# UNIVERSITY OF THE PUNJAB

**First Semester    2018**
**Examination: B.S. 4 Years Programme**

**PAPER: Statistics-I**
**Course Code: STAT-101**

**TIME ALLOWED: 2 hrs. & 30 mins.**
**MAX. MARKS: 50**

*Attempt this Paper on Separate Answer Sheet provided.*

## SHORT QUESTIONS

**Q2.**    Explain the following:                                                        *(4 marks each)*

i)    Properties of Arithmetic Mean
ii)   Properties of Variance
iii)  Properties of Correlation Coefficient
iv)   Index Numbers
v)    Components of Time Series.

## SUBJECTIVE

**Q3**  **(a)**  From a random sample of voters in Rawalpindi, Islamabad, voters are classified by     *(06)*
age group, as shown by the following data.

| Age Group | 18-24 | 24-30 | 30-36 | 36-42 | 42-48 | 48-54 | 54-60 | 60-66 | 66-72 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Frequency | 170   | 220   | 260   | 350   | 330   | 300   | 320   | 210   | 150   |

Calculate Mean and Geometric Mean of the above data.

**(b)**  Following table gives the birth rates and death rates per hundred thousand of a few     *(04)*
districts in Punjab. Represent them in a Component bar chart.

| Country    | Birth Rate | Death Rate |
|------------|------------|------------|
| Lahore     | 330        | 240        |
| Multan     | 300        | 190        |
| Faisalabad | 180        | 80         |
| Gujrat     | 250        | 160        |

**Q4**    Compute first four Mean moments of the following Data. Also compute coefficient of     *(10)*
skewness and kurtosis.

| Weight (grams) | 65-84 | 85-104 | 105-124 | 125-144 | 145-164 | 165-184 | 185-204 |
|----------------|-------|--------|---------|---------|---------|---------|---------|
| Frequency      | 90    | 100    | 170     | 100     | 50      | 40      | 50      |

**Q5.**                                                                                      *(05)*
**(a)**  The following are the results of height (X) and weight (Y) of 12 persons.

$$\sum x = 766, \qquad \sum y = 1700, \qquad \sum xy = 109380, \qquad \sum x^2 = 49068, \qquad \sum y^2 = 246100$$

Estimate a regression line of Y on X from the above data. Also find the correlation
coefficient between height and weight.

**(b)**  Calculate the seasonal indices by ratio to moving average method from the following     *(05)*
data. Also deseasonalize the data for the year 1944.

| Year | Jan | Feb | Mar | April | May | June | July | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-------|-----|------|------|-----|-----|-----|-----|-----|
| 1943 | 6.2 | 1.8 | 0.9 | 1.4   | 3.2 | 2.3  | 2.2  | 3.2 | 3.4 | 2.7 | 2.1 | 2.1 |
| 1944 | 3.3 | 1.7 | 0.5 | 2.2   | 1.5 | 2.5  | 2.8  | 3.2 | 4.2 | 4.5 | 6.1 | 2.8 |

# UNIVERSITY OF THE PUNJAB

Roll No. .................

**First Semester    2018**
**Examination: B.S. 4 Years Programme**

| | |
|---|---|
| **PAPER: Statistics-I** | **TIME ALLOWED: 30 mins.** |
| **Course Code: STAT-101** | **MAX. MARKS: 10** |

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE

**Q1.    Read the following items carefully and encircle the correct option listed below at each item.**

*(One mark for each)*

1.  The section of statistics which involves the collection, organization, summarizing, and presentation of data relating to some population or sample is

    a)     Inferential statistics.                              c) Descriptive statistics.

    b)     An example of a frequency distribution.    d) The study of statistics.

2.  The number of occurrences of a data value is called

    a) The class limits     b) The frequency     c) Class interval     d) The relative frequency.

3.  A large collection of data may be condensed by constructing

    a) Classes.          b) A frequency polygon.     c) Class limits.     d) A frequency distribution.

4.  The cumulative relative frequency for a given class is defined to be

    a)    The proportion of values preceding the given class.
    b)    The proportion of values up to and including the given class.
    c)    The proportion of values for the given class.
    d)    The proportion of values below the given class.

5.  A student has seven statistics books open in front of him. The page numbers are as follows:

    231,423,521,139,347,400,345. The median for this set of numbers is

    a) 139.          b) 347.          c) 346.          d) 373.5

6.  A cyclist recorded the number of miles per day that she cycled for 5 days. The recordings were as follows: 13, 10, 12, 10, and 11. The mean number of miles she cycled per day is

    a) 13.          b) 11.          c) 10.          d) 11.2

7.  An instructor recorded the following quiz scores (out of a possible 10 points) for the 12 students present: 7, 4, 4,7, 2,9,10, 6,7,3,8,5. The mode for this set of scores is

    a) 9.5.          b) 7.          c) 6.          d) 4.

8.  Which of the following is not a measure of central tendency?

    a) Mode          b) Variance          c) Median          d) Mean

9.  The most frequently occurring value in a data set is called the

    a) Spread.          b) Mode.          c) Skewness.          d) Maximum value.

10. The correlation coefficient provides:
    a) a measure of the extent to which changes in one variable cause changes in another variable.
    b) a measure of the strength of the linear association between two categorical variables.
    c) a measure of the strength of the association (not necessarily linear) between two categorical variables.
    d) a measure of the strength of the linear association between two quantitative variables.

# UNIVERSITY OF THE PUNJAB

Roll No. ..................

### First Semester    2018
### Examination: B.S. 4 Years Programme

PAPER: Statistics-I                                    TIME ALLOWED: 30 mins.
Course Code: STAT-101                                  MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE

Q1.    Read the following items carefully and encircle the correct option listed below at each item.

*(One mark for each)*

1. The section of statistics which involves the collection, organization, summarizing, and presentation of data relating to some population or sample is

    a)    Inferential statistics.                       c) Descriptive statistics.

    b)    An example of a frequency distribution.       d) The study of statistics.

2. The number of occurrences of a data value is called

    a) The class limits      b) The frequency      c) Class interval      d) The relative frequency.

3. A large collection of data may be condensed by constructing

    a) Classes.       b) A frequency polygon.      c) Class limits.      d) A frequency distribution.

4. The cumulative relative frequency for a given class is defined to be

    a)    The proportion of values preceding the given class.
    b)    The proportion of values up to and including the given class.
    c)    The proportion of values for the given class.
    d)    The proportion of values below the given class.

5. A student has seven statistics books open in front of him. The page numbers are as follows:

    231,423,521,139,347,400,345. The median for this set of numbers is

    a) 139.       b) 347.       c) 346.       d) 373.5

6. A cyclist recorded the number of miles per day that she cycled for 5 days. The recordings were as

    follows: 13, 10, 12, 10, and 11. The mean number of miles she cycled per day is

    a) 13.       b) 11.       c) 10.       d) 11.2

7. An instructor recorded the following quiz scores (out of a possible 10 points) for the 12 students

    present: 7, 4, 4,7, 2,9,10, 6,7,3,8,5. The mode for this set of scores is

    a) 9.5.       b) 7.       c) 6.       d) 4.

8. Which of the following is not a measure of central tendency?

    a) Mode       b) Variance       c) Median       d) Mean

9. The most frequently occurring value in a data set is called the

    a) Spread.       b) Mode.       c) Skewness.       d) Maximum value.

10. The correlation coefficient provides:
    a) a measure of the extent to which changes in one variable cause changes in another variable.
    b) a measure of the strength of the linear association between two categorical variables.
    c) a measure of the strength of the association (not necessarily linear) between two categorical variables.
    d) a measure of the strength of the linear association between two quantitative variables.

# UNIVERSITY OF THE PUNJAB

## First Semester  2018
### Examination: B.S. 4 Years Programme

Roll No. .....................

**PAPER: Statistics-I**

**Course Code: STAT-101**

**TIME ALLOWED: 2 hrs. & 30 mins.**

**MAX. MARKS: 50**

*Attempt this Paper on Separate Answer Sheet provided.*

## SHORT QUESTIONS

**Q2.** Explain the following: *(4 marks each)*

i) Properties of Arithmetic Mean
ii) Properties of Variance
iii) Properties of Correlation Coefficient
iv) Index Numbers
v) Components of Time Series.

## SUBJECTIVE

**Q3** **(a)** From a random sample of voters in Rawalpindi, Islamabad, voters are classified by age group, as shown by the following data. *(06)*

| Age Group | 18-24 | 24-30 | 30-36 | 36-42 | 42-48 | 48-54 | 54-60 | 60-66 | 66-72 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Frequency | 170 | 220 | 260 | 350 | 330 | 300 | 320 | 210 | 150 |

Calculate Mean and Geometric Mean of the above data.

**(b)** Following table gives the birth rates and death rates per hundred thousand of a few districts in Punjab. Represent them in a Component bar chart. *(04)*

| Country | Birth Rate | Death Rate |
|---------|-----------|-----------|
| Lahore | 330 | 240 |
| Multan | 300 | 190 |
| Faisalabad | 180 | 80 |
| Gujrat | 250 | 160 |

**Q4** Compute first four Mean moments of the following Data. Also compute coefficient of skewness and kurtosis. *(10)*

| Weight (grams) | 65-84 | 85-104 | 105-124 | 125-144 | 145-164 | 165-184 | 185-204 |
|----------------|-------|--------|---------|---------|---------|---------|---------|
| Frequency | 90 | 100 | 170 | 100 | 50 | 40 | 50 |

**Q5.** **(a)** The following are the results of height (X) and weight (Y) of 12 persons. *(05)*

$$\sum x = 766, \quad \sum y = 1700, \quad \sum xy = 109380, \quad \sum x^2 = 49068, \quad \sum y^2 = 246100$$

Estimate a regression line of Y on X from the above data. Also find the correlation coefficient between height and weight.

**(b)** Calculate the seasonal indices by ratio to moving average method from the following data. Also deseasonalize the data for the year 1944. *(05)*

| Year | Jan | Feb | Mar | April | May | June | July | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-------|-----|------|------|-----|-----|-----|-----|-----|
| 1943 | 6.2 | 1.8 | 0.9 | 1.4 | 3.2 | 2.3 | 2.2 | 3.2 | 3.4 | 2.7 | 2.1 | 2.1 |
| 1944 | 3.3 | 1.7 | 0.5 | 2.2 | 1.5 | 2.5 | 2.8 | 3.2 | 4.2 | 4.5 | 6.1 | 2.8 |

# UNIVERSITY OF THE PUNJAB

**Second Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. ....................

**PAPER: Business Statistics**

**Course Code: STAT-121 / BUS-12133  Part – I (Compulsory)**

**TIME ALLOWED: 15 Mints.**

**MAX. MARKS: 10**

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

Q.1  Encircle the correct answer in the following.                    **(10x1=10)**

I.  A relative frequency distribution includes:
   (a)  Whole numbers             (b)  Percentages
   (c)  Fractions                   (d)  None of these

II.  The value that occurs most often in a set of data is
   (a)  Mean                   (b)  Median
   (c)  Mode                   (d)  Weighted mean

III.  For normal data the measure of skewness is
   (a)  3                      (b)  Zero
   (c)  Positive number        (d)  Negative number

IV.  The consumer price index is basically
   (a)  A fixed weight index      (b)  A laspeyres index
   (c)  Both of the above       (d)  None of the above

V.  If X and Y are two independent random variables then Var $(X - Y)$ is equal to
   (a)  Var $(X)$ – Var $(Y)$      (b)  Var $(X)$ + Var $(Y)$
   (c)  Var $(X)$ + Var $(Y)$ – 2 Cov $(X, Y)$      (d)  None of the above

VI.  For $\hat{Y} = a + bX$, which of the following statement is correct for "a"
   (a)  It is symmetrical with respect to X and Y      (b)  It is independent of origin
   (c)  It is independent of scale      (d)  None of the above

VII.  Chi-square is used to check the independence of variable when
   (a)  Both variables are quantitative      (b)  Both variables are qualitative
   (c)  One quantitative and other is qualitative      (d)  All of above

VIII.  The standard deviation is
   (a)  Two times the variance      (b)  The square of the variance
   (c)  Half the variance       (d)  The square root of the variance

IX.  The test-statistic used in ANOVA is
   (a)  Chi-square statistic      (b)  F-statistic
   (c)  t-statistic           (d)  Z-statistic

X.  In case of an open-end frequency distribution
   (a)  A median cannot be computed      (b)  The arithmetic mean and the median will always be exactly equal
   (c)  A mean cannot be computed      (d)  The distribution is always positively skewed

# UNIVERSITY OF THE PUNJAB

**Second Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. ......................

**PAPER: Business Statistics**

**Course Code: STAT-121 / BUS-12133  Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**

**MAX. MARKS: 50**

## Attempt this Paper on Separate Answer Sheet provided.

Q.2 Define the following: **(10x2=20)**

    (i)    Discrete variable

    (ii)    Diagrams

    (iii)    Arithmetic Mean

    (iv)    Mean-deviation

    (v)    Simple Regression

    (vi)    Simple price index number

    (vii)    Probability distribution function

    (viii)    Chi-square variable

    (ix)    Two way ANOVA

    (x)    Median

Q.3 Consider the following frequency distribution of lives in years of car batteries. **(09)**

| Classes: | 1.5 – 1.9 | 2.0 – 2.4 | 2.5 – 2.9 | 3.0 – 3.4 | 3.5 – 3.9 | 4.0 – 4.4 |
|---|---|---|---|---|---|---|
| No. of batteries: | 4 | 2 | 8 | 30 | 20 | 11 |

Compute Bowley's Co-efficient of skewness.

Q.4 A study was made on the amount of converted sugar in a certain process at various temperature. The coded data were as follows: **(07)**

| Temperature: | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sugar: | 8.1 | 7.8 | 8.5 | 9.8 | 9.5 | 8.9 | 8.6 | 10.2 | 9.3 | 9.2 |

Estimate the linear regression to find amount of sugar produced when the coded temperature is 1.75.

Q.5 The following data represent the number of packets of popular brands of cigarettes (A, B, C, D) sold in a store on five randomly selected days. **(07)**

| A | 21 | 35 | 32 | 28 | 14 | 47 |
|---|---|---|---|---|---|---|
| B | 35 | 12 | 27 | 41 | 19 | 23 |
| C | 45 | 60 | 33 | 36 | 31 | 40 |
| D | 32 | 53 | 29 | 42 | 40 | 23 |

Apply ANOVA and draw conclusion.

Q.6 Compute the consumer price index number for the following data for 2007 with 2000 as base. Use as weights (i) the quantities consumed in the base year, (ii) the value in the base year, **(07)**

| Article | Quantity 2000 | Price (Rs.) in 2000 | Price (Rs.) in 2007 |
|---|---|---|---|
| Food | 10 maunds | 18.00 | 26.50 |
| Cloth | 40 meters | 2.60 | 2.80 |
| Electricity | 100 units | 0.25 | 0.30 |
| Rent | 3 rooms | 30.00 | 27.50 |
| Miscellaneous | 34 units | 0.50 | 0.60 |

# UNIVERSITY OF THE PUNJAB

### Second Semester - 2018
### Examination: B.S. 4 Years Programme

**PAPER: Statistics-II**
**Course Code: STAT-103, STT-12314  Part – I (Compulsory)**

**TIME ALLOWED: 15 Mints.**
**MAX. MARKS: 10**

## Attempt this Paper on this Question Sheet only.

### Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.

Q.1    Encircle the correct answer in the following.                                    **(10)**

I.    A compound event includes
- (a)   at least four outcomes
- (b)   one and only one outcome
- (c)   at least two outcomes
- (d)   all the outcomes of an experiment

II.    Which of the following values cannot be the probability of an event?
- (a)   .82
- (b)   0
- (c)   1.75
- (d)   0.36

III.    The two events A and B are mutually exclusive. Which one of the following statements must be true?
- (a)   $P(A \cap B) = 0.$
- (b)   $P(A \cap B) = 1$
- (c)   $P(A \cup B) = 0.$
- (d)   $P(A \cup B) = 1.$

IV.    If X and Y are two independent random variables, then $Var(X - Y)$ is equal to
- (a)   $Var(X) - Var(Y)$
- (b)   $Var(X) + Var(Y)$
- (c)   $Var(X) + Var(Y) - 2\,COV(X,Y)$
- (d)   None of above

V.    If X and Y are two random variables, then E(X + Y) is equal to
- (a)   $E(X) + E(Y)$
- (b)   $E(X) + Y$
- (c)   $E(X) - E(Y)$
- (d)   None of above

VI.    For a binomial distribution, the mean and variance are related by:
- (a)   $\mu < \sigma^2$
- (b)   $\mu = \sigma^2$
- (c)   $\mu > \sigma^2$
- (d)   $\mu < \sqrt{\sigma^2}$

VII.    For a Poisson distribution, the mean and variance are related by:
- (a)   $\mu = \sigma^2$
- (b)   $\mu < \sigma^2$
- (c)   $\mu > \sigma^2$
- (d)   None of above

VIII.    A binomial distribution may be approximated by a Poisson distribution when
- (a)   n is large and p is small
- (b)   n is small and p is large
- (c)   n is small and p is small
- (d)   n is large and p is large

IX.    The middle area under the normal curve with $\mu \pm 2\sigma$ is
- (a)   0.6827
- (b)   1.0000
- (c)   0.9545
- (d)   0.9973

X.    In a normal distribution, mean deviation is equal to
- (a)   $\sigma$
- (b)   $0.8\,\sigma$
- (c)   $0.6745\,\sigma$
- (d)   $2.0\,\sigma$

# UNIVERSITY OF THE PUNJAB

**Second Semester - 2018**

**Examination: B.S. 4 Years Programme**

**PAPER: Statistics-II**

**Course Code: STAT-103, STT-12314 Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**

**MAX. MARKS: 50**

## Attempt this Paper on Separate Answer Sheet provided.

Q.2     Define the following:                                                                    $2 \times 10 = (20)$

       (i)     Simple event

       (ii)    Addition Law

       (iii)   Probability distribution function

       (iv)    Mathematical expectation

       (v)     Moment generating function

       (vi)    Binomial distribution

       (vii)   Multinomial distribution

       (viii)  Cumulative distribution function

       (ix)    Mean and variance of Hypergeometric distribution

       (x)     M.G.F. of Normal distribution

Q.3     The probability that an automobile being filled with gasoline will also need an oil **(07)** change is 0.35; the probability that it needs a new oil filter is 0.45; and the probability that both the oil and filter need changing is 0.15.

       (a) if the oil had to be changed, what is the probability that a new oil filter is needed?

       (b) What is the probability that at least one of these need change?

Q.4     Derive the variance of the hypergeometric distribution.                        **(08)**

Q.5     Derive upper and lower quartiles and the quartile deviation of the Normal **(08)** distribution.

Q.6     If X is b (x; 20, 0.45) Find P (X = 12). Then find the approximation to this **(07)** probability using

       (a) the Poisson distribution

       (b) the normal distribution

# UNIVERSITY OF THE PUNJAB

**Third Semester 2018**

Roll No. .....................

| | |
|---|---|
| **PAPER:** Statistics-III | **TIME ALLOWED: 2 hrs. & 30 mins.** |
| **Course Code: STAT-201/STT-21314** | **MAX. MARKS: 50** |

*Attempt this Paper on Separate Answer Sheet provided.*

## Part-2 (SUBJECTIVE)

**Note:** Attempt ALL questions.

| | | Marks |
|---|---|---|
| **Q.3.** | **Write short answers to each of the followings.** | |
| i) | Sampling is superior than complete enumeration, discuss. | **(04)** |
| ii) | In population census cluster sampling is preferred over stratified random sampling, why? | **(02)** |
| iii) | Compare random and simple random sampling. | **(02)** |
| iv) | Which is more dangerous, sampling error or non-sampling error? Why? | **(03)** |
| v) | What is meant by mean square error of an estimator? | **(02)** |
| vi) | What is the purpose of goodness- of- fit test? | **(02)** |
| vii) | Why is the z-test usually inappropriate as a test statistic when sample size is small? | **(03)** |
| viii) | What do you mean by p-value? | **(02)** |

**Q.4.** Suppose a population of N =- 9 is stratified into 3 strata with the following measurements: **(06)**

| Stratum I | $X_{11} = 1,\ X_{12} = 2,\ X_{13} = 4$ |
|---|---|
| Stratum II | $X_{21} = 6,\ X_{22} = 8$ |
| Stratum III | $X_{31} = 11, X_{32} = 15, X_{33} = 16, X_{34} = 19$ |

If two measurements are drawn from each stratum for the sample, state how many samples of size 6 could be chosen from this population? List these samples and compute the mean for each sample

**Q.5.** Let $X_1, X_2, X_3$, and $X_4$ be a random sample of size n = 4 from a $N(\mu, \sigma^2)$. A statistician wishes to estimate the mean by using either of the following two estimators of the mean $\mu$ **(06)**

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{4},\qquad \text{(the sample mean, } \bar{X})$$

$$T_2 = \frac{X_1 + 2X_2 + 3X_3 + X_4}{7},\qquad \text{(a weighted mean, } \bar{X}_\omega)$$

Which estimator should be preferred?

**Q.6.** A sample of readings from a normal population with unknown mean $\mu$ and unknown variance $\sigma^2$ gave the following data: **(06)**

| x | 17.4 | 17.5 | 17.6 | 17.7 | 17.8 |
|---|---|---|---|---|---|
| f | 12 | 16 | 19 | 23 | 10 |

A second sample of readings taken from the same population gave $n_2 = 72,\ \sum X = 1267.2,\ \sum X^2 = 22536$. Combine the two samples to give estimates $\mu$ and $\sigma^2$, and give the approximate 90% confidence interval for $\mu$.

**P.T.O.**

**Q.7.** A random sample of size 40 from non- normal population yielded the $\overline{X}_1 = 70.4$, $S_1^2 = 31.40$ another random sample of size 50 from second non normal population yielded $\overline{X}_2 = 65.3$, $S_2^2 = 44.82$, test the hypothesis that the difference between two means is equal to 2. **(06)**

**Q.8.** Ten young recruits were put through a strenuous physical training program by the Army. Their weights were recorded before and after the training with the following results: **(06)**

| Recruit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Weight before | 125 | 195 | 160 | 171 | 140 | 201 | 170 | 176 | 195 | 139 |
| Weight after | 136 | 201 | 158 | 184 | 145 | 195 | 175 | 190 | 190 | 145 |

Using $\alpha = 0.05$, would you say that the program affects the average weight of recruits? Assume the distribution of weights before and after to be approximately normal.

PAPER: Statistics-III

Course Code: STAT-201/STT-21314

TIME ALLOWED: 30 mins.

MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*

## Part-1 (OBJECTIVE)

**Note: Attempt ALL questions. Cutting and overwriting is not allowed in MCQs**
**All questions carry equal marks.**

Q.1. Some possible answers are given for each of the followings, encircle (10x1) (or tick) the correct answer.

i) A Population should be
   A) large enough          B) Small enough
   C) such that its elements possess similar characteristics      D) A&C

ii) A random sample is most suitable when population is
   A) homogenous          B) heterogeneous
   C) Finite               D) A&C

iii) Selecting a cricket team from Punjab is a type of ------- sampling
   A) random              B) non- random
   C) purposive           D) B&C

iv) Probability distribution of mean for large sample is
   A) normal              B) binomial
   C) poisson             D) none of the above

v) Inference consists of two parts(testing and estimation)
   A) first we test then estimate      B) first we estimate then test
   C) testing & estimation are done simultaneously D) only estimation is done

vi) In the comparison of type-I and type-II errors
   A) Type-I error is more dangerous   B) Type-II error is more dangerous
   C) both are dangerous              D) B&C

vii) Null hypothesis is designed for possible--------
   A) acceptance          B) rejection
   C) both for acceptance and rejection    D) none of the above

viii) Equality of more than two means may be tested by
   A) Z-test              B) t- test
   C) F-test              D) All of the above.

ix) The critical region for composite hypothesis is always --- ---
   A) one- sided          B) two-sided
   C) A &B                D) none of the above

x) F-distribution is
   A) uni-directional     B) two tails
   C) unimodal            D) A&C

# UNIVERSITY OF THE PUNJAB

**PAPER: Elementary Statistics**      TIME ALLOWED: 2 hrs. & 30 mins.
**Course Code: STAT-211/GEN-21129**      MAX. MARKS: 50

*Attempt this Paper on Separate Answer Sheet provided.*

## SUBJECTIVE

**Q.2: Write short answers to the following questions:**      (20 marks)

i) Differentiate between Primary Data and Secondary Data.
ii) Differentiate between Parameter and Statistic.
iii) Define Relative dispersion.
iv) Define Tabulation.
v) Define Positive and Negative Correlation.
vi) Define Testing of Hypothesis.
vii) Write down two properties of Arithmatic Mean.
viii) What do you mean by Classification of data.
ix) Describe regression coefficient.
x) Define Probability Sampling.

## QUESTIONS WITH BRIEF ANSWERS      (30 marks)

**Q. No.3:** Prepare a frequency table for the following data taking classes as 11-20, 21-3, ..........

100, 96, 92, 88, 86, 84, 82, 70, 62, 58, 36, 56, 25, 45, 89, 45, 78, 54, 62, 42, 73, 46, 24,
39, 47, 58, 65, 43, 15, 54, 70, 56, 80, 41, 64, 26, 39, 44, 28, 33, 84, 66, 54, 22,  33, 87,
25, 55, 73, 56, 55, 70, 49, 37, 28, 46, 61, 37, 83, 47, 59, 67, 43,  29, 34, 76.

**Also draw the Histogram.**      (10 Marks)

**Q. No.4:** The following frequency distribution gives the ages of 100 college Students:

| Ages | 14-15 | 16-17 | 18-19 | 20-21 | 22-23 | 24-25 | Total |
|------|-------|-------|-------|-------|-------|-------|-------|
| Number of students | 6 | 16 | 20 | 31 | 15 | 12 | 100 |

**Compute Coefficient of Variation (C.V.).**      (10 Marks)

**Q. No.5:** The data of heights and weights is given below:

| Height (X) | 65 | 64 | 73 | 63 | 66 | 65 | 60 | 70 | 71 | 69 |
|------------|----|----|----|----|----|----|----|----|----|----|
| Weight (Y) | 57 | 59 | 68 | 64 | 68 | 72 | 67 | 73 | 69 | 72 |

**Find Regression Line of X on Y and Correlation Coefficient between X and Y.**      (10 Marks)

# UNIVERSITY OF THE PUNJAB

Roll No. ..................

### Third Semester    2018
### Examination: B.S. 4 Years Programme

**PAPER: Elementary Statistics**
**Course Code: STAT-211/GEN-21129**

**TIME ALLOWED: 30 mins.**
**MAX. MARKS: 10**

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE

Q.1: Each Question has four possible answers. Encircle the correct answer.

(i) Number of students in class room is an example of:
  (a) Discrete variable
  (b) Qualitative variable
  (c) Attribute
  (d) Continuous variable

(ii) The mode of the following data 8, 5, 8, 7, 3, 5, 8, 4, 8, 9, 8, 2, 4, 3, 8 is:

  (a) 8
  (b) 4
  (c) 7
  (d) 3

(iii) A numerical quantity calculated from population data is called:
  (a) Statistic
  (b) Parameter
  (c) Constant
  (d) None of above

(iv) A branch of statistics by which obtained data is organized and summarized in order to describe its nature is called as:
  (a) Explanatory statistics
  (b) Descriptive statistics
  (c) Inferential statistics
  (d) None of these

(v) Sampling error can be reduced by:
  (a) Increasing sample size
  (b) Decreasing sample size
  (c) Fixing the size
  (d) None

(vi) For a Positively Skewed distribution:
  (a) Mean, mode, median are equal
  (b) Mean, mode, median are not equal
  (c) Mean is greater than mode
  (d) Mean is less than mode

(vii) Weights of students in Punjab University are best shown by:
  (a) Simple bar chart
  (b) Histogram
  (c) Pie chart
  (d) Frequency Polygon

(viii) Mean and Standard Deviation of -5,-5, -5, -5, -5 is:
  (a) -5, -5
  (b) 0, -5
  (c) -5, 0
  (d) 5, 5

(ix) The range of Normal Distribution is:
  (a) 0 to $\infty$
  (b) $-\infty$ to 0
  (c) $-\infty$ to $\infty$
  (d) -1 to 1

(x) A non-deserving candidate is passed in an interview is an example of:
  (a) Correct decision
  (b) Type –II error
  (c) Type –I error
  (d) Level of Significance

**PAPER: Statistics-IV**
**Course Code: STAT-203 / STT-22314  Part – I (Compulsory)**

**TIME ALLOWED: 15 Mints.**
**MAX. MARKS: 10**

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

Q1.     Read the following items carefully and encircle the correct option listed below at each item.
*(One mark for each)*

1. **Which of the following is not a necessary assumption underlying the use of the Analysis of Variance technique?**
   a) The samples are independent and randomly selected.
   b) The populations are normally distributed.
   c) The variances of the populations are the same.
   d) The means of the populations are equal.

2. **Which of the following is NOT CORRECT about a randomized complete block experiment?**
   a) Every treatment must appear at least once in every block.
   b) Blocking is used to remove the effects of another factor (not of interest) from the comparison of levels of the primary factor.
   c) The ANOVA table will have another line in it for the contribution to the variability from blocks.
   d) Blocks should contain experimental units that are as different as possible from each other.

3. **What is the purpose of a multiple regression?**
   a) To predict scores on an independent variable from scores on a single dependent variable
   b) To predict scores on a dependent variable from scores on multiple independent variables
   c) To predict scores on an independent variable from scores on multiple dependent variables
   d) To predict scores on a dependent variable from scores on a single independent variable.

4. **Which of the following points are not true when conducting a multiple regression?**
   a) Multiple regression can be used to assess quadratic relationships
   b) Data must be homogeneous for a multiple regression
   c) Multiple regression can be used to assess linear relationships
   d) Data must be normally distributed for multiple regression.

5. **What are residuals?**
   a) Residuals are the differences between the observed and expected dependent variable scores
   b) Uncontrolled variables
   c) Left over scores
   d) Confidence intervals.

6. **The F-Ratio in a completely randomized ANOVA is:**
   a) MST/MSE
   b) MSE/MSTR
   c) MSE/MST
   d) MSTR/MSE

7. **An ANOVA procedure is applied to data obtained from 5 samples, where each sample contains 9 observations. The degrees of freedom for the critical value of F are:**
   a) 5 numerator and 9 denominator degrees of freedom
   b) 4 numerator and 8 denominator degrees of freedom
   c) 45 degrees of freedom
   d) 4 numerator and 40 denominator degrees of freedom

8. **In the ANOVA, treatment refers to.**
   a) experimental units
   b) c) different levels of a factor
   c) a factor
   d) d) None of these

9. **Multiple comparisons tests are applicable when:**
   a) Null Hypothesis about equality of means is rejected
   b) Null Hypothesis about equality of means is accepted
   c) Does not depend upon the rejection or acceptance of Null Hypothesis
   d) None of these

10. **The efficiency of two experimental designs can simply be measured through ..... of error variances.**
    a) Addition
    b) Subtraction
    c) Multiplication
    d) Ratio

# UNIVERSITY OF THE PUNJAB
### Fourth Semester - 2018
### Examination: B.S. 4 Years Programme

Roll No. .....................

**PAPER: Statistics-IV**
**Course Code: STAT-203 / STT-22314  Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**
**MAX. MARKS: 50**

## Attempt this Paper on Separate Answer Sheet provided.

**Q2.** Differentiate between the following terms: *(4 marks each)*

i) Regression and Correlation
ii) One-way classification and Two-way Classification
iii) Completely Randomized Design and Completely Randomized Block Design
iv) Mean Prediction Interval Estimates and Individual Prediction Interval Estimates
v) Multiple Correlation Coefficient and Partial Correlation Coefficient.

## SUBJECTIVE

**Q3.** Three sections of the same statistics course are taught by three teachers. The final *(06)* grades given by the three teachers (independently) were recorded as follows:

| Teacher A | 75 | 91 | 84 | 45 | 82 |
|-----------|----|----|----|----|----|
| Teacher B | 59 | 83 | 99 | 77 | 65 |
| Teacher C | 66 | 77 | 51 | 90 | 73 |

Test the hypothesis that the average grades given by the three teachers are same.

**Q4.** Is the type of beverage ordered with lunch at a restaurant independent of the age of *(06)* the customer? A random poll of 55 lunch customers is taken, resulting in the following contingency table of observed values. Use $\alpha = .01$ to determine whether the two variables are independent.

| | | Preferred Beverage | | |
|---|---|---|---|---|
| | | Coffee/Tea | Soft Drink. | Other (Milk etc.) |
| Age | 1 – 34 | 8 | 14 | 9 |
| | 35 – 55 | 5 | 4 | 2 |
| | > 55 | 6 | 4 | 3 |

**Q5.** Five varieties of Rice are compared in five randomized complete blocks. The SS for *(06)* Total, Blocks and Varieties are 282.30, 26.96, and 155.44 respectively.

i) Test the significance of difference between varietal means

ii) Compare the efficiency of this design with the design in which blocks are ignored.

**Q6.** Given that: *(12)*

| $X_1$ | 3 | 5 | 6 | 8 | 12 | 14 |
|-------|---|---|---|---|----|----|
| $X_2$ | 16 | 10 | 7 | 4 | 3 | 2 |
| $X_3$ | 90 | 72 | 54 | 42 | 30 | 12 |

i) Determine $r_{12}$, $r_{12.34}$ and $R_{1.23}$

ii) Test each one of these correlation coefficients for a significance at 5% level.

# UNIVERSITY OF THE PUNJAB

**Fourth Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. ..................

PAPER: Probability and Statistics

Course Code: STAT-221 / IT-22407  Part – I (Compulsory)

TIME ALLOWED: 15 Mints.

MAX. MARKS: 10

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

Q.1                                                                                        (1x10=10)

I. The number of occurrences of a data value is called
   - (a) Frequency
   - (b) Cumulative frequency
   - (c) Relative frequency
   - (d) All of above

II. The value that occurs most often in a set of data is called
   - (a) Mean
   - (b) Mode
   - (c) Geometric mean
   - (d) Harmonic mean

III. The standard deviation is
   - (a) The square of the variance
   - (b) Two times the standard deviation
   - (c) Half the variance
   - (d) The square root of the variance

IV. The probability of an event is always
   - (a) Greater than zero
   - (b) In the range zero to 1
   - (c) Less than 1
   - (d) Greater than 1

V. Two mutually exclusive events:
   - (a) Always occur together
   - (b) Cannot occur together
   - (c) Can sometimes occur together
   - (d) Can never occur together

VI. A discrete probability distribution may be represented by
   - (a) A table
   - (b) A mathematical function
   - (c) A graph
   - (d) All of above

VII. For a binomial distribution
   - (a) Mean < Variance
   - (b) Mean > Variance
   - (c) Mean = Variance
   - (d) Mean > Standard deviation

VIII. A binomial distribution may be approximated by a Poisson distribution when
   - (a) n is large and P is small
   - (b) n is small and P is large
   - (c) n is small and P is small
   - (d) n is large and P is large

IX. The area under normal curve between $\mu - \sigma$ and $\mu + \sigma$ is
   - (a) 68.27%
   - (b) 95.45%
   - (c) 99.73%
   - (d) 100%

X. In a normal distribution, mean deviation is equal to
   - (a) $0.6745\,\sigma$
   - (b) $0.8\,\sigma$
   - (c) $\sigma$
   - (d) $2.0\,\sigma$

# UNIVERSITY OF THE PUNJAB

**Fourth Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. .....................

**PAPER: Probability and Statistics**
**Course Code: STAT-221 / IT-22407  Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**
**MAX. MARKS: 50**

=========================================================

## Attempt this Paper on Separate Answer Sheet provided.

**Q.2** Define the following:                                                                 **(4x5=20)**

    (i)     Descriptive statistics

    (ii)    Gamma Distribution

    (iii)   Regression

    (iv)   Correlation

    (v)    Conditional Probability

    (vi)   Addition law of Probability

    (vii)   Random Variable

    (viii)  Moment Ratios

    (ix)   Histogram

    (x)    Bayes theorem

**Q.3** The following is the frequency distribution of neck circumferences of college **(06)** students (X).

| X: | 12.5 | 13.0 | 13.5 | 14.0 | 14.5 | 15.0 | 15.5 | 16.0 |
|---|---|---|---|---|---|---|---|---|
| No. of students: | 08 | 19 | 30 | 63 | 66 | 29 | 18 | 04 |

Compute Co-efficient of variation.

**Q.4** In a certain college, 4% of the men and 1% of the women are taller than 6 feet. **(06)** Furthermore, 60% of the students are women. Now if a student is selected at random and is taller than 6 feet, what is the probability that the student is a woman?

**Q.5** Suppose that the death rate from Malaria is 20% find the probability that the **(08)** number of deaths in a particular village is equal to 6 out of next 20 deaths.

    (i)    Using Binomial distribution

    (ii)   Using Poisson distribution

    (iii)  Using Normal distribution

**Q.6** The data given below represent the heights (X) and the weights (Y) of different **(10)** men.

| X: | 60 | 62 | 65 | 70 | 72 | 68 | 74 |
|---|---|---|---|---|---|---|---|
| Y: | 130 | 135 | 158 | 170 | 185 | 160 | 190 |

Find the linear regression and predict the weight of a man whose height is 73. Also compute correlation Co-efficient.

# UNIVERSITY OF THE PUNJAB

Roll No. .................

**Fifth Semester    2018**
**Examination: B.S. 4 Years Programme**

**PAPER: Parametric and Nonparametric Tests (Theory)    TIME ALLOWED: 30 mins.**
**Course Code: STAT-301                                    MAX. MARKS: 10**

*Attempt this Paper on this Question Sheet only.*
## OBJECTIVE TYPE

I.  To perform a Run test for randomness, the data must be:
(a) Qualitative                          (b) Quantitative
(c) Divided into at least two categories  (d) Divided into exactly two categories

II.  If the critical region is located on one side of the sampling distribution of test-statistic, the test is called:
(a) One tailed      (b) Two tailed        (c) both (a) & (b)      (d) none of these

III.  Level of significance α lies between:
(a) -1 and +1      (b) 0 and 1            (c) 0 and n            (d) -∞ to +∞

IV.  The Mann-Whitney U test is preferred to a t-test when
a) Data are paired                        b) Sample sizes are small
c) The assumption of normality is not met  d) Samples are dependent

V.  What is maximum number of runs possible in a sequence of length 7 using two symbols?
a) 6                b) 4                  c) 3                  d) 7

VI.  A hypothesis that specifies all the values of parameter is called:
a) Simple hypothesis                      b) Composite hypothesis
c) Statistical hypothesis                 d) None of the above

VII.  In a Chi-square distribution no expected frequency should be:
a) More than 5      (b) Less than 5       (c) Equal to 5         (d) Both (a) and (b)

VIII.  Which of the following is not true of parametric statistics?
a) They are inferential tests
b) They assume certain characteristics of population parameters
c) They assume normality of the population
d) They are distribution-free

IX.  The formula for Kruskal Wallis test is based upon
(a) Means      (b) Deviations              (c) Ranks              (d) Categories

X.  The range of F variable is
a) $0 \le F^2 \le +\infty$                 b) $-\infty \le F^2 \le 0$
c) $-\infty \le F^2 \le +\infty$           d) None of these

**PAPER: Parametric and Nonparametric Tests (Theory)    TIME ALLOWED: 2 hrs. & 30 mins.**
**Course Code: STAT-301                                   MAX. MARKS: 50**

*Attempt this Paper on Separate Answer Sheet provided.*
## SUBJECTIVE TYPE

Q1.  Write short answers on the following questions
   a)  What are non-parametric alternative of the ANOVA, when samples are dependent?
   b)  What is the Power of the test.
   c)  Explain briefly, with example, what do you understand by goodness of fit test.
   d)  Outline the procedure of the sign test
   e)  State the important application of $\chi^2$ distribution.

                                                                    (4+4+4+4+4)

Q2.  The government awarded grants to nine different experimental stations of the agricultural department to test the yield capabilities of two varieties of wheat. Five acres of each variety are planted each station and the yield, in kilograms per acre recorded as follows:

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------|-----|-----|-----|-----|------|-----|-----|-----|------|
| Variety 1 | 900 | 700 | 850 | 950 | 1025 | 750 | 875 | 790 | 900 |
| Variety 2 | 1050 | 750 | 790 | 900 | 1150 | 800 | 825 | 825 | 1000 |

Test the hypothesis at $\alpha = 0.05$, that the average yields of the two varieties of wheat are equal against the alternative hypothesis that they are unequal, assuming the distribution of yields to be approximately normal.                                                      (10)

Q3.  Two varieties of tomato were experimented with concerning their fruit-producing abilities, measured in pounds. The following data were obtained:

| Location | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|------|------|------|------|------|------|------|------|------|------|
| Variety A | 3.03, | 3.10, | 2.35, | 3.86, | 3.91, | 1.72, | 2.65, | 2.30, | 2.70, | 3.60 |
| Variety B | 2.28, | 3.68, | 2.17, | 3.56, | 3.73, | 1.85, | 1.48, | 1.86, | 2.76, | 2.68 |

Apply (i) the sign test, (ii) the Wilcoxon signed-rank test at the 0.05 level of significance, to test the hypothesis that there is no difference in fruit-producing abilities of the two varieties.
                                                                    (10)

Q4.  The various sums for two sets of data, each of 4 observations are as follows:

| Set: | $\sum X$ | $\sum Y$ | $\sum XY$ | $\sum X^2$ | $\sum Y^2$ |
|------|------|------|------|------|------|
| A | 8 | 37 | 76 | 18 | 349 |
| B | 15 | 47 | 179 | 59 | 557 |

   i)  Find the estimates of $\beta_1$ and $\beta_2$, the regression coefficients of two linear regression lines.
   ii) Construct the 95% confidence interval for $\beta_1 - \beta_2$ and test the hypothesis that $\beta_1 = \beta_2$ against $H_1 : \beta_1 \neq \beta_2$ at the 0.05 level of significance.
                                                                    (10)

PAPER: Design and Analysis of Experiments (Theory)    TIME ALLOWED: 30 mins.
Course Code: STAT-303    MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*
## OBJECTIVE TYPE

**Q1.**  Read the following items carefully and encircle the correct option listed below at each item.

*(One mark for each)*

**i)**  The F-Ratio in a completely randomized ANOVA is:
a) MST/MSE        b) MSE/MSTR        c) MSE/MST        d) MSTR/MSE

**ii)**  An ANOVA procedure is applied to data obtained from 5 samples, where each sample contains 9 observations. The degrees of freedom for the critical value of F are:
a) 5 numerator and 9 denominator degrees of freedom
b) 4 numerator and 8 denominator degrees of freedom
c) 45 degrees of freedom
d) 4 numerator and 40 denominator degrees of freedom

**iii)**  In the ANOVA, treatment refers to.
a) experimental units        c) different levels of a factor
b) a factor        d) None of these

**iv)**  The assumptions under analysis of variance consist of:
a) Normality and Independence        c) Linearity and Additivity
b) Both (a) and (b)        d) None of these

**v)**  The following design provides the maximum number of degrees of freedom for error sum of squares:
a) Completely Randomized Design        c) Completely Randomized Block Design
b) Latin Square Design        d) None of these

**vi)**  Multiple comparisons tests are applicable when:
a) Null Hypothesis about equality of means is rejected
b) Null Hypothesis about equality of means is accepted
c) Does not depend upon the rejection or acceptance of Null Hypothesis
d) None of these

**vii)**  One can estimate the missing observation through covariance technique by simply changing the sign of....
a) b        b) r        c) Correction Factor        d) None of these

**viii)**  The efficiency of two experimental designs can simply be measured through ..... of error variances.
a) Addition        b) Subtraction        c) Multiplication        d) Ratio

**ix)**  Two Latin squares are ........................ if each letter of one square design occurs exactly once with every letter of the other square when they are superimposed.
a) Orthogonal        b) Factorial Designs   c) Efficient        d) None of these

**x)**  A contrast is ....combination of treatments.
a) Linear        b) Exponential        c) Quadratic        d) None of these

# UNIVERSITY OF THE PUNJAB

**Fifth Semester 2018**
**Examination: B.S. 4 Years Programme**

Roll No. .....................

PAPER: Design and Analysis of Experiments (Theory)   TIME ALLOWED: 2 hrs. & 30 mins.
Course Code: STAT-303                                 MAX. MARKS: 50

*Attempt this Paper on Separate Answer Sheet provided.*
## SUBJECTIVE TYPE

## SHORT QUESTIONS

**Q2.   Differentiate between the following terms:**                    *(4 marks each)*

i)    Random Effects and Mixed Effects
ii)   Contrast and Orthogonal Contrasts
iii)  Completely Randomized and Completely Randomized Block Designs
iv)   Replication and Repetition
v)    Latin Square and Graeco Latin Square Designs.

## SUBJECTIVE

Q3. a)  For a RCB experiment

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad \begin{aligned} i &= 1,2,\ldots,r \\ j &= 1,2,\ldots,k \end{aligned}$$

Let $\mu, \alpha_i$ and $\beta_j$ are least square estimates of $\mu, \alpha_i$ and $\beta_j$ respectively. Develop expected Mean Squares indicating the assumptions.

b)  An agriculturist approaches you with five varieties of a crop to be compared for their yields in five replicates on a piece of land which varies in fertility from South to North and in water level from East to West. Suggest a suitable layout plan with proper randomization.          (6+4)

Q4. a)  Seven treatments arranged in six randomized complete blocks gave the following sum of squares and products

| S.O.V. | YY | XY | XX |
|--------|------|-----|-----|
| Blocks | 1200 | 600 | 200 |
| Treatments | 800 | 300 | 100 |
| Error | 1400 | 700 | 600 |

i)    Is the regression of Y on X significance at 5% level.
ii)   Construct ANOVA and write the inference.

b)  Five varieties of Rice are compared in five randomized complete blocks. The SS for Total, Blocks and Varieties are 282.30, 26.96, and 155.44 respectively.
(i) Test the significance of difference between varietal means
(ii) Compare the efficiency of this design with the design in which blocks are ignored.      (6+4)

Q5.   Derive formula for estimating two missing observations in a LS Design when values are missing in different rows, different columns and different treatments.          *(10)*

# UNIVERSITY OF THE PUNJAB

Roll No. .................

### Fifth Semester    2018
### Examination: B.S. 4 Years Programme

PAPER: Sampling Techniques (Theory)                 TIME ALLOWED: 30 mins.
Course Code: STAT-305                                MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*
## OBJECTIVE

**Q. No. 1: Encircle the correct option.**

1. If ------------- sampling is done, each item in the population has a chance of being chosen.
   A. Non-probability                   B. Convenience
   C. Probability                       D. None of above

2. What sample statistic is used to estimate a population parameter?
   A. Point estimate                    B. Sampling error
   C. Bias                              D. Non-sampling error

3. The type of sampling technique where each person in the sampling frame has an equal chance of being selected is best described as ---------------
   A. Systematic Sampling               B. Stratified Random Sampling
   C. Simple Random Sampling            D. None of the above

4. To determine the proportion of students at a university who favor the hiring of a new dean, every tenth student is surveyed. What type of sampling is this?
   A. Random                            B. Cluster
   C. Stratified                        D. Systematic

5. _____ is a set of elements taken from a larger population according to certain rules.
   A. Sample                            B. Population
   C. Statistic                         D. Element

6. In sampling with replacement, an element can be chosen:
   A. Less than once                    B. More than once
   C. Once                              D. Difficult to tell

7. People who are available, volunteer, or can be easily recruited are used in the sampling method called _____.
   A. Simple Random Sampling            B. Cluster Sampling
   C. Stratified Random Sampling        D. Convenience Sampling

8. Which of the following will give a more "accurate" representation of the population from which a sample has been taken?
   A. A large sample based on the convenience sampling technique
   B. A small sample based on simple random sampling
   C. A large sample based on simple random sampling
   D. A small cluster sample

9. Probability distribution of a statistic is called:
   A. Sampling unit                     B. Population unit
   C. Sampling distribution             D. Sampling

10. The difference between a statistic and the parameter is called:
    A. Standard error                   B. Error
    C. Bias                             D. Sampling error

# UNIVERSITY OF THE PUNJAB

**Fifth Semester    2018**
**Examination: B.S. 4 Years Programme**

Roll No. .......................

**PAPER: Sampling Techniques (Theory)**
**Course Code: STAT-305**

**TIME ALLOWED: 2 hrs. & 30 mins.**
**MAX. MARKS: 50**

*Attempt this Paper on Separate Answer Sheet provided.*
## SUBJECTIVE

**Attempt all questions.**

**Q. No. 2.** Write a short note on the following:                                                                 (5 each)
  i.   Sources of non-sampling errors
  ii.  Inverse Sampling
  iii. Advantages of systematic sampling
  iv.  Optimum Allocation in Stratified Random Sampling

**Q. No. 3.** For simple random sampling, show that the sample mean $\bar{y}$ is an unbiased       (07)
estimator of the population mean $\bar{Y}$ and variance of sample mean is given by

$$V(\bar{y}) = \left(\frac{N-1}{Nn}\right)S^2$$

Use Cornfield approach to prove this.

**Q. No. 4.**                                                                                                   (07)
Compare the values obtained for $V(p_{st})$ under proportional allocation and
optimum allocation for fixed sample size in the following two populations.
Each stratum is of equal size. The f.p.c may be ignored.

| Population 1 | | Population 2 | |
| --- | --- | --- | --- |
| Stratum | $P_h$ | Stratum | $P_h$ |
| 1 | 0.1 | 1 | 0.01 |
| 2 | 0.5 | 2 | 0.05 |
| 3 | 0.9 | 3 | 0.10 |

What general result is illustrated by these two populations?

**Q. No. 5.** Given the results of a stratified random sample, an unbiased estimator of       (09)
$V_{ran}$ which is the variance of the mean of a SRS from the same population is;

$$v_{ran} = \frac{N-n}{n(N-1)}\left[\frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{j=1}^{n_h} y_{hj}^2 - \bar{y}_{st}^2 + v_{\bar{y}_{st}}\right]$$

where $v_{\bar{y}_{st}}$ is the usual unbiased estimator of $V_{\bar{y}_{st}}$ i.e. $E\left(v_{\bar{y}_{st}}\right) = V_{\bar{y}_{st}}$

**Q. No. 6.** Show that the variance of the mean of the systematic sample is       (07)

$$V(\bar{y}_{sy}) = \left(\frac{N-1}{N}\right)\frac{S^2}{n}[1 + (n-1)\rho_w]$$

where $\rho_w$ is the correlation coefficient between pairs of units that are in the
same systematic sample.

# UNIVERSITY OF THE PUNJAB

Roll No. ...............

**Fifth Semester    2018**

**Examination: B.S. 4 Years Programme**

PAPER: Probability Theory (Theory)                    TIME ALLOWED: 30 mins.

Course Code: STAT-307                                  MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE TYPE

**Q.1    Tick the correct answer for the following multiple choice questions.**

I.    If $X$ is number of trials for negative binomial distribution with parameters p and k then its minimum value is :
   a.   0
   b.   k
   c.   k+1
   d.   k-1

II.   For a given binomial distribution with n fixed if p=0.5 then?
   a.   Poisson distribution will provide a bad approximation
   b.   Poisson distribution will provide a good approximation
   c.   Binomial distribution will be skewed left
   d.   Binomial distribution will be skewed right

III.  A fair coin is tossed 100 times, the expected number of heads are :
   a.   100
   b.   50
   c.   15
   d.   None of these

IV.   Which of the following is a characteristic of the probability distribution for any random variable?
   a.   A probability is provided for every possible value
   b.   The sum of all probabilities is one
   c.   No given probability occurs more than once
   d.   Both (a) and (b)

V.    A binomial distribution may be approximated by a Poisson distribution if:
   a.   n is small and p is small
   b.   n is large and p is large
   c.   n is small and p is large
   d.   none of these

VI.   If a letter is chosen at random from the 10 letters of the word STATISTICS, what is probability that it is a vowel?
   a.   0.20
   b.   0.23
   c.   0.30
   d.   0.40                                                **P.T.O.**

VII. A and B are two independent events. The probability of A is 1/4 and B is 1/3. The neither probability of A nor B is
 a. 5/12
 b. 1/3
 c. 3/4
 d. 11/12

VIII. The number of ways to select to persons from 6, ignoring order of selection;
 a. 64
 b. 15
 c. 36

IX. A cricket team captain wins the toss for three consecutive matches. What is the probability that he will call correctly for the fourth march?
 a. 1/16
 b. 1/4
 c. 1/8
 d. 1/2

X. Which of the following statement is not true?
 a. Mutually exclusive events are statistically dependent
 b. Complementary events have probabilities that sum to one?
 c. Opposite events ate statistical independent.
 d. An experiment's elementary events are collectively exhaustive and mutually exclusive

# UNIVERSITY OF THE PUNJAB

**Fifth Semester    2018**

**Examination: B.S. 4 Years Programme**

Roll No. .....................

---

**PAPER: Probability Theory (Theory)**          **TIME ALLOWED: 2 hrs. & 30 mins.**
**Course Code: STAT-307**                        **MAX. MARKS: 50**

---

## *Attempt this Paper on Separate Answer Sheet provided.*
## SUBJECTIVE TYPE

**Q.2**   Write short answer on the following questions.                    **(20)**

a)   What are the differences between a priori probability and empirical probability?

b)   If $P(A) = 0.7, P(B) = 0.6$, and events A and B are independent, find $P(A \text{ and } B)$

c)   What is the difference between mutually exclusive events and collectively exhaustive events?

d)   A coin is known to be unbalanced in such a way that heads only comes up 0.4 of the time.

   (i)   What is the probability that first head appears on $4^{th}$ toss?

   (ii)   How many toss would it take, on average, to flip two heads?

e)   What are the first four properties that must be present in order to use the negative binomial distribution?

**Q.3.a)** A sample of 500 respondents was selected in a large metropolitan area to study consumer behavior. Among the question asked was" Do you enjoy shopping for clothing?" Of 240 males, 136 answered yes. Of 260 females, 224 answered yes. What is the probability that a respondent chosen at random

**a) enjoys shopping for clothing?**
**b) is a female and enjoys shopping for clothing?**
**c) is a female or enjoy shopping for clothing?**
**d) is a male or a female?**

**b)**   The probability that a person chosen has a certain disease is 0.03. Medical diagnostic tests are available to determine whether the person actually has the disease. If the disease is actually present, the probability that the medical diagnostic test will give a positive result is 0.90. If the disease is not actually present, the probability of a positive test result is 0.02. Suppose that the medical diagnostic test has given a positive result. What is the probability that the disease is actually present? What is the probability of a positive test result?          **(10)**

**Q.4**   Compute the mean, variance, coefficient of skewness and kurtosis of geometric distribution.          **(10)**

**Q.5.a)** Derive the Poisson distribution as the limiting form of the binomial distribution, stating clearly the assumptions you make.

**b)**   Three people each toss a coin and the odd man pays for the coffee. If the coins all show heads or all show tails, they are tossed again. What is the probability that a decision is reached in five tosses or fewer?

                                                                              **(10)**

PAPER: Statistical Computer Packages
Course Code: STAT-309

TIME ALLOWED: 30 mins.
MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*
## OBJECTIVE TYPE

Q1.    Read the following items carefully and encircle the correct option listed below at each item.

*(One mark for each)*

i.    **In SPSS, what is the 'Data Viewer'?**
a) A table summarizing the frequencies of data for one variable.
b) A spreadsheet into which data can be entered.
c) A dialog box that allows you to choose a statistical test
d) screen in which variables can be defined and labelled.

ii.    **How is a variable name different from a variable label?**
a) It is shorter and less detailed.
c) It is longer and more detailed
b) It is abstract and unspecific.
d) It refers to codes rather than variables.

iii.    **What does the operation 'Recode Into Different Variables' do to the data?**
a) Replaces missing data with some random scores.
b) Reverses the position of the independent and dependent variable on a graph.
c) Redistributes a range of values into a new set of categories and creates a new variable.
d) Represents the data in the form of a pie chart.

iv.    **How would you use the drop-down menus in SPSS to generate a frequency    table?**
a) Open the Output Viewer and click: Save As → Pie Chart
b) Click on: Analyze → Descriptive Statistics → Frequencies.
c) Click on: Graphs → Frequencies → Pearson.
d) Open the Variable Viewer and recode the value labels.

v.    **In which sub-dialog box can the Chi Square test be found?**
a) Frequencies: Percentages.
c) Crosstabs: Statistics.
b) Bivariate: Pearson.
d) Gender: Female.

vi.    **There are ......... basic windows in SPSS.**
a) 1            b) 2            c) 3            d) 4

vii.    **In spss we can use ........ types of variables.**
a) 4            b) 6            c) 7            d) 8

viii.    **........ menu allows you to open, print, and save data files and results, to close files and to quit SPSS.**
a) Data            b) File            c) Analyze            d) Transform

ix.    **The SPSS data editor provides ..... number of spread sheets.**
a) 2            b) 3            c) 4            d) None of the above

x.    **The recode command is available in SPSS under ....menu bar.**
a) File            b) Analyze            c) Transform            d) Data

# UNIVERSITY OF THE PUNJAB

**Fifth Semester   2018**
**Examination: B.S. 4 Years Programme**

**PAPER: Statistical Computer Packages**
**Course Code: STAT-309**

**TIME ALLOWED: 2 hrs. & 30 mins.**
**MAX. MARKS: 50**

### *Attempt this Paper on Separate Answer Sheet provided.*
## SUBJECTIVE TYPE
## SHORT QUESTIONS

**Q2.   Define the following:**                                   *(5 marks each)*

   i)     Scales of Measurement and their Types
   ii)    Rules of defining a variable name in SPSS
   iii)   Components of a Computer
   iv)    Types of Windows in SPSS

## SUBJECTIVE

Note: You are required to justify your answers in this section by stating hypothesis, statistics, reference of graphs etc. on your answer sheet.

**Q3.   a)   The following are some output generated through SPSS from a data set containing 56 observations.**

| | | Statistic | Std. Error |
|---|---|---|---|
| Mean | | 6.7080 | .11773 |
| 95% Confidence Interval for Mean | Lower Bound | 6.4721 | |
| | Upper Bound | 6.9440 | |
| 5% Trimmed Mean | | 6.7601 | |
| Median | | 6.6350 | |
| Variance | | .776 | |
| Std. Deviation | | .88102 | |
| Minimum | | 4.12 | |
| Maximum | | 8.26 | |
| Range | | 4.14 | |
| Interquartile Range | | .99 | |
| Skewness | | -.901 | .319 |
| Kurtosis | | 1.710 | .628 |

Normal Q-Q Plot of VAR00001
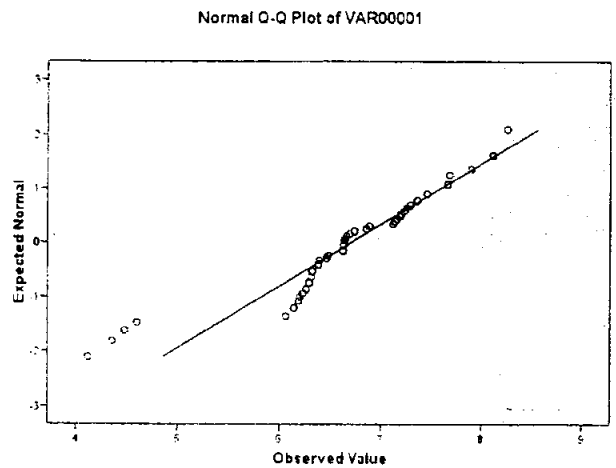


| | Kolmogorov-Smirnov | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| VAR00001 | .170 | 56 | .000 | .905 | 56 | .000 |

| | VAR00001 |
|---|---|
| Chi-Square | 14.000 |
| df | 39 |
| Asymp. Sig. | 1.000 |

| | | | | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Lower | Upper |
| VAR00001 | -2.480 | 55 | .016 | -.5279 | -.0560 |

| | VAR00001 |
|---|---|
| Test Value(a) | 6.64 |
| Cases < Test Value | 28 |
| Cases >= Test Value | 28 |
| Total Cases | 56 |
| Number of Runs | 21 |
| Z | -2.158 |
| Asymp. Sig. (2-tailed) | .031 |

Use the above output and answer the following questions.

   i)     Does the data used in analysis follows a normal distribution?
   ii)    Can we say that the data is randomly distributed?
   iii)   Does the data significantly differing from 7.0?                    (2+2+2)

b) Three sections of the same elementary mathematics course are taught by three teachers A, B and C. The final grades were recorded and the following output is generated through SPSS:

Grades

| Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|
| 1.956 | 2 | 42 | .154 |

| | Grades |
|---|---|
| Chi-Square | .601 |
| df | 2 |
| Asymp. Sig. | .740 |

a Kruskal Wallis Test

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 142.178 | 2 | 71.089 | .217 | .805 |
| Within Groups | 13728.133 | 42 | 326.860 | | |
| Total | 13870.311 | 44 | | | |

i) Test the Homogeniety of Grades among three Teachers.
ii) Can we conclude that there is no significant difference among the Grades of three Teachers? (2+2)

Q4. a) Sixty candidates for scientific posts gave particulars of their university degrees and their hobbies. The degrees were in mathematics, statistics or economics, and the hobbies could be classified roughly as music, craftswork, reading or drama. The following output is generated from the recorded data using SPSS:

| | | | Subject | | | Total |
|---|---|---|---|---|---|---|
| | | | Mathematics | Statistics | Economics | |
| Hobbies | Music | Count | 5 | 2 | 3 | 10 |
| | | Expected Count | 3.5 | 3.2 | 3.3 | 10.0 |
| | | % within Hobbies | 50.0% | 20.0% | 30.0% | 100.0% |
| | | % within Subject | 23.8% | 10.5% | 15.0% | 16.7% |
| | | % of Total | 8.3% | 3.3% | 5.0% | 16.7% |
| | Craftswork | Count | 6 | 8 | 7 | 21 |
| | | Expected Count | 7.4 | 6.7 | 7.0 | 21.0 |
| | | % within Hobbies | 28.6% | 38.1% | 33.3% | 100.0% |
| | | % within Subject | 28.6% | 42.1% | 35.0% | 35.0% |
| | | % of Total | 10.0% | 13.3% | 11.7% | 35.0% |
| | Book Reading | Count | 4 | 6 | 5 | 15 |
| | | Expected Count | 5.3 | 4.8 | 5.0 | 15.0 |
| | | % within Hobbies | 26.7% | 40.0% | 33.3% | 100.0% |
| | | % within Subject | 19.0% | 31.6% | 25.0% | 25.0% |
| | | % of Total | 6.7% | 10.0% | 8.3% | 25.0% |
| | Drama | Count | 6 | 3 | 5 | 14 |
| | | Expected Count | 4.9 | 4.4 | 4.7 | 14.0 |
| | | % within Hobbies | 42.9% | 21.4% | 35.7% | 100.0% |
| | | % within Subject | 28.6% | 15.8% | 25.0% | 23.3% |
| | | % of Total | 10.0% | 5.0% | 8.3% | 23.3% |
| Total | | Count | 21 | 19 | 20 | 60 |
| | | Expected Count | 21.0 | 19.0 | 20.0 | 60.0 |
| | | % within Hobbies | 35.0% | 31.7% | 33.3% | 100.0% |
| | | % within Subject | 100.0% | 100.0% | 100.0% | 100.0% |
| | | % of Total | 35.0% | 31.7% | 33.3% | 100.0% |

| | | Value | Approx. Sig. |
|---|---|---|---|
| Nominal by Nominal | Contingency Coefficient | .218 | .810 |
| N of Valid Cases | | 60 | |

i) What is the percentage of statistics students having craftswork as a hobby from all statisticians?
ii) What percentage of Math students like book reading from all students like book reading?
iii) What is the residual value of Economics students with Drama as a hobby?
iv) Can we conclude that selection of hobbies does not depend upon the subject? (1+1+1+3)

b) A study was conducted to examine the effect of blood group on individual's weight. Two groups of individuals with +ve and –ve blood group were selected randomly from a large population and their weights (kg) were recorded. Some output generated from the recorded data is given below:

| | Group | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Weight | +ve | 20 | 37.3000 | 8.44861 | 1.88917 |
| | –ve | 20 | 39.9000 | 12.11306 | 2.70856 |

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Weight | Equal variances assumed | 3.246 | .080 | -.787 | 38 | .436 | -2.60000 | 3.30231 | -9.28518 | 4.08518 |
| | Equal variances not assumed | | | -.787 | 33.948 | .437 | -2.60000 | 3.30231 | -9.31148 | 4.11148 |

| | Weight |
|---|---|
| Mann-Whitney U | 180.000 |
| Wilcoxon W | 390.000 |
| Z | -.542 |
| Asymp. Sig. (2-tailed) | .588 |
| Exact Sig. [2*(1-tailed Sig.)] | .602(a) |

| | Group - Weight |
|---|---|
| Z | -.691(a) |
| Asymp. Sig. (2-tailed) | .490 |

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

| | | Paired Differences | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | | | | |
| | | Mean | Std. Deviation | Std. Error Mean | Lower | Upper | t | df | Sig. (2-tailed) |
| Pair 1 | Weight - Group | -2.60000 | 14.53634 | 3.25043 | -9.40322 | 4.20322 | -.800 | 19 | .434 |

i) Test the Homogeniety of Variance between the two blood groups.

ii) Can we conclude that there is no significant difference in the average weight between +ve blood group and –ve blood group persons? (2+2)

Q5. An investigator believes that individual's age ($X_1$), income ($X_2$) and savings ($X_3$) are the major factors of insurance policies purchase (Y). The following output is generated from the recorded data using SPSS:

| | | Y | X1 | X2 | X3 |
|---|---|---|---|---|---|
| Y | Pearson Correlation | 1 | .892 | .906 | .925 |
| | Sig. (2-tailed) | | .000 | .000 | .000 |
| | N | 27 | 27 | 27 | 27 |
| X1 | Pearson Correlation | .89 | 1 | .990 | .988 |
| | Sig. (2-tailed) | .000 | | .000 | .000 |
| | N | 27 | 27 | 27 | 27 |
| X2 | Pearson Correlation | .906 | .990 | 1 | .993 |
| | Sig. (2-tailed) | .000 | .000 | | .000 |
| | N | 27 | 27 | 27 | 27 |
| X3 | Pearson Correlation | .925 | .988 | .993 | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | |
| | N | 27 | 27 | 27 | 27 |

| Control Variables | | | Y | X1 |
|---|---|---|---|---|
| X2 & X3 | Y | Correlation | 1.000 | -.313 |
| | | Significance (2-tailed) | . | .128 |
| | | df | 0 | 23 |
| | X1 | Correlation | -.313 | 1.000 |
| | | Significance (2-tailed) | .128 | . |
| | | df | 23 | 0 |

| Control Variables | | | Y | X3 |
|---|---|---|---|---|
| X2 & X1 | Y | Correlation | 1.000 | .552 |
| | | Significance (2-tailed) | . | .004 |
| | | df | 0 | 23 |
| | X3 | Correlation | .552 | 1.000 |
| | | Significance (2-tailed) | .004 | . |
| | | df | 23 | 0 |

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | .936(a) | .877 | .861 | 15.51364 | .763 |

a Predictors: (Constant), X3, X1, X2
b Dependent Variable: Y

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 39349.189 | 3 | 13116.396 | 54.499 | .000(a) |
| | Residual | 5535.477 | 23 | 240.673 | | |
| | Total | 44884.667 | 26 | | | |

a Predictors: (Constant), X3, X1, X2
b Dependent Variable: Y

**Coefficients(a)**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Tolerance | VIF |
| 1 | (Constant) | 387.370 | 44.192 | | 8.766 | .000 | | |
| | X1 | -4.496 | 2.849 | -.859 | -1.578 | .128 | .018 | 55.259 |
| | X2 | -1.211 | 3.096 | -.264 | -.391 | .699 | .012 | 85.076 |
| | X3 | 38.664 | 12.179 | 2.036 | 3.175 | .004 | .013 | 76.687 |

a Dependent Variable: Y

i) Is the correlation between Income and Policies Purchase significant?
ii) Is the correlation between savings and Policies Purchase significant when age and income is controlled?
iii) Fit a regression model $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e$ and interpret the results. Also test the significance of the regression coefficients and the intercept term.
iv) Test the significance of the regression model given in (iii).          (2+2+4+2)

# UNIVERSITY OF THE PUNJAB

**Sixth Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. .....................

**PAPER: Advanced Experimental Design (Theory)**     **TIME ALLOWED: 2 Hrs. & 45 Mints.**

**Course Code: STAT-310   Part – II**                              **MAX. MARKS: 50**

**Attempt this Paper on Separate Answer Sheet provided.**

## SHORT QUESTIONS

**Q2.** Define the following:                                                                     (5x4=20)

    i.     Factorial Design

    ii.    Confounding

    iii.   Youden Square Design

    iv.   Incomplete Block Design

## SUBJECTIVE

**Q3.** Suppose a study is performed to determine the effect of 3 factors (A, B, C). Suppose that each batch of raw material is only large enough to test only 4 treatment combinations. Thus, each replicate of $2^3$ design must be run in 2 blocks. Two replicates are run with ABC confounded in replicate-I and AB confounded in replicate-II. The data is as follows:                              (08)

| Replicate-I | | Replicate-II | |
|---|---|---|---|
| ABC Confounded | | AB Confounded | |
| Block 1 | Block 2 | Block 1 | Block 2 |
| (1)= -3 | a= 0 | (1)= -1 | a= 1 |
| ab= 2 | b=-1 | c= 0 | b= 0 |
| ac= 2 | c=-1 | ab=3 | ac= 1 |
| bc= 1 | abc=6 | abc= 5 | bc= 1 |

Analyze the experiment.

**Q4.** Consider the $2^5$ design in four blocks of eight run each with ACDE and BCE as independent effects chosen to be confounded with blocks. Generate the design. Find the other effects confounded with blocks.                                                                                        (06)

**Q5.** Complete and analyze the following ANOVA table:                                   (08)

| S.O.V | d.f | SS | MS | F-Ratio |
|---|---|---|---|---|
| Varieties (V) | 2 | 0.1781 | - | - |
| Blocks | - | 4.1499 | - | - |
| Error (a) | 10 | 1.3622 | - | - |
| Dates (D) | - | 1.9625 | - | - |
| V X D | - | 0.2105 | 0.0351 | - |
| Error (b) | - | 1.2583 | - | |

**Q6.** Consider the following partially balanced incomplete block design          (08)

| Blocks | Treatment Combinations | | |
|---|---|---|---|
| 1 | 1 | 2 | 3 |
| 2 | 3 | 4 | 5 |
| 3 | 2 | 5 | 6 |
| 4 | 1 | 2 | 4 |
| 5 | 3 | 4 | 6 |
| 6 | 1 | 5 | 6 |

Verify the following relationships among the parameters

$$p_{11}^1 + p_{12}^2 = n_1$$
$$n_1 p_{12}^1 = n_2 p_{11}^2$$
$$p_{21}^1 + p_{22}^1 = n_2$$
$$n_1 p_{11}^1 = n_2 p_{12}^2$$

$$p_{11}^1 + p_{12}^1 = n_1 - 1$$
$$n_1 \lambda_1 + n_2 \lambda_2 = r(k - 1)$$
$$p_{21}^1 + p_{22}^2 = n_2 - 1$$
$$n_1 + n_2 = a - 1$$

# UNIVERSITY OF THE PUNJAB

## Sixth Semester - 2018
## Examination: B.S. 4 Years Programme

Roll No. ..................

PAPER: Advanced Experimental Design (Theory)  TIME ALLOWED: 15 Mints.
Course Code: STAT-310   Part – I (Compulsory)   MAX. MARKS: 10

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

Q1.   Read the following items carefully and encircle the correct option listed below at each item.

**(1x10=10)**

i)   A .... Experiment is concerned with examining simultaneously a number of basic treatments or factors, each of which take a number of possible forms or levels:
   a) Factorial       b) Youden Square       c) Latin Square       d) None of these

ii)   In a random effect model under factorial design with 'n' observations and 'a' levels of factor A and 'b' levels of factor B, then E(MSA)= :
   a) $\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2$   b) $\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2$   c) $\sigma^2 + n\sigma_{\alpha\beta}^2$   d) $\sigma^2$

iii)   In a random effect model under factorial design with 'n' observations and 'a' levels of factor A and 'b' levels of factor B, then E(MSB)= :
   a) $\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2$   b) $\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2$   c) $\sigma^2 + n\sigma_{\alpha\beta}^2$   d) $\sigma^2$

iv)   In a random effect model under factorial design with 'n' observations and 'a' levels of factor A and 'b' levels of factor B, then E(MSAB)= :
   a) $\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2$   b) $\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2$   c) $\sigma^2 + n\sigma_{\alpha\beta}^2$   d) $\sigma^2$

v)   In a random effect model under factorial design with 'n' observations and 'a' levels of factor A and 'b' levels of factor B, then E(MSE)= :
   a) $\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2$   b) $\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2$   c) $\sigma^2 + n\sigma_{\alpha\beta}^2$   d) $\sigma^2$

vi)   In a design of four factors confounded in 8-blocks, the number of interactions confounded with blocks are:
   a) 2           b) 3           c) 7           d) None of these

vii)   In PBIB design, pair of treatments i and j that occur together in the same block are called:
   First Associates       b) Second Associates   c) Parameters of first kind       d) None of these

viii)   Consider 26-2 design. Suppose we choose P=ABCE and Q=ACDF then the aliases for A will be:
   a) BCE,CDF,ABDEF       b) BCE,CDF   c) ABCE,ACDF       d) None of these

ix)   When the factors were analyzed according to their importance in the design, the following design is useful:
   a) Factorial       b) Fractional Replication       c) Split Plot       d) Incomplete Block Design

x)   In.......... Design, blocks do not contain full set of treatments in each and every block.
   a) Factorial       b) Fractional Replication       c) Split Plot       d) Incomplete Block

# UNIVERSITY OF THE PUNJAB

### Sixth Semester - 2018
### <u>Examination: B.S. 4 Years Programme</u>

**PAPER: Advanced Sampling Techniques (Theory)**   **TIME ALLOWED: 2 Hrs. & 45 Mints.**
**Course Code: STAT-312   Part – II**   **MAX. MARKS: 50**

## Attempt this Paper on Separate Answer Sheet provided.

**Q. No. 2.**   Write a short note on the following:   (5x4=20)
- i.    Advantages of Cluster Sampling
- ii.   Combined Regression Estimator
- iii.  Probability Proportional to Size Sampling
- iv.   Two-Phase Sampling.

**Q. No. 3.**   In ratio estimator, show that the co-efficient of variation of $\hat{R}$, $\hat{Y}_R$, $\widehat{\bar{Y}}_R$ are all equal.   (08)

**Q. No. 4.**   In regression estimator, show that the leading term in bias of $\bar{y}_{lr}$ turns out to be   (07)

$$\frac{-(1-f)}{n}\frac{E e_i (x_i - \bar{X})^2}{S_x^2}$$

**Q. No. 5.**   Calculate an expression for the optimum value of $M$, the size of cluster, by using the cost function   (08)

$$C = c_1 nM + c_2 \sqrt{n}$$

where $C$ is the total cost of the survey.

**Q. No. 6.**   If $n$ units and $m$ subunits from each chosen unit are selected by simple random sampling, prove that $v(\bar{\bar{y}}) = \frac{1-f_1}{n}s_1^2 + \frac{f_1(1-f_2)}{mn}s_2^2$ is an unbiased estimate of $V(\bar{\bar{y}})$.   (07)

| | |
|---|---|
| **PAPER: Advanced Sampling Techniques (Theory)** | **TIME ALLOWED: 15 Mints.** |
| **Course Code: STAT-312   Part – I (Compulsory)** | **MAX. MARKS: 10** |

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

**Q. No. 1: Encircle the correct option.** (10)

1. The ---------- is the ratio of the variance of the estimate obtained from the (more complex) sample to the variance of the estimate obtained from a Simple random sample of the same number of units.
   - A. Design effect
   - B. Ratio estimator
   - C. Regression estimator
   - D. Efficiency

2. Which of the following will give a more "accurate" representation of the population from which a sample has been taken?
   - A. A large sample based on the convenience sampling technique
   - B. A small sample based on simple random sampling
   - C. A large sample based on simple random sampling
   - D. A small cluster sample

3. The standard deviation of the sampling distribution is called---------.
   - A. Sampling Error
   - B. Standard Error
   - C. Response Error
   - D. Non-response Error

4. Sampling Error is reduced by -------------.
   - A. reducing incomplete surveys
   - B. increasing response rates
   - C. increasing the sample size
   - D. conducting callbacks

5. Suppose that, in cluster sampling $s_w^2$ represents the variance within the clusters and $s_b^2$ between clusters. What is the relation between $s_w^2$ and $s_b^2$?
   - A. $s_w^2 = s_b^2$
   - B. $s_w^2 \geq s_b^2$
   - C. $s_w^2 \leq s_b^2$
   - D. None of the above

6. In ratio estimator the relation between $y_i$ and $x_i$ is straight line passes through
   - A. Origin
   - B. X-axis
   - C. Y-axis
   - D. None of above.

7. In general the Ratio estimate has a bias of order -------------------------------
   - A. $\frac{1}{n^2}$
   - B. $\frac{1}{n}$
   - C. $\sqrt{\frac{1}{n}}$
   - D. $\frac{1}{\sqrt{n}}$

8. It is helpful to use a multi-stage cluster sample when:
   - A. The population is widely dispersed geographically.
   - B. You have limited time and money available for travelling.
   - C. You want to use a probability sample in order to generalize the results.
   - D. All of the above.

9. Non-response in surveys mean:
   - A. Non-availability of respondent
   - B. Non-return of questionnaire by the respondents
   - C. Refuse to give information by the respondents
   - D. All of the above

10. Double sampling is also called
    - A. Two-phase sampling
    - B. Two-stage sampling
    - C. Multistage sampling
    - D. None of above.

# UNIVERSITY OF THE PUNJAB

**Sixth Semester - 2018**

**Examination: B.S. 4 Years Programme**

**PAPER: Multivariate Techniques (Theory)**
**Course Code: STAT-314   Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**
**MAX. MARKS: 50**

---

## Attempt this Paper on Separate Answer Sheet provided.

**Q2.** Define the following:  (5x4=20)

   i)    Decomposition of a Matrix
   ii)   Multivariate Normal Distribution
   iii)  Generalized Variance
   iv)   Data Matrix

**Q3.** If $X$ denotes a $(p \times 1)$ column vector of random variables, $\mu$ is a column vector of constants and $\Sigma$ is a positive definite matrix then find the value of $k$ such that

$$f(X) = k \exp\left[-\frac{1}{2}(X-\mu)' \Sigma^{-1}(X-\mu)\right]$$ is a pdf $X$. Also find $E(X)$.   *(08)*

**Q4.** Let $W \sim W_P(f, \Sigma, M)$. If $C$ is any $(p \times 1)$ vector of constants, then show that

$$C'WC \sim \sigma^2 \chi^2\left(f, \delta^2\right).$$   *(06)*

**Q5.** Find the Spectral decomposition of the following matrix.   *(08)*

$$\begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$$

**Q6.** The random vector $X' = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix}$ has a Multivariate Normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$ given by:

$$\mu = \begin{bmatrix} 5 \\ 5 \\ 10 \\ 10 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 & 4 & 4 \\ & 20 & 10 & 10 \\ & & 20 & 10 \\ & & & 20 \end{bmatrix}$$

Suppose $Y_1' = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and $Y_2' = \begin{bmatrix} X_3 & X_4 \end{bmatrix}$ are the sub-vectors of $X$ then find

a)  $E\left(Y_1/Y_2\right)$       b)  $Cov\left(Y_1/Y_2\right)$   *(08)*

**PAPER: Multivariate Techniques (Theory)**
**Course Code: STAT-314    Part – I (Compulsory)**

**TIME ALLOWED: 15 Mints.**
**MAX. MARKS: 10**

### Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

Q1. Read the following items carefully and encircle the correct option listed below at each item.

**(1x10=10)**

i. **What type of statistical techniques focus upon and bring out the structure of simultaneous relationships among three or more variables?**
   a) Bivariate        b) Multivariate        c) Parametric        d) Non-parametric

ii. **A data recorded with 'p' variables on 'n' objects in array shape is known as:**
   a) Data Matrix        b) Covariance Matrix        c) Correlation Matrix    d) None of these

iii. **A relation between population covariance matrix and correlation matrix is?**
   a) $\rho = D^{-1} \Sigma D^{-1}$        b) $\rho = D^{-1} \Sigma^{-1} D^{-1}$        c) $\rho = D \Sigma^{-1} D$        d) None of these

iv. **Let X be a p-dimensional random vector with $E(X) = \mu$ and $Cov(X) = \Sigma$. Also A is a non-zero matrix and b is a vector of constants then:**
   a) $Cov(AX+b) = A\Sigma A'$   b) $Cov(AX+b) = A\Sigma A' + b$   c) $Cov(AX+b) = A\Sigma^{-1} A'$    d) None of these.

v. **The covariance matrix $\Sigma$ contains 'p' variances and ....covarinces.**
   a) $\frac{1}{2} p(p+1)$        b) $\frac{1}{2} p(p-1)$        c) $p(p+1)$        d) $(p+1)$

vi. **A single numerical value used to describe the variation of $\Sigma$ is called:**
   a) Sample Variance        b) Variance-Covariance        c) Generalized Variance        d) None of these

vii. **Generalized Variance may be computed as:**
   a) $\frac{1}{2} p(p-1)$        b) $|\Sigma|$        c) $(n-1)\Sigma$        d) None of these

viii. **Wishart Distribution is a generalization of:**
   a) Normal Distribution   b) t-distribution        c) Chi-Square        d) None of these

ix. **If A is $W_p(n, I_{p \times p})$ and B is a $(q \times p)$ matrix, such that $BB' = I_{q \times q}$ then BAB' is:**
   a) $W_p(n, I_{p \times p})$        b) $W_q(n, I_{q \times q})$        c) $W_q(n, I_{p \times q})$        d) None of these

x. **If A is $W_p(n, \Sigma, M)$ then $E(A) =$**
   a) $n\Sigma$        b) $n\Sigma + M$        c) $n\Sigma + M'M$        d) None of these

# UNIVERSITY OF THE PUNJAB

**Sixth Semester - 2018**

**Examination: B.S. 4 Years Programme**

**PAPER: Probability Distributions (Theory)**
**Course Code: STAT-316   Part – II**

**TIME ALLOWED: 2 Hrs. & 45 Mints.**
**MAX. MARKS: 50**

**Attempt this Paper on Separate Answer Sheet provided.**

## SUBJECTIVE

Q.2.  a)  Find the mean and variance of Pareto distribution.  **(5)**

b)  If $X_1, X_2 .....Xn$ are 'n' independent r.v's having exponential distribution having same parameter '$\theta$' i.e. $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$  **(5)**

c)  Find the pdf of $Y = X_1 + X_2 + ...... + Xn$

If f is a F distributed random variable with $n$ & $n_2$ d.f then show that  **(5)**

$$y = \left(1 + \frac{n_1}{n_2} F\right)^{-1}$$ has beta distribution of first kind.

d)  If X is a random variable with p.d.f as  **(5)**

$$f(x) = \frac{1}{2} \qquad -1 \leq x \leq +1$$

Find the p.d.f of $Y = X^2$

Q.3.  Derive $\chi^2$ – distribution.  **(10)**

Q.4.  Find the coefficient of skewness and kurtosis of lognormal distribution.  **(10)**

Q.5.  Let $X_1$ and $X_2$ be two independent standard normal variables. Then obtain the distribution of $Z = \frac{X_1}{X_2}$.  **(10)**

**PAPER: Probability Distributions (Theory)**   **TIME ALLOWED: 15 Mints.**
**Course Code: STAT-316   Part – I (Compulsory)**   **MAX. MARKS: 10**

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

**Q.1.   Tick the correct option.**                                   **(10)**

a)  Probability distribution having shape of bell and in which values of mean lies in center of probability distribution is classified as

   i.   continuous distribution
   ii.  normal distribution
   iii. discrete distribution
   iv.  hyper geometric distribution

b)  Successful life of product, time, weight and height are classified as

   i.   continuous random variable
   ii.  discrete random variable
   iii. continuous waiting time variable
   iv.  continuous hyper geometric variable

c)  Which of the following do the normal distribution and the exponential density function have in common?
   i.   both are bell-shaped
   ii.  both are symmetrical distributions
   iii. both approach zero as x approaches infinity
   iv.  all of the above are features common to both distributions

d)  Which of the following statement is not true for an exponential distribution with parameter $\lambda$?
   i.   mean = $1 / \lambda$
   ii.  standard deviation = $1 / \lambda$
   iii. the distribution is completely determined once the value of $\lambda$ is known
   iv.  the distribution is a two-parameter distribution since the mean and standard deviation are equal

e)  _____ is referred to rectangular distribution.

   i.   Normal distribution
   ii.  Uniform distribution
   iii. Weibull distribution
   iv.  Lognormal distribution

f)  If all values move towards one tail of a distribution then this scenario results in

   i.   width of distribution
   ii.  height of distribution
   iii. lengthening tail
   iv.  shortening tail

g)  The Joint Cumulative Density Function (CDF) _____
   i. Is a non-negative function
   ii. Is a non-decreasing function of x & y planes
   iii. Is always a continuous function in x y plane
   iv. All of the above

h)  What is the value of an area under the conditional PDF ?
   i    Greater than '0' but less than '1'
   ii.  Greater than '1'
   iii  Equal to '1'
   iv.  Infinite

i)  Which of the following distributions is suitable to measure the length of time that elapses between the arrival of cars at a petrol station pump?
   i.   normal
   ii.  uniform.
   iii. Weibull
   iv   exponential

j)  F=_____
   i.   Square of t
   ii.  Square root of t
   iii. Cube of t
   v.   Cube root of t

# UNIVERSITY OF THE PUNJAB

**Sixth Semester - 2018**

**Examination: B.S. 4 Years Programme**

Roll No. .....................

PAPER: FORTRAN Computer Language

Course Code: STAT-318   Part – II

TIME ALLOWED: 2 Hrs. & 45 Mints.

MAX. MARKS: 50

**Attempt this Paper on Separate Answer Sheet provided.**

Q.2.  a)  Find the final value of M for each of the following program segment   **(5)**
if     J = 2, K = 5.

i)
```
M = 5
IF (J.LT.K) then
J = J + 5
M = M + 3
IF (J.LT.8) then
M = M + 10
ELSE
M = M + 20
ENDIF
```

ii)
```
M = 5
IF (J.LT.K) then
J = J + 5
M = M + 3
ELSEIF (J.LT.8) then
M = M + 10
ELSE
M = M + 20
ENDIF
```

b)  Find error if any in the following FTN statements.   **(5)**

i)  3.14= PI

ii)  X * Y = Z

iii)  −K = I * * 2

iv)  X = SQRT (A + B)

v)  IF (X = Y) GOTO 100

c)  Write a note on FTN worksheet   **(5)**

d)  Write the following mathematical expression into Fortran   **(5)**
expressions.

i) $\dfrac{Sin\,x}{|y| + Cos\,Z}$    ii) $e^{|x-y|} + x$    iii) $\dfrac{e^{x+y}}{x+y}$

Q.4.  a)  Write a FTN program to read in an integer n > 2 and determine if n   **(5+5)**
is a prime number

b)  Write a FTN programme to find sum of the following series

$$\frac{2}{3} + \frac{4}{5} + \frac{6}{7} + \ldots\ldots\ldots\ldots + \frac{N-1}{N}$$

Q.5.  a)  Write a FTN which computers and print the sum of odd integers   **(6+4)**
and also prints the product of even integers using Do-loop.

b)  Differentiate between conditional GoTo statement and computed
GoTo statement.

Q.6.  a)  Write a FTN program to calculate mean, variance and S.D. of a set   **(5)**
of N numbers ($x_1, x_2, \ldots\ldots, x_n$).

b)  Write a FTN program to calculate sum of the diagonal elements of   **(5)**
a square matrix.

# UNIVERSITY OF THE PUNJAB
## Sixth Semester - 2018
### Examination: B.S. 4 Years Programme

Roll No. ................

PAPER: FORTRAN Computer Language
Course Code: STAT-318    Part – I (Compulsory)

TIME ALLOWED: 15 Mints.
MAX. MARKS: 10

## Attempt this Paper on this Question Sheet only.

**Please encircle the correct option. Each MCQ carries 1 Mark. This Paper will be collected back after expiry of time limit mentioned above.**

**Q.1.**    Tick the correct option.    (10)

a)  Suppose A = 2.7, B = 3.5, J = 3 and K = –2, then the value of FTN statement
L = A + J * K * + 2 + B will be
   i)    18                    iii) 18.2
   ii)   42.2                  iv) None of the above

b)  Suppose J = 3 and K = 5. The final value of J
      IF (J – K) 10, 10, 20
   10    J = K
   20    J = J + 2
   is
   i)    5                     ii)  7
   iii)  20                    iv)  15

c)  F-specification is used for
   i)    Real number           ii)  Integer numbers
   iii)  both 'a' and 'b'       iv)  None of the above

d)  To terminate a block IF statement, _____ statement is used
   i)    ENDIF                 ii)  Continue
   iii)  GOTO                  iv)  STOP

e)  If x=2.0, a=2.0 and b=4.0 what is the value of y is_____ where
$y = a*x + b**2/x$
   i)    8.0                   ii)  16.0
   iii)  12.0                  iv)  32.0

f)  " / " in FTN means
   i) Skip one column          ii) Skip one row
   iii) Both a. and b.         iv) None of the above

g)  What is the purpose of 6$^{th}$ column in FTN worksheet
   i)    used for continuation of previous line
   ii)   just a jump label
   iii)  just a format label
   iv)   None of the above

h)  After compiling the FTN statement $y = c/d + a*x**2.5$ which operation is performed first by computer
   i)    /                     ii)  *
   iii)  +                     iv) **

i)  In which column does the code start in FTN fixed format
   i)    6                     ii)  9
   iii)  1                     iv)  7

j)  If x=1.0, y=2.0 and w=3.0 the value of $z = 2.0(x(y + 3.0) + w)$ will be
   i)    12.0                  ii)  16.0
   iii)  34.0                  iv)  not determined cue to o FTN syntax error

# UNIVERSITY OF THE PUNJAB

Roll No. ..................

## Seventh Semester 2018
## Examination: B.S. 4 Years Programme

PAPER: Statistical Inference-I (Theory)
Course Code: STAT-401

TIME ALLOWED: 30 mins.
MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*

**Q. No. 1: Encircle the correct option.**

1. The $1 - \alpha$ is the probability of ---------------------
   A. Type-I Error
   C. Acceptance Region
   B. Rejection Region
   D. Type-II Error

2. -------------------- is the property by which we compare an estimator with the other on the basis of its variances.
   A. Unbiasedness
   C. Efficiency
   B. Consistency
   D. Sufficiency

3. What would happen (other things equal) to a confidence interval if you calculated a 99 percent confidence interval rather than a 95 percent confidence interval?
   A. It will be narrower
   C. The sample size will increase
   B. It will not change
   D. It will become wider

4. The estimator which has the variance equal to -------------------- is most efficient estimator.
   A. Mean Square Error
   C. Population mean
   B. Cramer's Rao Lower bound
   D. None of the above

5. As a general rule, researchers tend to use ---------- percent confidence intervals.
   A. 99%
   C. 50%
   B. 95%
   D. None of above

6. When the researcher rejects a true null hypothesis, a ------------- error occurs
   A. Type-I Error
   C. Type-III Error
   B. Type-II Error
   D. Type-B Error

7. The statistic whose conditional distribution is -------------------- of parameter is called sufficient statistics.
   A. Independent
   C. Conditionally dependent
   B. Not independent
   D. None of the above

8. Criteria to check a point estimator to be good are ---------------------
   A. Unbiasedness
   C. Efficiency
   B. Consistency
   D. All above

9. Which of the following is the researcher usually interested in supporting when he or she is engaging in hypothesis testing?
   A. The alternative hypothesis
   C. Both the alternative and null hypothesis
   B. The null hypothesis
   D. Neither the alternative or null hypothesis

10. The mean of the sample means is ------------------
    A. A biased estimator of the population
    C. Neither biased nor unbiased
    B. An unbiased estimator of the population mean
    D. None of the above

# UNIVERSITY OF THE PUNJAB

## *Attempt this Paper on Separate Answer Sheet provided.*

**Attempt all questions.**

**Q. No. 2.** Write a short note on the following: (5 each)
    i.    Weak and strong consistency
    ii.   Regularity conditions
    iii.  Loss Function
    iv.  Bayes Estimator

**Q. No. 3.** Let $X_1, X_2, \ldots \ldots, X_6$ is an identically independently distributed random sample (07) from normal distribution with mean $\mu$ and variance $\sigma^2$. Determine 'c' such that $c[(X_1 - X_2)^2 + (X_3 - X_4)^2 + (X_5 - X_6)^2]$ is an unbiased estimator of $\sigma^2$.

**Q. No. 4.** Let $y_1 < y_2 < \cdots < y_n$ denote the order statistic of a random sample $X_1, X_2, \ldots. X_n$ (07) from the distribution that has p.d.f

$$f(X; \theta) = e^{-(x-\theta)} \qquad \theta < x < \infty, \qquad -\infty < \theta < +\infty$$

Assume $n = 3$ and show that $\max(X_i)$ is not a sufficient statistic for $\theta$.

**Q. No. 5.** If '$t$' is a Minimum Variance Unbiased Estimator of $\tau(\theta)$ with variance $v_1$ (08) and $T_1 \& T_2$ are any two unbiased estimators of $\tau(\theta)$ with finite variances.

If either $T_1 or T_2$ is minimum variance estimator. Then show that correlation co-efficient between $T_1 and T_2$ is $E^{1/2}$ (where '$E$' is the relative efficiency of $T_1 and T_2$).

**Q. No. 6.** Show that if minimum variance estimator exists, it would essentially be unique (08) irrespective of whether any bound is attained.

**PAPER: Basic Econometrics (Theory)**     **TIME ALLOWED: 2 hrs. & 30 mins.**
**Course Code: STAT-403**                  **MAX. MARKS: 50**

## *Attempt this Paper on Separate Answer Sheet provided.*

**Q.2.** Discuss the following.                                               **(20)**

    (i) Properties of MLE of error variance in G.L. Regression.

    (ii) Stepwise regression

    (iii) Gauss Markov's theorem

    (iv) Functions of Econometrics

    (v) Restricted L.S. estimation

**Q.3.** For GL Regression $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$, develop the testing procedure to test some **(08)** regression co-efficient (not all).

**Q.4.** Parameters of the model $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ follow the linear relations $R\underline{\beta} = \underline{\gamma}$. Obtain least **(08)** square estimate of $\underline{\beta}$ under the given relation. Also find variance covariance matrix of the derived estimators.

**Q.5.** We are given the following calculations such that variables follow the relation **(08)**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

$n = 5,$     $\Sigma X_1 = 15,$     $\Sigma X_2 = 25,$     $\Sigma Y = 20,$     $\Sigma X_1^2 = 55,$     $\Sigma X_2^2 = 129,$

$\Sigma Y^2 = 108,$     $\Sigma X_1 Y = 76,$     $\Sigma X_2 Y = 109,$     $\Sigma X_1 X_2 = 81$

Test the hypothesis that Co-efficient of $X_1$ and $X_2$ are equal in magnitude but opposite in sign.

**Q.6.** If the sample values of X in the linear model $Y = \alpha + \beta X + u$, have zero mean, show **(06)** that the covariance of the least squares estimates of $\alpha$ and $\beta$ is zero.

# UNIVERSITY OF THE PUNJAB

Roll No. .............

### Seventh Semester 2018
### Examination: B.S. 4 Years Programme

**PAPER: Basic Econometrics (Theory)**　　　　**TIME ALLOWED: 30 mins.**
**Course Code: STAT-403**　　　　　　　　　　**MAX. MARKS: 10**

## *Attempt this Paper on this Question Sheet only.*

Q.1.　State whether the following statements are true, false or uncertain.　　**(10)**

(i)　OLS estimation requires that error variances are equal.

(ii)　MLE of error variances is BLUE.

(iii)　Total sum of squares follows Normal distribution.

(iv)　Ridge regression estimates have smaller variances as compare to OLS estimators.

(v)　For comparison of two error variances of G.L. regressions, CHOW test is applied.

(vi)　Restricted estimators are more efficient as compare to unrestricted estimators.

(vii)　Co-efficient of determination lies between -1 and +1.

(viii)　$H_0 : \beta = 0$ in S.L. Regression $Y = \alpha + \beta X + \in$ can be tested by ANOVA method.

(ix)　For comparison of two G.L. regressions, chi-square test is applied.

(x)　In stepwise Regression, all the parameters are estimated at a time.

# UNIVERSITY OF THE PUNJAB

Roll No. ...................

## Seventh Semester 2018
### Examination: B.S. 4 Years Programme

PAPER: C++ Computer Programming Language     TIME ALLOWED: 30 mins.
Course Code: STAT-405     MAX. MARKS: 10

*Attempt this Paper on this Question Sheet only.*

Question1.    Objective Part

    a.   Fill in the Blanks

      i.   Let
        int variable = 0;
        variable *= 5
        cout<< variable %17;       output will be _____

      ii.   Let
        int variable ;
        if(variable >10)
             cout<< "i am good today";
        else
             cout<<"i am sick today"
        output will be _____

      iii.   for(int i = 0; i < 5 ; i--)
             cout<<"wao";
        How many time "Wao" will be printed   output will be _____
        After loop what is value of i       output will be _____

      iv.   If we have a function prototype as following:
        int * prime( int );
        name of function is _____

    b.   Choose whether following statement is true or false

      i.   Float can be added to integers       (True/False)

      ii.   function can call another function       (True/False)

      iii.   Union can act as if variable having different data-types       (True/False)

      iv.   (float) variable will make variable float       (True/False)

      v.   Accessing RAM is faster as compared to Access of HardDisk       (True/False)

# UNIVERSITY OF THE PUNJAB

**Seventh Semester 2018**
**Examination: B.S. 4 Years Programme**

Roll No. .......................

**PAPER: C++ Computer Programming Language**       **TIME ALLOWED: 2 hrs. & 30 mins.**
**Course Code: STAT-405**       **MAX. MARKS: 50**

## *Attempt this Paper on Separate Answer Sheet provided.*

Question2.   Write short note of following                                    (20) Marks

    a.   % of negative number

    b.   String

    c.   Macros

    d.   break keyword

    e.   while loop vs do-while loop


Question3.   Write code to implement following programs                ( 7+9+7+7) Marks

    a.   Write a program to make calculator

    b.   Write a function that inputs array of 50 integers and return their mean and sum using structure

    c.   Write code to check whether a number is even or odd

    d.   Write a code to find length of string

# UNIVERSITY OF THE PUNJAB

Roll No. ..............

## Seventh Semester 2018
## Examination: B.S. 4 Years Programme

**PAPER: Research Methodology**

**Course Code: STAT-406**

**TIME ALLOWED: 30 mins.**

**MAX. MARKS: 10**

---

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE

Q1.     Select the suitable answer.

i.      Research is_____
   a)  A lab experiment
   b)  A report
   c)  A systematic Enquiry
   d)  A procedure

ii.     When Ranks are assigned to objects based on their properties or characteristics then the level of measurement is described as
   a)  Ordinal measurement
   b)  Nominal measurement
   c)  Ratio measurement
   d)  interval measurement

iii.    Which of the following is not an essential element of report writing?
   a)  Research Methodology
   b)  Reference
   c)  Conclusion
   d)  None of the above

iv.     Which of the following is non-probability sampling?
   a)  Snowball
   b)  Random
   c)  Cluster
   d)  Stratified

v.      You wish to make precise estimate about the characteristics of a population of individuals. You should use
   a)  Quota Sampling
   b)  Probability Sampling
   c)  Accidental Sampling
   d)  snowball Sampling

**(P.T.O.)**

vi.     Statistics is used by researchers to
- a) Analyze the empirical data collected in a study
- b) Operationally defines their varialbes
- c) Ensure the study comes out the way it was intended

vii.     A literature review is based on the assumption that
- a) Copy from the work of others
- b) Knowledge accumulates and learns from the work of others
- c) Knowledge dis accumulates
- d) None of the above

viii.     Which of the following is not a major method of data collection
- a) Questionnaires
- b) Focus groups
- c) Correlation method
- d) Secondary data

ix.     Question that consists of two or more questions joined together is called a:
- a) Double barreled question
- b) General question
- c) Accurate question
- d) Confusing question

x.     The number of questionnaires returned or completed divided by the total number of eligible people who were contacted or asked to participate in the survey is called the
- a) Response rate
- b) Participation rate
- c) Inflation rate
- d) Exchange rate

# UNIVERSITY OF THE PUNJAB

**Seventh Semester 2018**
**Examination: B.S. 4 Years Programme**

Roll No. ....................

**PAPER: Research Methodology**
**Course Code: STAT-406**

**TIME ALLOWED: 2 hrs. & 30 mins.**
**MAX. MARKS: 50**

*Attempt this Paper on Separate Answer Sheet provided.*

## SUBJECTIVE

Q1. Write short notes on **(20)**
   a) Stratified sampling
   b) Measurement error
   c) Processing of survey data
   d) Types of data collection

Q2. What is a questionnaire? Which factors should be considered in designing of a questionnaire? Explain in detail. **(15)**

Q3. Define Scales and its types. Also explain different tupes of rating scales by giving at least two examples each. **(15)**

# UNIVERSITY OF THE PUNJAB

## *Attempt this Paper on Separate Answer Sheet provided.*

**Note: Attempt all questions.**

**Question No.2.** Describe the following. ( 5 × 4 = 20 )

a) Derive the condition of stationarity for an AR(1) process.
b) Define autocorrelation and write down the properties of autocorrelation coefficient.
c) Find autocorrelation function of MA(2) process given by $Y_t = Z_t + 0.7Z_{t-1} - 0.2Z_{t-2}$ .
d) Show that for an ARMA($p, q$) process, $\varphi(B) \rho_k = 0$     ; for k≥ $q + 1$, where q is the order of MA component.

**Question No. 3. a)** The first ten sample autocorrelation coefficients of 400 random numbers are given below; ( 05 )

| K | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|----|
| $r_k$ | 0.02 | 0.05 | -0.09 | 0.08 | -0.02 | 0.0 | 0.12 | 0.06 | 0.02 | -0.08 |

Is there any evidence of non-randomness? Give reasons for your answer.

**b)** Define Seasonal ARIMA model, write down the model for **SARIMA ( 1,0,1)× (0, 0, 2)₆** ( 05 )

**Question No. 4.** Show that the ACF of the following AR(2) stationary process ( 10 )

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t$$

is given by,

$$\rho_K = \left(\frac{1}{\sqrt{2}}\right)^{|K|} \left|Cos\frac{|K\pi|}{4} + \frac{1}{3}Sin\frac{|K\pi|}{4}\right| \qquad ; for\ K = 0, \pm 1, \pm 2, \pm 3, ....$$

**Question No. 5.** A time series model is given below , ( 10 )

$$X_t = 0.6X_{t-1} + Z_t - 0.7Z_{t-1}$$

a) Identify the model.
b) Check stationarity and invertibility of the model.
c) Obtain first three $\psi$ – weights of the model when expressed as an MA model.
d) Obtain first three $\pi$ - weights of the model when expressed as an AR model.

# UNIVERSITY OF THE PUNJAB

Roll No. ..............

**Seventh Semester 2018**
**Examination: B.S. 4 Years Programme**

**PAPER: Time Series Analysis-I**                    **TIME ALLOWED: 30 mins.**
**Course Code: STAT-411**                                **MAX. MARKS: 10**

### *Attempt this Paper on this Question Sheet only.*

**Question No.1.** Four options are given against each statement, encircle the correct option.     ( 10×1 )

1) What does autocovariance measure?

    **a)** Linear dependence between multiple points on the different series observed at different times
    **b)** Quadratic dependence between two points on the same series observed at different times
    **c)** Linear dependence between two points on different series observed at same time
    **d)** Linear dependence between two points on the same series observed at different times

2) Consider the following AR(1) model with the disturbances having zero mean and unit variance.
$Y_t = 0.4 + 0.2Y_{t-1} + Z_t$. Variance of $Y_t$ will be.

    **a)** 1.042
    **b)** 1.442
    **c)** 1.202
    **d)** 2

3) Second differencing in time series can help to eliminate which trend?

    **a)** Quadratic trend
    **b)** Linear trend
    **c)** Both a) & b)
    **d)** None of the above

4) Which of the following process is stationary?

    **a)** An MA(1) process with $\theta = -1.4$
    **b)** $Y_t = 12.3 + 1.2Y_{t-1} + Z_t$
    **c)** $(1 - 0.2B)Y_t = (1 - 0.5B)Z_t$
    **d)** Both a) & c)

5) Which statement about AR(2) process is always true?

    **a)** The process is invertible
    **b)** The process is stationary
    **c)** The theoretical ACF $\rho_k = 0$, for all K>2
    **d)** The theoretical PACF $\rho_{kk}$ decays exponentially or according to a sinusoidal pattern.

**P.T.O.**

6) In the acronym "SARIMA," what does the "S" stands for?

a) Seasonal

b) Stationary

c) Special

d) Specified

7) What is the name of the process defined by,

$$(1 + 0.6B)(1\text{-}B)Y_t = (1 - 0.9B)^2 Z_t$$

a) $ARIMA(1,1,2)$

b) $ARIMA(2,1,1)$

c) $ARIMA(1,2,1)$

d) None of the above

8) Square root transformations are usually applied to,

a) Stabilize Mean

b) Stabilize variance

c) Remove Trend

c) None of the above.

9) Under normality, what is a valid interpretation of the partial autocorrelation $\rho_{kk}$?

a) It measures the auto correlation in the data $Y_1, Y_2, \dots, Y_n$ after taking kth differences

b) It equals the variance of the large-sample distribution of $r_k$

c) It is the correlation of the first k residuals in an ARIMA model fit

d) It is the correlation between $Y_t$ & $Y_{t-k}$, after removing the linear effects of variables between $Y_t$ & $Y_{t-k}$

10) The first difference of a Random walk process results as,

a) An invertible MA process

b) A stationary AR process.

c) A purely random process.

d) None of the above.

# UNIVERSITY OF THE PUNJAB

Roll No. ..................

### Seventh Semester 2018
### Examination: B.S. 4 Years Programme

**PAPER: Operation Research**          **TIME ALLOWED: 30 mins.**
**Course Code: STAT-412**              **MAX. MARKS: 10**

*Attempt this Paper on this Question Sheet only.*

## OBJECTIVE

**Q.1    Tick the correct answer for the following multiple choice questions.**

I.    The objective function in stochastic model have ................. variables.
   a. Fixed    b.    random    c.    control    d. both a and b

II.    The decision variables controlled by the decision maker also known as
   a. Dependent    b. Independent    c.    unknown    d. both a and b

III.    The physical models are used in operation research for
   a. Manipulation   b. prediction   c. graphical representation d. extent of generality

IV.    Flight simulators are used by airlines to train pilot in operation research by the following model.
   a. Analogue model    b. physical model    c. iconic models d. both b and c

V.    Preparation of flood action plan is an example of practical problem in
   a. Operation research    b. Design of experiment    c. Optimization
   d. Computer simulation model

VI. In simplex method, slack, surplus and artificial variables are restricted to be
   a. multiplied
   b. negative
   c. non-negative
   d. divided

VII. In simplex method, feasible basic solution must satisfy the
   a. non-negativity constraint
   b. negativity constraint
   c. basic constraint
   d. common constraint

VIII. According to algebra of simplex method, slack variables are assigned zero coefficients because
   a. no contribution in objective function
   b. high contribution in objective function
   c. divisor contribution in objective function
   d. base contribution in objective function

IX. In linear programming, most popular non-graphical procedure is classified as
   a. linear procedure
   b. non-graphical procedure
   c. graphical procedure
   d. simplex method

X. In linear programming problems, set of basic variables which are appeared   in linear problem consists of
   a. slack and real variables
   b. slack and artificial variables
   c. departing basic variable
   d. departing non basic variable

# UNIVERSITY OF THE PUNJAB

| PAPER: Operation Research | TIME ALLOWED: 2 hrs. & 30 mins. |
|---|---|
| Course Code: STAT-412 | MAX. MARKS: 50 |

## *Attempt this Paper on Separate Answer Sheet provided.*

**Q.2** Answer briefly and to the point, each short question carry four marks. **(20)**

    a. Objective function
    b. Decision variable
    c. Uncontrollable variables.
    d. Constraints
    e. Deterministic model

**Q.3** The cut-Right knife company sells sets of kitchen knives. The Basic Set consists of 2 utility knives and 1 chef's knife. The Regular Set consists of 2 utility knives and 1 chef's knife and 1 bread knife. The Deluxe Set consists of 3 utility knives, 1 chef's knife and 1 bread knife. Their profit is \$30 on Basic Set, \$40 on a Regular Set, and \$60 on a Deluxe Set. The factory has on hand 800 utility knives, 400 chef's knives and 200 bread knives. Assuming all sets are sold, how many of set should be sold to maximize the profit. What is the maximum profit? **(10)**

**Q.4** Give a graphical solution for the following problem. **(10)**

Maximize $Z = 80X_1 + 100X_2$

Subject to: $X_1 + 2X_2 \leq 720$; $\quad 5X_1 + 4X_2 \leq 1800$; $\quad 3X_1 + X_2 \leq 900$; $\quad X_1, X_2 \geq 0$

**Q.5** Provide simplex solution to following program. **(10)**

    a) Minimize: cost $= 4X_1 + 5X_2$
        Subject to: $X_1 + 2X_2 \geq 2.7$
                 $3X_1 + X_2 \geq 75$
                 $x_1 \geq 0, x_2 \geq 0$

    b) Maximize: $Z = 3X_1 + 2X_2$
        With $\quad X_1 + 2X_2 \leq 6$
                $2X_1 + X_2 \leq 8$
                $-X_1 + 2X_2 \leq 1$
                $X_2 \leq 2$
                $X_1, X_2 \geq 0$