|| Jai Sri Gurudev ||

S.J.C Institute of Technology, Chickballapur

Department of Mathematics

Lecture Notes

Complex Analysis, Probability and Statistical Methods [18MATH41]

Prepared By:

Purushotham.P

Assistant Professor

SJC Institute of technology

Email id: ppurushothamp48@gmail.com

Phone: 8197481658

MODULE-05 : JOINT PROBABILITY DISTRIBUTION AND SAMPLING THEORY

Joint Probability :

Let $X=\{x_1,x_2,x_3,\cdots x_m\}$ and $Y=\{y_1,y_2,y_3,\cdots y_n\}$ are two discrete random variables, then the joint probability function of X and Y is defined as

$P(X=x_i, Y=y_j) = P(x_i,y_j) = f(x_i,y_j) = P_{ij} = f_{ij}$

where the function $f(x,y)$ satisfy the conditions

i) $f(x,y) \geq 0$ ii) $\sum_i \sum_j f(x_i,y_j) = 1$

The joint probability table as shown below.

X \ Y	y_1	y_2	y_3	....	y_n	$f(x)$
x_1	P_{11}	P_{12}	P_{13}	....	P_{1n}	$f(x_1)$
x_2	P_{21}	P_{22}	P_{23}	....	P_{2n}	$f(x_2)$
x_3	P_{31}	P_{32}	P_{33}	....	P_{3n}	$f(x_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	....	$\vdots$	$\vdots$
x_m	P_{m1}	P_{m2}	P_{m3}	....	P_{mn}	$f(x_m)$
$g(y_i)$	$g(y_1)$	$g(y_2)$	$g(y_3)$	....	$g(y_n)$	1.

Marginal probability distribution:

In the joint probability table $f(x_1), f(x_2), f(x_3) \ldots f(x_m)$ and $g(y_1), g(y_2), g(y_3) \ldots g(y_n)$ are called the marginal probability distributions respectively and represents the sum of all entries in all the rows and columns.

Independent random Variables:

The discrete random variables X and Y are said to be independent if $P(X = x_i, Y = y_j) = P(X = x_i) . P(Y = y_j)$ for every i, j and it is equivalent to $P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) = f(x_i) \cdot g(y_j)$

or $COV(X, Y) = 0$

Expectation, Variance and Covariance:

Let X be the random variable taking the random values $x_1, x_2, x_3 \ldots x_m$, having the probability function $f(x)$.
Then

a) The expectation of X is denoted by $E(X)$ and is defined as
$\mu_x = E(X) = \sum_i x_i f(x_i)$

b) The expectation of Y is denoted by $E(Y)$ and is defined as
$\mu_Y = E(Y) = \sum_j y_j g(y_j)$

c) The Variance of X is denoted by σ_x^2 and is defined by
$\sigma_x^2 = E(X^2) - [E(X)]^2$
$\Rightarrow \sigma_x^2 = \sum_i x_i^2 f(x_i) - \mu_x^2$

d) The variance of Y is denoted by σ_Y^2 and is defined by
$\sigma_Y^2 = E(X^2) - [E(X)]^2$
$\Rightarrow \sigma_Y^2 = \sum_i y_i^2 g(y_i) - \mu_Y^2$

e) The covariance of X and Y is defined as $COV(X, Y) = E(XY) - E(X) . E(Y)$
$COV(X, Y) = \sum_i \sum_j x_i y_j f(x_i, y_j) - \mu_x . \mu_y$

f) The Correlation between X and Y is $\rho(X, Y) = \dfrac{COV(X, Y)}{\sigma_X \sigma_Y}$

Problems :

1) The joint distribution of two random variables X and Y as follows

X \ Y	-4	2	7
1	1/8	1/4	1/8
5	1/4	1/8	1/8

Compute the following

i) E(X) and E(Y)

ii) E(XY)

iii) σ_X and σ_Y

iv) $\rho(X, Y)$

Soln Given $x_1 = 1$, $x_2 = 5$

$y_1 = -4$, $y_2 = 2$, $y_3 = 7$

And the probabilities are

$P_{11} = \frac{1}{8}$, $P_{12} = \frac{1}{4}$, $P_{13} = \frac{1}{8}$

$P_{21} = \frac{1}{4}$, $P_{22} = \frac{1}{8}$, $P_{23} = \frac{1}{8}$

Given the joint probability distribution is as follows

X \ Y	-4	2	7	$f(x_i)$
1	1/8	1/4	1/8	1/2
5	1/4	1/8	1/8	1/2
$g(y_i)$	3/8	3/8	1/4	1

The marginal distribution of X and Y are

x_i	1	5
$f(x_i)$	1/2	1/2

y_i	-4	2	7
$g(y_i)$	3/8	3/8	1/4

i) $\mu_X = E(X) = \sum_i x_i f(x_i) = \left[1 \times \frac{1}{2}\right] + \left[5 \times \frac{1}{2}\right] = 3$

$\mu_Y = E(Y) = \sum_j y_j g(y_j) = \left[-4 \times \frac{3}{8}\right] + \left[2 \times \frac{3}{8}\right] + \left[7 \times \frac{1}{4}\right] = 1$

$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$

ii) $E(XY) = \left[1 \times (-4) \times \frac{1}{8}\right] + \left[1 \times 2 \times \frac{1}{4}\right] + \left[1 \times 7 \times \frac{1}{8}\right] + \left[5 \times (-4) \times \frac{1}{4}\right] + \left[5 \times 2 \times \frac{1}{8}\right] + \left[5 \times 7 \times \frac{1}{8}\right]$

$E(XY) = 3/2.$

iii) $\sigma_x^2 = E(X^2) - \mu_x^2$

$\Rightarrow \sigma_x^2 = \sum_i x_i^2 f(x_i) - \mu_x^2$

$\Rightarrow \sigma_x^2 = \left[1^2 \times \frac{1}{2}\right] + \left[5^2 \times \frac{1}{2}\right] - 9 = 13 - 9 = 4 \Rightarrow \sigma_x = 2$

$\sigma_y^2 = E(Y^2) - \mu_y^2$

$\Rightarrow \sigma_y^2 = \sum_j y_j^2 g(y_j) - \mu_y^2$

$\Rightarrow \sigma_x^2 = \left[(-4)^2 \times \frac{3}{8}\right] + \left[2^2 \times \frac{3}{8}\right] + \left[7^2 \times \frac{1}{4}\right] - 1^2 \Rightarrow \sigma_y^2 = \frac{75}{4} \Rightarrow \sigma_y = 4.33$

iv) $\operatorname{cov}(X,Y) = E(XY) - \mu_x \mu_y$

$\Rightarrow \operatorname{cov}(X,Y) = \frac{3}{2} - (3)(1) = -\frac{3}{2}$

v) $\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y}$

$\Rightarrow \rho(X,Y) = \frac{-3/2}{2 \times 4.33} = -0.1732$

2) Determine i) marginal distribution ii) Covariance between the discrete random variables X and Y, using the joint probability distribution:

X \ Y	3	4	5
2	1/6	1/6	1/6
5	1/12	1/12	1/12
7	1/12	1/12	1/12

Soln Given $x_1 = 2, x_2 = 5, x_3 = 7$

$y_1 = 3, y_2 = 4, y_3 = 5$

And the probabilities are

$P_{11} = \frac{1}{6}, P_{12} = \frac{1}{6}, P_{13} = \frac{1}{6}$

$P_{21} = \frac{1}{12}, P_{22} = \frac{1}{12}, P_{23} = \frac{1}{12}$

$P_{31} = 1/12, P_{32} = 1/12, P_{33} = 1/12$

The joint distribution table is as follows.

X \ Y	3	4	5	$f(x_i)$
2	1/6	1/6	1/6	1/2
5	1/12	1/12	1/12	1/4
7	1/12	1/12	1/12	1/4
$g(y_j)$	1/3	1/3	1/3	1

The marginal distributions of X and Y are

x_i	2	5	7
$f(x_i)$	1/2	1/4	1/4

y_j	3	4	5
$g(y_j)$	1/3	1/3	1/3

$$\mu_X = E(X) = \sum_i x_i f(x_i)$$

$$\Rightarrow \mu_X = \left[2 \times \frac{1}{2}\right] + \left(5 \times \frac{1}{4}\right) + \left(7 \times \frac{1}{4}\right)$$

$$\Rightarrow \mu_X = 4$$

$$\mu_Y = E(Y) = \sum_j y_j g(y_j)$$

$$\Rightarrow \mu_Y = \left(3 \times \frac{1}{3}\right) + \left[4 \times \frac{1}{3}\right] + \left[5 \times \frac{1}{3}\right]$$

$$\Rightarrow \mu_Y = 4$$

$$E(XY) = \sum_{i,j} x_i y_j p_{ij}$$

$$\Rightarrow E(XY) = \left(2 \times 3 \times \frac{1}{6}\right) + \left[2 \times 4 \times \frac{1}{6}\right] + \left[2 \times 5 \times \frac{1}{6}\right] + \left[5 \times 3 \times \frac{1}{12}\right] + \left(5 \times 4 \times \frac{1}{12}\right) + \left[5 \times 5 \times \frac{1}{12}\right]$$

$$+ \left(7 \times 3 \times \frac{1}{12}\right) + \left[7 \times 4 \times \frac{1}{12}\right] + \left[7 \times 5 \times \frac{1}{12}\right] = 16$$

$$\therefore \text{COV}(X, Y) = E(XY) - \mu_X \mu_Y$$

$$\Rightarrow \text{COV}(X, Y) = 16 - 4 \times 4$$

$$\text{COV}(X, Y) = 0$$

3) The joint probability distribution of discrete random variables X and Y is given below:

X \ Y	1	3	6
1	1/9	1/6	1/18
3	1/6	1/4	1/12
6	1/18	1/12	1/36

Determine i) marginal distribution of X and Y ii) Are X and Y statistically independent?

Given $x_1 = 1, x_2 = 3, x_3 = 6$

$y_1 = 1, y_2 = 3, y_3 = 6$

And the probabilities are

$P_{11} = \frac{1}{9}, P_{12} = \frac{1}{6}, P_{13} = \frac{1}{18}$

$P_{21} = \frac{1}{6}, P_{22} = \frac{1}{4}, P_{23} = \frac{1}{12}$

$P_{31} = \frac{1}{18}, P_{32} = \frac{1}{12}, P_{33} = \frac{1}{36}$

The joint distribution table is as follows

X \ Y	1	3	6	$f(x_i)$
1	1/9	1/6	1/18	1/3
3	1/6	1/4	1/12	1/2
6	1/18	1/12	1/36	3/18
$g(y_j)$	1/3	1/2	3/18	1

The marginal distributions of X and Y are

x_i	1	3	6
$f(x_i)$	1/3	1/2	3/18

y_j	1	3	6
$g(y_j)$	1/3	1/2	3/18

$P(X = x_1, Y = y_1) = \frac{1}{9}, P(X = x_1) \times P(Y = y_1) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

$P(X = x_1, Y = y_1) = P(X = x_1) \times P(Y = y_1)$ and similarly

$P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j)$

$\therefore$ X and Y statistically independent.

4) Determine i) marginal distribution ii) Covariance between the discrete random variables X and Y along the Correlation using the joint probability distribution

X \ Y	1	3	9
2	1/8	1/24	1/12
4	1/4	1/4	0
6	1/8	1/24	1/12

Soln Given $x_1 = 2, x_2 = 4, x_3 = 6$

$y_1 = 1, y_2 = 3, y_3 = 9$

And the probabilities are

$P_{11} = \frac{1}{8}$, $P_{12} = \frac{1}{24}$; $P_{13} = \frac{1}{12}$

$P_{21} = \frac{1}{4}$, $P_{22} = \frac{1}{4}$, $P_{23} = 0$

$P_{31} = \frac{1}{8}$, $P_{32} = \frac{1}{24}$, $P_{33} = \frac{1}{12}$

The joint distribution table is as follows.

X \ Y	1	3	9	$f(x_i)$
2	1/8	1/24	1/12	1/4
4	1/4	1/4	0	1/2
6	1/8	1/24	1/12	1/4
$g(y_i)$	1/2	1/3	1/6	1

The marginal distributions of X and Y are

x_i	2	4	6
$f(x_i)$	1/4	1/2	1/4

y_i	1	3	9
$g(y_i)$	1/2	1/3	1/6

$$\mu_X = E(X) = \sum_i x_i f(x_i)$$

$$\Rightarrow \mu_X = \left[2 \times \frac{1}{4}\right] + \left[4 \times \frac{1}{2}\right] + \left[6 \times \frac{1}{4}\right]$$

$$\Rightarrow \mu_X = 4$$

$$\mu_Y = E(Y) = \sum_j y_j g(y_j)$$

$$\Rightarrow \mu_Y = \left[1 \times \frac{1}{2}\right] + \left[3 \times \frac{1}{3}\right] + \left[9 \times \frac{1}{6}\right]$$

$$\Rightarrow \mu_Y = 3$$

$$E(XY) = \sum_{i,j} x_i y_j P_{ij}$$

$$\Rightarrow E(XY) = \left[2 \times 1 \times \frac{1}{8}\right] + \left[2 \times 3 \times \frac{1}{24}\right] + \left[2 \times 9 \times \frac{1}{12}\right] + \left[4 \times 1 \times \frac{1}{4}\right] + \left[4 \times 3 \times \frac{1}{4}\right] + 4 \times 9 \times 0$$
$$+ \left[6 \times 1 \times \frac{1}{8}\right] + \left[6 \times 3 \times \frac{1}{24}\right] + \left[6 \times 9 \times \frac{1}{12}\right] = 12$$

$$\therefore \operatorname{cov}(X, Y) = E(XY) - \mu_X \mu_Y$$

$$\Rightarrow \operatorname{cov}(X, Y) = 12 - 4 \times 3$$

$$\Rightarrow \operatorname{cov}(X, Y) = 0$$

The Correlation $\rho(X, Y) = \dfrac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = 0$

5) The joint probability distribution for two random variables X and Y is as given below

X \ Y	-2	-1	4	5
1	0.1	0.2	0	0.3
2	0.2	0.1	0.1	0

Find the marginal distribution of X and Y also find the covariance of X and Y.

Soln

X \ Y	-2	-1	4	5	$f(x_i)$
1	0.1	0.2	0	0.3	0.6
2	0.2	0.1	0.1	0	0.4
$g(y_i)$	0.3	0.3	0.1	0.3	1

Marginal distribution of X and Y are

x_i	1	2
$f(x_i)$	0.6	0.4

y_i	-2	-1	4	5
$g(y_i)$	0.3	0.3	0.1	0.3

$E(X) = \mu(X) = x_1 f(x_1) + x_2 f(x_2)$

$= 1(0.6) + 2(0.4)$

$= 1.4$

$E(Y) = \mu Y = y_1 g(y_1) + y_2 g(y_2) + y_3 g(y_3) + y_4 g(y_4)$

$= -2(0.3) + (-1)(0.3) + 4(0.1) + 5(0.3)$

$= -0.6 - 0.3 + 0.4 + 1.5 = 1$

$\sigma_x^2 = x_1^2 f(x_1) + x_2^2 f(x_2) - \mu_x^2$

$= 1^2(0.6) + 2^2(0.4) - 1.96$

$= 0.6 + 1.6 - 1.96 = 0.24$

$\sigma_y^2 = y_1^2 g(y_1) + y_2^2 g(y_2) + y_3^2 g(y_3) + y_4^2 g(y_4) - \mu_{y^2}$

$= (-2)^2(0.3) + (-1)^2(0.3) + 4^2(0.1) + 5^2(0.3) - 1$

$= 1.2 + 0.3 + 1.6 + 7.5 - 1 = 9.6$

$E(XY) = \sum_{i=1}^{2} \sum_{j=1}^{4} x_i y_j P(x_i y_j)$

$= 1(-2)(0.1) + 1(-1)(0.2) + 1(4)(0) + 1(5)(0.3) + 2(-2)(0.2) + 2(-1)(0.1)$
$+ 2(4)(0.1) + 2(5)(0)$

$= -0.2 - 0.2 + 0 + 1.5 - 0.8 - 0.2 + 0.8 + 0 = 0.9$

$COV(X,Y) = E(XY) - E(X).E(Y) = 0.9 - 1.4$

$= -0.5$

6) Determine i) marginal distribution ii) Covariance between the discrete random variables X and Y along with Correlation using the joint probability distribution.

X \ Y	-3	2	4
1	0.1	0.2	0.2
2	0.3	0.1	0.1

Sol^n Given $x_1 = 1$, $x_2 = 2$

$y_1 = -3$, $y_2 = 2$, $y_3 = 4$

And the probabilities are

$P_{11} = 0.1$ $P_{12} = 0.2$ $P_{13} = 0.2$

$P_{21} = 0.3$ $P_{22} = 0.1$ $P_{23} = 0.1$

The joint probability distribution table is as follows

X \ Y	-3	2	4	$f(x_i)$
1	0.1	0.2	0.2	0.5
2	0.3	0.1	0.1	0.5
$g(y_j)$	0.4	0.3	0.3	1

The marginal distributions of X and Y are

x_i	1	2
$f(x_i)$	0.5	0.5

y_j	-3	2	4
$g(y_j)$	0.4	0.3	0.3

$\Rightarrow \mu_X = E(X) = \sum_i x_i f(x_i)$

$\mu_X = [1 \times 0.5] + 2[0.5]$

$= 1.5$

$\mu_Y = E(Y) = \sum_j y_j g(y_j)$

$= [-3 \times 0.4] + [2 \times 0.3] + [4 \times 0.3]$

$= -1.2 + 0.6 + 1.2$

$= 0.6$

$E(XY) = \sum_{ij} x_i y_j P_{ij}$

$\Rightarrow E(XY) = (1 \times (-3) \times 0.1) + (1 \times 2 \times 0.2) + (1 \times 4 \times 0.2) + (2 \times (-3) \times 0.3) +$
$(2 \times 2 \times 0.1) + (2 \times 4 \times 0.1)$

$\Rightarrow E(XY) = -0.3 + 0.4 + 0.8 - 1.8 + 0.4 + 0.8$
$= 0.3$

$\Rightarrow COV(X,Y) = E(XY) - \mu_X \mu_Y$

$\Rightarrow COV(X,Y) = 0.3 - 1.5 \times 0.6$

$\Rightarrow COV(X,Y) = 0.3 - 0.9$

$= -0.6$

The Correlation $\rho(X,Y) = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$.

$\sigma_x^2 = E(X^2) - \mu_X^2$

$\sigma_x^2 = \sum_i x_i^2 f(x_i) - \mu_X^2$

$= [1^2 \times 0.5] + [2^2 \times 0.5] - (1.5)^2$

$= 0.5 + 2 - 2.25$

$\sigma_x^2 = 0.25 \Rightarrow \sigma_X = 0.5$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$

$\sigma_Y^2 = \sum_j y_j^2 g(y_j) - \mu_Y^2$

$\sigma_Y^2 = [(-3)^2 \times 0.4] + [2^2 \times 0.3] + [4^2 \times 0.3] - (0.6)^2$

$= [3.6 + 1.2 + 4.8] - 0.36$

$\sigma_Y^2 = 9.24 \quad \Rightarrow \sigma_Y = 3.0397$

Correlation $\rho(X,Y) = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$

$= \dfrac{-0.6}{0.5 \times 3.0397}$

$= \dfrac{-0.6}{1.51985}$

$= -0.39477$

SAMPLING THEORY

Population:

A large collection of individuals or attributes or numerical data can be understood as population or universe.

Sample:

A finite subset of the population is called as sample.

Sample size:

The number of individuals in a sample is called a sample size. If the sample size n is less than or equal to 30, then the sample is said to be small, otherwise it is called a large sample.

Sampling:

The process of selecting a sample from the population is called as sampling.

Sampling with and without replacement:-

Sampling where a number of the population may be selected more than once is called as sampling with replacement. If a number cannot be chosen more than once is called as sampling without replacement.

Standard Error:

The standard Error is a statistical term that measures the accuracy with which a sample distribution represents a population by using standard deviation. In statistics, a sample mean deviates from the actual mean of a population; this deviation is the standard error of the mean. (or) The Standard deviation of a sampling distribution is called the Standard Error (SE).s

The reciprocal of the standard error is called precision.

Null hypothesis:

A null hypothesis is a type of hypothesis used in statistics that proposes that there is no difference between certain characteristics of a population.

Example: Given the test scores of two random samples, one of men and one of women, does one group differ from the other? A possible null hypothesis is that the mean male score is the same as the mean female score:

$H_0 : \mu_1 = \mu_2$

where:

H_0 = the null hypothesis

μ_1 : the mean of population 1, and

μ_2 = the mean of population 2.

A stronger null hypothesis is that the two samples are drawn from the same population, such that the variances and shapes of the distributions are also equal.

Confidence Intervals:

A Confidence interval is a type of estimate computed from the statistics of the observed data. This gives a range of values for an unknown parameter. The interval has an associated Confidence level that gives the probability with which an estimated interval will contain the true values of the parameter.

Type I error:

The first kind of error is the rejection of a true null hypothesis as the result of a test procedure. This kind of error is called a type I error and is sometimes called an error of the first kind. In terms of the courtroom example, a type I error corresponds to convicting an innocent defendant.

Type II error:

The second kind of error is the failure to reject a false null hypothesis as the result of a test procedure. This sort of error is called a type II error and is also referred to as an error of the second kind.

Level of significance:

The significance level is the probability of rejecting the null hypothesis when it is true.

Degree of freedom:

Degrees of freedom refers to the maximum number of logically independent values, which are values that have freedom to vary, in the data sample.

Test of hypothesis:

Let x be the observed number of successes in a sample size of n and $\mu = np$ be the expected number of successes. Then the standard normal variate z is defined as

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$$

Test of hypothesis for means:

Let μ_1, μ_2 be the means, σ_1, σ_2 be the standard deviations of two populations and $\bar{x}_1, \bar{x}_2$ are the means of the sample, then

$$z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ If the samples are drawn from the same population,

then $\sigma_1 = \sigma_2 = \sigma$ we have $z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Student's t-distribution:

Let μ be the mean of population, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the mean and $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ be the standard deviation of a sample, then the Student's t-distribution is defined as $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{\bar{x}-\mu}{s}\sqrt{n}$

Another formula for t-test of two samples is

$$t = \frac{(\bar{x}_2 - \bar{x}_1)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2}$

Chi-square distribution:

Let $O_i (i = 1,2,3,\dots n)$ and $E_i (i = 1,2,3,\dots n)$ be the set of observed frequencies and expected frequencies respectively, then the Chi-square distribution is defined as

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \dots + \frac{(O_n - E_n)^2}{E_n}$$

$$\Rightarrow \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Problems:

1) A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin in unbiased.

Soln Let us assume that the coin is unbiased

and let p = the probability of getting a head in one toss $= 1/2 = 0.5$

Since $p+q=1, q=1-p=1/2=0.5$

Expected number of heads in 1000 tosses $= np = 1000 \times 0.5 = 500$, $npq = 250$

$\therefore$ The difference is $x-\mu = 540-500 = 40$

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$$

$$\Rightarrow z = \frac{40}{\sqrt{250}} = 2.53 < 2.58$$

Thus we can say that the coin is unbiased.

2) In 324 throws of a six faced 'die', an odd number turned up 181 times. Is it possible to think that the 'die' is an unbiased one?

Soln Let us assume that the die is unbiased

and let p = the probability of the turn up of an odd number $= 3/6 = 1/2 = 0.5$

Since $p+q=1, q=1-p \Rightarrow 1/2 = 0.5$

Expected number of successes $= np = 324 \times 0.5 = 162$, $npq = 81$

$\therefore$ The difference is $x-\mu = 181-162 = 19$

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$$

$$\Rightarrow z = \frac{19}{\sqrt{81}} = \frac{19}{9} = 2.11 < 2.58$$

Thus we can say that the die is unbiased.

3) A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Show that the die cannot be regarded as an unbiased one.

Soln The probability of getting 3 or 4 in a single throw is

$p = 2/6 = 1/3$

And $q = 1-p = 1-1/3 = 2/3$

$\therefore$ Expected number of success $= \frac{1}{3} \times 9000 = 3000$

$\therefore$ The difference $= 3240 - 3000 = 240$

$$z = \frac{x - np}{\sqrt{npq}}$$

$$\Rightarrow z = \frac{(3240) - \left(9000 \times \frac{1}{3}\right)}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$$

$$z = \frac{240}{\sqrt{2000}}$$

$$\Rightarrow z = 5.37$$

Since $z = 5.37 > 2.58$, we conclude that the die is biased.

4) A manufacturer claimed that atleast 95% of the euipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 182 of them were faulty. Test his claim at a significance level of 1% and 5%.

Soln: Let p be the probability of success which being the probability of the equipment supplied to the factory conformal to the specifications.

$\therefore$ By the data $p = 0.95$ and hence $q = 0.05$

$H_0: p = 0.95$ and the claim is correct

$H_1: p < 0.95$ and the claim is false.

We choose the one tailed test to determine whether the supply is conformal to the specification.

$\therefore \mu = np = 200 \times 0.95 = 190$

$\sigma = \sqrt{npq} = \sqrt{200 \times 0.95 \times 0.05} = 3.082$

Expected number of equipments according to the specification $= 190$

Actual number $= 182$, since 182 out of 200 were faulty.

$\therefore$ Difference $= 190 - 182 = 8$

$$z = \frac{x - np}{\sqrt{npq}} = \frac{8}{3.082} = 2.6$$

The value of z is greater than the critical value 1.645 at 5% level and 2.33 at 1% level of significance.

The claim of the manufacturer is rejected at 5% as well as 1% level of significance in accordance with one tailed test.

5) A Certain stimulus administered to each of the 12 patients resulted in the following change in the blood pressures 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase blood pressure? (Note: $t_{0.05}$ for 11 d.f is 2.201)

Soln Given the change in blood pressure

$$x: 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4$$

$$\therefore \bar{x} = \frac{1}{n}\Sigma x = \frac{31}{12} = 2.5833$$

Variance

$$s^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2$$

$$\Rightarrow s^2 = \frac{1}{11}\left\{\begin{array}{l}(5-2.58)^2+(2-2.58)^2+(8-2.58)^2+(-1-2.58)^2+(0-2.58)^2+(6-2.58)^2 \\ +(-2-2.58)^2+(1-2.58)^2+(5-2.58)^2+(0-2.58)^2+(4-2.58)^2\end{array}\right\}$$

$$\Rightarrow s^2 = 9.538 \Rightarrow s = 3.088$$

Let us assume that the stimulus administration is not accompanied with increase in blood pressure, we can take $\mu = 0$

we have

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$\Rightarrow t = \frac{2.5833-0}{\left(\frac{3.088}{\sqrt{12}}\right)}$$

$$\Rightarrow t = 2.8979 \approx 2.9 > 2.201$$

Hence the hypothesis is rejected at 5% level of significance. We conclude with 95% confidence that the stimulus in general is accompanied with increase of blood pressure.

6) A random sample of 10 boys had the following IQ: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does this data support the assumption of a population mean I.Q of 100 at 5% level of significance? (Note: $t_{0.05} = 2.262$ for 9 d.f).

Soln Given the I.Q of 10 boys

$$x: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100$$

$$\therefore \bar{x} = \frac{1}{n}\Sigma x = \frac{972}{10} = 97.2$$

Variance

$$s^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2$$

$$\Rightarrow s^2 = \frac{1}{9} \times 1833.6$$

$$\Rightarrow s^2 = 203.73333$$

$$\Rightarrow s = 14.2735$$

Given the mean of population $\mu = 100$

we have

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$\Rightarrow t = \frac{97.2 - 100}{\left(\frac{14.2735}{\sqrt{10}}\right)}$$

$$\Rightarrow t = \frac{-2.8}{4.5136} \approx -0.6203 < 2.262$$

7) Ten individuals are chosen at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis that the mean height of the universe is 66 inches ($t_{0.05} = 2.262$ for 9 d.f)

Soln Given the heights of the population in inches

$$x : 63, 63, 66, 67, 68, 69, 70, 70, 71, 71.$$

$$\therefore \bar{x} = \frac{1}{n} \Sigma x = \frac{678}{10} = 67.8$$

Variance

$$s^2 = \frac{1}{n-1} \Sigma (x - \bar{x})^2$$

$$\Rightarrow s^2 = \frac{1}{9} \begin{Bmatrix} (63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + \\ (69-67.8)^2 + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2 \end{Bmatrix}$$

$$\Rightarrow s^2 = 9.067 \Rightarrow s = 3.011$$

And given the mean of population $\mu = 66$

we have.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t = \frac{67.8 - 66}{\left[\frac{3.011}{10}\right]}$$

$$\Rightarrow t = 1.8979 \approx 1.89 < 2.262$$

Thus the hypothesis is accepted at 5% level of significance.

8) Two types of batteries are tested for their length of life and the following results are obtained:

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs, $\sigma_1^2 = 100$

Battery B : $n_2 = 10$, $\bar{x}_2 = 560$ hrs, $\sigma_2^2 = 121$

Compute student's t and test whether there is a significant difference in the two means.

Soln Given

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs., $\sigma_1^2 = 100$

Battery B : $n_2 = 10$, $\bar{x}_2 = 560$ hrs., $\sigma_2^2 = 121$

W.K.T $\quad S^2 = \dfrac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2}$

$$\Rightarrow S^2 = \frac{(10 \times 100) + (10 \times 121)}{10 \times 10 - 2}$$

$$\Rightarrow S^2 = 122.78$$

$$\Rightarrow S = 11.0805$$

we have $\quad t = \dfrac{(\bar{x}_2 - \bar{x}_1)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$\Rightarrow t = \frac{560 - 500}{11.0805\sqrt{0.1 + 0.1}}$$

$$\Rightarrow t = 12.1081 \approx 12.11$$

The value of t is greater than the table value of t for 18 d.f at all levels of significance.

9) A group of boys and girls were given an intelligent test. The mean score, SD score and numbers in each group are as follows.

	Boys	Girls
Mean	74	70
SD	8	10
n	12	10

Is the difference between the means of the two groups significant at 5% level of significance ($t_{0.05} = 2.086$ for 20 d.f)

Soln

Given, $\bar{x}_1 = 74$, $\sigma_1 = 8$, $n_1 = 12$ (Boys)

$\bar{x}_2 = 70$, $\sigma_2 = 10$, $n_2 = 10$ (Girls)

W.K.T $S^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2-2}$

$$\Rightarrow S^2 = \frac{(12\times 64)+(10\times 100)}{12+10-2}$$

$$\Rightarrow S^2 = \frac{1768}{20} = 88.4$$

$$\Rightarrow S = 9.402 \approx 9.4$$

we have $t = \frac{|(\bar{x}_2 - \bar{x}_1)|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$\Rightarrow t = \frac{74-70}{9.4\sqrt{\frac{1}{12}+\frac{1}{10}}}$$

$$\Rightarrow t = 0.994 < t_{0.05} = 2.086$$

Thus the hypothesis that there is a difference between the means of the two groups is accepted at 5% level of significance.

10) Two horses A and B were tested according to the time (in seconds) to run a particular race with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	-

test whether you can discriminate between the two horses.

Soln Let the variables X and Y respectively correspond to Horse A and B

x : 28, 30, 32, 33, 33, 29, 34

y : 29, 30, 30, 24, 27, 29

$\therefore \bar{x} = \frac{1}{n_1}\sum_i x_i = \frac{219}{7} = 31.30$, $\therefore \bar{y} = \frac{1}{n_2}\sum_i y_i = \frac{169}{6} = 28.20$

$$\Sigma(x-\bar{x})^2 = (28-31.3)^2 + (30-31.3)^2 + (32-31.3)^2 + (33-31.3)^2 + (33-31.3)^2 + (29-31.3)^2 + (34-31.3)^2 = 31.4$$

$$\Sigma(y-\bar{y})^2 = (29-28.2)^2 + (30-28.2)^2 + (30-28.2)^2 + (24-28.2)^2 + (27-28.2)^2 + (29-28.20)^2 = 26.84$$

$$\therefore S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right]$$

$$S^2 = \frac{31.4+26.84}{7+6-2} = 5.2973$$

$$\Rightarrow S = 2.3016$$

we have $t = \dfrac{|(\bar{x}_2 - \bar{x}_1)|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$\Rightarrow t = \frac{31.30 - 28.20}{2.3016\sqrt{\frac{1}{7}+\frac{1}{6}}}$$

$$\Rightarrow t = 2.42 \begin{cases} > t_{0.05} = 2.2 \\ < t_{0.02} = 2.72 \end{cases}$$

11) Fit a binomial distribution for the data and test the goodness of fit given that $x_{0.05}^2 = 7.815$ for 3 d.f.

x	0	1	2	3	4
f	122	60	15	2	1

Soln Given

Observed frequencies are 122, 60, 15, 2, 1

The expected frequencies by using Binomial distribution are 121, 61, 15, 3, 0

O_i	122	60	15	2+1=3
E_i	121	61	15	3+0=3

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\Rightarrow \chi^2 = \frac{1}{121} + \frac{1}{61} + 0 + 0$$

$$\Rightarrow \chi^2 = 0.025 < \chi^2_{0.05} = 7.815$$

Hence the fitness is good.

12) Four coins are tossed 100 times and the following results were obtained

No of Heads	0	1	2	3	4
Frequency	5	29	36	25	5

Fit a binomial distribution for the data and test the goodness of fit $x_{0.05}^2 = 9.49$ for 4 d.f.

Soln Given the 4 coins are tossed 100 times

The probability of getting head is $p = 0.5, q = 0.5$

The probability mass function of a binomial distribution is

$P(X=x) = {}^4C_x (0.5)^x (0.5)^{4-x}$

$P(1) = {}^4C_1 (0.5)^1 (0.5)^{4-1} = 0.25$

$P(2) = {}^4C_2 (0.5)^2 (0.5)^{4-2} = 0.375$

$P(3) = {}^4C_3 (0.5)^3 (0.5)^{4-3} = 0.25$

$P(4) = {}^4C_4 (0.5)^4 (0.5)^{4-4} = 0.0625$

$\therefore E_0 = 100 \times 0.0625 = 6.25$

$E_1 = 100 \times 0.25 = 25$

$E_2 = 100 \times 0.375 = 37.5$ where 100 is the sum of frequency

$E_3 = 100 \times 0.25 = 25$

$E_4 = 100 \times 0.0625 = 6.25$

O_i	5	29	36	25	5
E_i	6.25	25	37.5	25	6.25

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\Rightarrow \chi^2 = \frac{1.5625}{6.25} + \frac{16}{25} + \frac{2.25}{37.5} + 0 + \frac{1.5625}{6.25}$$

$$\Rightarrow \chi^2 = 0.25 + 0.64 + 0.06 + 0.25$$

$$\Rightarrow \chi^2 = 1.2 < \chi^2_{0.05} = 9.49$$

Hence the fitness is good.

13) A dice thrown 264 times and the number appearing on the face (x) follows the following frequency (f) distribution.

x	1	2	3	4	5	6
f	40	32	28	58	54	60

Calculate the value of χ^2.

Soln The frequencies in the given data are the observed frequencies, assuming that dice is unbiased, the expected number of frequencies for the numbers 1, 2, 3, 4, 5, 6 to appear on the face is $\frac{264}{6} = 44$ each.

Now the data is as follows.

x	1	2	3	4	5	6
O_i	40	32	28	58	54	60
E_i	44	44	44	44	44	44

$$\therefore \chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\Rightarrow \chi^2 = \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(54-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$\Rightarrow \chi^2 = \frac{1}{44}\left[16+144+256+196+100+256\right] = \frac{968}{44}$$

$$\chi^2 = 22$$

14) Ten individuals are chosen at random from the population and their heights are found to be inches 63, 63, 64, 65, 66, 69, 69, 70, 70, 71. Discuss the suggestion that the mean height in the universe is 65 inches given that for 9 degree of freedom the value of student's 't' at 0.05 level of significance is 2.262.

Soln $x_i = 63, 63, 64, 65, 66, 69, 69, 70, 70, 71$ and $n = 10$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{670}{10} = 67$$

Variance

$$S^2 = \frac{1}{n-1}\sum (x-\bar{x})^2$$

$$\Rightarrow S^2 = \frac{1}{9}\left\{(63-67)^2+(63-67)^2+(64-67)^2+(65-67)^2+(66-67)^2+(69-67)^2+(69-67)^2+(70-67)^2+(70-67)^2+(71-67)^2\right\}$$

$$S^2 = \frac{88}{9} \Rightarrow S = \sqrt{\frac{88}{9}} = 3.13$$

The alternative hypothesis $H_\alpha: \mu \neq 65$ inches

The $df = n-1 = 10-1 = 9$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

$$= \frac{67-65}{\left[\frac{3.13}{\sqrt{10}}\right]} = 2.02$$

But $t_{0.05}$ at $df = 9 = 2.02$

$$\therefore t = 2.02 < 2.262$$

The difference is not significant at at 0.05 level thus H_0 is accepted and we conclude that the mean height is 65 inches.

15) The theory predicts the proportion of beans in the four groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beans, the number in the four groups were 882, 313, 287 and 118. The x^2 value of above data is approximately equal to?

Soln

Total number of beans:

$\Rightarrow = 882 + 313 + 287 + 118$

$= 1600$

Sum of ratios $= 9+3+3+1$

$E(A) = 1600 \times \frac{9}{16} = 900$

$E(B) = 1600 \times \frac{3}{16} = 300$

$E(C) = 1600 \times \frac{3}{16} = 300$

$E(D) = 1600 \times \frac{1}{16} = 100$

O_i	882	313	287	118
E_i	900	300	300	100
$(O_i - E_i)^2$	324	169	169	324

$$x^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{324}{900} + \frac{169}{300} + \frac{169}{300} + \frac{324}{100}$$

$$= 4.72$$

16) The nine item of a sample have the following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differ significantly from the assumed mean of 47.5?

Soln $\bar{x} = \frac{\Sigma x_i}{n} = \frac{442}{9} = 49.11$

Variance

$$S^2 = \frac{1}{n-1} \Sigma (x - \bar{x})^2$$

$$s^2 = \frac{1}{8}\left\{(45-49.11)^2+(47-49.11)^2+(50-49.11)^2+(52-49.11)^2+(48-49.11)^2 \right.$$
$$\left. +(47-49.11)^2+(49-49.11)^2+(53-49.11)^2+(51-49.11)^2\right\}$$

$$s^2 = \frac{54.9}{8} \Rightarrow s = \sqrt{\frac{54.9}{8}} = 2.6196$$

The null hypothesis $H_0: \mu = 47.5$. Alternative hypothesis $H_1: \mu \neq 47.5$

$$t = \frac{(\bar{x}-\mu)}{(s/\sqrt{n})} = \frac{(49.11-47.5)}{\left(\frac{2.6196}{\sqrt{9}}\right)} = 1.84$$

Level of significance: $\alpha = 0.05$

Critical value: the value of t_α at 5% level of significance for $v = 9-1 = 8$ degrees of freedom is 2.306

Since the calculated value of $|t| = 1.84$ is less than $t_\alpha = 2.306$.

∴ The null hypothesis is accepted.

Thanks for the patience...

By: Purushotham P @ SJCIT