8/14/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

5

Candidate Roll No. ____________ (In figures)

Name: ____________

Test Exam.

Date: ____________ 20

Junior Supervisor's full Signature with Date

Examination: ____________ Branch/Semester ____________

Subject: ____________

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 generated by tossing of coin 10 times

T T H H H T H H T T
2 3 1 2 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.

Scanned with CamScanner

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 0.1

→ 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 these two extrems.

→ 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 ~ independently.

H_1 : R_i ≁ 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.

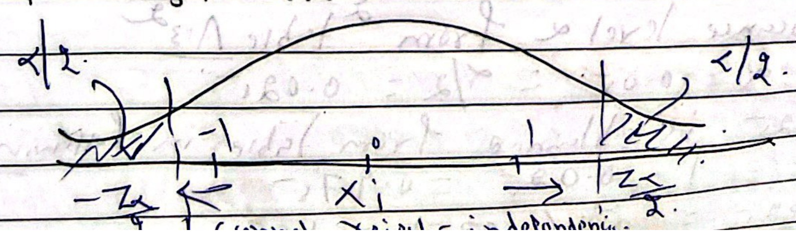
$$\mu_a = \frac{2N-1}{3} \text{ and } \sigma_a^2 = \frac{16N-29}{90}$$

5) Compute std normal statistics.

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

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 $-Z_{\frac{\alpha}{2}} \le Z_0 \le Z_{\frac{\alpha}{2}}$, 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
0.97 0.44 0.43 0.94 0.82 0.60 0.73 0.69 0.21
0.04 0.81 0.85 0.30 0.47 0.96 0.17 0.72 0.62
0.10 0.60 0.34 0.65 0.79 0.44 0.02 0.37 0.4
Based on runs up and runs down determine whether the hypothesis of independence can be rejected, where $\alpha = 0.05$.

→ Solution 1) Define hypothesis.
H_0: $R_i \sim$ independently.
H_1: $R_i \nsim$ 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.
$\mu_a = 2N-1/3 = 2(40)-1/3 = 26.33$
$\sigma_a^2 = 16N-29/90 = 16(40)-29/90 = 6.79$

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 $Z_{\alpha/2}$ and $-Z_{\alpha/2}$ 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 CamScanner

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 0.01

→ 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 deve[...] and it is compared with the critical value.

→ The following algorithm is used to test the independence on basis of runs above and bel[ow] 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.

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

5) Compute std normal statistics.

$Z_0 = \frac{b - \mu_b}{\sigma_b}$ 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.

7) If $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$ $\therefore$ 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 0.32
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 : $R_i \sim$ independently.

H_1 : $R_i \not\sim$ independently.

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

− − − + + + − − −
+ − + + − + − + −
+ + − + − − − + + −
+ − + − − + − − − +.

3) The no of observ^n 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 run 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 down or runs above and runs below the mean are 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

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

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

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

Question No.													
Marks Obtained													

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

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

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.

Scanned with CamScanner

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. classes

9) Determine the critical value for the specific 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
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79
0.18	0.21	0.97	0.88	0.64	0.47	0.60	0.11	0.29

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

Scanned with CamScanner

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	10.	7.1	9.58.	0.0184.
3	2		1.98		
≥ 4	0.		0.5		

→. 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 adjacient class interval.

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. 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.

Scanned with CamScanner

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 $\alpha = 0.05$?

⇒ 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

Scanned with CamScanner

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.

$$E(Y_i) = \frac{NW_i}{E(I)}$$

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

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

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

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

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

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	1.5625		
≥ 5	01	1.5828		
	04		6.2603	0.8161

→ Since class 3, 4 and 5 has expected freq. less than 5. So it is combined which leads to 6.2603. Similarly the observed freq 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

$$\chi_0^2 = \sum_{i=1}^{n} \frac{[O_i - E(y_i)]^2}{E(y_i)}$$
$$= (3.4593 + 0.5094 + 0.8161)$$
$$= \underline{4.7848}.$$

Scanned with CamScanner

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.

Scanned with CamScanner

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 autocorrelation betwn every m nos (lag) starting with the ith number.

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

→ A nonzero autocorrelation implies a lack of independence.

→ If $\rho_{im} > 0$ then the subsequence has +ve autocorrelation 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.

50 · Mistake

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

0.19	0.11	0.82	0.13	0.04	0.16	0.30	0.22	0.88	0.98
0.29	0.56	0.44	0.05	0.81	0.38	0.59	0.37	0.71	0.43
0.92	0.45	0.57	0.99	0.20	0.74	0.64	0.50	0.73	0.15
0.02	0.49	0.86	0.24	0.90	0.74	0.41	0.09	0.40	0.42
0.11	0.23	0.77	0.08	0.69	0.46	0.39	0.18	0.21	0.98

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 $N = 40$, 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.

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

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

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

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

$$= \frac{1}{9}[1.6144] - 0.25 = -0.0794 \quad -0.0706$$

5) The standard deviation of the estimator.

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

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

6) The test statistics.

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

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

$$= -0.1767 \qquad = -0.7278$$

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

$\therefore$ 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.

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 theoretical 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 ~ independently.
   H_1: R_i ≁ 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 distribution table for the sample of gaps and apply K-S test.

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

4) Compute the test statistic D which is the maximum deviation betwn $F(x)$ 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
1 3 7 4 8 6 2 5 1 6 4 4 3 3 4 2 1 5 8
0 8 8 2 6 7 8 1 3 5 3 8 4 0 9 0 3 0 9
4 6 9 9 8 5 6 0 1 7 6 7 0 3 1 0 2 4 2
1 1 2 6 7 6 3 7 5 9 3 6 6 7 8 2 3 5 9
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 with which gap occurs.
Use $\alpha = 0.05$.

$F(x) = 1-0.9^{x+1} = 1-0.9^{3+1} = 1-0.9^4 = 0.3439.$

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.

$= 120 - 10$

$= \underline{110}$.

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, 2, 1, 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, 2, 0, 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 Theoritical 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.

Scanned with CamScanner

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 $N = 110$

$\therefore D_{0.05} = \frac{1.36}{\sqrt{110}}$

$= 0.1297$

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.

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.

$P = P$ (2nd digit diff from 1st) P(3rd digit diff from 1st, 2nd).

$= (0.9)(0.8) = \underline{0.72}$

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

$P = P$ (2nd digit same as 1st) P(3rd same as 1st)

$= (0.1)(0.1) = \underline{0.01}$

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

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

$= 1 - (0.72 + 0.01) = \underline{0.27}$

→ 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 combination and apply chi square test.

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

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 $\chi_0^2 > \chi^2_{\alpha, n-1} \Rightarrow$ H_0 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 Freq^n, 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 $\alpha = 0.05$, $n-1 = 3-1 = 2$.

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

5) Since $\chi^2_0 = 330.38 > \chi^2_{0.05, 2} = \underline{5.99}$

$\therefore$ Ho is rejected