Answer . The normal equations of the parabolic fit of the curve are

$$\begin{aligned} a\Sigma x^2 + b\Sigma x + nc &= \Sigma y \\ a\Sigma x^3 + b\Sigma x^2 + c\Sigma x &= \Sigma xy \\ a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 &= \Sigma x^2 y \end{aligned}$$

Question 97. Write the normal equations to fit a curve of the form $y = ax^b$ by the method of least squares. **Nov.'11**

Answer . The given fit of the curve is $y = ax^b$. ...(1)

Taking logarithm on both sides of (1), we get

$$\begin{aligned} \log_{10} y &= \log_{10}\left(ax^b\right) \\ &= \log_{10} a + \log_{10} x^b \\ \therefore \quad \log_{10} y &= \log_{10} a + b\log_{10} x \\ \text{i.e., } Y &= A + bX \qquad \text{...(2)} \end{aligned}$$

Then we determine the parameters A and B by using method of least squares. The normal equations are

$$\begin{aligned} nA + b\Sigma X &= \Sigma Y \\ A\Sigma X + b\Sigma X^2 &= \Sigma XY \end{aligned}$$

Question 98. Define 'Population'. **Nov.'14**

Solution . A population consists of collection of individuals, which may be person or experimental outcomes, whose characteristic are to be studied.

Question 99. Define sample.

Solution . A sample is a portion of the population that is studied to learn about the characteristics of the population.

Question 100. Define 'Large and Small samples. **Nov.13, Nov.'14, Nov.'15**

Solution . If the number of elements (or) items (say n) in the sample is greater than or equal to 30 (i.e., $n \geq 30$), then the sample is called as large samples. If the number of elements (or) items (n) in the sample is less than 30 (i.e.,$n < 30$), then it is called as small samples.

Question 101. Define a residual. **Nov.'13**

Solution . The residual (of an observed value) is the difference between the observed value and the estimated function value.

Question 102. Define standard error.

Solution . The standard deviation of the sampling distribution of a statistic is called standard error. Standard error is used to access the difference between the expected and observed values.

Question 103. What is the formula for finding the difference between the sample proportion and population proportion? **Apr.'13**

Answer . Let P is the population proportion and p is the sample proportion then

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \text{ and } Q = 1 - P$$

Question 104. Define hypothesis.

Solution . A hypothesis is a statement of a relationship between two or more variables. A statistical hypothesis is simply a particular kind of hypothesis.

Question 105. Define Statistical hypothesis.

Solution . A statistical hypothesis is either

1. A statement about the value of a population parameter (e.g., mean, median, mode, variance, standard deviation etc), or

2. A statement about the kind of probability distribution that a certain variable obeys.

Question 106. Define "test statistic."

Solution . The test statistic is a mathematical formula to determine the likelihood of obtaining sample outcomes if the null hypothesis were true. The value of the test statistic is used to make a decision regarding the null hypothesis.

Question 107. What are the null and alternative hypothesis?
Nov.'10, Nov.'11, Apr.'12, Apr.'14

Solution . Null hypothesis always predicts that there is no relationship between the variables being studied and it is denoted by H_0. The null hypothesis typically corresponds to a general or a default position. Making this assertion will make no difference and hence cannot be proven positively.

An alternative hypothesis (H_1) is a statement that directly contradicts a null hypothesis by stating that that the actual value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis.

Question 108. Define parameters and statistics.

Solution . Parameters are the measures such as mean, standard deviation etc which describe a population and the word statistic is used is used to indicate various measures relating to the sample such as mean, S.D., etc. A statistic is any function of observations in a random sample.

Question 109. Define sampling distribution

Solution . The probability distribution of a statistic is called as a sampling distribution

Question 110. Define critical region or region of rejection.

Solution . A region corresponding to a statistic, in the sample space which amounts to rejection of null hypothesis H_0 is called as critical region. The region of the sample space which amounts to the acceptance of null hypothesis is called acceptance region.

Question 111. Define critical value (or) significant value.

Solution . The value of the test statistic which separates the acceptance region and critical region is called the critical value.

Question 112. What is one tailed test? **Apr.'14**

Solution . A one-tailed test uses an alternate hypothesis that states either $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$, but not both. One tailed test is also classified as (i) right tailed test (ii) left tailed test.

Question 113. Define right-tail test and left-tail test.

Solution . If the observed test values fall to the right of the critical value then the test is called as right-tail test. In a right-tail test, we accept the null hypothesis if the observed test value is less than or equal to the critical value.

It the observed test values fall to the left of the critical value then the test is called as left-tail test. In a left-tail test, we accept the null hypothesis if the observed test value is greater than or equal to the critical value.

Question 114. Define two tailed test.

Solution . When two tails of the sampling distribution of the normal curve are used, then the test is called as two-tailed test. In the two tailed test the alternative hypothesis is stated as $\neq$.

Question 115. Define Test of significance. **Apr.'15**

Solution . Significance testing (also called hypothesis testing) is a mathematical method which is used to decide whether or not an hypothesis (an assumption made about a population parameter) should be accepted or rejected based upon our sample statistic.

Question 116. Define 'Level of Significance'. **Apr.'11, Nov.'15**

Solution . The level of significance is the maximum probability of committing a type I error. This probability is symbolized by α. That is, P(type I error) $= \alpha$. and P(type II error) $= \beta$.

It is the probability level below which the null hypothesis is rejected. Generally 5% and 1% level of significance are used.

Question 117. What are the Type I and Type II errors in sampling theory? **Nov.'11, Apr.'11, Nov.'15**

Solution . Type I error is the probability of rejecting a null hypothesis (H_0) when it is actually true. Type II error is the probability of accepting a null hypothesis (H_0) when it is false.

Question 118. A sample of 900 items has mean 3.4 and standard deviation 2.61. Can the sample regarded as drawn from a population with mean 3.25 at 5% level of significance? **Apr.'11**

Solution . Given $n = 900$, $\overline{x} = 3.4$, $\mu = 3.25$ and $\sigma = 2.61$
Null hypothesis (H_0) : $\mu = 3.25$.
Alternative hypothesis (H_1) : $\mu \neq 3.25$.

$$\therefore \; z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{2.61/\sqrt{900}} = 1.724$$

Tabulated value of z at 5% level of significance $= 1.96$
At 5% level, the calculated value of $z <$ tabulated value of z.
$\therefore$ Accept H_0. Difference is not significant.

Question 119. A sample of 900 members from a normal population with SD 2.61 cms has a mean 3.5 cms. Find the 95% fiducial limits for the population mean. **Nov.'10**

Solution . Given $n = 900$, $\overline{x} = 3.5$ and $\sigma = 2.61$

If the population mean μ is unknown, 95% confidence limits of μ are

$$= \overline{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 3.5 \pm 1.96 \times \frac{2.61}{\sqrt{900}}$$
$$= 3.5 \pm 0.17 = (3.33,\ 3.67)$$

Question 120. Write down the test statistic for large sample test for difference of means. **Nov.'11**

Solution . If the samples are drawn from the same population with known S.D σ then the test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

.

Question 121. A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased **Nov.'12**

Solution . Given that, sample size $n = 400$. Number of heads $= 216$.

$p =$ sample proportion of heads $= \frac{216}{400} = 0.54$

$P =$ Population proportion $= \frac{1}{2} = 0.5$. $\therefore\ Q = 1 - P = 0.5$

Null hypothesis (H_0) : Coin is unbiased. or $P = \frac{1}{2}$

Alternative hypothesis (H_1) : $P \neq \frac{1}{2}$.

$$\therefore\ z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 1.6$$

$\therefore$ The calculated value of $z = 1.6$

Tabulated value of z at 5% level of significance $= 1.96$

Since calculated value of $z <$ tabulated value of z.

$\therefore$ Accept H_0. Thus coin is unbiased.

5.6 Part – A : Two Marks Questions

Question 122. Define the $'t'$ statistic. **Apr.'11, Apr.'14**

Answer . If $x_1, x_2, ...x_n$ be the small samples (i.e., $n < 30$) drawn from the normal population with mean μ and variance S^2 then the sample mean $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and standard deviation $S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ then the statistic is

$$t = \frac{\overline{x} - \mu}{S/\sqrt{n}} \text{ (or) } \frac{\overline{x} - \mu}{s/\sqrt{n-1}}.$$

Question 123. Write down the 95% and 99% of confidence or fiducial limits of μ. **(OR)** Give the 95% and 99% confidence interval of the population mean interms of mean and SD of small sample **Nov.'15**

Answer . 1. 95% confidence limits of mean $\mu = \overline{x} \pm t_{0.05} \frac{S}{\sqrt{n}}$.
2. 99% confidence limits of mean $\mu = \overline{x} \pm t_{0.01} \frac{S}{\sqrt{n}}$.

Question 124. State the properties of t-distribution. **Nov.'10, Apr.'13**

Answer . 1. The pdf of t distribution is $f(t) = \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$, where $t = \frac{\overline{x}-\mu}{S/\sqrt{n}} = \frac{\overline{x}-\mu}{s/\sqrt{n-2}}$, $\int_{-\infty}^{\infty} f(t)\, dt = 1$.

2. The values of t varies from $-\infty$ to ∞.
3. It has symmetrical about $t = 0$ and mean zero.
4. The variance of the $t-$distribution is greater than 1 but tends to 1 as $n \to \infty$. As $\nu \to \infty$ then $t-$distribution becomes normal.
5. The variance $= \frac{\nu}{\nu-2}$, if $\nu > 2$ and $\mu_2 > 1$ always.

Question 125. State some applications (or) uses of $t-$distribution. **Apr.'15**

Answer . 1. To test the significance of a single mean.
2. To test the significance difference between the two sample means.

3. To test the significance of the correlation coefficient.
4. To test the significance of observed partial and multiple correlation coefficients.

Question 126. State the assumptions of t–distribution. **Apr.'11**

Answer . 1. The parent population from which the sample is drawn is normal.

2. The sample observations are independent, i.e., the sample is random.
3. The population standard deviation is unknown.
4. Sample size $n < 30$.

Question 127. State the properties of χ^2 distribution. **Nov.'14**

Answer . 1. The sum of independent chi-square variables is also chi-square variable.

2. Chi-square distribution approaches to normal distribution when $n \to \infty$.
3. The Mgf of chi-square distribution is $(1-2t)^{-\nu/2}$.
4. If k is the number of linearly independent constraints then $\nu = n - k$.
5. For $p \times q$ contigency table, the degrees of freedom is $(p-1)(q-1)$.

Question 128. State some applications or Uses of χ^2 distribution **Nov.'10, Nov.'12**

Answer . 1. To test the 'goodness of fit'.

2. To test the 'independence of attributes'.
3. To test the homogeneity of independent estimates of the population variance.
4. To test the hypothetical value of the population variance is σ^2.

Question 129. What the conditions for application of chi-square test. (OR) State the conditions under which chi-square test of goodness fit is valid. **Nov.'11, Apr.'12, Apr.'15, Nov.'15**

Answer . This is an appropriate test for large values of n. For the validity of chi-squáre test of 'goodness of fit' between the theory and experiment, the following conditions must be satisfied.

1. The sample observations should be independent.
2. Constraints on cell frequencies must be linear.
3. N, the total frequency, should be reasonably large say > 50.
4. No theoretical frequencies is less than 5. If it is 5, then the application of χ^2 test, it is pooled with the succeeding or preceding so that the pooled frequency is more than 5.

Question 130. A random sample of size 16 has 53 as mean. The sum of squares of the deviations taken from mean is 135. Can this sample be regarded as taken from the populations having 56 as mean?

Solution . Step 1: Given $n = 16$, $\overline{x} = 53$, $\mu = 56$ and $\sum(x - \overline{x})^2 = 135$.

$$\therefore \quad s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3.$$

Step 2 : Null hypothesis H_0 : $\mu = 56$.

Alternative hypothesis H_1 : $\mu \neq 56$.

Step 3: Calculated value of z

$$\text{Test statistic } z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{53 - 56}{3/\sqrt{16}} = 4$$

Step 4: Tabulated value of z at 5% level of significance $= 2.58$

Step 5: Conclusion :

At 5% level, the calculated value of $z <$ tabulated value of z.
$\therefore$ Accept H_0. There is no significance difference of the population mean and sample mean.

Question 131. Write the uses of $F-$ distribution.

Nov.'13, Nov.'14, Nov.'15

Answer . 1. It is used to test whether two independent samples have been drawn from the normal populations with same variance σ^2

2. It is used to test whether the two independent estimates of the population variance are homogeneous or not.

Question 132. What are the applications of $F-$ test. **Apr.'12**

Answer . 1. It is used to test whether two independent samples have been drawn from the normal populations with same variance σ^2.

2. It is used to test whether the two independent estimates of the population variance are homogeneous or not.

Question 133. Define a 'Attributes'. **Nov.'13**

Answer . A mathematical process used to analyze the characteristics of a given population of subjects. Attributes have only two possible ratings (negative or positive) expressed as acceptable or unacceptable, desirable or undesirable, good or bad, etc.

Question 134. A random sample of size 16 has 53 as mean. The sum of squares of the deviation from mean is 135. Can this sample be regarded as taken from the population having 56 as mean? **Nov.'12**

Solution . Given $n = 16$, $\overline{x} = 53$, $\mu = 56$ and $\sum(x - \overline{x})^2 = 135$.

$\therefore \quad s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3.$

Null hypothesis H_0 : $\mu = 56$. Alternative hypothesis H_1 : $\mu \neq 56$.

$$\text{Test statistic } z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{53 - 56}{3/\sqrt{16}}$$

$$\therefore z = 4$$

Tabulated value of z at 5% level of significance $= 2.58$

At 5% level, the calculated value of $z <$ tabulated value of z.

$\therefore$ Accept H_0. There is no significance difference of the population mean and sample mean.

Question 135. Write the probability density function of $F-$ test. **Apr.'13**

Answer . The probability distribution of F is given by

$$f(F) = k(F)^{\frac{\nu_1}{2}-1}\left[1+\frac{\nu_1}{\nu_2}\right]^{-(\nu_1+\nu_2)/2}, \quad 0 \le F < \infty.$$

Question 136. Define Student's $t-$ test for difference of means of two samples. **Nov.'11**

Answer . Consider two independent small samples of sizes n_1 and n_2. Let $\overline{x}_1$ and $\overline{x}_2$ be the sample means and s_1^2 and s_2^2 be the sample variances. To test the difference between the means are significant by using test statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}. \quad \text{Where } S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

The degrees of freedom of this statistic is $\nu = n_1 + n_2 - 2$.

Question 137. Define standard error. **Apr.'11; Apr.'15**

Answer . The standard deviation of the sampling distribution of a statistic is known as standard error.

Question 138. Define chi-square test for goodness of fit. **Apr.'11, Apr.'12**

Answer . If O_i and E_i $(i = 1, 2, ...n)$ be the set of observed (experimental) and expected (theoretical) frequencies then χ^2 is defined as

$$\chi^2 = \sum_{n=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Question 139. What do you mean by degrees of freedom **Apr.'15**

Answer . The number of degrees of freedom (ν) is the total number of observations(n) less than the number of independent constraints(k) imposed on the observations. Thus $\nu = n - k$.

Question 140. Write down the chi-square statistic for the contingency table.

a	b
c	d

Apr.'14

Answer . The value of χ^2 test for 2×2 contingency table can be calculated as follows:

Table of Observed Frequencies and Expected frequencies

	A_1	A_2	Total
B_1	a	b	$a+b$
B_2	c	d	$c+d$
Total	$a+c$	$b+d$	N

	A_1	A_2
B_1	$\frac{(a+b)(a+c)}{N}$	$\frac{(a+b)(b+d)}{N}$
B_2	$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(c+d)}{N}$

Where $N = a+b+c+d$.

Then the χ^2 statistic is

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

Question 141. Write the assumptions of F–test.

Answer . 1. The populations for each sample must be normally distributed.

2. The samples must be random also independent.
3. The ratio of the variances must be greater than 1. So keep always larger variance of the sample in the numerator.