1/16/15,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

①

Random Number

1) Use linear congruential method to generate a sequence of three two-digit-random integers. Let $X_0 = 27$, $a = 8$, $c = 47$ and $m = 100$.

$\Rightarrow$ $X_0 = 27$ $a = 8$, $c = 47$ and $m = 100$.

$X_1 = (8 \times 27 + 47) \bmod 100 = 63$ $R_1 = 63/100 = 0.63$

$X_2 = (8 \times 63 + 47) \bmod 100 = 51$, $R_2 = 51/100 = 0.51$

$X_3 = (8 \times 51 + 47) \bmod 100 = 55$, $R_3 = 55/100 = \underline{0.55}$

2) Use Multiplicative congruential method to generate a sequence of four three digit random integers. Let $X_0 = 117$, $a = 43$ and $m = 1000$.

$\Rightarrow$ $X_1 = [43(117)] \bmod 1000 = 31$

$X_2 = [43(31)] \bmod 1000 = 333$.

$X_3 = [43(333)] \bmod 1000 = 319$.

$X_4 = [43(319)] \bmod 1000 = 717$.

3) Determine whether the linear congruential generators shown below can achive a maximum period. Also state restrictions on X_0 to obtain this period.

a) The Mixed congruential method

$a = 2814,749,767,109$

$c = 59,482,661,568,307$, $m = 2^{48}$.

b) The multiplicative congruential generator with $a = 69{,}069$, $c = 0$, $m = 2^{32}$.

c) The mixed congruential generator with $a = 4951$, $c = 247$, $m = 256$.

d) The multiplicative congruential generator with $a = 6507$, $c = 0$ and $m = 1024$.

Ans (c) $a = 1 + 4k$ $\therefore k = 1237.5$, which is not an integer $\therefore$ maximum period cannot be achived.

4) Use mixed congruential method to generate a sequence of three two digit random numbers with $X_0 = 37$, $a = 7$, $c = 29$ and $m = 100$.

$\Rightarrow$ $X_1 = [7 \times 37 + 29] \bmod 100 = \underline{88}$

$R_1 = 0.88$

$\rightarrow$ $X_2 = [7 \times 88 + 29] \bmod 100 = \underline{45}$

$\therefore R_2 = 0.45$

$\rightarrow$ $X_3 = [7 \times 45 + 29] \bmod 100 = \underline{44}$

$R_3 = \underline{0.44}$

5) Use mixed congruential method to generate a sequence of three two digit random integers betwn 0 and 24 with $X_0 = 13$, $a = 9$ and $c = 35$.

$\Rightarrow$ Use $m = 25$

$X_1 = [9 \times 13 + 35] \bmod 25 = 2$.

$X_2 = [9 \times 2 + 35] \bmod 25 = 3$.

$X_3 = [9 \times 3 + 35] \bmod 25 = 12$.

5) Consider the multiplicative congruential generator under the following circumstances.

a) $a = 11$, $m = 16$, $X_0 = 7$.

b) $a = 11$, $m = 16$, $X_0 = 8$.

c) $a = 7$, $m = 16$, $X_0 = 7$.

d) $a = 7$, $m = 16$, $X_0 = 8$.

Generate enough values in each case to complete a cycle. What inferences can be drawn? Is maximum period achived?

i	case(a) X_i	case(b) X_i	case(c) X_i	case(d) X_i
0	7	8	7	8
1	13	8	1	8
2	15		7	8.
3	5			
4	7			

Inference: Maximum Period $P=4$, occurs when X_0 is odd and $a = 3+8k$ where $k=1$. Even seeds have the minimal possible period regardless of a.

6). For 16-bit computers, recommends combining three multiplicative generators with $m_1 = 32363$, $a_1 = 157$, $m_2 = 31727$, $a_2 = 146$, $m_3 = 31657$ and $a_3 = 142$. The period of this generator is approximately 8×10^{12}. Generate 5 random numbers with combined generator using initial seeds, $X_{i,0} = 100, 300, 500$ for individual generators $i = 1, 2, 3$.

$\Rightarrow$ $X_{1,0} = 100$, $X_{2,0} = 300$, $X_{3,0} = 500$.

The generator is.

$X_{1,j+1} = 157 X_{1,j} \mod 32363$.

$X_{2,j+1} = 146 X_{2,j} \mod 31727$.

$X_{3,j+1} = 142 X_{3,j} \mod 31657$.

$X_{j+1} = (X_{1,j+1} - X_{2,j+1} + X_{3,j+1}) \mod 32362$.

$$R_{j+1} = \begin{cases} \dfrac{X_{j+1}}{32363}, & \text{if } X_{j+1} > 0 \\ \dfrac{32362}{32363} = 0.999, & \text{if } X_{j+1} = 0. \end{cases}$$

Scanned with OKEN Scanner

1/16/15,000

(2)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Random numbers

Candidate Roll No. ______________
(In figures)

Test Exam.

Name: ______________

Date: ______________ 20

Examination: ______________ Branch/Semester ______________

Subject: ______________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

The first 5 random numbers are

$X_{1,1} = [157 \times 100] \bmod 32363 = 15700.$

$X_{2,1} = [146 \times 300] \bmod 31727 = 12073.$

$X_{3,1} = [142 \times 500] \bmod 31657 = 7686.$

$X_1 = [15700 - 12073 + 7686] \bmod 32362 = 11313$

$R_1 = 11313/32363 = 0.3496$

$X_{1,2} = 5312$

$X_{2,2} = 17673.$

$X_{3,2} = 15074$

$X_2 = 2713.$

$R_2 = 0.0838$

$X_{1,3} = 24909$

$X_{2,3} = 10371$

$X_{3,3} = 19489.$

$X_3 = 1665.$

$R_3 = 0.0515.$

$X_{1,4} = 27153.$

$X_{2,4} = 22997.$

$X_{3,4} = 13279.$

$X_4 = 17435.$

$R_4 = 0.5387.$

$X_{1,5} = 23468.$
$X_{2,5} = 26227.$
$X_{3,5} = 17855.$
$X_5 = 15096.$
$R_5 = 0.4665.$

Hypothesis Testing:-

→ Many real world problems require making decision about populations on the basis of limited data.

→ In order to reach at a decision we generally make assumption about the behavior of underlying population.

→ Such assumption, which may or may not be true, is called Statistical hypothesis.

→ It is also a statement about one or more probability distribution associated with the population.

→ Methods that help us to decide whether to accept or reject hypothesis on the basis of inforn available are called as Statistical tests.

→ Here our objective is to test uniformity and independence properties of random numbers.

→ Following hypothesis are defined for testing uniformity

H_0: $R_i \sim$ independently.
H_1: $R_i \nsim$ independently.

→ The null hypothesis, H_0, states that the numbers are independent.

→ If test fails to reject the null hypothesis means that no evidence of departure from independence has been detected.

→ While hypothesis testing two types of errors may occur :-

1) Type I error (α).
→ Reject Ho when in fact it is true.
→ The Probability of type I error is denoted by α.
→ For each test a level of significance α must be state
→ The level of significance α is the probability of reject the null hypothesis given that null hypothesis is true.
→ $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$.
→ Usually, α is set to 0.01 or 0.05

2) Type II error (β)
→ Accept Ho when it is false (we reject ~~to reject~~ Ho
→ The Probability of type II error is denoted by β and $(1-\beta)$ is known as power of the test.

→ If several tests are conducted on the same set of numbers the probability of rejecting null hypothesis on at least one test increases.
→ If one test is conducted on many sets of numbers from a generator, the probability of rejecting the null hypothesis on at least one test increases as more ~~test~~ of sets of numbers are tested.
→ An error of type I or II leads to a wrong decision, hence we must minimize both errors.
→ Consider a sample size of N, a decrease in one type of error may lead to increase

in other type of error. Also a cost factor is always associated with any type of decision and wrong decision always lead to increase in cost.
→ We can reduce the cost by decreasing both types of errors simultaneously which inter can be done by increasing the sample size N.

Test for uniformity

Frequency Test:-

→ There are two methods available to test the uniformity.

→ Those are Kolmogorov – Smirnov test and chi – square test.

→ Both measure the degree of agreement betwn the distribution of sample of generated random numbers and theoretical uniform distribution

The Kolmogorov Smirnov Test:-

→ In statistics the K-S test is used to determine whether two underlying probability distributions differ or whether an underlying probability distribution differs from a hypothesized distribution in either case based on finite samples.

→ The K-S- test compares the continuous CDF $F(x)$ of uniform distribution specified by null hypothesis with empirical CDF, $S_N(x)$ of sample of N observations.

→ We know that.

The continuous CDF of uniform distribution is –

$F(x) = x, \quad 0 \leq x \leq 1.$

The empirical CDF of sample random numbers $R_1, R_2 --- R_N$ is.

$$S_N(x) = \frac{\text{no of } R_1, R_2 --- R_N \text{ which are} \leq x}{N}$$

8/14/30,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

(3).

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

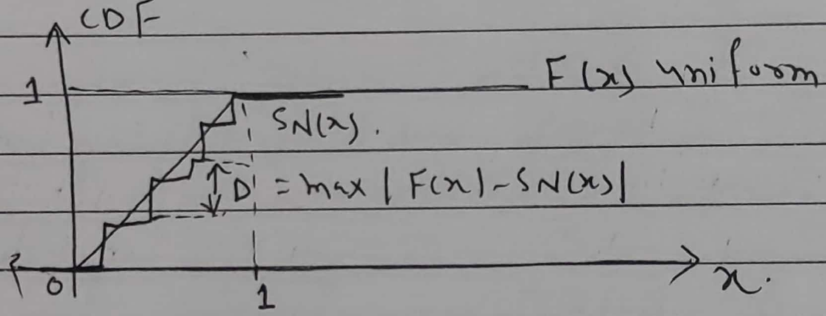
Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

→ The test is based on largest absolute deviation between $F(x)$ and $S_N(x)$ over the range of random variable.

→ Plot $F(x)$ and $S_N(x)$ and get the statistic $D = \max|F(x) - S_N(x)|$ which is maximum vertical distance as shown in fig below.



→ The computed value of D is compared with tabulated value of sampling distribution (critical value) D_α for specified value of significance level 'α and sample size N.

→ If compared value is greater than critical value then reject the null hypothesis. The value of D can be computed by using following algorithm also.

Algorithm

1) Define the hypothesis for testing the uniformity as.

$H_0 : R_i \sim U[0,1]$.

$H_1 : R_i \nsim U[0,1]$.

2) Arrange the data in increasing order.

Scanned with OKEN Scanner

Let $R_{(i)}$ denote the ith smallest no then
$R_{(1)} \leq R_{(2)} \leq \cdots \leq R_{(N)}$.

3) Compute D^+ and D^- where

$$D^+ = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)} \right\} \text{ and}$$

$$D^- = \max_{1 \leq i \leq N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

4) ~~Compute D^+ and D^- where~~ Compute $D = \max(D^+, D^-)$

5) Determine the critical value, D_α for specified significance level α and the given sample size N from table A.8.

6) If $D > D_\alpha \Rightarrow H_0$ is rejected, else no difference has been detected betwn the sample distribution and uniform distribution.

→ The Kolmogorov - Smirnov test is more powerful and can be used for small sample sizes.

Example (June - 2011, May 2012)

The sequence of numbers 0.63, 0.49, 0.24, 0.89, 0.57 and 0.71 has been generated. Use the Kolmogorov Smirnov test with $\alpha = 0.05$ to determine if hypothesis that nos are uniformly distributed on the interval $[0,1]$ can be rejected.

⇒ Solution Method - 1

1) Define the hypothesis for testing the uniformity as

H_0: $R_i \sim U[0,1]$.

H_1: $R_i \nsim U[0,1]$.

2) Rank data in increasing order.

0.24 ≤ 0.49 ≤ 0.57 ≤ 0.63 ≤ 0.71 ≤ 0.89.

3) Compute D^+ and D^-.

i	1	2	3	4	5	6
$R_{(i)}$	0.24	0.49	0.57	0.63	0.71	0.89
i/N	0.17	0.33	0.50	0.67	0.83	1.00
$i/N - R_{(i)}$	–	–	–	0.04	0.12	0.11
$R_{(i)} - \frac{(i-1)}{N}$	0.24	0.32	0.24	0.13	0.04	0.06

$D^+ = \max(0.04, 0.12, 0.11)$

$= \underline{0.12}$

and $D^- = \max\{0.24, 0.32, 0.24, 0.13, 0.04, 0.06\}$

$= \underline{0.32}$

4) Compute D

$D = \max(D^+, D^-)$.

$= \max(0.12, 0.32) = \underline{0.32}$

5) Determine the critical value D_α for specified level of significance $\alpha = 0.05$ and sample size $N = 6$.

$D_{0.05} = \underline{0.521}$ (obtained from table A.8)

6) Since $D = 0.32 < D_{0.05} = 0.521$

$\therefore$ H_0 is not rejected

From this we can say that given set of random numbers are uniformly distributed.

Method - 2

For getting the value of D, plot the CDF of empirical distribution ($SN(x)$) and uniform distribution ($F(x)$). The largest vertical distance between $F(x)$ and $SN(x)$ is the D.

1) Define hypothesis for testing the uniformity as,

$H_0 : R_i \sim U[0,1]$.

$H_1 : R_i \nsim U[0,1]$

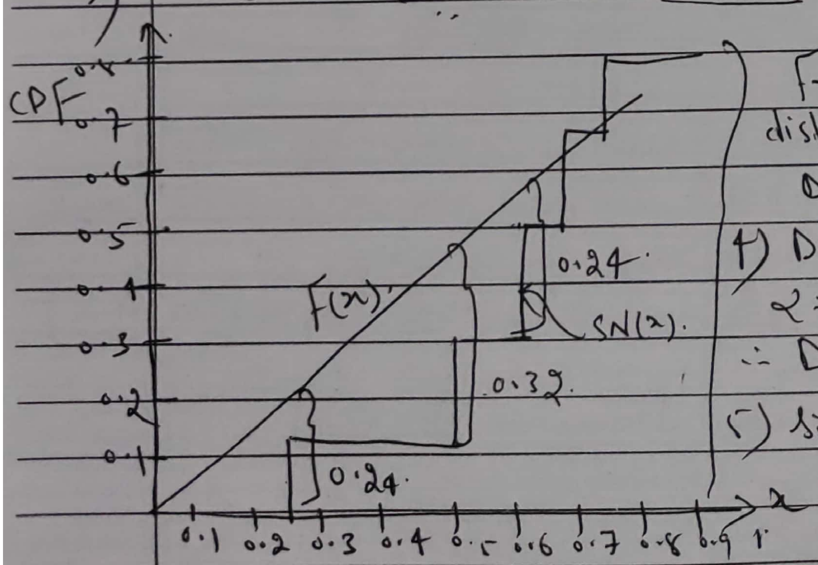
2) Compute the uniform CDF, ($F(x)$) and empirical CDF, ($SN(x)$).

X	0.63	0.49	0.24	0.89	0.57	0.71
$F(x) = x$	0.63	0.49	0.24	0.89	0.57	0.71

$SN(x) = \frac{\text{no of } R_1, R_2 \cdots R_N \text{ which are} \leq x}{N}$

→ 0.67 0.33 0.17 1.00 0.50 0.83

3) Plot the CDF of empirical and uniform distribution



From Fig largest vertical distance is 0.32

D = 0.32

4) Determine D_α from table.

$\alpha = 0.05$, $N = 6$.

$\therefore D_{0.05} = 0.521$

5) Since $D = 0.32 < D_{0.05} = 0.521$ ∴ H_0 is not rejected

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

4

Candidate Roll No. ______________ (In figures)

Test Exam.

Name : ______________

Date : ______________ 20

Examination : ______________ Branch/Semester ______________

Subject : ______________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

* The chi-Square Test:-

→ Pearson was first to recognize the Practical value of chi-square test.

→ The chi-square test can be used to test the uniformity (or homogeneity) within a data set or to determine the probability that there is a dependency relationship betwn. two or more distinct data sets. (e.g in industry, the chi square test could be used to determine the probability that a particular machine in a group breaks down too often or that Production increases when a raw material is changed).

→ This is statistical ~~Produ~~ Procedures whose results are evaluated by ref to chi-square distribution. It tests a null hypothesis that the relative frequencies of occurence of observed events follow a specified frequency distribution.

→ The Problem is to decide whether N independent Samples can be belonging to a common Population. The algorithm for chi-square test is given below

1) Define the hypothesis for testing the uniformity as,

$H_0 : R_i \sim U[0,1]$.

$H_1 : R_i \nsim U[0,1]$

2) Divide the total no of observations (N) into mutually exclusive equally numbered classes. (n) $[(a_0, a_1)\ (a_1, a_2) \dots (a_{n-1}, a_n)]$. The n be chosen such a way that each $E_i \geq 5$.

3) Compute the sample test statistics.

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where

$\chi_0^2 \rightarrow$ approximate chi square distribution with n-1 degrees of freedom.

$O_i \rightarrow$ observed number in ith class.

$E_i \rightarrow$ expected number in ith class which is N/n

4) Determine the critical value for specified significance level α with n-1 degrees of freedom from table A.6

5) If $\chi_0^2 > \chi_{\alpha, n-1}^2 \Rightarrow$ Ho is rejected;

else no difference has been detected between the sample distribution and uniform distribution

→ This test is valid only for large samples, say, $N \geq 50$.

Example Consider the following sequence of 100 nos

0.43	0.09	0.52	0.98	0.78	0.44	0.21	0.12	0.64	0.76
0.38	0.67	0.97	0.46	0.07	0.18	0.49	0.47	0.22	0.47
0.69	0.91	0.77	0.76	0.65	0.14	0.25	0.37	0.99	0.20
0.74	0.03	0.71	0.28	0.65	0.10	0.54	0.13	0.87	0.10
0.97	0.17	0.32	0.91	0.28	0.39	0.56	0.73	0.93	0.24
0.99	0.71	0.99	0.64	0.50	0.66	0.01	0.24	0.81	0.74
0.73	0.15	0.45	0.10	0.18	0.82	0.96	0.43	0.57	0.94
0.27	0.34	0.65	0.79	0.03	0.49	0.69	0.85	0.37	0.50
0.60	0.93	0.48	0.42	0.04	0.46	0.04	0.91	0.97	0.26
0.81	0.62	0.79	0.88	0.46	0.74	0.06	[illegible]	0.12	0.17

Use the chi-square test with $\alpha = 0.05$, to test the hypothesis that the nos are uniformly distributed on the interval $[0,1]$ can be rejected.

Solution 1) Define the hypothesis for testing the uniformity as

$H_0 :$ $R_i \sim U[0,1]$

$H_1 =$ $R_i \nsim U[0,1]$.

2) Choose the value of n such that $E_i \geq 5$.

Since $N = 100$ and $E_i = N/n$.

$\therefore 100/n \geq 5 \Rightarrow 100/5 \geq n \Rightarrow n \leq 20$

Let $n = 10$ intervals of equal length, namely $[(0, 0.1), (0.2, 0.3), \ldots (0.9, 1.0)]$.

3) Compute the test statistics

Interval	O_i	$E_i = N/n$	$\frac{(O_i - E_i)^2}{E_i}$
(0, 0.1) - 1	8	10	0.4
(0.1, 0.2) - 2	9	10	0.1
(0.2, 0.3) - 3	10	10	0.6
(0.3, 0.4) - 4	6	10	1.6
(0.4, 0.5) - 5	13	10	0.9
(0.5, 0.6) - 6	8	10	0.4
(0.6, 0.7) - 7	11	10	0.1
(0.7, 0.8) - 8	12	10	0.4
(0.8, 0.9) - 9	7	10	0.9
(0.9, 1.0) - 10	16	10	3.6

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$= (0.4 + 0.1 + 0.0 + 1.6 + 0.9 + 0.4 + 0.1 + 0.4 + 0.9 + 3.6)$

$= 8.4$

→ Determine critical value $\chi^2_{\alpha, n-1}$ for specified significance level α with $n-1$ degree of freedom from table A.6

Since $\alpha = 0.05$, degree of freedom $= 10-1 = 09$

$\chi^2_{0.05, 9} = 16.9$

5) Since $\chi_0^2 = 8.4 < \chi^2_{0.05, 9} = 16.9$

$\therefore$ H_0 is not rejected

8/14/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

5

Candidate Roll No. ____________ (In figures)

Test Exam.

Name: ____________

Date: ____________ 20

Examination: ____________ Branch/Semester ____________

Subject: ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

Test for Independence

1) Runs Test:-

→ The runs test analyzes an orderly grouping of nos in a sequence to test the hypothesis of independence.

→ A run is defined as succession of similar events preceded and followed by different event.

→ The length of the run is the number of events that occur in the run.

→ e.g consider the following sequence generating by tossing of coin 10 times

$$\underbrace{T\ T}_{2}\ \underbrace{H\ H\ H}_{3}\ \underbrace{T}_{1}\ \underbrace{H\ H}_{2}\ \underbrace{T\ T}_{2}$$

→ In above example there are five runs

→ The length of each run is 2, 3, 1, 2 and 2.

→ There are two concerns in a runs test. namely no of runs and length of runs.

→ Three types of runs test are used namely runs up and runs down, runs above and runs below the mean and length of runs.

→ In all three test the actual values are compared with expected values using the chi square test.

① * Runs up and runs down *

→ An up run is a sequence of nos each of which is followed by a larger number.

→ Similarly a down run is a sequence of nos each of which is followed by smaller no.

→ E.g consider following sequence of 10 nos
0.57 0.64 0.34 0.45 0.22 0.01 0.23 0.78 0.88 0.19.

→ The nos are given a "+" or a "−" depending on whether they are followed by a larger no or a smaller number.

→ Since there are 10 nos and all of them are different so there will be 9 +'s and −'s.

→ The last no is followed by "no event" and hence it will not get neither + nor a −.

→ The sequence of 9 +'s and −'s is given.
\+ − + − − + + + −

→ Here each succession of +'s and −'s forms a run.

→ In above example there are six runs and length of each run is 1, 1, 1, 2, 3 and 1.

→ There can be a situation of few runs (only one run) or too many runs (N−1, if N is total no of observations).

→ Both are unlikely case for a valid random generator.

→ The more likely case is that the no of runs will be somewhere betw^n these two extremes.

→ Let 'a' be the total no of runs found in the sequence.

→ For N > 20 'a' is approximated by a normal distribution $N(\mu_a, \sigma_a^2)$ which can be used to test the independence of nos. from a generator.

→ Finally a standardize is used to test the independence on the basis of runs up and runs down in given sequence of nos.

Algorithm

1) Define hypothesis for testing.

   H_0 : $R_i \sim$ independently.

   H_1 : $R_i \nsim$ independently.

2) write down sequence of runs up and down.

3) Count the total no of runs (a), present in sequence.

4) compute mean and variance of a.

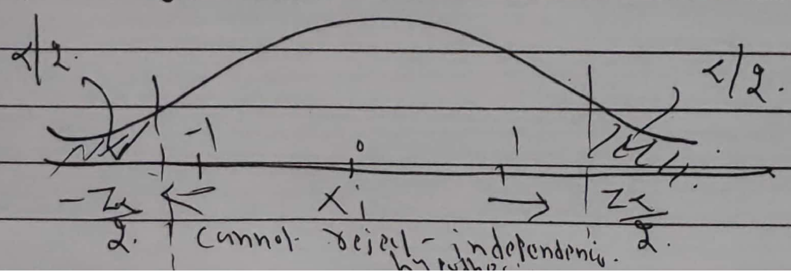
$$\mu_a = \frac{2N-1}{3} \quad \text{and} \quad \sigma_a^2 = \frac{16N-29}{90}$$

5) Compute std normal statistics.

$$Z_0 = \frac{a-\mu_a}{\sigma_a} \quad \text{where } Z_0 \sim N(0,1)$$

6) Determine the critical value $Z_{\frac{\alpha}{2}}$ and $-Z_{\frac{\alpha}{2}}$ for specified significance level α from table A.3.

7) If $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$, H_0 is not rejected for significance level α.



7) Since $-Z_{0.025} = -1.96 \leq Z_0 = -0.51 \leq Z_{0.025} = 1.96$.

$\therefore$ H_0 is not rejected.

Example:-

Consider the following sequence of 40 numbers.

0.52 0.99 0.46 0.58 0.64 0.25 0.48 0.11 0.20 0.18

0.97 0.44 0.43 0.94 0.82 0.60 0.73 0.69 0.21 0.05

0.04 0.81 0.85 0.30 0.47 0.96 0.17 0.72 0.62 0.27

0.10 0.60 0.34 0.65 0.79 0.44 0.02 0.37 0.18 0.80

Based on runs up and runs down determine whether the hypothesis of independence can be rejected, where $\alpha = 0.05$:

$\Rightarrow$ Solution 1) Define hypothesis.

H_0: $R_i \sim$ independently.

H_1: $R_i \not\sim$ independently.

2) The sequence of runs up and down is.

+ − + + − + − + − +

− − + − − + − − − +

+ + − + + − + − − −

+ − + + − − + + +

3) The total number of runs $a = 25$

4) Mean and Variance of 'a' is.

$\mu_a = \frac{2N-1}{3} = 2(40)-1/3 = 26.33$

$\sigma_a^2 = 16N-29/90 = 16(40)-29/90 = 6.79$

5) The std normal statistics :-

$Z_0 = \frac{a - \mu_a}{\sigma_a} = 25 - 26.33/\sqrt{6.79}$

$= -0.51$

6) Determine critical value $Z_{\frac{\alpha}{2}}$ and $-Z_{\frac{\alpha}{2}}$ for specified significance level α from table A.3

Since $\alpha = 0.05$ $\therefore \alpha/2 = 0.025$

Now $Z_{0.025}$ is obtained from Table in following manner

$\phi(Z_{0.025}) = 1 - 0.025 = 0.975$

$Z_{0.025} = \phi^{-1}(0.975) = 1.96$

Scanned with OKEN Scanner

8/14/30,000

(6)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name: ____________

Date: ____________ 20

Examination: ____________ Branch/Semester ____________

Subject: ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

(2) Runs above and below the mean

→ The test for runs up and runs down is not completely adequate to evaluate the independence of sequence of numbers.

→ So runs are redefined as being above the mean or below the mean.

→ A "+" sign will be used to denote a no above the mean and "−" sign will denote a no below mean

→ consider following sequence of 10 nos

0.57 0.64 0.34 0.45 0.22 0.01 0.23 0.78 0.88 0.19

→ The nos are given a "+" or a "−" depending on whether they are greater than or smaller than the expected mean which is $(0+0.99)/2 = 0.495$

→ The sequence of 10 +'s and −'s is given below.

+ + − − − − − + + −

→ In above example there are four runs.

→ The length of each run is 2, 5, 2 and 1.

→ Let n_1 and n_2 be the no of observations above and below mean respectively.

→ The maximum no of runs is $N = n_1 + n_2$ whereas minimum is one.

→ Let 'b' is total no of runs in a sequence.

→ For either $n_1 > 20$ or $n_2 > 20$, b is approximated by a normal distribn $N(\mu_b, \sigma_b^2)$ which can be used to test the independence of nos from a generator.

→ Finally Std. normal test Statistics Z_0 is developed and it is compared with the critical value.

→ The following algorithm is used to test the independence on basis of runs above and below mean in the given sequence of nos.

Algorithm :-

1) Define hypothesis for testing the independence.
   H_0 : $R_i \sim$ independently
   H_1 : $R_i \not\sim$ independently

2) Write down sequence of runs above and runs below the mean.

3) Count the no of observations above mean (n_1), no of observations below mean (n_2) and total no. of runs (b) present in the sequence.

4) compute mean and variance of b.
   $$\mu_b = \frac{2n_1n_2}{N} + \frac{1}{2} \quad \text{and} \quad \sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}$$

5) Compute std normal statistics.
   $$Z_0 = \frac{b-\mu_b}{\sigma_b} \quad \text{where } Z_0 \sim N(0,1)$$

6) Determine the critical value $Z_{\alpha/2}$ and $-Z_{\alpha/2}$ for specified significance level α from table.

7) If $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$ ∴ H_0 is not rejected for the significance level α

Example consider following sequence of 40 nos

0.09 0.41 0.23 0.68 0.89 0.72 0.12 0.45 0.08 0.39
0.53 0.13 0.65 0.97 0.14 0.49 0.55 0.46 0.77 0.28
0.81 0.63 0.40 0.57 0.02 0.16 0.33 0.86 0.99 0.22
0.76 0.48 0.61 0.39 0.43 0.78 0.20 0.35 0.17 0.93.

Determine whether there is an excessive no of runs above or below the mean. Use $\alpha = 0.05$

⟹ 1) Define hypothesis for testing independence as.

H_0 : R_i ~ independently.

H_1 : R_i ≁ independently.

2) The sequence of runs above and below the mean (0.495) is.

$- - - + + + - - - -$
$+ - + + - - + - + -$
$+ + - + - - - + + -$
$+ - + - - + - - - +.$

3) The no of observⁿ above mean $n_1 = 17$ and $n_2 = 23$ → below mean.

Total no of runs. $b = 24$

4) Mean and variance of b

$$\mu_b = \frac{2n_1n_2}{N} + \frac{1}{2} = \frac{2(17)(23)}{40} + \frac{1}{2} = 20.05$$

$$\sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)} = \frac{2(17)(23)(2(17)(23)-40)}{40^2(40-1)}.$$

$$= 9.3$$

5) The std normal statistics

$$Z_0 = \frac{b - \mu_b}{\sigma_b} = 24 - 20.05 / \sqrt{9.3} = 1.295$$

6) Since $\alpha = 0.05$ ∴ $\alpha/2 = 0.025$

∴ $Z_{0.025} = 1.96$

7) Since $-Z_{0.025} = -1.96 \leq Z_0 = 1.295 \leq Z_{0.025} = 1.96$. ∴ H_0 is not rejected.

(3) Length of Runs:-

→ Length of runs is another concern of runs test and it is expected that length of runs should not be a constant.

→ Let y_i be the number of runs of length 'i' in a sequence of N numbers.

→ The expected value of y_i for runs up and run down or runs above and runs below the mean is determined.

→ Then the chi square test is applied to compare expected value with the observed one.

Algorithm:-

1) Define hypothesis for testing the independence.
2) Write down the sequence of runs up and down or above and below mean.
3) Find the length of runs in the sequence.
4) Prepare the table for no of observed runs of each length.

Run length (i)	1	2		- - -
observed runs (O_i)	-	-	-	. - .

5) Compute the expected value of y_i.

i) For runs up and runs down

$$E(y_i) = \begin{cases} \dfrac{2}{(i+3)!}\left[N(i^2+3i+1)-(i^3+3i^2-i-4)\right], & i \le N-2 \\ \dfrac{2}{N!}, & i = N-1 \end{cases}$$

ii) For runs above and runs below mean:-

$$E(y_i) = \frac{N w_i}{E(I)}, \quad N > 20.$$

Where

w_i is approximate probability that a run has length i and it is

$$w_i = \left(\frac{n_1}{N}\right)^i\left(\frac{n_2}{N}\right) + \left(\frac{n_1}{N}\right)\left(\frac{n_2}{N}\right)^i, \quad N > 20.$$

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

7

Candidate Roll No. ______________ (In figures)

Test Exam.

Name : ______________

Date : ______________ 20

Examination : ______________ Branch/Semester ______________

Subject : ______________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

E(I) is the approximate expected length of a run and it is

$$E(I) = \frac{n1}{n_2} + \frac{n2}{n_1}, \quad N > 20.$$

6) Compute the mean or expected total no of runs (of all lengths) in a sequence

i, For runs up and runs down

$$\mu a = \frac{2N-1}{3}.$$

ii) For runs above and runs below mean:-

$$E(A) = \frac{N}{E(I)}, \quad N > 20.$$

7) compute the expected no of runs of length greater than or equal to the maximum length of observed run.

i, For runs up and runs down

$\mu a - \sum_{i=1}^{m} E(y_i)$, where m is equal to the max length of observed run.

ii) For runs above and runs below mean:-

$E(A) - \sum_{i=1}^{m} E(y_i)$, where m is equal to the maximum length of observed run.

8) Apply chi square test-

Run length, i	observed No of Runs, O_i	Expected No of runs, $E(y_i)$	$\frac{[O_i - E(y_i)]^2}{E(y_i)}$

The test statistics is.

$$\chi_0^2 = \sum_{i=1}^{n} \frac{[O_i - E(y_i)]^2}{E_i}$$, where n is the no of classes.

9) Determine the critical value for the specified significance level α with $n-1$ degrees of freedom.

10) If $\chi_0^2 < \chi^2_{\alpha, n-1} \Rightarrow$ Ho is not rejected.

Example 1-

Consider the following sequence of 40 nos

0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

Can the hypothesis that the nos are independent be rejected on the basis of length of runs up and down, where $\alpha = 0.05$?

$\Rightarrow$ 1) Define hypothesis for testing independence.

2) The sequence of runs up and down is.

+ + − + − + − + − +
− + + − + + − + − +
+ + − + + − − + − −
+ + − − − + − + +

3) The length of run in the sequence is.

2 1 1 1 1 1 1 1 1 1
1 2 1 2 1 1 1 3
1 2 2 1 2
2 3 1 1 2.

4) The number of observed runs of each length is.

Run length, i	1	2	3
observed Runs o_i	17	8	2

5) The expected no of runs of length one, two and three are,

$E(Y_1) = \frac{2}{(1+3)!}\left[40(1+3+1) - (1+3-1-4)\right] = \underline{16.75}$

$E(Y_2) = \frac{2}{(2+3)!}\left[40(4+6+1) - (8+12-2-4)\right] = \underline{7.1}$

$E(Y_3) = \frac{2}{(3+3)!}\left[40(9+9+1) - (27+27-3-4)\right] = \underline{1.98}$

6) The mean or expected total no of runs of all lengths in a sequence is.

$\mu_a = 2N-1/3 = 2(40)-1/3 = \underline{26.33}$

7) The expected no of runs of length greater than or equal to 4 is.

$\mu_a - \sum_{i=1}^{3} E(Y_i) = 26.33 - (16.75 + 7.1 + 1.98)$

$= \underline{0.5}$

8) Apply chi square test:-

Run length i	Observed No of Runs, o_i	Expected No of Runs $E(Y_i)$	$\frac{[o_i - E(Y_i)]^2}{E(Y_i)}$
1	17	16.75	3.73×10^{-3}
2	8	7.1	
3	2	1.98	
≥ 4	0	0.5	
2, 3, ≥ 4 (combined)	10	9.58	0.0184

→ As, it is suggested, the minimum value of expected frequency is 5 in case of chi square test. If it is less than 5 it can be combined with the expected frequency of an adjacent class interval.

Scanned with OKEN Scanner

The corresponding observed frequencies would also be combined accordingly and the value of n (no of classes) would be reduced.

→ Here class 3 and 4 has expected freqn less than 5 so it is combined with class 2. Similarly combining the observed freqn of class 3 and 4 with class 2. Reduce the no. of classes by 2 which leads $n = 4-2 = 2$

The test statistics is:-

$$\chi_0^2 = \sum_{i=1}^{n} \frac{[o_i - E(y_i)]^2}{E(y_i)}$$

$$= (3.73 \times 10^{-3} + 0.0184)$$

$$= 0.02213$$

9) The critical value for specified significance level $\alpha = 0.05$ with $n-1 = (2-1) = 1$ degree of freedom is.

$$\chi^2_{0.05,1} = 3.84$$

Since $\chi_0^2 = 0.02213 < \chi^2_{0.05,1} = 3.84$

∴ H_0 is not rejected.

8/14/30.000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

(8)

Candidate Roll No. ____________ (In figures)

Test Exam.

Name: ____________

Date: ____________ 20

Examination: ____________ Branch/Semester ____________

Subject: ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

* **Example** Consider the following sequence of 50 nos

0.89	0.17	0.99	0.46	0.05	0.66	0.10	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.21	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

Can the hypothesis that the nos. are independent be rejected on the basis of length of runs above and below the mean, where $\alpha = 0.05$?

$\Rightarrow$ 1) Define hypothesis. H_0: - - - - .

2) The sequence of runs above the. below the mean (0.495) is

\+ − + − − + − − − −
− + + − + − + − + −
\+ − + + − + + + + −
− − + − − + − − + +
− − + + + − + − − +

3) The length of run in sequence is.

1 1 1 2 1 5
2 1 1 1 1 1 1 1
1 1 2 1 4 3
1 2 1 2 2
2 3 1 1 2 1

4) The no observed runs of each length is:-

Run length, i :	1	2	3	4	5
Observed Runs, Oi :	19	8	2	1	1

5) The expected no of runs of length one, two, three, four and five are.

Since no of observations above mean $= n_1 = 24$

and no of observations below mean $= n_2 = 26$.

i) The approximate probability that a run has length 'i' is

$$W_i = \left(\frac{n_1}{N}\right)^i\left(\frac{n_2}{N}\right) + \left(\frac{n_1}{N}\right)\left(\frac{n_2}{N}\right)^i$$

$$W_1 = \left(\frac{24}{50}\right)^1\left(\frac{26}{50}\right) + \frac{24}{50}\left(\frac{26}{50}\right)^1 = \underline{0.4992}$$

$$\therefore W_2 = \left(\frac{24}{50}\right)^2\left(\frac{26}{50}\right) + \frac{24}{50}\left(\frac{26}{50}\right)^2 = 0.2496.$$

$\therefore W_3 = 0.1250.$

$W_4 = 0.0627.$

$W_5 = 0.0315.$

ii) The approximate expected length of run is

$$E(I) = \frac{n_1}{n_2} + \frac{n_2}{n_1} = \frac{24}{26} + \frac{26}{24}$$

$$= \underline{2.0064}$$

iii) The expected runs of various lengths.

$$E(X_i) = \frac{NW_i}{E(I)}.$$

$$E(Y_1) = \frac{NW_1}{E(I)} = \frac{50(0.4992)}{2.0064} = \underline{12.44}$$

$$E(Y_2) = \frac{NW_2}{E(I)} = \underline{6.22}$$

$$E(Y_3) = \frac{NW_3}{E(I)} = \underline{3.115}$$

$E(Y_4) = \underline{1.5625}$

$E(Y_5) = \underline{0.785}$

6). The mean or expected total no of runs of all length in a sequence is.

$$E(A) = \frac{N}{E(I)} = \frac{50}{2.0064} = \underline{24.9203}.$$

7) The expected no of runs of length greater than or equal to 5 is

$$E(A) - \sum_{i=1}^{4} E(y_i) = 24.9203 - 23.3375$$
$$= \underline{1.5828}.$$

8) Apply chi-square test

Run length, i	O_i	$E(y_i)$		$\frac{[O_i - E(y_i)]^2}{E(y_i)}$
1	19	12.44		3.4593
2.	08	6.22		0.5094
3	02	3.115		
4	01 } 04	1.5625	6.2603	0.8161.
≥5	01.	1.5828.		

→ Since class 3, 4 and 5 has expected freqn. less than 5. So it is combined which leads to 6.2603. Similarly the observed freqn of class 3, 4 and 5 is also combined which leads to 4. Reduce the no of classes by 2 which leads in $= 5-2 = 3$.

The test statistics is

$$\chi_0^2 = \sum_{i=1}^{n} \frac{[O_i - E(y_i)]^2}{E(y_i)}.$$

$$= \underline{(3.4593 + 0.5094 + 0.8161)}$$

$$= \underline{4.7848}.$$

9) The critical value for specified significance level $\alpha = 0.05$ with $n-1 = (3-1) = 2$ degrees of freedom is.

$$\chi^2_{0.05, 2} = 5.99$$

10) Since $\chi^2_{0, 2} = 4.7848 < \chi^2_{0.05, 2} = 5.99$.

$\therefore$ H_0 is not rejected.

8/14/30,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

9

Candidate Roll No. ________ (In figures)

Test Exam.

Name : ________

Date : ________ 20

Examination : ________ Branch/Semester ________

Subject : ________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

* Autocorrelation Test :- *

→ It deals with depedence (relation) betwn nos in a sequence.

→ The nos in the sequence might be related.

→ E.g the no at position 2, 7, 12 ... has larger or smaller values.

→ The test computes the auto correlation betwn every m nos (lag) starting with the ith number.

→ Finally it compares the sample correlation to the expected correlation of zero.

→ A nonzero auto correlation implies a lack of independence.

→ If $\rho_{im} > 0$ then the subsequence has +ve auto correlation whereas if $\rho_{im} < 0$ then the subsequence has -ve auto correlation.

→ The following algorithm is used to test the auto correlation

Algorithm

1) Define hypothesis for testing the independence is.
$H_0 : \rho_{im} = 0$ if nos are independent.
$H_1 : \rho_{im} \neq 0$, if nos are dependent.

2) Find out the value of 'i' and lag 'm' using given data.

3) Using i, m, and N, estimate the value of M where M is the largest integer such that $i+(M+1)m \leq N$, and N is the total no of values in the sequence.

4) For large values of M, the distribution of estimator of ρ_{im}, denoted by $\hat{\rho}_{im}$ is approximately normal if the nos R_i, R_{i+m}, R_{i+2m} ---, $R_{i+(M+1)m}$ are uncorrelated where

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} R_{i+(k+1)m}\right] - 0.25$$

5) Find standard deviation of the estimator.

$$\sigma_{\hat{\rho}_{im}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

6) Compute the test statistics.

$$Z_0 = \frac{\hat{\rho}_{im}}{\sigma_{\hat{\rho}_{im}}}$$

7) Determine the critical values $Z_{\frac{\alpha}{2}}$ and $-Z_{\frac{\alpha}{2}}$ for specified significance level α from table A-3.

8) If $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}} \Rightarrow H_0$ is not rejected for significance level α.

9) For small values of M, the test is not sensitive when nos are tested on low side.

Scanned with OKEN Scanner

50 · Mis

Example:- Consider the following sequence of (40) nos.

					0.16	0.30	0.22	0.88	0.4
0.19	0.11	0.82	0.63	0.04	0.38	0.59	0.37	0.71	0.4
0.29	0.56	0.44	0.05	0.81	0.74	0.64	0.50	0.73	0.
0.92	0.45	0.57	0.99	0.20	0.74	0.41	0.09	0.80	0.4
0.02	0.49	0.86	0.24	0.90	0.46	0.39	0.18	0.21	0.9
0.11	0.23	0.77	0.08	0.69					

Test whether 2nd, 7th, 12th --- nos in the sequence are autocorrelated where $\alpha = 0.05$.

⇒ 1) Define hypothesis for testing the independence.

$H_0: \rho_{im} = 0$

$H_1: \rho_{im} \neq 0$.

→ 2) Here the value of $i = 2$ (starting with second number) and lag $m = 5$ (every five numbers).

3) Given that $N = 40$ 50, using i, m and N estimate M which is the largest integer such that.

$i + (M+1)m \leq N$.

$\Rightarrow 2 + (M+1)5 \leq 40$ 50.

$\Rightarrow (M+1)5 \leq 38$ 48

$M+1 \leq 7.6$ 9.6

$M \leq 6.6$ 8.6.

$\therefore M = 6$ 8.

4) The distribution of estimator

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} \cdot R_{i+(k+1)m}\right] - 0.25$$

$$\hat{\rho}_{2,5} = \frac{1}{(6)+1}\left[\sum_{k=0}^{6} R_{2+5k} \cdot R_{2+5(k+1)}\right] - 0.25$$

$$= \frac{1}{9\cancel{7}}\left[R_2 \cdot R_7 + R_7 \cdot R_{12} + R_{12} R_{17} + R_{17} \cdot R_{22} + R_{22} \cdot R_{27} + R_{27} \cdot R_{32} + R_{32} \cdot R_{37}\right] - 0.25$$

$$= \frac{1}{9\cancel{7}}\left[(0.16)(0.30) + (0.30)(0.56) + (0.56)(0.59) + (0.59)(0.45) + (0.45)(0.64) + (0.64)(0.49) + (0.49)(0.41)\right] - 0.25$$

$$= \frac{1}{\cancel{7}}[1.6144] - 0.25 = -0.\cancel{0194} \quad -0.0706$$

5) The standard deviation of the estimator.

$$\sigma_{\hat{f}_{im}} = \sqrt{\frac{13M+7}{12(M+1)}} \quad \sqrt{\frac{13(8)+7}{12(9)}} = 0.097$$

$$= \sqrt{\frac{13(6)+7}{12(6+1)}} = 0.1098$$

6) The test statistics.

$$Z_0 = \frac{\hat{f}_{im}}{\sigma_{\hat{f}_{im}}} = \frac{\hat{f}_{25}}{\sigma_{\hat{f}_{25}}}$$

$$= \frac{-0.0194}{0.1098} \quad Z_0 = \frac{-0.0706}{0.097}$$

$$= -0.1767 \quad = -0.7278$$

7) Determine the critical value $Z_{\frac{\alpha}{2}}$ and $-Z_{\frac{\alpha}{2}}$ for specified significance level α from Table A.3

Since $\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$\therefore Z_{0.025} = 1.96$

-0.7278

8) Since $-Z_{0.025} = -1.96 \leq Z_0 = -0.1767 \leq Z_{0.025} = 1.96$

$\therefore$ Ho is not rejected.

1/16/15,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

10

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

* Gap Test * (Refer - 3rd edition - Jerry Banks)

→ The gap test is used to count the no of digits betwn successive occurrences of the same digit.

→ Here we are interested in the freq of the gaps.

→ The probability of gap is determined by
P(m followed by exactly x non m digits)
$= (0.9)^x (0.1), \; x = 0, 1, 2 \ldots$ {Refer 3rd edit Jerry banks}

→ Every digit 0, 1, 2, ... 9 must be analyzed to test the nos are independent using gap test.

→ The observed frequencies of various gap sizes for all the digits are recorded and compared with the theoretical freq using K-S test.

→ Where as the CDF of theoritical freq distribution based on selected class interval width is given by

$$F(x) = 0.1 \sum_{n=0}^{x} (0.9)^n$$

$$= 1 - 0.9^{x+1} = P(\text{gap} \le x).$$

→ The following algorithm is used to test the independence on the basis of length of gaps associated with every digit.

Algorithm

1) Define hypothesis for testing the independence.

   H_0: R_i is independently.

   H_1: R_i is not independently.

2) Determine the no of gaps and length of each gap associated with each digit (0, 1 -- 9).

3) Select the interval width based on the no of gaps and generate the freq^n distribution table for the sample of gaps and apply K-S test.

Gap length	Freq^n	Relative Freq^n	Cumulative Relative freq^n $S_N(x)$	CDF of Theoretical Distr^n $F(x)$	$\lvert F(x) - S_N(x) \rvert$

4) Compute the test statistic D which is the maximum deviation betw^n $F(x)$ and $S_N(x)$.

   $D = \max \lvert F(x) - S_N(x) \rvert$.

5) Determine the critical value D_α for the specified value of significance level α and the sample size N from table A.8.

6) If $D > D_\alpha \Rightarrow H_0$ is rejected.

Example :- Consider the following sequence of 120 digits.

2	3	6	5	6	0	0	1	3	4	5	6	7	9	4	9	3	1	8	3
1	3	7	4	8	6	2	5	1	6	4	4	3	3	4	2	1	5	8	7
0	8	8	2	6	7	8	1	3	5	3	8	4	0	9	0	3	0	9	2
4	6	9	9	8	5	6	0	1	7	6	7	0	3	1	0	2	4	2	0
1	1	2	6	7	6	3	7	5	9	3	6	6	7	8	2	3	5	9	6
6	4	0	3	9	3	6	8	1	5	0	7	6	2	6	0	5	7	8	0

Test whether these digits can be assumed to be independent based on the freq^n with which gap occurs.

Use $\alpha = 0.05$.

Similarly $1-0.9^{7+1} = 1-0.9^{8} = 0.5695$

$\rightarrow 1-0.9^{11+1} = 1-0.9^{12} = 0.7176$

$\Rightarrow$ Define hypothesis for testing the independence:

H_0 : $R_i \sim$ independently.

H_1 : $R_i \nsim$ independently.

2) Given that - No of digits = 120.

$\therefore$ Total no of gaps = No of digits − No of distinct digits.

= 120 − 10

= 110.

The no of gaps and length of each gap associated with each digit is.

Digit	Length of each gap	No of gaps
0	0, 33, 12, 1, 1, 9, 4, 2, 3, 2, 2, 7, 4, 3	13.
1	9, 2, 27, 7, 10, 20, 5, 5, 0, 26.	10
2	25, 8, 7, 15, 16, 13, 12, 17.	09
3	6, 7, 21, 10, 0, 14, 1, 5, 16, 12, 3, 5, 6, 1.	15
4	4, 8, 6, 0, 2, 17, 7, 16, 23.	09
5	6, 16, 9, 11, 15, 22, 8, 11, 6.	09
6	1, 6, 13, 3, 14, 16, 4, 3, 12, 1, 5, 0, 6, 0, 5, 5, 1.	17
7	9, 16, 5, 23, 1, 12, 2, 5, 17, 5.	10
8	5, 13, 20, 3, 4, 12, 29, 12, 10.	10
9.	1, 38, 3, 3, 0, 25, 8, 5.	08.

3) Select the interval width based on the no of gaps and generate the freq distribution table for the sample of gaps and apply K-S test.

Gap length	Frequency	Relative Frequency	Cumulative Relative Freq $S_N(x)$	CDF of Theoretical freq distribution $F(x) = 1-0.9^{x+1}$	$\lvert S_N(x) - F(x)\rvert$
0-3	34	0.3091	0.3091	0.3439	0.0348
4-7	30	0.2727	0.5818	0.5695	0.0123
8-11	13	0.1182	0.7000	0.7176	0.0176
12-15	13	0.1182	0.8182	0.8147	0.0035
16-19.	09	0.0818	0.9000	0.8784	0.0216
20-23	05	0.0454	0.9454	0.9202	0.0252
24-27	03	0.0273	0.9727	0.9497	0.0230
28-31	01	0.0091	0.9818	0.9657	0.0161
32-35	01	0.0091	0.9909	0.9775	0.0134
36-39.	01.	0.0091	1.0000	0.9852.	0.0148

4) The test statistics

$$D = \text{Max} |F(x) - S_N(x)|$$
$$= 0.0348$$

5) Determine the critical value, D_α for specified value of significance level α and the sample size N from table A.8. Since $\alpha = 0.05$ and $N = 110$

$$\therefore D_{0.05} = \frac{13.6}{\sqrt{110}}$$
$$= 0.1297$$

6) Since $D = 0.0348 < D_{0.05} = 0.1297$

$\therefore$ H_0 is not rejected.

* Poker Test :- *

→ It tests the freq of certain digits in a series of numbers.

→ Here we show only the three digit version for testing the independence property.

→ In this case the generated random numbers are rounded to three digits. Then each random number is classified into one of following categories.

→ 1) All digits are different from each other.

→ 2) All digits are identical.

→ 3) There is exactly one pair of identical digits.

→ The focus of test is not the sequence from one random number to another. Rather this test focuses on the internal digits within a given number.

Scanned with OKEN Scanner

(11)

1/16/15,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

→ If digits are truely generated randomly, the following results hold.

1) The individual numbers can all be different and which has a probability.

$P = P(2^{nd}$ digit diff from $1^{st}) \, P(3^{rd}$ digit diff from $1^{st}, 2^{nd})$

$= (0.9)(0.8) = 0.72$

2). The individual numbers can all be same and which has a probability.

$P = P(2^{nd}$ digit same as $1^{st}) \, P(3^{rd}$ same as $1^{st})$

$= (0.1)(0.1) = 0.01$

3) There can be one pair of like digits and which has a probability.

$P = 1 - [P(\text{three diff digits}) + P(\text{three like digits})]$

$= 1 - (0.72 + 0.01) = 0.27$

→ The observed value is then compared with expected value using the chi square test.

Algorithm

→ 1) Define hypothesis for testing independence ...

→ 2) Generate freq distrib table for above three combinatio and apply chi square test.

Combination	Observed Freq, O_i	Expected Freq $E_i = P \times N$	$\frac{(O_i - E_i)^2}{E_i}$

3) Compute the sample test statistics.

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

4) Determine the critical value for specified significance level α with $n-1$ degrees of freedom from table A.6

5) If $\chi_0^2 > \chi^2_{\alpha, n-1} \Rightarrow$ Ho is rejected, else no difference has been detected betwn the sample distribution and uniform distribution.

Example A sequence of 1000 three digit numbers has been generated and an analysis indicates that 560 have three different digits, 380 contain exactly one pair of like digits and 60 contain three like digits. Based on Poker test, test whether these nos are independent. Use $\alpha = 0.05$

$\Rightarrow$ 1) Define hypothesis ....

2) Generate freq distribn table and apply chisquare test.

Combination, i	Observed Freqn O_i	Expected Freq $E_i = P \times N$	$\frac{(O_i - E_i)^2}{E_i}$
Three dist digits 1	560	720	35.56
Three like digits, 2	60	10	250.00
Exactly one pair, 3	380	270	44.82

3) The sample test statistics-

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= (35.56 + 250.00 + 44.82) = 330.38$$

4) Determine the critical value for specific significance level α and $n-1$ degree of freedom from table A.6

Since $\alpha = 0.05$, $n-1 = 3-1 = 2$

$\therefore \chi^2_{0.05, 2} = 5.99$

5) Since $\chi_0^2 = 330.38 > \chi^2_{0.05, 2} = 5.99$

$\therefore$ Ho is rejected

1/16/15,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

(12)

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date	

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

(M-2007)

Random Number Problems

1) Test the following random numbers for independence by run test Take $\alpha = 0.05$ and critical value $Z_{0.025} = 1.96$.

$\{37, 59, 63, 07, 92, 48, 12, 86\}$

⇒ Runs up and Runs Down

1) Define hypothesis.

2) The sequence of runs up and down is.

$+ + \quad - \quad + \quad - - \quad +$.

3) Total no of runs $= a = 5$

4) Mean and variance of a is.

$\mu_a = \frac{2N-1}{3}$ $\quad \sigma_a^2 = \frac{16N-29}{90}$

$= \frac{2(8)-1}{3} = \frac{15}{3} = 5$ $\quad = \frac{16(8)-29}{90} = 1.1$

5) Get the Std normal Statistics.

$Z_0 = \frac{a-\mu_a}{\sigma_a} = \frac{5-5}{\sqrt{1.1}} = 0$

6) Get the critical value for $\alpha = 0.05$

$\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$

7) Since $-Z_{0.025} = -1.96 \leq Z_0 = 0 \leq Z_{0.025} = 1.96$

∴ H_0 is Accepted

* Runs above and below mean $\bar{x}$.

1) The sequence of runs above and below mean (49.5) is

$- \quad ++ \quad - \quad + \quad -- \quad +$.

2) No of observations above mean $= n_1 = 4$.
No of observations below mean $= n_2 = 4$.
Total no of runs $= \underline{b = 6}$.

3) Mean and variance of b.

$$\mu_b = \frac{2n_1 n_2}{N} + \frac{1}{2} \qquad \sigma_b^2 = \frac{2n_1 n_2(2n_1n_2 - N)}{N^2(N-1)}$$

$$= \frac{2(4)(4)}{8} + \frac{1}{2} = \underline{4.5}, \qquad = \frac{2(4)(4)[2\cdot4\cdot4 - 8]}{8^2(8-1)} = \underline{\underline{12}}$$

4) Det. the std. normal statistics.

$$Z_0 = \frac{b - \mu_b}{\sigma_b} = \frac{6 - 4.5}{\sqrt{1.71}} = \underline{1.15}$$

5) Det. the critical value for $\underline{\alpha = 0.05}$

$\pm Z_{\alpha/2} = \pm Z_{0.025} = \underline{\pm 1.96}$

6) Since $-Z_{0.025} = -1.96 < Z_0 = 1.15 < Z_{0.025} = 1.96$

$\therefore$ $\underline{H_0 \text{ is Accepted}}$.

2) (May 2009)

Test the following random numbers for independence by runs up and down test.

Take $\alpha = 0.50$ and critical value $Z_{0.025} = 1.96$

$\{0.12, 0.01, 0.23, 0.28, 0.89, 0.31, 0.64, 0.28, 0.33, 0.93\}$

⇒ 1) Define hypothesis

2) Sequence of runs up and down is.

$- \; + + \; + \; - \; + \; - \; + +$

3) Total no of runs $= a = 6$.

4) Mean and variance of a is.

$\mu_a = \frac{2N-1}{3}$ $\quad \sigma_a^2 = \frac{16N-29}{90}$

$= \frac{2(10)-1}{3} = 6.33$, $\quad \frac{16(10)-29}{90} = 1.46$

5) Get the std normal statistics

$Z_a = \frac{a - \mu_a}{\sigma_a} = \frac{6 - 6.33}{1.46} = -0.23$

6) Get the critical value $\alpha = 0.05$

$\pm Z_{\frac{\alpha}{2}} = \pm Z_{0.025} = \pm 1.96$

7) Since $-Z_{0.025} = -1.96 \leq Z_0 = \leq Z_{0.025} = 1.96$.

∴ H_0 is Accepted.

3) Test the following random numbers for independence by poker test. {Dec, 2009, 2011}

$\{0.594, 0.928, 0.515, 0.055, 0.507, 0.351, 0.262, 0.797, 0.788, 0.442, 0.097, 0.798, 0.227, 0.127, 0.474, 0.825, 0.007, 0.182, 0.929, 0.852\}$ $\alpha = 0.05$, $\chi^2_{0.05, 2} = 5.99$

⇒ 1) Define hypothesis --

2) Generate freq distrib table and apply chi square test.

Combination i	Observed freq O_i		Expected freq $E_i = P \times N$		$\frac{(O_i - E_i)^2}{E_i}$
Three diff digits, 1	10		$= 0.72 \times 20 = 14.4$		1.34
Three like digits, 2	0	10	$= 0.01 \times 20 = 0.2$	5.6	3.46
Exactly one pair = 3.	10		$= 0.27 \times 20 = 5.4$		

3) The sample test statistics

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = [1.34 + 3.46] = 4.80$$

4) Determine the critical value for specified significance level α and $n-1 = 2-1 = 1$.

$\therefore \chi^2_{0.05, 1} = 3.84$

5) Since $\chi_0^2 = 4.8 > \chi^2_{0.05, 1} = 3.84$

$\therefore$ Ho is rejected

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

4) Consider the following sequence of 40 numbers. (Dec-2010)

0.67	0.31	0.53	0.91	0.80	0.27	0.61	0.49	0.76	0.85
0.62	0.28	0.55	0.77	0.38	0.65	0.29	0.55	0.83	0.92
0.09	0.33	0.24	0.07	0.30	0.54	0.43	0.66	0.71	0.52
0.11	0.36	0.12	0.78	0.95	0.44	0.50	0.19	0.22	0.38

Based on runs up and runs down determine whether the hypothesis of independence can be rejected where $\alpha = 0.05$

⇒ 1) Define hypothesis.

2) The sequence of runs up and down as,

- + + + - + - + + - - + + - + - + + + - + - - - + + - + + - - + - + + - + + +.

3) The total number of runs $= a = 26$.

4) Mean and Variance of a is.

$\mu_a = 26.33 \qquad \sigma_a^2 = 6.79$.

5) The std normal statistics :-

$Z_0 = -0.1266$.

6) The critical value for $\alpha = 0.05$ is

$\pm Z_{0.025} = \pm 1.96$.

7) Since $-Z_{0.025} = -1.96 < Z_0 = 0.1266 < Z_{0.025} = 1.96$

$\therefore H_0$ is rejected.

5) Consider the sequence of 40 digits. (Dec-2010)

Consider the sequence				0.43	0.56	0.97	0.75	0.49
0.81	0.62	0.27	0.88	0.72	0.43	0.56	0.97	0.75
0.06	0.39	0.89	0.23	0.02	0.66	0.40	0.19	0.73
0.53	0.30	0.92	0.96	0.58	0.47	0.60	0.11	0.29
0.87	0.28	0.69	0.37	0.64	0.39	0.57	0.54	0.01

Can the hypothesis that the numbers are independent be rejected on the basis of runs above and below the mean, where $\alpha = 0.05$?

⟹ 1) Define hypothesis for testing the independence as,

2) The sequence of runs above and below the mean (0.495) is.

++ - ++ - +++ - - - + - - + - - +++ - +++ - + - - ++ - + - + - ++ - -

3) The number of observations above mean $= n_1 = 22$.
The number of observations below mean $= n_2 = 18$.
The total number of runs $= b = 24$.

4) Mean and variance of b:-
$\mu_b = 20.3$ and $\sigma_b^2 = 9.5446$.

5) The standard normal statistics:-
$Z_0 = 1.1976$.

6) Let the critical value for $\alpha = 0.05$.
$\pm Z_{0.025} = \pm 1.96$

7) Since $-Z_{0.025} = -1.96 < Z_0 = 0.1976 < Z_{0.025} = 1.96$

$\therefore$ Ho is accepted