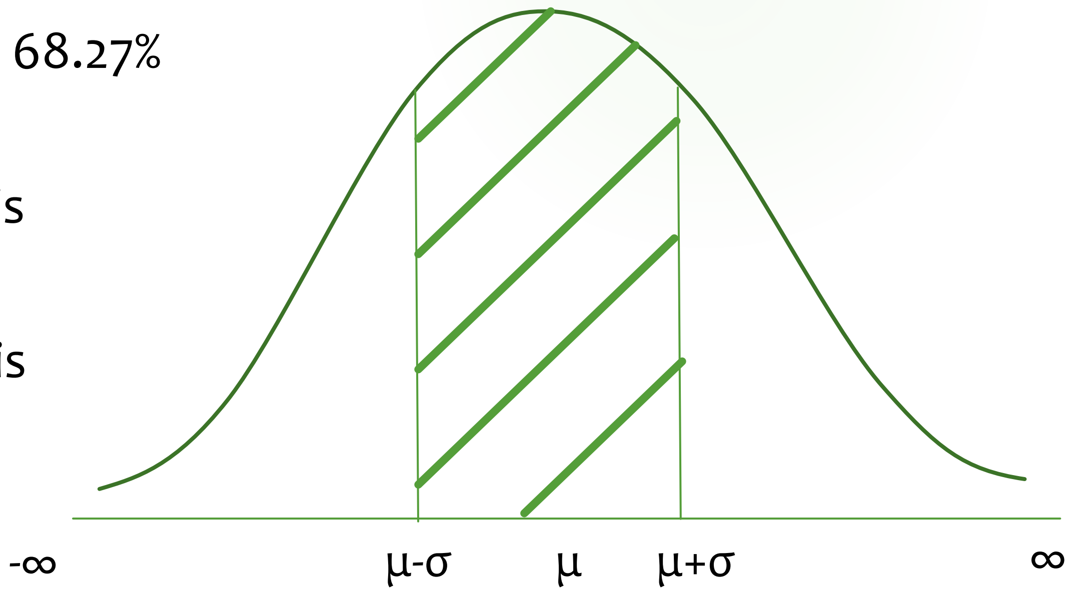
$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \; ; \; -\infty < z < \infty$$



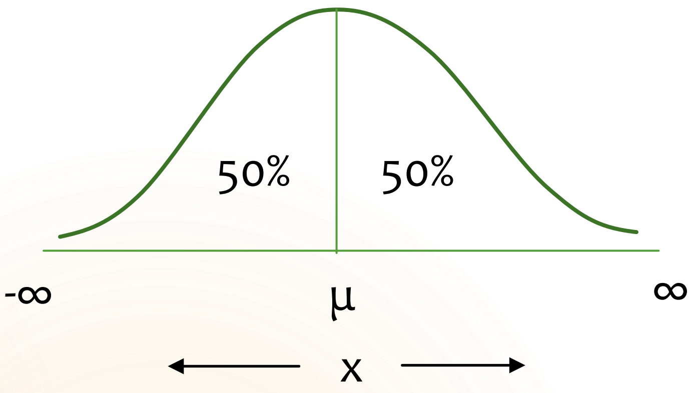
We write it as, $Z \sim N\,(0, 1)$

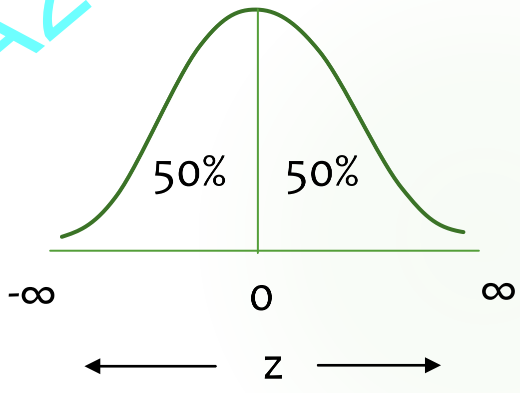
Characteristics of a Normal Distribution

1. Mean= Median = Mode
2. Symmetric and Mesokurtic
3. Bell-shaped curve
4. The area under the curve lying between $\mu \pm \sigma$ is 68.27% of the total area
5. The area under the curve lying between $\mu \pm 2\sigma$ is 95.45% of the total area
6. The area under the curve lying between $\mu \pm 3\sigma$ is 99.73% of the total area



Characteristics of a Normal Distribution





$P[X<\mu] = P[X>\mu] = 0.5$

$P[Z<0] = P[Z>0] = 0.5$

$P[X<-x] = P[X>x]$

$P[Z<-z] = P[Z>z]$

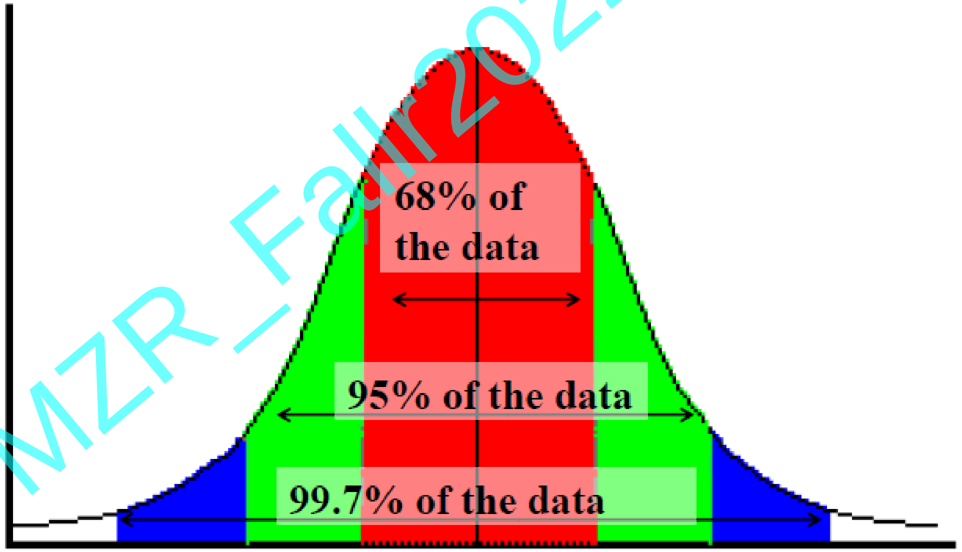
Normal Distribution Table

Z-table

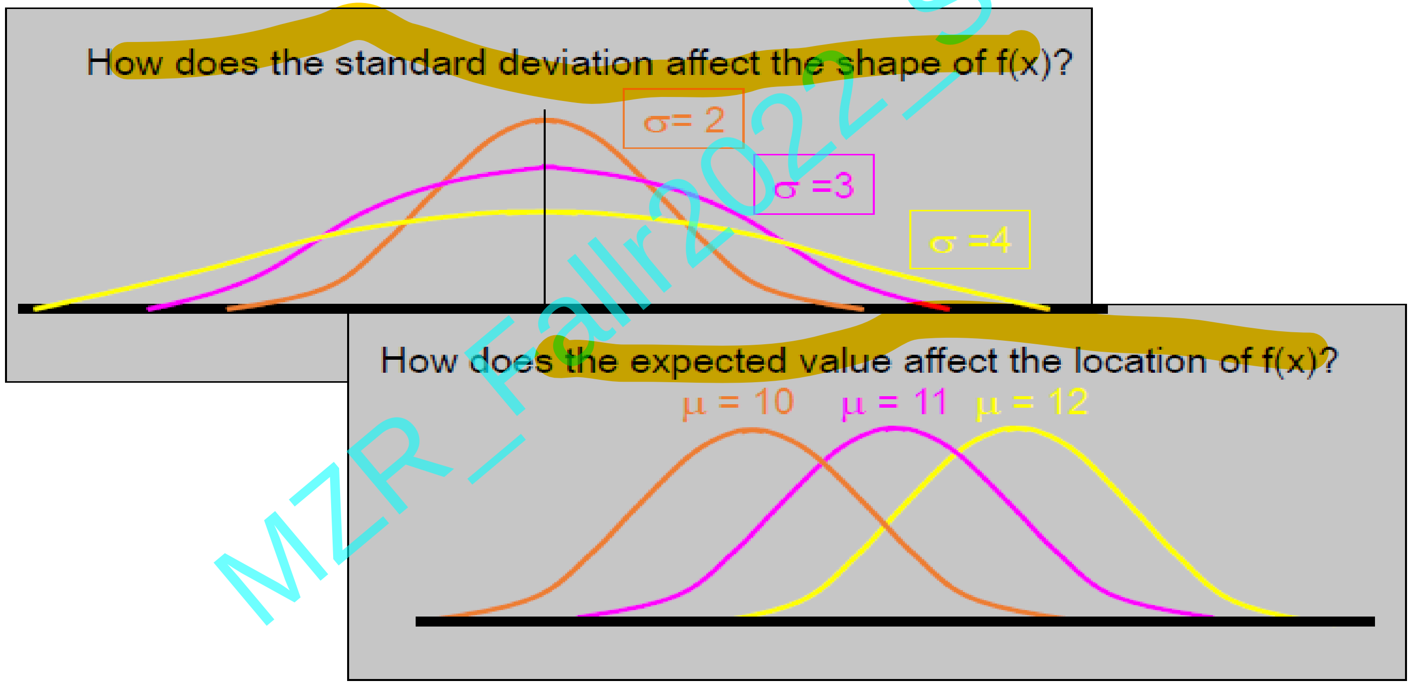
- Normal distribution table provides probabilities for N(0,1) i.e. for standard normal distribution
- Usually, normal table gives P[0 < Z < z] for positive values of Z.
- For other values, we can use the property of symmetry with median 0 of standard normal distribution
- To find probabilities for a normal random variable X, we can transform the probability statement about X in terms of probability statement for Z and then calculate the probability using the standard normal distribution table or Z-table

$$P[X < a] = P\left[\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right] = P\left[Z < \frac{a-\mu}{\sigma}\right]$$

68-95-99.7 Rule



The effects of μ and σ



Example: Given a normal distribution with $\mu = 50, \sigma = 10,$ find the probability that X assumes a value between 45 and 62

Solution: The Z values corresponding to x_1=45 and x_2=62

$$\therefore Z_1 = \frac{45-50}{10} = -0.5 \qquad Z_2 = \frac{62-50}{10} = 1.2$$

P(45<x<62) = P(-0.5<Z<1.2)

= P(Z<1.2) – P(Z<-0.5)

= P(Z<1.2) –{1- P(Z<0.5)}

= ϕ(1.2) - {1- ϕ(0.5)}

= 0.8849 – {1 – 0.6915} [USING TABLE] = 0.5764

Example: The weekly incomes of the bankers of a bank follow normal distribution with a mean of \$ 1,000 and std. of \$100.

What is the livelihood of selecting a banker whose weekly income is between \$1000 and \$100?

$$P(1,000<X<1,100) = P\left(\frac{1000-1000}{100} < Z < \frac{1100-1000}{100}\right)$$

$$= P(0<Z<1)$$

$$= 0.3413$$

$P(X<1,100) = P(Z<1) = 0.5 + 0.3413 = 0.8413$

$P(790<X<1,000) = (-2.10<Z<0) = 0.4821$

$P(X<790) = P(Z<-2.10) = 0.5 - 0.4821 = 0.0179$

$P(X>790) = P(Z>-2.10) = 0.4821 + 0.5 = 0.9821$

$$\# P(X>482) = P[Z> \frac{482-400}{50}] = P(Z>1.64) = 0.5 - 0.4495 = 0.0505$$

Finding Area Under the Normal Curve using Z-table

Example 6:

The number of viewers of a TV show per week has a mean of 29 million with a standard deviation of 5 million. Assume that, the number of viewers of that show follows a normal distribution.

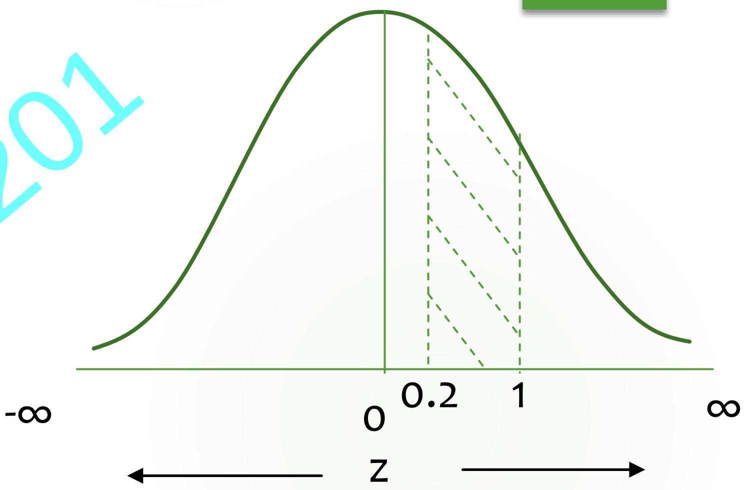
What is the probability that, next week's show will-

a. Have between 30 and 34 million viewers?

b. Have at least 23 million viewers?

c. Exceed 40 million viewers?

Finding Area Under the Normal Curve using Z-table



Solution:

Let, X= Number of viewers of the show per week (in million)

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

a. the probability that, next week's show will have between 30 and 34 million viewers-

$$P[30 \le X \le 34] = P\left[\frac{30-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{34-\mu}{\sigma}\right] = P\left[\frac{30-29}{5} \le \frac{X-\mu}{\sigma} \le \frac{34-29}{5}\right]$$

$$= P[0.20 \le Z \le 1] = P[0 \le Z \le 1] - P[0 \le Z \le 0.2] = 0.3413 - 0.0793$$

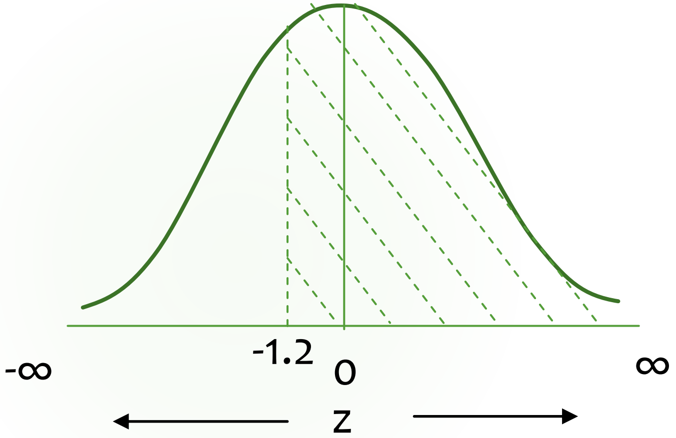
$$= 0.262$$

Finding Area Under the Normal Curve using Z-table

Solution (contd.)**:**

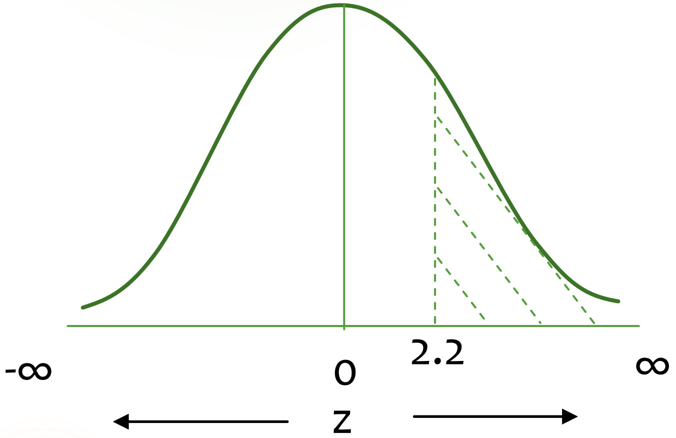
b. the probability that, next week's show will have at least 23 million viewers-

$$P[X \geq 23] = P\left[\frac{X-\mu}{\sigma} \geq \frac{23-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} \geq \frac{23-29}{5}\right]$$
$$= P[Z \geq -1.2] = P[-1.2 \leq Z \leq 0] + P[Z \geq 0] = 0.3849 + 0.5$$
$$= 0.8849$$



c. the probability that, next week's show will exceed 40 million viewers-

$$P[X > 40] = P\left[\frac{X-\mu}{\sigma} > \frac{40-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} > \frac{40-29}{5}\right]$$
$$= P[Z > 2.2] = P[Z \geq 0] - P[0 \leq Z \leq 2.2] = 0.5 - 0.4861 = 0.0139$$



4. A large group of students took a test in Physics and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the students

a) scored higher than 80?

b) should pass the test (grades≥60)?

c) should fail the test (grades<60)?

5. The annual salaries of employees in a large company are approximately normally distributed with a mean of $50,000 and a standard deviation of $20,000.
a) What percent of people earn less than $40,000?
b) What percent of people earn between $45,000 and $65,000?
c) What percent of people earn more than $70,000?

6. Suppose that the height of UCLA female students has normal distribution with mean 62 inches and standard deviation 8 inches.

a. Find the height below which is the shortest 30% of the female students.
b. b. Find the height above which is the tallest 5% of the female students.

4. a) For x = 80, z = 1
Area to the right (higher than) z = 1 is equal to 0.1586 = 15.87% scored more that 80.
b) For x = 60, z = -1
Area to the right of z = -1 is equal to 0.8413 = 84.13% should pass the test.
c)100% - 84.13% = 15.87% should fail the test.

5. a) For x = 40000, z = -0.5
Area to the left (less than) of z = -0.5 is equal to 0.3085 = 30.85% earn less than $40,000.
b) For x = 45000 , z = -0.25 and for x = 65000, z = 0.75
Area between z = -0.25 and z = 0.75 is equal to 0.3720 = 37.20 earn between $45,000 and $65,000.
c)For x = 70000, z = 1
Area to the right (higher) of z = 1 is equal to 0.1586 = 15.86% earn more than $70,000.

6. We are given $X \sim N(62, 8)$. a. We want to find the height h such that $P(X < h) = 0.30$. From the standard normal table this corresponds to $z = -0.525$. Therefore $-0.525 = h-62\ 8$ $\Rightarrow h = 57.8$ inches.

b. We want to find the height h such that $P(X > h) = 0.05$. From the standard normal table this corresponds to $z = 1.645$. Therefore $1.645 = h-62\ 8 \Rightarrow h = 75.16$ inches.

Finding Area Under the Normal Curve using Z-table

Example 7:

a. For what value of 'a', P[Z≤a] = 0.95?

b. For what value of 'a', P[Z≥a] = 0.05?

c. For what value of 'a', P[Z≤a] = 0.975?

Solution:

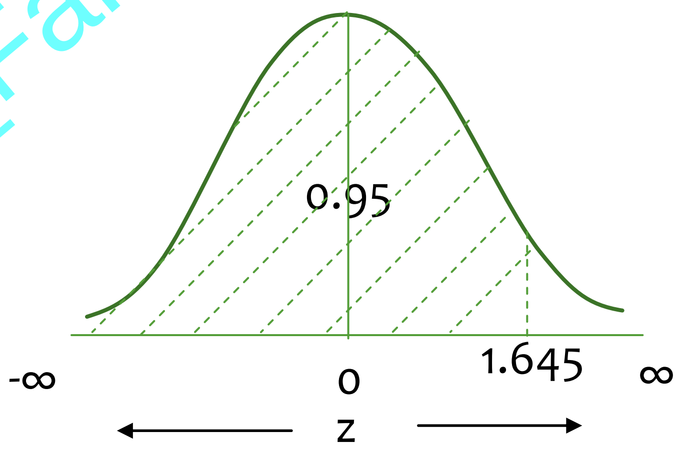
a. P[Z≤a] = 0.95

Or, P[Z≤0] + P[0<Z≤a] = 0.95

Or, 0.5+ P[0<Z≤a] =0.95

Or, P[0<Z≤a] =0.95-0.5= 0.45

For a= 1.645, P[0<Z≤a] =0.45



Finding Area Under the Normal Curve using Z-table

Solution (contd.)**:**

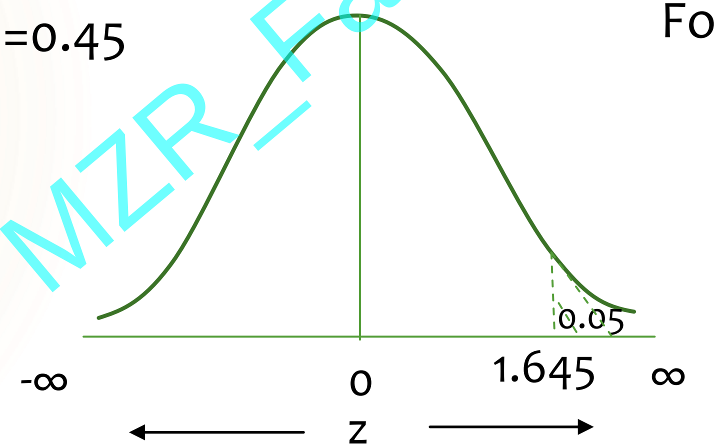
b. $P[Z \geq a] = 0.05$

Or, $P[Z \geq 0] - P[0<Z\leq a] = 0.05$

Or, $0.5 - P[0<Z\leq a] = 0.05$

Or, $P[0<Z\leq a] = 0.5-0.05 = 0.45$

For $a = 1.645$, $P[0<Z\leq a] = 0.45$



Solution (contd.)**:**

c. $P[Z\leq a] = 0.975$

Or, $P[Z\leq 0] + P[0<Z\leq a] = 0.975$

Or, $0.5 + P[0<Z\leq a] = 0.975$

Or, $P[0<Z\leq a] = 0.975-0.5 = 0.475$

For $a = 1.96$, $P[0<Z\leq a] = 0.475$