STAT 105

Chapter 1
Some Discrete and Continuous Probability Distributions

Dr. Mona Elwakeel

Discrete Probability Distributions:

Definition:

The set of ordered pairs **(x, f(x))** is a probability function, probability mass function or probability distribution of the discrete random variable **X** if for each possible outcome **x,**

1. $f(x) \geq 0$

2. $\sum_{\forall x} f(x) = 1$

3. $P(X = x) = f(x)$

The Cumulative distribution Function:

The cumulative distribution function, denoted by **F(x)** of a discrete random variable **X** with probability distribution **f(x)** is given by:

$$F(x) = P(X \le x) = \sum_{X<x} f(x) \qquad for \ -\infty < x < \infty \qquad (1)$$

For example $F(2) = P(x \le 2)$

$$P(a \le X \le b) = F(b) - F(a-1) \qquad (2)$$

For example $F(3 \le x \le 7) = F(7) - F(2)$

Theorem:

The mean and variance of the discrete distributions are given by:

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n} \quad ,$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X-\mu)^2}{n} = \frac{\sum_{i=1}^{n} X_i^2 - n\mu^2}{n}$$

Continuous Probability Distributions

The function **f(x)** is a probability density function for the continuous random variable **X** defined over the set of real numbers **R**, if:

1. $f(x) \geq 0 \qquad for\ all\ x \in R$

2. $\int_{-\infty}^{\infty} f(x)\ dx = 1$

3. $P(a < X < b) = \int_{a}^{b} f(x)\, dx$

The Cumulative Distribution

The cumulative distribution **F(x)** of a continuous random variable **X** with density function **f(x)** is given by:

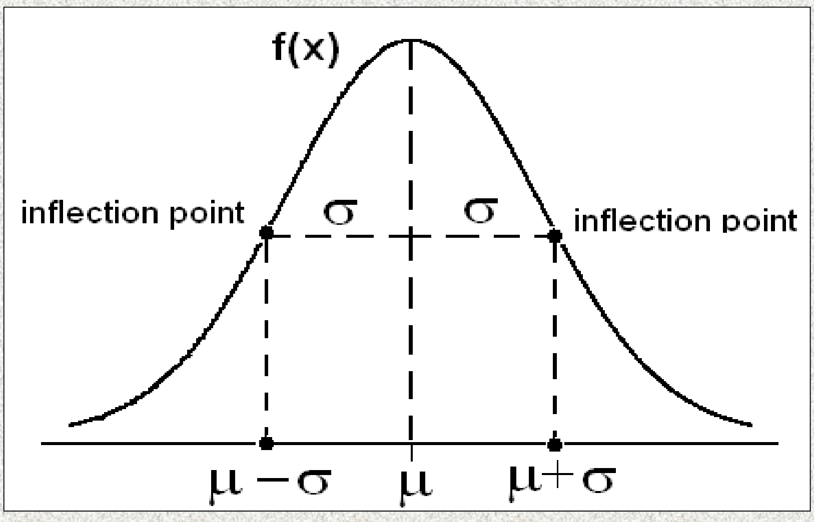
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx \quad for \quad -\infty < x < \infty$$

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$$

Normal Distribution:

- The probability density function of the normal random variable **X,** with mean μ and variance σ² is given by:

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



Properties of The Normal Curve :

- The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at X= μ , (Mode = Median = Mean).
- The curve is symmetric about a vertical axis through the mean μ .

- The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
- The total area under the curve and above the horizontal axis is equal to1.

Standard Normal Distribution:

- The distribution of a normal random variable with mean **zero** and variance **one** is called a standard normal distribution denoted by *Z≈N(0,1)*

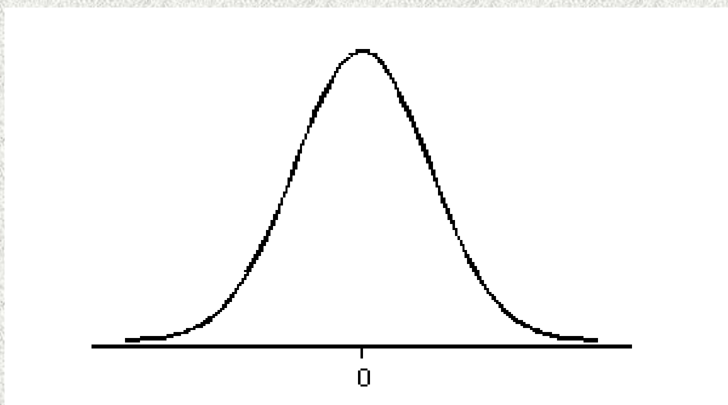
- **Areas under the Normal Curve:**

$$X \approx N(\mu, \sigma)$$

$$Z = \frac{X - \mu}{\sigma} \approx N(0,1)$$

- Using the standard normal tables to find the areas under the curve.

The pdf of Z~N(0,1) is given by:



EX (1):

Using the tables of the standard normal distribution, find:

$$(a) P(Z < 2.11)$$

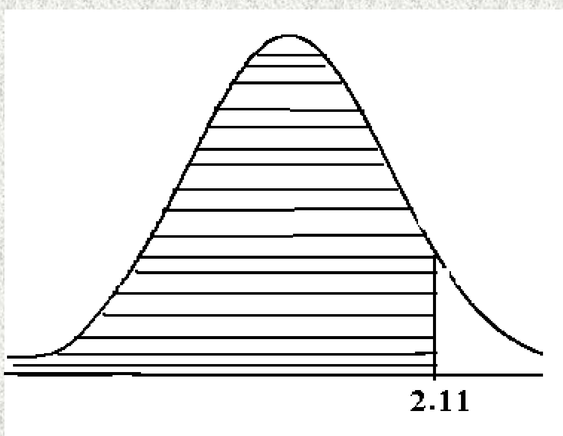
$$(b) P(Z > -1.33)$$

$$(c) P(Z = 3)$$

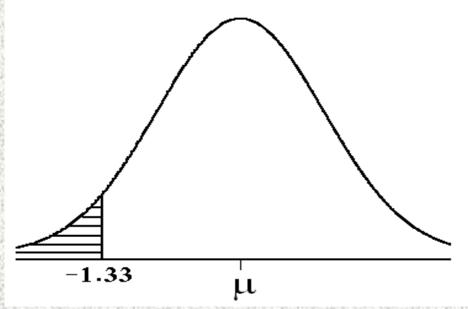
$$(d) P(-1.2 < Z < 2.1)$$

Solution:

$$(a) P(Z < 2.11) = 0.9826$$



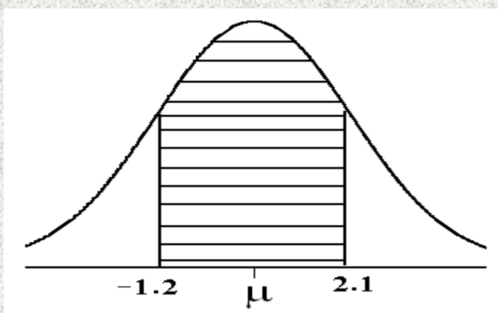
$$(b)\,P(Z > -1.33) = 1 - 0.0918 = 0.9082$$



15

$$(c)\,P(Z = 3) = 0$$

$$(d)\,P(-1.2 < Z < 2.1) = 0.9821 - 0.1151 = 0.867$$



16

EX (2):

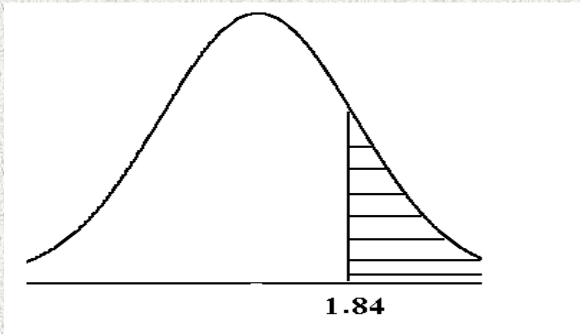
Using the standard normal tables, find the area under the curve that lies:

A. to the right of Z=1.84
B. to the left of z=2.51
C. between z=-1.97 and z=0.86
D. at the point z= -2. 15

Solution:

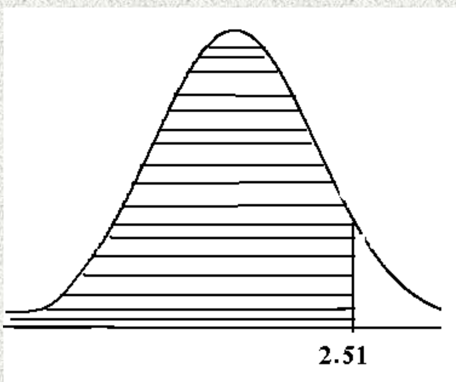
A. to the right of Z=1.84

$$P(Z > 1.84) = 1 - 0.9671 = 0.0329$$



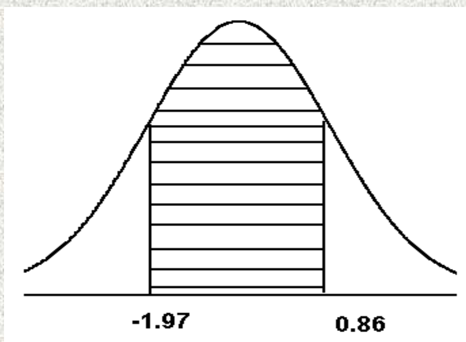
B. to the left of z=2.51

$$P(Z < 2.51) = 0.9940$$



C. between z=-1.97 and z=0.86

$$P(-1.97 < Z < 0.86) = 0.8051 - 0.0244 = 0.7807$$



D. at the point z= -2. 15

$$P(Z = -2.15) = 0$$

EX (3):

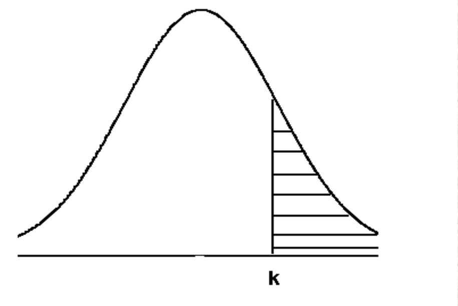
Find the constant **K** using the tables such that:

(a) $P(Z > K) = 0.3015$

(b) $P(K < Z < -0.18) = 0.4197$

Solution:

(a) $P(Z > K) = 0.3015$

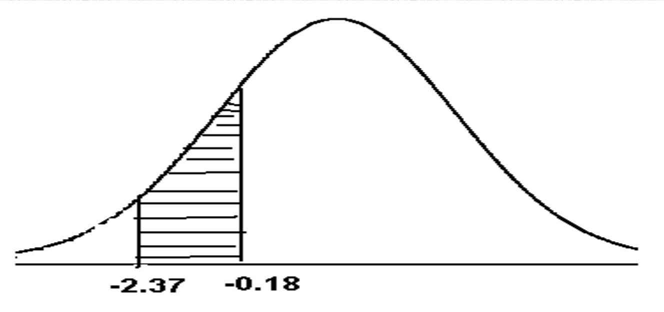


$$P(Z > K) = 0.3015 \Rightarrow 1 - 0.3015 = 0.6985 \Rightarrow k = 0.52$$

(b) $P(K < Z < -0.18) = 0.4197$

$$\Rightarrow 0.4286 - 0.4197 = 0.0089$$

$$\Rightarrow k = -2.37$$



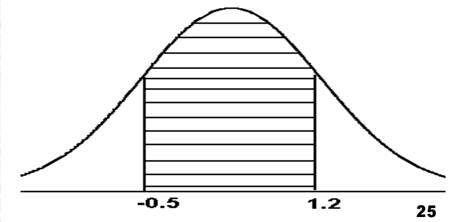
EX (4):

Given a normal distribution with μ=50 , σ=10 . Find the probability that **X** assumes a value between **45** and **62**.

Solution:

$$P(45 < X < 62) = P(\frac{45-50}{10} < Z < \frac{62-50}{10}) = P(-0.5 < Z < 1.2)$$

$$= 0.8849 - 0.3085 = 0.5764$$



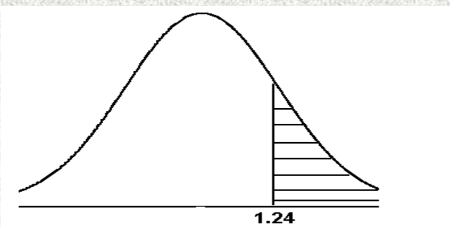
EX(5) :

Given a normal distribution with μ=300 , σ=50, find the probability that **X** assumes a value greater than **362**.

Solution:

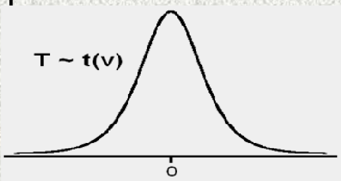
$$P(X > 362) = P(Z > \frac{362-300}{50}) = P(Z > 1.24)$$

$$= 1 - 0.8925 = 0.1075$$



t – Distribution:

* *t* distribution has the following properties:



1. It has mean of zero.
2. It is symmetric about the mean.
3. It ranges from $-\infty \ \ to \ \ \infty$.
4. Compared to the normal distribution, the *t* distribution is less peaked in the center and has higher tails.
5. It depends on the degrees of freedom **(n-1)**.
6. The t-distribution approaches the normal distribution as **(n-1)** approaches ∞.

EX(6):

Find:

$$(a) t_{0.025} \quad when \quad v = 14$$

$$(b) t_{0.01} \quad when \quad v = 10$$

$$(c) t_{0.995} \quad when \quad v = 7$$

solution

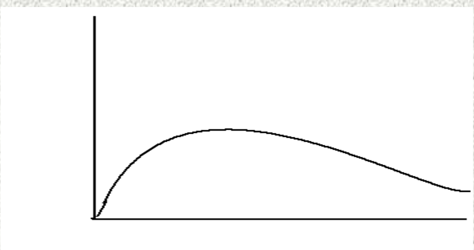
$$(a)t_{0.025} \quad at \quad v = 14 \rightarrow t = -2.1448$$

$$(b)t_{0.01} \quad at \quad v = 10 \rightarrow t = -2.764$$

$$(c)t_{0.995} \quad at \quad v = 7 \rightarrow t = 3.4995$$

The Chi- Square Distribution

The chi-square distribution is important because it is the basis for a number of procedures in statical inference. The central role played by the chi-square distribution in inference springs from its relationship to normal distribution. We will discuss this distribution in more detail in later chapters.



Let ν be a positive integer. Then a random variable X is said to have a chi-square distribution with parameter ν if the pdf of X is the gamma density with $\alpha = \nu\backslash 2$ *and* $\beta = 2$. The pdf of a chi-square rv is:

$$f(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\,\Gamma(\nu/2)} x^{(\nu/2)-1}\, e^{-x/2} & , \quad x \geq 0 \\ 0 & , otherwise \end{cases}$$

The parameter ν is called the number of degrees of freedom (df) of X. The symbol χ^2 is often used in place of "chi-square"

EX(9):

By using table of chi- square distribution, Find:

a) $\chi^2_{0.995}$ *when* $\nu = 19$

b) $\chi^2_{0.025}$ *when* $\nu = 15$

c) $\chi^2_{0.95}$ *when* $\nu = 2$

solution

a) $\chi^2_{19,\ 0.995} = 38.582$

b) $\chi^2_{15,\ 0.025} = 6.262$

c) $\chi^2_{2,\ 0.95} = 5.991$

Note that:

When the degree of freedom (df) not exist in the table of chi square , we have to use the following rule :

$$\chi^2_{A,df} = \chi^2_{A,df(L)} + \frac{df - df(L)}{df(H) - df(L)}\left[\chi^2_{A,df(H)} - \chi^2_{A,df(L)}\right]$$

EX(10):

By using table of chi- square distribution, Find:

$\chi^2_{42,0.975}$

Solu:

$$\chi^2_{42,0.975} = \chi^2_{40,0.975} + \frac{42-40}{45-40}\left[\chi^2_{45,0.975} - \chi^2_{40,0.975}\right]$$

$$=59.342+\frac{2}{5}[65.410 - 59.342] = 61.7692$$

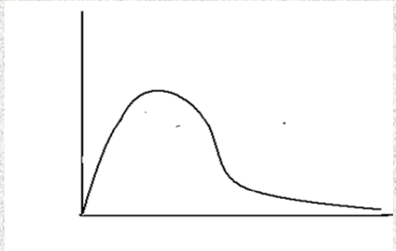
35

Stat 105 Dr. Mona Fouad Elwakeel

F – Distribution:

The F- distribution

For tests involving two variances, it is necessary to have independent samples($of\ size\ n_1, n_2$) from two normal distributions. Under these conditions, we can get a new random variable has F-distribution with degrees of freedom$(n_1 - 1), (n_2 - 1)$ and denoted by$F_{(n_1-1),(n_2-1)}$ which can be obtained from its tables at a different values of α. Note that a variable with F distribution can only have positive values.



36

EX(11):

From the tables of F-distribution ,Find:

a) $F_{0.995,15,24}$

b) $F_{0.005,15,24}$

c) $F_{0.9,10,8}$

solution

a) $F_{0.995,15,24} = 3.25$

b) $F_{0.005,15,22} = \frac{1}{F_{0.995,24,15}} = \frac{1}{3.79} = 0.2639$

c) $F_{0.9,10,8} = 2.54$