8.19 Can we say that education depends on sex at $\alpha = 0.05$ on the basis of a random sample of 300 persons classified in the following (2×3) contingency table:

Education / Sex	middle	SSC	college
male	30	45	75
female	75	30	45

8.20 Find the value of chi-square from the following table and test the hypothesis that there is no relation between the attributes A and B. Use $\alpha = 0.05$

Attributes	B_1	B_2	B_3
A_1	600	225	180
A_2	300	660	320
A_3	200	275	600

8.21 Discuss coefficient of mean square contingency for an $(r \times c)$ contingency table.

8.22 Find the Pearson's coefficient of mean square contingency form the following table of frequencies showing the gender of customer and mode of payment for purchasing articles on a store.

Attributes	cash	loan
male	431	5
female	291	9

8.23 Calculate the coefficient of contingency from the following data showing resemblance between hair colour and eye colour:

Hair colour / Eye colour	black	brown	red	grey
grey	1768	807	47	189
blue	946	1287	53	746
brown	115	438	6	288

Unit - 9

Design of Experiment

After studying this unit, the students will be able to

- Describe the meaning of the design of experiment.
- Explain the elements involved in designing an experiment: the experimental unit, the treatment, the replication and the response.
- Define randomization, completely randomized design.
- Give layout plan of completely randomized design.
- Identify the merits and demerits of completely randomized design.
- Know the meaning of analysis of variance.
- State the assumptions of analysis of variance.
- Describe and calculate: the total sum of squares, the treatment sum of squares, and the error sum of squares.
- Describe and calculate the degrees of freedom for: the total sum of squares, the treatment sum of squares and the error sum of squares.
- Describe and calculate the treatment mean square, the error mean square. Test the equality of means of several normal populations.

9.1 Introduction to design of experiment

In our daily life, we compare things directly by using the results, which is not good because these results are not only due to the effect of those things which we are comparing but are also influenced by some hidden factors. For example, if we want to compare the yield of two varieties of wheat, one is sown in the hot area of D.I. Khan and the second in the cold area of swat and on the basis of their production we say that one is better than the other is not justifiable because the production is surely also affected by soil, amount of water, weather, cultivation techniques, fertilizer, pesticide etc. To find the real difference between the two varieties it is reasonable to make a plan where both the varieties are given a fair chance to show their effects in the form of yield while all other factors, mentioned above are kept under control. Such a plan of experiment in which all situations except that of treatments are kept under control as much as possible, is known as design of experiments. Experimental designs are widely used for comparing treatments under identical conditions.

9.1.1 Explanation of basic terms

- **Treatment**

A thing whose effect is measured and is compared with others is called treatment. For example, to test the effect of feeds on cows, medicines on patients, dates of sowing on the yield of crop, teaching methods on students, etc. Here feeds, medicines, dates of sowing, teaching methods are treatments.

- **Experimental material**

The total material or objects on which the experiment is done is known as experimental material. For example, cows, patients, soil, students etc.

- **Experimental unit**

The smallest division of an experimental material to which a treatment is applied is called experimental unit. For example, a cow, a patient, a plot of land, a student etc.

- **Yield**

The results obtained from the experimental units after applying the treatment are called yield or responses.

- **Block**

A group of homogenous experimental units e.g. land of same fertility, students of same age, weight, I.Q, etc. to which all treatments are assigned at random is called a block. For soil fertility blocks are always made perpendicular to variation. Generally, blocks are made by uniformity trial. Each treatment is given a chance in each block and each appears in a block only once.

- **Uniformity trial**

It is a trial or an experiment in which same treatment is given to all experimental units and then experimental units of the same performance is consider as a block. For example, same test is given to whole class then all the students who get first division is a block, all the students getting second division is a block, all third division holders is a block. Uniformity trial is used only for making blocks.

- **Extraneous factor**

The responses from all experimental units receiving the same treatment may not be identical even if the experiment is performed under similar conditions. These responses are influenced not only by treatments but also by other factors as well, some of which can be controlled while there are some over which there is no control or very little control. For example, natural differences between the experimental units like heterogeneity of soil, I.Q of students, climatic factors etc. All such factors which are not in the control of researcher are called uncontrolled or extraneous factors.

- **Experimental error**

The error caused by extraneous factors which are beyond the control of human approach is known as experimental error. It is a major problem for experimenter and is a mask on the true effect of the treatments because the observed difference in treatments is a sum of the true difference of the treatments plus due to the experimental error. Design of experiment is actually a strategy for controlling the experimental error in order to bring out the real difference among the treatments.

9.1.2 Basic requirements or principles of a good experimental design

An experimental design can reduce the experimental error only if it follows the three requirements of a good design namely, randomization, replication and local control.

(i) Randomization

The allocation of treatments to experimental units in such a manner that an experimental unit has equal chance of receiving any of the treatments is called randomization. It is usually done by (a) using lottery method (b) using random number table method (c) using any computer package which may perform randomization.

- Advantages of randomization are:
  a) Eliminates the human biases.
  b) Introduces the independence in the assignment of treatments to the experimental units which in turn creates independence amongst the observations, required for the validity of F-test.

(ii) Replication

Repetition of a treatment on a number of experimental units in an experiment is known as replication of the treatment. Replications are essential because in all experiments, there are great variations in fertility of experimental units and all the treatments do not get equal chance of experiencing every type of environment in the field. If a treatment is allotted only once and it goes to more fertile experimental unit, will be in more advantageous position than those which are applied to less fertile experimental units. This type of variation can be eliminated by using replication process which improve the precision of an experiment and provides a valid estimate of the experimental error.

(iii) Local control

Local control is a procedure which maintains greater homogeneity of experimental units within a block of an experiment. For example, soil fertility of field is a factor which affects the plant growth and yield. All the plots having the same soil fertility should constitute a block. The soil fertility of land can be assessed by conducting a uniformity trial on the field prior to actual field experiment. The treatments should be assigned to the blocks in equal number of times. Local control also called error control reduces the experimental error. Note that all these three principles contribute a lot in increasing the efficiency of design.

9.2 Completely randomized design

Experimental design, in which the treatments are allocated completely at random to the experimental units, is called completely randomized design (CRD). In this design all the experimental units which are homogeneous, as much as possible, are taken as a single group. Any number of treatments and any number of units per treatment may be used. It is the simplest design.

9.2.1 Experimental layout for CRD

The layout of an experiment is the actual placement of the treatments on the experimental units. To explain the procedure of randomization let us consider four treatments, each replicated three times.

Step 1: Determine the total number of plots $n = t \times r = 4 \times 3 = 12$

Step 2: Assign a plot number to the experimental plots or units as follows:

1	2	3	4
5	6	7	8
9	10	11	12

Step 3: Assign treatments to the experimental units. The randomization both by lottery and random number table methods is explained as follows:

(i) Lottery method

Make 12 slips of paper. Write T_1 on 3 slips, T_2 on 3 slips, T_3 on 3 slips, T_4 on 3 slips and place them in a bowl and mix them thoroughly. Draw the slips one at a time without replacement and allot to the above experimental units serially from 1 to 12

(ii) Random number table method

Take any random number table, start from anywhere and take twelve two-digit random numbers. Exclude the number which is greater than 12 or which is repeated. Let the random numbers selected are: 03 02 04 06 01 07 11 08 10 12 09 05, so allot treatment T_1 to (03, 02, 04) experimental units, T_2 to (06, 01, 07) experimental units, T_3 to (11, 08, 10) and T_4 to (12, 09, 05). The layout will appear as:

1 T_2	2 T_1	3 T_1	4 T_1
5 T_4	6 T_2	7 T_2	8 T_3
9 T_4	10 T_3	11 T_3	12 T_4

- **Merits of completely randomized design**

i. Its layout is very easy.

ii. There is complete flexibility in this design i.e. any number of treatments and replicates for each treatment can be taken.

iii. Whole experimental material can be utilized in this design.

iv. This design yields maximum degrees of freedom for experimental error.

v. The analysis of data, both for equal and unequal number of replications is simplest as compared to any other design.

vi. Missing observation(s) creates no problem in analysis of data, nor efficiency of the design affected.

vii. CRD is suitable in situations where a fraction of the units is likely to be destroyed during experimentation or is likely to fail to respond.

- **Demerits of completely randomized design**

i) It is difficult to find homogeneous experimental units in all respects.

ii) The design is suitable for a small number of treatments.

iii) CRD is seldom suitable for field experiments as compared to other experimental designs.

9.3 Analysis of variance

We have learnt that Z-test and t-test are used for testing hypothesis about population mean or equality of two population means. Analysis of variance (ANOVA) is also a hypothesis testing procedure that is used to compare three or more populations or treatments means. ANOVA procedure provides greater flexibility in designing experiments, analysis and interpretation of experimental results to the researchers. This procedure was developed by the British statistician Sir Ronald A. Fisher.

Definition

Measurements are taken on each experimental unit according to characteristic of interest and its variance is calculated. This total variance is due to various factors involved in the experiment. "ANOVA is a technique which split the total variance in to meaningful component variances; each gives an estimate of the population variance. The ratio of two component variances is distributed as F with corresponding degrees of freedom." Thus ANOVA enables one to know whether the variance due to a component factor is significantly more than the variance due to experimental error or not and hence decision is made about the null hypothesis.

9.3.1 Assumptions for analysis of variance

Analysis of variance technique is based on the following assumptions.

i) Independence

The samples are drawn randomly and are independent of other samples.

ii) Normality

The populations under study have distributions which are assumed to be normal having means $\mu_1, \mu_2, \ldots \mu_k$.

iii) Homogeneity

All populations under study have equal variance i.e. $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$

iv) Additivity

The random observations X_{ij} can be expressed as the sum of means μ_j and the error terms ε_{ij} as $X_{ij} = \mu_j + \varepsilon_{ij}$

Note that in practice minor deviations from the assumptions is tolerated but serious violation of these assumptions will damage the ANOVA procedure.

9.3.2 One way analysis of variance (equal sample size case)

The observed data of k samples collected for k treatments, each of size r such that $rk = n$ can be arranged in tabular form as:

Treatments / Observations	1	2	...	j	...	k	Total
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1k}	
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2k}	
⋮							
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ik}	
⋮							
r	x_{r1}	x_{r2}	...	x_{rj}	...	x_{rk}	
Total $T_{.j}$	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.k}$	$T_{..}$ (Grand total)
Means $\bar{x}_{.j}$	$\bar{x}_{.1}$	$\bar{x}_{.2}$	...	$\bar{x}_{.j}$	...	$\bar{x}_{k}$	$\bar{x}_{..}$ (Grand mean)
Sum of squares	$\sum_{i=1}^{r}(\bar{x}_{i1}-\bar{x}_{1})^2$	$\sum_{i=1}^{r}(\bar{x}_{i2}-\bar{x}_{2})^2$		$\sum_{i=1}^{r}(\bar{x}_{ij}-\bar{x}_{j})^2$		$\sum_{i=1}^{r}(x_{ik}-\bar{x}_{k})^2$	$\sum_{j=1}^{k}\sum_{i=1}^{r}(x_{ij}-\bar{x}_{..})^2$ (Total sum of square)

9.3.3 Components of the total sum of squares (equal sample size)

In one way analysis of variance the total variation present in the observed data may be due to two meaningful components i.e. (i) variation among treatments (samples) (ii) variation with in the samples, so the T.S.S is divided in to two components as:

$$\text{T.S.S} = \sum_{j=1}^{k}\sum_{i=1}^{r}(x_{ij} - \bar{x}_{..})^2$$

$$= \text{within SS} + \text{Treatment SS}$$

To obtain unbiased estimates for σ^2 divide both sides by respective degrees of freedom to have

$$S_T^2 = S_e^2 + S_{Tr}^2$$

Or MST = MSE + MSTr

The F-Test is defined as $F = \frac{\text{MSTr}}{\text{MSE}}$ with ($v_1 = k-1$ and $v_2 = n-k$) df.

which is used for testing H_0: $\mu_1 = \mu_2 = ... = \mu_k$ verses

H_1: Not all means are equal.

H_0 is rejected when, $F_c \geq F_{\alpha,(v_1,v_2)}$, whereas $F_{\alpha,(v_1,v_2)}$ is obtained from the percentage points of the F-distribution, Table 9.1.

9.3.4 Components of the total degrees of freedom

When the total sum of squares is partitioned into component parts then the total degrees of freedom can also be partitioned into component parts accordingly as below:

$(n-1) = (n-k) + (k-1)$

Total df = within df + between df

9.3.5 ANOVA table

For convenience all sources of variation (S.O.V), degrees of freedom (df), sum of squares (S.S), mean squares (M.S) are presented in tabular form, called ANOVA table given below:

ANOVA table

S.O.V	df	S.S	M.S	F_{cal}
Treatment	$k-1$	T_rSS	$\frac{TrSS}{k-1}=MSTr$	$F_c=\frac{MSTr}{MSE}$
Error	$n-k$	ESS	$\frac{ESS}{n-k}=MSE$	—
Total	$n-1$	TSS	$\frac{TSS}{n-1}=MST$	—

This table is used for testing hypothesis H_0: $\mu_1=\mu_2=\cdots=\mu_k$ in one way analysis of variance.

- **Formulas for sum of squares**

Correction factor: $C.F=\frac{T_{..}^2}{n}$

Total sum of squares: $TSS=\sum_{j=1}^{k}\sum_{i=1}^{r}x_{ij}^2-C.F$

Treatment sum of squares: $TrSS=\frac{\sum_{j=1}^{k}T_{.j}^2}{r}-C.F$

Error sum of squares: $ESS=TSS-T_rSS$

9.3.6 General procedure for testing hypothesis in case of one way ANOVA (equal sample size case)

i. H_0: $\mu_1=\mu_2=\cdots=\mu_k$

H_1: Not all means are equal

ii. Choose an appropriate level of significance

iii. Test statistic to be used in this case is

$$F_c=\frac{MSTr}{MSE}\text{ with }(v_1=k-1, v_2=n-k)\text{ d.f}$$

iv. Computations.

v. Critical region: reject H_0 if $F_{cal}\geq F_{tab}$

vi. Conclusion

Table 9.1 (Percentage points of the F-distribution)

At $\alpha = \begin{cases} 5\% \\ 1\% \\ 10\% \end{cases}$ Level of Significance

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11
1	161.4 4052 39.1	199.5 4999 49.5	215.7 5403 53.6	224.6 5625 55.8	230.2 5859 58.2	234.0 5859 58.2	236.8 5928 58.9	238.9 5981 59.4	240.5 6022 59.9	241.9 6056 60.2	243 6081 60.4
2	18.51 98.50 8.53	19.00 99.00 9.0	19.16 99.17 9.16	19.25 99.25 9.24	19.30 99.30 9.29	19.33 99.33 9.33	19.35 99.36 9.35	19.35 99.36 9.35	19.38 99.39 9.38	19.40 99.40 9.39	19.4 99.4 9.40
3	10.13 32.12 5.54	9.522 30.82 5.46	9.27 29.46 5.39	9.12 28.71 5.34	9.013 28.24 5.31	8.94 27.91 5.28	8.89 27.67 5.27	8.89 27.67 5.27	8.81 27.35 5.24	8.79 27.23 5.23	8.75 27.1 5.22
4	7.709 21.20 4.45	6.94 18.00 4.32	6.59 16.69 4.19	6.39 15.98 4.11	6.26 15.52 4.05	6.16 15.21 4.01	6.09 14.98 3.98	6.09 14.98 3.98	5.99 14.66 3.94	5.96 14.55 3.92	5.94 14.4 3.91
5	6.608 16.26 4.06	5.79 13.27 3.78	5.41 12.06 3.62	5.19 11.39 3.52	5.05 10.97 3.45	4.95 10.67 3.40	4.88 10.46 3.37	4.82 10.29 3.34	4.77 10.16 3.32	4.47 10.05 3.30	4.70 9.96 3.29
6	5.987 13.75 3.78	5.14 10.92 3.46	4.76 9.78 3.29	4.53 9.15 3.18	4.39 8.75 3.11	4.29 8.47 3.05	4.21 8.26 3.01	4.15 8.10 2.98	4.09 7.98 2.96	4.06 7.87 2.94	4.03 7.79 2.92
7	5.59 12.25 3.59	4.74 9.55 3.26	4.35 8.45 3.07	4.12 7.85 2.96	3.97 7.46 2.88	3.87 7.19 2.83	3.79 6.969 2.78	3.73 6.84 2.75	3.68 6.72 2.72	3.64 6.62 2.70	3.60 6.54 2.69
8	5.318 11.26 3.46	4.46 8.65 3.11	4.07 3.59 2.92	3.84 7.01 2.81	3.69 6.63 2.73	3.58 6.37 2.67	3.50 6.18 2.62	3.44 6.03 2.59	3.39 5.91 2.56	3.35 5.81 2.54	3.31 5.73 2.52
9	5.117 10.56 3.26	4.26 8.02 3.01	3.86 6.99 2.81	3.36 6.42 2.69	3.48 6.06 2.61	3.37 5.80 2.55	3.29 5.61 2.51	3.23 5.47 2.47	3.18 5.35 2.44	3.14 5.26 2.42	3.10 5.18 2.40

10	4.965 10.04 3.28	4.10 7.56 2.92	3.71 6.55 2.73	3.48 5.99 2.61	3.33 5.64 2.52	3.22 5.39 2.46	3.14 5.20 2.41	3.07 5.06 2.38	3.02 4.94 2.35	2.98 4.85 2.32	2.94 4.77 2.30
11	4.84 9.65 3.23	3.98 7.21 2.86	3.59 6.22 2.66	3.36 5.67 2.54	3.20 5.32 2.45	3.09 5.07 2.39	3.01 4.89 2.34	2.95 4.74 2.30	2.89 4.63 2.27	2.85 4.54 2.25	2.82 4.46 2.23
12	4.75 9.33 3.18	3.89 6.93 2.81	3.49 5.95 2.61	3.26 5.41 2.48	3.11 5.06 2.39	2.99 4.82 2.33	2.91 4.64 2.28	2.85 4.49 2.24	2.79 4.39 2.21	2.75 4.29 2.19	2.72 4.22 2.17
13	4.67 9.07 3.14	3.81 6.70 2.76	3.41 5.74 2.56	3.18 5.21 2.43	3.03 4.86 2.35	2.91 4.62 2.28	2.83 4.44 2.23	2.77 4.30 2.20	2.71 4.19 2.16	2.67 4.10 2.14	2.63 4.02 2.12
14	4.60 8.86 3.80	3.74 6.52 2.73	3.34 5.56 2.52	3.11 5.04 2.39	2.96 4.69 2.31	2.85 4.46 2.24	2.76 4.28 2.19	2.69 4.14 2.15	2.65 4.03 2.12	2.60 3.94 2.10	2.57 3.88 2.08
15	4.54 8.68 3.07	3.68 6.36 3.07	3.29 5.42 2.70	3.06 4.89 2.49	2.90 4.96 2.36	2.79 4.32 2.27	2.71 4.14 2.21	2.64 4.00 2.16	2.58 3.89 2.12	2.54 3.81 2.06	2.51 3.73 2.04
16	4.49 8.53 3.05	3.63 6.23 2.67	3.24 5.29 2.46	3.01 4.77 2.33	2.85 4.44 2.24	2.74 4.20 2.18	2.66 4.03 2.13	2.59 3.89 2.09	2.54 3.78 2.06	2.49 3.69 2.03	2.46 3.62 2.01
17	4.45 8.40 3.03	3.59 6.11 2.64	3.20 5.18 2.44	2.96 4.67 2.31	2.81 4.34 2.22	2.70 1.10 2.15	2.61 3.93 2.10	2.55 3.79 2.06	2.49 3.68 2.03	2.45 3.59 2.00	2.41 3.52 1.98
18	4.41 8.29 3.01	3.55 6.01 2.62	3.16 5.09 2.42	2.93 4.58 2.29	2.77 4.25 2.20	2.66 4.01 2.13	2.58 3.84 2.08	2.51 3.71 2.04	2.46 3.60 2.00	2.41 3.51 1.98	2.37 3.43 1.96
19	4.38 8.18 2.99	3.52 5.93 2.61	3.13 5.01 2.40	2.90 4.50 2.27	2.74 4.17 2.18	2.63 3.94 2.11	2.54 3.77 2.06	2.48 3.63 2.02	2.42 3.52 1.98	2.38 3.43 1.96	2.34 3.36 1.94
20	4.35 8.10 2.97	3.49 5.85 2.59	3.10 4.94 2.38	2.87 4.43 2.25	2.71 4.10 2.16	2.60 3.87 2.09	2.51 3.70 2.09	2.45 3.56 2.00	2.39 3.46 1.96	2.35 3.37 1.94	2.31 3.29 1.92
25	4.24 7.77 2.92	3.39 5.57 2.53	2.99 4.68 2.32	2.76 4.18 2.18	2.60 3.86 2.09	2.49 3.63 2.02	2.40 3.46 1.97	2.34 3.32 1.93	2.28 3.22 1.89	2.24 3.13 1.87	2.20 3.06 1.85

30	4.17 7.56 2.88	3.32 5.39 2.49	2.92 4.51 2.28	2.69 4.02 2.14	2.53 3.70 2.05	2.42 3.47 1.98	2.33 3.30 1.93	2.27 3.17 1.88	2.21 3.07 1.85	2.16 2.98 1.82	2.13 2.91 1.80
35	4.12 7.42 2.86	3.27 5.27 2.47	2.88 4.40 2.26	2.64 3.91 2.12	2.49 3.59 2.03	2.37 3.37 1.96	2.29 3.20 1.90	2.22 3.07 1.86	2.16 2.97 1.82	2.12 2.87 1.79	2.08 2.81 1.77
40	4.08 7.31 2.84	3.23 5.18 2.44	2.84 4.31 2.23	2.61 3.83 2.09	2.45 3.54 2.00	2.34 3.29 1.93	2.25 3.12 1.87	2.18 2.99 1.83	2.12 2.89 1.79	2.08 2.80 1.76	2.04 2.73 1.74
45	4.06 7.23 2.82	3.21 5.11 2.43	2.82 4.25 2.22	2.58 3.77 2.08	2.43 3.46 1.39	2.31 3.23 1.92	2.23 3.07 1.86	2.16 2.94 1.82	2.10 2.83 1.78	2.05 2.74 1.75	2.01 2.73 1.73
50	4.03 7.17 2.81	3.18 5.06 2.41	2.79 4.20 2.20	2.56 3.72 2.06	2.40 3.41 1.97	2.29 3.19 1.90	2.20 3.02 1.84	2.13 2.89 1.80	2.07 2.79 1.76	2.03 2.70 1.73	1.99 2.63 1.71
60	4.00 7.08 2.79	3.15 4.98 2.39	2.76 4.13 2.18	2.53 3.65 2.04	2.37 3.34 1.95	2.25 3.12 1.87	2.17 2.95 1.82	2.10 2.82 1.77	2.04 2.72 1.74	1.99 2.63 1.71	1.95 2.56 1.69
70	3.98 7.01 2.78	3.13 4.92 2.38	2.74 4.08 2.17	2.50 3.60 2.03	2.35 3.29 1.94	2.23 3.07 1.86	2.14 2.91 1.81	2.07 2.78 1.76	2.02 2.67 1.73	1.97 2.59 1.69	1.93 2.51 1.67
80	3.96 6.96 2.77	3.11 4.88 2.37	2.72 4.04 2.15	2.19 3.56 2.02	2.33 3.26 1.92	2.21 3.01 1.85	2.13 2.87 1.79	2.06 2.74 1.75	2.00 2.64 1.71	1.95 2.55 1.68	1.91 2.48 1.65
90	3.95 6.93 2.77	3.10 4.85 2.37	2.71 4.01 2.15	2.47 3.54 2.01	2.32 3.23 1.92	2.20 3.01 1.84	2.11 2.84 1.79	2.04 2.72 1.74	1.99 2.61 1.71	1.94 2.52 1.67	1.90 2.45 1.65
100	3.94 6.60 2.76	3.09 4.82 2.36	2.70 3.98 2.14	2.46 3.51 2.00	2.31 3.21 1.91	2.19 2.99 1.83	2.10 9.82 1.78	2.03 2.69 1.73	1.97 2.59 1.70	1.93 2.50 1.66	1.89 2.43 1.64

Example 9.1

The attention spans in minutes were randomly observed by a teacher from a group of 15 first year students during a morning reading period and are shown in the following table.

No breakfast	Light breakfast	Full breakfast
8	14	10
7	16	12
9	12	16
13	17	15
10	11	12

Do the data indicate that the average attention spans of the three treatments are equal? Use $\alpha = 0.05$

Solution:

i. H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: Not all means are equal

ii. $\alpha = 0.05$

iii. Test statistic to be used is

$$F_c = \frac{MST_r}{MSE} \sim F\text{-distribution with } v_1 = k-1, v_2 = n-k \text{ d.f}$$

iv. Calculation:

	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	Total
	8	64	14	196	10	100	
	7	49	16	256	12	144	
	9	81	12	144	16	256	
	13	169	17	289	15	225	
	10	100	11	121	12	144	
$T_{.j}$	47	↓	70	↓	65	↓	$182 = T_{..}$
$T^2_{.j}$	2209		4900		4225		$11334 = \sum_{j=1}^{3} T_j^2$
$\sum_{i=1}^{5} x_{.j}^2$		463		1006		869	$\rightarrow 2338 = \sum_{i=1}^{5}\sum_{j=1}^{3} x_{ij}^2$

$$C.F = \frac{T^2_{..}}{n} = \frac{(182)^2}{15} = 2208.2667$$

$$TSS = \sum_{j=1}^{3}\sum_{i=1}^{5} x_{ij}^2 - C.F = 2338 - 2208.2667 = 129.7333$$

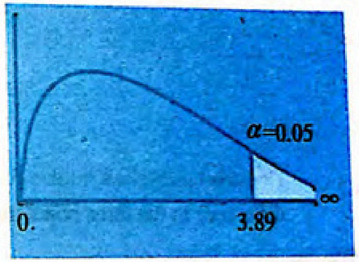
$$T_rSS = \frac{\sum_{j=1}^{3} T_{.j}^2}{r} - C.F = \frac{11334}{5} - 2208.2667 = 58.5333$$

$$ESS = TSS - T_rSS = 129.7333 - 58.5333 = 71.2$$

ANOVA table

S.O.V	d.f	S.S	M.S	F
Treatments	$3-1=2$	58.5333	$\frac{58.5333}{2}=29.27$ MSTr	$\frac{MSTr}{MSE}=\frac{29.27}{5.93}=4.93$
Error	$15-3=12$	71.2	$\frac{71.2}{12}=5.93$MSE	—
Total	$15-1=14$	129.7333	—	—

v. Critical region



Reject H_0 if $F_c \geq F_{tab}$ whereas from F-table 9.1

$$F_{tab} = F_{\alpha,(v_1,v_2)} = F_{0.05,(2,12)} = 3.89$$

vi. Conclusion:

Since $F_c = 4.93$ falls in the rejection region, therefore, we reject H_0 and conclude that average attention span of students is different for the three treatments (types of breakfast).

9.3.7 One way ANOVA (unequal sample sizes)

Practically sample sizes may not always be equal. One way ANOVA for equal sample sizes will still be used with minor changes in the sum of square formulas as follows:

$$n = r_1 + r_2 + \ldots\ldots + r_k = \sum_{j=1}^{k} r_j$$

$$TSS = \sum_{j=1}^{k}\sum_{i=1}^{rj} x^2_{ij} - C.F$$

$$T_rSS = \sum_{j=1}^{k} \frac{T^2_{.j}}{r_j} - C.F$$

Example 9.2

The following observations are collected using a completely randomized design:

Sample1	Sample 2	Sample3
3	4	2
2	3	0
4	5	2
3	2	1
2	5	

Construct an ANOVA table for the data and determine whether there is a difference in the three population means. Use $\alpha = 0.01$

Solution:

i. H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: Not all means are equal

ii. $\alpha = 0.01$

iii. Test statistic to be used is

$$F = \frac{MSTr}{MSE} \text{ with } (v_1 = k-1, v_2 = n-k) \text{ d.f}$$

iv. Calculation:

	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	Total
	3	9	4	16	2	4	
	2	4	3	9	0	0	
	4	16	5	25	2	4	
	3	9	2	4	1	1	
	2	4	5	25		↓	
$T_{.j}$	14	↓	19	↓	5	↓	$38 = T_{..}$
$T^2_{.j}$	196		361		25		
$\sum_{i=1}^{5} x^2_{.j}$		42		79		9	$130 = \sum_j\sum_i x^2_{ij}$

$n = r_1 + r_2 + r_3 = 5 + 5 + 4 = 14$

$$C.F = \frac{T_{..}^2}{n} = \frac{(38)^2}{14} = 103.1429$$

$$TSS = \sum_{j=1}^{3}\sum_{i=1}^{r_j} x^2_{ij} - C.F = 130 - 103.1429 = 26.8571$$

$$T_rSS = \sum_{j=1}^{3} \frac{T^2_{.j}}{r_j} - C.F$$

$$= \left[\frac{T^2_{.1}}{r_1} + \frac{T^2_{.2}}{r_2} + \frac{T^2_{.3}}{r_3}\right] - C.F$$

$$=\left[\frac{196}{5}+\frac{361}{5}+\frac{25}{4}\right]-C.F$$

$$=[39.2+72.2+6.25]-103.1429=14.5071$$

$$ESS = TSS - T_rSS$$

$$=26.8571-14.5071=12.35$$

Put all these values in ANOVA table we have

ANOVA table

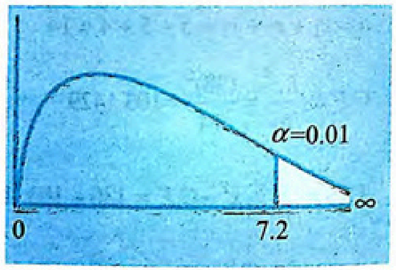
S.O.V	df	S.S	M.S	F
Treatments	2	14.5071	7.2536	6.461
Error	11	12.35	1.1227	—
Total	13	26.8571	—	—

v. Critical region

Reject H_0 if $F_c \geq F_{tab}$ whereas,

$F_{tab} = F_{\alpha,(\nu_1,\nu_2)} = F_{0.01,(2,11)} = 7.2$

(From F- table 9.1)



vi. Conclusion:

Since our computed value of $F_c = 6.461$ lies in the acceptance region, therefore, we accept H_0.

Key points

- A plan of experiment in which all situations except that of treatments are kept under control as much as possible, is known as design of experiment.
- Anything whose effect is measured and is compared with others is called treatment.
- The total material or objects on which the experiment is done is known as experimental material.
- The smallest division of an experimental material to which a treatment is applied is called experimental unit.
- The results obtained from the experimental units are called yield or responses.
- A group of homogenous experimental units e.g. land of same fertility, students of same age, weight, I.Q, etc. to which all treatment are assigned at random is called a block.
- Uniformity trial is used only for making blocks.
- All factors which are not in the control of a researcher are called uncontrolled or extraneous factors.
- The error which arises due to the extraneous factors is called experimental error.
- Design of experiment is actually a strategy for controlling the experimental error in order to bring out the real difference among the treatments.
- The allocation of treatments to experimental units in such a manner that an experimental unit has equal chance of receiving any of the treatments is called randomization.
- Repetition of a treatment on a number of experimental units in an experiment is known as replication of the treatment.
- Local control also called error control reduces the experimental error.
- Experimental design in which the treatments are allocated randomly to the experimental units is called completely randomized design
- ANOVA is a technique which split the total variance into meaningful component variances; each gives an estimate of the population variance. The ratio of two component variances is distributed as F with corresponding degrees of freedom.

Exercise

9.1 Read the following statements carefully and indicate which statement is true or false.

i. F-distribution is a ratio of sample variances.
ii. ANOVA is a useful technique for testing hypothesis about several means.
iii. In ANOVA sample sizes must always be equal.
iv. In ANOVA it is assumed that the populations from which samples are drawn are normally distributed
v. F-test is used in ANOVA to test the null hypothesis about several means.
vi. The range of F-distribution is from 0 to ∞.
vii. The plan of an experiment which controls all factors, as much as possible except the treatment is called design of experiment.
viii. Larger the experimental error, more efficient is the design.
ix. A completely randomized design is used when all experimental units are homogeneous.
x. Missing observation in a CRD creates a serious problem.

9.2. Fill in the blanks.

i. A subject receiving a treatment in an experiment is called ________.
ii The allocation of treatments to experimental units with equal probability is known as ________.
iii. The number of times a treatment is repeated in an experiment is called its ________.
iv. Experimental error is the error caused by________.
v. Smaller the experimental error, more ________is the design.
vi. ANOVA splits the total variance is into ________.
vii. Missing observation in a completely randomized design creates ________.
viii. Error sum of squares ________ be negative.
ix. The average performance of a treatment is better reflected through ________.
x. The designs of experiments were originated mainly for ________.

9.3 Choose the correct answer.

i. An experimental design is
(a) An architect (b) a map
(c) a plan of experiment (d) all of the above

ii. Randomization is a process in which the treatments are allocated to the experimental units:
(a) In a sequence (b) at the will of the investigator
(c) With equal probability (d) with unequal probability

iii. Replication in an experiment means:
(a) The number of times a treatment occurs in an experiment
(b) The numbers of blocks
(c) Total number of treatments
(d) The reduction of blocks

iv. Local control is a method to maintain
(a) Homogeneity among blocks (b) Homogeneity within blocks
(c) both (a) and (b) (d) all of the above

v. Experimental error is due to
(a) Variation in treatment effects (b) extraneous factors
(c) Experimenter's mistake (d) lack of experience

vi. Completely randomized designs are mostly used in
(a) Pot experiments (b) Experiments on animals
(c) Field experiments (d) all of the above

vii. An experimental unit in a research work is
(a) A patient (b) A field plots
(c) An animal (d) All of the above

viii. Factors fertilizer, date of sowing and breeds are called

(a) Replicates (b) experimental unit

(c) Treatments (d) All of the above

ix. Local control in the field is maintained through

(a) Natural factors (b) Randomization

(c) Replication (d) Uniformity trials

x. Randomization in an experiment helps the researcher to eliminate

(a) Dependence among observations

(b) Systematic influences

(c) Human biases

(d) all of the above

9.4 Describe the design of an experiment in your own words.

9.5 Define the following.

i. Experimental material

ii. Experimental unit

iii. Treatment

iv. Uniformity trial

v. Block

9.6 What do you understand by randomization, replication and local control in experimental design?

9.7 i. What is meant by experimental design?

ii. What are the basic principles of design of experiment?

9.8 Discuss and define (a) Extraneous factor (b) experimental error.

9.9 Discuss the need and utility of planning a statistical experiment.

9.10 What is a completely randomized design? What are the merits and demerits of a completely randomized design?

9.11 Explain the experimental layout for a completely randomized design using 3 treatment and 15 experiment plots.

9.12 Write a short note on analysis of variance.

9.13 What is meant by analysis of variance? What are the assumptions under which this technique is applied?

9.14 What do you understand by?

i. Variance among samples

ii. Variance within samples

iii. Total variation

9.15 Given the data below, perform the analysis of variance and test the hypothesis that the means of the three populations are equal. Let $\alpha = 0.05$

X_1	X_2	X_3
13	18	17
14	19	20
16	12	8
17	15	11

9.16 Four salesmen were posted in different areas by a company. The numbers of units of commodity "x" sold by them are as follow:

A	28	23	20	29
B	30	32	25	21
C	35	28	23	18
D	19	21	15	25

Is there a significant difference in the performance of these salesmen?

9.17 The following data gives the figures of production of rice of three varieties A, B, C of rice shown in 12 plots

A	18	15	24	23
B	19	23	24	18
C	16	19	31	22

Carry out the analysis of variance and test 5% level of significance that is there a significant difference among varieties?

9.18 The three samples below have been obtained from normal populations with equal variances. Test the hypothesis at 5% level of significance that the population means are equal.

X_1	X_2	X_3
10	5	9
8	7	12
7	10	13
14	9	12
11	9	14

9.19 A test was given to 5 students chosen at random from the first year statistics class each of the three colleges in Peshawar. Their scores were found as follow.

A	50	80	90	70	60
B	40	50	70	40	50
C	70	60	60	50	60

Perform analysis of variance and show if there is any significant difference among the score of students in the three colleges. Use $\alpha = 0.05$

9.20 Test the hypothesis that no differences exist among the four treatments at $\alpha = 0.05$.

Sample 1	Sample 2	Sample 3	Sample 4
5	3	4	8
4	6	11	18
4	4	8	14
11	6	6	27

9.21 Given the following data obtained from a completely randomized design with four treatments;

X_1	X_2	X_3	X_4
12.4	14.4	10.2	6.1
20.9	9.0	13.2	5.8
10.1	23.7	5.1	4.8
4.2			1.5

Analyse the given data and draw conclusion about the equality of treatment effects. Use $\alpha = 0.05$

9.22 The following data were obtained by using completely randomized design

Sample 1	6	8	10	8
Sample 2	8	10	9	
Sample 3	10	8		
Sample 4	9	10	7	8
Sample 5	8	10	12	

Construct an ANOVA table for the data. Test $\alpha = 0.05$ that five population means are equal.

9.23 Compare the following random samples.

Sample 1: 17 19 4 9 10 11

Sample 2: 12 15 6 8 10 11 12

Sample 3: 20 23 9 13 15

Perform ANOVA and test the hypothesis at $\alpha = 0.05$ that the samples came from populations having same means.

9.24 The results shown in the following table were obtained through completely randomized design.

A	2	1	0	2	
B	2	5	3	4	5
C	3	2	3	2	4

Test at 1% level of significance that there is no difference in the means of the three populations.

ANSWERS

Exercise-1

1.1	(i)	T	(ii)	F	(iii)	F	(iv)	T	(v)	T
	(vi)	F	(vii)	T	(viii)	T	(ix)	T	(x)	F
1.2	(i)	impossible	(ii)	sample space	(iii)	permutations				
	(iv) 0 to 1		(v)	gambling	(vi)	addition				
	(vii)	$P(A)+P(B)$	(viii)	independent	(ix)	$1-P(A)$	(x)	$\frac{6}{36}$		
1.3	(i)	c	(ii)	d	(iii)	c	(iv)	d	(v)	b
	(vi)	c	(vii)	d	(viii)	b	(ix)	c	(x)	b
1.5	(i)	40320	(ii) 6375600		(iii)	10626	(iv)	3876		
	(v)	167960								
1.7	40320, 720									
1.8	60									
1.9	(i)	4989600	(ii)	15135120	(iii)	19958400				
	(iv)	4989600	(v)	50400						
1.10	330									
1.15	0.50									
1.16	$\frac{5}{36}$									
1.17	(i)	$\frac{1}{8}$	(ii)	$\frac{3}{8}$	(iii)	$\frac{7}{8}$				
1.18	0.0045									
1.19	(i)	$\frac{13}{52}$	(ii)	$\frac{26}{52}$	(iii)	$\frac{4}{52}$				
1.20	(i)	$\frac{1}{3}$	(ii)	0.133						

1.21 (b) 0.92

1.22 $\frac{8}{36}$

1.24 $\frac{1}{4}$

-1.25 (i) 0.0059 (ii) 0.0045

1.26 (b) $\frac{1}{3}$

1.27 (i) 0.12 (ii) 0.6 (iii) 0.68

1.28 $\frac{3}{4}$

1.29 (i) $\frac{1}{210}$ (ii) $\frac{209}{210}$

1.30 (a) $p=\frac{1}{2}, d=1$ (b) $p=\frac{1}{4}, d=\frac{1}{3}$

Exercise-2

2.1 (i) T (ii) F (iii) T (iv) F (v) T
(vi) F (vii) T (viii) T (ix) F (x) T

2.2 (i) finite (ii) unity (iii) $\frac{5}{21}$ (iv) random variable
(v) discrete & continuous (vi) joint distribution (vii) one
(viii) $E(X)E(Y)$ (ix) S.D(X)+S.D(Y) (x) $\frac{2}{3}$

2.3 (i) b (ii) b (iii) b (iv) c (v) d
(vi) c (vii) c (viii) b (ix) a (x) d

2.6 (i) continuous (ii) continuous (iii) discrete
(iv) continuous (v) continuous (vi) discrete
(vii) continuous (viii) continuous (ix) discrete
(x) discrete

2.8

X	0	1	2
$p(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

2.9

X	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

2.10 (i) 0.3 (ii) 0.6 (iii) 0.8

2.11

X	−3	−1	1	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

2.12

X	−5	−4	−3	−2	−1	0	1	2	3	4	5
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.13 No, because sum of probabilities is not equal to one

2.14 (i) $\frac{3}{8}$ (ii) $\frac{1}{2}$ (iii) $\frac{3}{4}$

2.15 (b) (i) 3 (ii) 0.125 (iii) 0.019

2.17 (i) 0.5 (ii) 3.5 (iii) 6

2.18 0.55, 1.35, 1.16

2.19 (i) 2.6 (ii) 1.34 (iii) 1.16

2.20 (i) 20 (ii) $\frac{2}{3}$ (iii) $\frac{2}{63}$ (iv) 0.178

2.21 Mean $=\frac{4}{9}$ Variance $=\frac{13}{162}$, S.D $= 0.2833$

2.22 $E(X)=2$, $E(Y)=0.5$, $E(X+Y)=2.5$, $E(XY)=1$

2.23 (i) Yes, X & Y are independent (ii) 35, 36.8, 71.8, 1288

2.24 (i) 0.41 (ii) 0.24 (iii) 0.65 (iv) 0.65

Exercise-3

3.1 (i) F (ii) T (iii) F (iv) T (v) T
(vi) F (vii) T (viii) T (ix) T (x) T

3.2 (i) two (ii) one (iii) n (iv) n and p
(v) np, npq (vi) n (vii) 0 and 1 (viii) random numbers
(ix) independent (x) varies/changes

3.3 (i) b (ii) c (iii) c (iv) d (v) b
(vi) c (vii) d (viii) a (ix) b (x) d

3.10 (i) 0 (ii) 0.29 (iii) 0 (iv) 0.936 (v) 0.352

3.11

X	0	1	2	3	4	5
p(x)	0.095	0.286	0.343	0.206	0.062	0.007

3.12 (i) 0.656 (ii) 0.891

3.13 (i) 0.07776 (ii) 0.2592 (iii) 0.92224

3.14 (i) 0.60501 (ii) 0.39499

3.15 0.227

3.16 0.62343

3.17 0.0005

3.19 (i) 40 bolts (ii) 6

3.20 $n = 12$, $p = 0.25$

3.21 $n = 41$, $p = 0.302$

3.22 No, because q =1.8 which cannot be greater than 1.

3.23 (i) 16 (ii) 0.9576

3.24 mean = 3, variance =1.6, S.D =1.25

3.25 (b) 52.08, 41.67, 12.5, 1.67, 0.08

3.26 mean = 5.42 variance = 2.4824

3.27 11.04 64.07 154.94 199.83 144.98 56.10 9.04

3.30

x	1	2	3	4	5
p(x)	1/66	2/11	5/11	10/33	1/22

Mean =3.187, Variance =0.706

3.31 0.275

3.32 0.213

Exercise-4

4.1 (i) F (ii) F (iii) T (iv) T (v) T
(vi) T (vii) F (viii) T (ix) F (x) T

4.2 (i) 4 (ii) Z~N(0,1) (iii) $X = \mu$ (iv) $\frac{4}{5}\sigma$
(v) zero (vi) normal (vii) 10 (viii) mean = μ
(ix) Mean = 0 (x) 95.4

4.3 (i) b (ii) a (iii) b (iv) b (v) c
(vi) c (vii) c (viii) b (ix) d (x) a

4.8 (i) 0.0548 (ii) 0.044 (iii) 0.103 (iv) 0.475 (v) 0.950
(vi) 0.682 (vii) 0.006

4.9 (i) 0.5987 (ii) 0.2119 (iii) 0.0668 (iv) 0.7734

4.10 (i) 0.0043 (ii) 0.0735 (iii) 0.0103 (iv) 0.1196 (v) 0.9777
(vi) 0.9850

4.11 (i) z =2.58 (ii) z =1.24 (iii) z =−1.75

4.12 (i) z =0.84 (ii) z =−0.25 (iii) z =−1.04 and +1.04

4.13 (i) 0.1587 (ii) 0.0228 (iii) 0.3721

4.14 (i) 0.9452 (ii) 0.0548 (iii) 0.3085 (iv) 0.6915 (v) 0.6898
(vi) 0.1832

4.15 0.1359

4.16 0.2515

4.17 (i) 0.9798 = 97.98% (ii) 0.3531 = 35.31%
(iii) 0.0559 = 5.59%

4.18 (ii) 0.0548 (ii) 0.4772

4.19 (i) 0.3446, 173 (ii) 0.3811, 191

4.20 (i) 758 (ii) 629 (iii) 206

4.21 0.9554 = 95.54%

4.22 0.1587, 159

4.23 (i) 0.6826 (ii) 0.0317

4.24 0.4706

4.25 (i) 0.0167 (ii) 0.3885 (iii) 0.0775

4.26 (i) 92.22 (ii) 4.44

4.27 M.D = 4

4.28 $\sigma = 20$

Exercise-5

5.1 (i) T (ii) T (iii) F (iv) F (v) F
(vi) T (vii) F (viii) T (ix) T (x) F

5.2
(i) finite (ii) infinite (iii) census (iv) sampling error
(v) standard error (vi) sampling frame (vii) more than once
(viii) sampling fraction (ix) fpc (x) $n < 5\%N$

5.3 (i) a (ii) b (iii) b (iv) d (v) d
(vi) c (vii) c (viii) c (ix) d (x) d

5.16 (b) $n_1=8$ $n_2=9$ $n_3=11$ $n_4=12$

5.17 $n_1=75$ $n_2=50$ $n_3=25$ $n_4=40$ $n_5=10$

5.19 (b) $\mu_{\bar{X}} = 6, \sigma_{\bar{X}} = 1$

5.20 $\mu_{\bar{X}} = 3.8,$ $\sigma_{\bar{X}} = 0.98$

5.21 $\mu_{\bar{X}} = 50$ $\sigma_{\bar{X}} = 6$

5.22 (i) z = 1 (ii) z = 2

5.23 (i) $\mu = 7,\ \sigma^2 = 5.2$ (ii) 0.87 (iii) 1.73

5.24 (i) $\mu_{\bar{X}} = \mu = 9$ (ii) $\sigma = 6.48,$ $\sigma_{\bar{X}} = 2.9$

5.26 $\mu_{\bar{X}_1-\bar{X}_2} = 1$ $\sigma_{\bar{X}_1-\bar{X}_2} = 1.732$

5.27 (b) $\mu_{\bar{X}_1-\bar{X}_2} = 2$ $\sigma^2_{\bar{X}_1-\bar{X}_2} = 2.1667$

5.29 $\mu_{\hat{p}} = 0.5,$ $\sigma^2_{\hat{p}} = 0.05$

5.30 $\hat{p} = \frac{3}{7}$ $\mu_{\hat{p}} = \frac{3}{7}$ $\sigma^2_{\hat{p}} = \frac{5}{49}$

5.31 $E(\hat{p}_1 - \hat{p}_2) = 0$ $V(\hat{p}_1 - \hat{p}_2) = \frac{1}{9}$

Exercise-6

6.1 (i) F (ii) T (iii) F (iv) F (v) T
(vi) F (vii) T (viii) T (ix) T (x) F

6.2 (i) random variable (ii) estimate (iii) estimator
(iv) point (v) unbiased (vi) efficient
(vii) shortest (viii) increasing (ix) two
(x) confidence coefficient

6.3 (i) b (ii) c (iii) b (iv) c (v) a
(vi) d (vii) d (viii) b (ix) c (x) a

6.9 (i) $\overline{X}=5.9$ (ii) $s^2 = 5.8178$ (iii) $s_{\overline{X}} = 0.7667$

6.10 (i) $E(\overline{X})=\mu=14$ (ii) $E(s^2)=\sigma^2=9.333$

6.14 $(9.84 < \mu < 12.16)$

6.15 $(0.774 < \mu < 0.826)$

6.16 $(788.24 < \mu < 811.76)$

6.17 $(-2.27 < \mu < 2.67)$

6.18 $(63.412 < \mu < 64.588)$

6.19 $(31.59 < \mu < 54.93)$

6.20 $(2.4 < \mu_1 - \mu_2 < 7.6)$

6.21 $(0.55 < \mu_2 - \mu_1 < 1.25)$

6.23 $(4.3371 < \mu < 4.4229)$

6.24 $(33.52 < \mu < 35.28)$

6.25 $(-15.3 < \mu_A - \mu_B < 1.3)$

6.26 $(0.45 < P < 0.75)$

6.27 $(0.62 < P < 0.74)$

6.28 $(-0.206 < P_1 - P_2 < -0.120)$

Exercise-7

7.1 (i) F (ii) T (iii) F (iv) F (v) T
(vi) T (vii) T (viii) F (ix) T (x) F

7.2 (i) assertion (ii) null (iii) alternative
(iv) composite (v) first, second (vi) level of significance
(vii) 0, 1 (viii) rejection region (ix) rule
(x) degrees of freedom

7.3 (i) c (ii) a (iii) b (iv) a (v) c
(vi) a (vii) c (viii) b (ix) b (x) b

7.10 $z = 2.61$, reject H_0

7.11 $z = 2.4$, reject H_0

7.12 $z = -2.88$, reject H_0

7.13 $z = 2.33$, reject H_0

7.14 $z = -1.93$, reject H_0

7.16 $t = 0.316$, accept H_0

7.17 $t = 1.844$, accept H_0

7.19 $z = 4.22$, reject H_0

7.20 $z = -1.334$, accept H_0

7.22 $t = 3.33$, reject H_0

7.23 $t = 3.05$, reject H_0

7.25 $z = 2.19$, reject H_0

7.26 $z = -0.73$, accept H_0

7.27 $z = -2.1004$, accept H_0

7.29 $z = -3.1623$, reject H_0

7.30 $z = 9.6825$, reject H_0

Exercise-8

8.1 (i) F (ii) T (iii) T (iv) F (v) T
(vi) F (vii) T (viii) F (ix) T (x) F

8.2 (i) dichotomous (ii) order (iii) second (iv) zero
(v) $\frac{(A)(B)}{n}$ (vi) zero (vii) not same
(viii) -1 to $+1$ (ix) equal (x) 20

8.3	(i)	c	(ii)	b	(iii)	d	(iv)	d	(v)	b
	(vi)	c	(vii)	b	(viii)	c	(ix)	c	(x)	b

8.9 Q =–0.76, the association between extravagant fathers and extravagant sons is negative.

8.10 Q = 0, attributes A and B are independent.

8.11 Q =0.46, there is positive association.

8.12 Q = 0.9, there is high degree of positive association between intelligent fathers and sons.

8.13 Q = −0.57, it means that vaccine and small-pox are negatively associated i.e. vaccine prevents the attack of small-pox.

8.14	χ^2= 48.24,	reject H_0
8.15	χ^2= 8.89,	reject H_0
8.16	χ^2= 60.183,	reject H_0
8.18	χ^2= 598.51,	reject H_0
8.19	χ^2= 29.79,	reject H_0
8.20	χ^2= 763.76,	reject H_0
8.22	χ^2= 3.279,	C =0.067
8.23	χ^2= 1048.5,	C =0.3681

Exercise-9

9.1	(i)	T	(ii)	T	(iii)	F	(iv)	T	(v)	T
	(vi)	T	(vii)	T	(viii)	F	(ix)	T	(x)	F

9.2	(i)	experimental unit	(ii)	randomization	(iii)	replication
	(iv)	extraneous factors	(v)	efficient		
	(vi)	component variances	(vii)	no problem		
	(viii)	can never	(ix)	replication		
	(x)	field experiments				

9.3	(i)	c	(ii)	c	(iii)	a	(iv)	b	(v)	b
	(vi)	a	(vii)	d	(viii)	c	(ix)	d	(x)	d

9.16	F_c = 0.277	accept H_0
9.17	F_c = 1.37	accept H_0
9.18	F_c = 0.055	accept H_0
9.19	F_c = 4	reject H_0
9.20	F_c = 3.333	accept H_0
9.21	F_c = 5.56	reject H_0
9.22	F_c = 2.56	accept H_0
9.23	F_c = 0.82	accept H_0
9.24	F_c = 2.10	accept H_0
9.25	F_c = 6.46	accept H_0

Glossary

Addition rule: Rule for determining the probability that, on a single trial, either event *A* occurs, or event *B* occurs, or they both occur.

Alternative hypothesis: Statement that is equivalent to the negation of the null hypothesis.

Analysis of variance: Method of analysing population variances in order to test hypotheses about means of populations.

Binomial experiment: Experiment with a fixed number of independent trials, where each outcome falls into exactly one of two categories.

Block: A group of subjects that is similar in the ways that might affect the outcome of an experiment.

Census: Collection of data from every element in a population.

Central limit theorem: Theorem stating that sample means tend to be normally distributed.

Chi-square distribution: A continuous probability distribution.

Classical approach to probability: Approach in which the probability of an event is determined by dividing the number of ways the event can occur by the total number of possible outcomes.

Combinations rule: Rule for determining the number of different combinations of selected items.

Complement of an event: All outcomes in which the original event does not occur.

Completely randomized design: An experiment where by each element is given the same chance of belonging to the different categories or treatments.

Compound event: Combination of simple events.

Conditional probability: The probability of an event, given that some other event has already occurred.

Confidence coefficient: Probability that a population parameter is contained within a particular confidence interval; also called confidence level or degree of confidence.

Confidence interval limits: Two numbers that are used as the high and low boundaries of a confidence interval.

Confidence interval: Range of values used to estimate some population parameter with a specific confidence level; also called an interval estimate.

Confidence level: Probability that a population parameter is contained within a particular confidence interval.

Contingency table: Table of observed frequencies where the rows correspond to one variable of classification and the columns correspond to another variable of classification; also called a two-way table.

Continuity correction: Adjustment made when a discrete random variable is being approximated by a continuous random variable.

Continuous data: Data resulting from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps.

Continuous random variable: A random variable with infinite values that can be associated with points on a continuous line interval.

Critical region: The set of all values of the test statistic that would cause rejection of the null hypothesis.

Critical value: Value separating the critical region from the values of the test statistic that would not lead to rejection of the null hypothesis.

Degree of confidence: Probability that a population parameter is contained within a particular confidence interval; also called level of confidence.

Degrees of freedom: Number of values that are free to vary after certain restrictions has been imposed on all values.

Dependent events: Events for which the occurrence of any one event affects the probabilities of the occurrences of the other events.

Dependent sample: Sample whose values are related to the values in another sample.

Discrete random variable: Random variable with either a finite number of values or a countable number of values.

Disjoint events: Events that cannot occur simultaneously.

Efficiency: It is a criterion for selection of an efficient estimator.

Estimate: Specific value or range of values used to approximate some population parameter.

Estimator: Sample statistic (such as the sample mean) used to approximate a population parameter.

Event: The collection of favorable outcomes to a happening from the sample space.

Expected value: For a discrete random variable, the mean value of the outcomes

Experiment: Application of some treatment followed by observation of its effects on the subjects.

Experimental units: Subjects in an experiment.

Factorial rule: Rule stating that n different items can be arranged $n!$ different ways.

Finite population correction factor: Factor for correcting the standard error of the mean when a sample size exceeds 5% of the size of a finite population.

Fundamental counting rule: Rule stating that, for a sequence of two events in which the first event can occurs m ways and the second can occur n ways, the events together can occur a total of mn ways.

Hypothesis: Statement or claim about some property of a population.

Hypothesis test: Method for testing claims made about populations; also called test of significance.

Independent events: Events for which the occurrence of any one of the events does not affect the probabilities of the occurrences of the other events.

Inferential statistics: Collection of methods that help make decisions about a population based on sample results.

Interval estimate: Range of values used to estimate some population parameter with a specific level of confidence; also called a confidence interval.

Left-tailed test: Hypothesis test in which the critical region is located in the extreme left area of the probability distribution.

Level of confidence: Probability that a population parameter is contained within a particular confidence interval; also called degree of confidence.

Multiplication rule: Rule for determining the probability that event A will occur on one trial and event B will occur on a second trial.

Mutually exclusive events: Events that cannot occur simultaneously.

Non sampling errors: Errors from external factors not related to sampling.

Null hypothesis: Claim made about some population characteristic, usually involving the case of no difference.

Odds against: Ratio of the probability of an event not occurring to the event occurring, usually expressed in the form of a: b where a and b are integers having no common factors.

Odds in favour: Ratio of the probability of an event occurring to the event not occurring, usually expressed as the ratio of two integers with no common factors.

One-way analysis of variance: Analysis of variance involving data classified into groups according to a single criterion only.

Paired samples: Two sample which are dependent in the sense, that the data values are matched by pair.

Parameter: A summary measure calculated for population data.

Point estimate: Single value that serves as an estimate of a population parameter.

Pooled estimate of variance: Estimate of the variance that is common to two populations, found by computing a weighted average of the two sample variances.

Population or target population: The collection of all elements whose characteristics are being studied.

Probability distribution: Collection of values of a random variable along with their corresponding probabilities.

Probability: Measure of the likelihood that a given event will occur expressed as a number between 0 and 1.

Qualitative or categorical variable: A variable that cannot assume numerical values but is classified into two or more categories.

Random sample: Sample selected in a way that allows every member of the population to have the same chance of being chosen.

Random selection: Selection of sample elements in such a way that all elements available for selection have the same chance of being selected.

Random variable: Variable (typically represented by X) that has a single numerical value (determined by chance) for each outcome of an experiment.

Representative sample: A sample that contains the same characteristics as the corresponding population.

Right-tailed test: Hypothesis test in which the critical region is located in the extreme right area of the probability distribution.

Sample: A portion of the population of interest.

Sample size: Number of items in a sample.

Sample space: Set of all possible outcomes or events in an experiment that cannot be further broken down.

Sample survey: A survey that includes elements of a sample.

Sampling distribution of proportion: The probability distribution of sample proportions, with all samples having the same sample size n.

Sampling distribution of sample means: Distribution of the sample means that is obtained when we repeatedly draw samples of the same size from the same population.

Sampling error: Difference between a sample result and the true population result; results from chance sample fluctuations.

Sampling variability: Variations of a statistic in different samples.

Significance level: Probability of making a type I error when conducting a hypothesis test.

Simple event: Experimental outcome that cannot be further broken down.

Simple random sample: Sample of a particular size selected so that every possible sample of the same size has the same chance of being chosen.

Standard normal distribution: Normal distribution with a mean of 0 and a standard deviation equal to 1

Standard score: Number of standard deviations that a given value is above or below the mean; also called z-score.

Statistic: A summary measure calculated for sample data.

Stratified sampling: The sampling method in which samples are drawn from each stratum.

Subjective probability: Guess or estimate of a probability based on knowledge of relevant circumstances.

Survey: Collection of data on the elements of a population or sample.

Systematic sampling: Sampling in which every k^{th} element is selected.

t-distribution: Bell-shaped distribution usually associated with sample data from a population with an unknown standard deviation.

Test of independence: Test of the null hypothesis that for a contingency table, the row variable and column variable are not related.

Test statistic: Sample statistic based on the sample data; used in making the decision about rejection of the null hypothesis.

Treatment: Property or characteristic that allows us to distinguish the different populations from one another; used in analysis of variance.

Tree diagram: Graphical depiction of the different possible outcomes in a compound event.

Two-tailed test: Hypothesis test in which the critical region is divided between the left and right extreme areas of the probability distribution.

Type I error: Rejecting the null hypothesis when it is true.

Type II error: Accept the null hypothesis when it is false.

Unbiased estimator: Sample statistic that tends to target the population parameter that it is used to estimate.

Uniform distribution: Probability distribution in which every value of the random variable is equally likely.

Variable: A characteristic under study or investigation that assumes different values for different elements.

Variance between samples: In analysis of variance, the variation among the different samples.

Variation within samples: In analysis of variance, the variation that is due to chance.

Z-score: Number of standard deviations that a given value is above or below the mean.

INDEX

A
Alternative hypothesis 233
Acceptance region 235
Attributes 267
Association of attributes 270
Analysis of variance 298
ANOVA table 301

B
Bivariate probability distribution 56
Bivariate probability function 57
Bernoulli trial 79
Binomial experiment 82
Binomial probability mass function 82
Binomial frequency distribution 93
Bias 156
Block 294

C
Combination 04
Compound event 07
Complementary event 07
Conditional probability 21
Continuous random variable 32
Continuous uniform distribution 117
Census 154
Confidence interval 209
Confidence limits 209
Confidence region 209
Confidence coefficient 209
Confidence level 209
Composite hypothesis 233
Class Frequency 268
Classification of attributes 268
Contingency table 275
Coefficient of contingency 282
Completely randomized design 296

D
Dependent events 07
Discrete random variable 32
Discrete uniform distribution 71
Degrees of freedom 213
Dichotomy classification 268

E
Event 06
Equally likely Events 07
Exhaustive events 07
Estimation 194
Estimator 195
Estimate 195
Efficiency 204
Experimental design 293
Experimental material 293
Experimental unit 293
Extraneous factor 294
Experimental error 294

F
Factorial 02
Finite population 152

G

Goldfish bowl method 159

H

Hypergeometric experiment 98
Hypergeometric probability distribution 98
Hypothesis testing 232

I

Impossible event 06
Independent events 08
Infinite population 152
Interval estimation 209

J

Joint distribution 56

L

Level of significance 235
Length of confidence interval 209
Local control 296

M

Mutually exclusive events 07
Mathematical expectation 41
Marginal probability distribution 57
Multiway classification 275

N

Normal distribution 120
Non-Sampling error 156
Non probability sampling methods 158
Null hypothesis 233
Negative attributes 267

O

Outcome 06
Odds 24
One tailed test 236
One way classification 268
Order of classes 269

P

Permutation 02
Probability 05
Probability distribution 32
Probability function 34
Probability mass function 34
Probability density function 51
Population 152
Parameter 155
Probability sampling methods 158
Proportion 176
Point estimation 195
Pooled estimator 207
Precision of confidence interval 210
Positive attributes 267

R

Random experiment 05
Random variable 32
Random digits 75

Random numbers 75
Rejection region 235
Randomization 295
Replication 295

S

Sample space 05
Simple event 06
Sure event 07
Standard normal variable 122
Standard normal distribution 123
Sampling survey 152
Sampling unit 152
Sample 152
Survey 153
Sampling frame 153
Sample design 153
Survey design 153
Sampling 154
Statistic 155
Sampling error 155
Sampling with replacement 156
Sampling without replacement 157
Simple random sampling 158
Stratified random sampling 160
Systematic random sampling
Sampling distribution 162
Standard error 162
Statistical inference 194
Significance level 209
Simple hypothesis 233

T

Type-I error 233
Type- II error 234
Test statistic 235
Test of significance 235
Two tailed test 236
Treatment 293
Two-way classification 268

U

Unbiasedness 197
Ultimate class 270
Ultimate class frequencies 270
Uniformity trial 294

W

Width of confidence interval 20

Y

Yule's coefficient of association 27
Yule's correction of continuity 28
Yield 29

About the Author

Jamal Shah is Professor of Statistics. He was born in 1964 in Mangah, District Mardan. He got his early education from Government higher secondary school Dargai Mangah district Charsadda. He did F.Sc. from Government postgraduate college Mardan, B.Sc. from Government postgraduate Jehanzeb College Saidu Sharif Swat. In 1988, from university of Peshawar, he received his M.Sc. Statistics degree and stood third in his batch.

He started his service in December, 1988 as Lecturer in statistics from Government Khushal Khan Khattak college Akora Khattak district Nowshera. He remained chairman department of Statistics, Government postgraduate college Mansehra and Abbottabad for many years. Presently he is working as Principal, Govt. Degree college, Zaida Swabi. The author has a vast teaching experience of 30 years in the subject statistics at Inter, Bachelor and Master level.

رشوت لینے والا اور رشوت دینے
والا دونوں جہنمی ہیں۔
(حدیث نبوی ﷺ)

قومی ترانہ

پاک سر زمین شاد باد کشورِ حسین شاد باد
تو نشانِ عزمِ عالی شان ارضِ پاکستان
مرکزِ یقین شاد باد

پاک سر زمین کا نظام قوتِ اخوتِ عوام
قوم، ملک، سلطنت پائندہ تابندہ باد
شاد باد منزلِ مراد

پرچمِ ستارہ و ہلال رہبرِ ترقی و کمال
ترجمانِ ماضی شانِ حال جانِ استقبال
سایۂ خدائے ذوالجلال

Printed by	Code No.	Quabtity
Etihad	[illegible]	[illegible]