Chapter-1

Sampling Distribution

1a What is a sampling distribution, and why is it important in statistics.

Ans: A Sampling distribution is a Probability distribution of a Statistics Obtained through a large number of Samples drawn from a specific Population.

Sample distributions are important in statistics because they Provide a major Simplification enroute to statistical inference.

more specifically, they allow analytical considerations to be based on the Probability distribution of a statistic, rather than on the joint Probability distribution of all the individual sample values.

For example: χ^2, F and t distribution are sampling distribution.

* The distribution of sample statistics called sampling distribution.

1.b: Difference between Sampling distribution and Parent distribution.

Ans:

Parent Distribution (Population)	Sampling Distribution
1. The distribution of the entire Population.	1. The distribution of a sample Statistics. (Sample mean).
2. The actual data of the Population.	2. Repeated samples taken from the Population.
3. Parameters are mean(μ), Standard deviation(σ).	3. Mean of Sample means (μ) Standard error($\sigma_{\bar{x}}$)
4. Can be any distribution. Shape.	4 Approaches normality as n increase.
5. Size fixed (entire Population)	5. Changes based on Sample size n.
6. Example: Heights of all adults in a city.	6. Example: The distribution of sample means from repeated height Samples (size 30)

1.c: Define the chi-square distribution and explain how it is derived from standard normal variables.

Ans: The sum of squares of n independent standard normal variables is called chi-squares (χ^2) ~~variable~~ distribution with n degrees of freedom.

Derived from standard normal variables:-

Let $z_1, z_2 \cdots z_n$ be n independent standard normal variables then chi-square denoted by χ^2, is defined as

$$\chi_n^2 = \sum_{i=1}^{n} z_i^2$$

$$= z_1^2 + z_2^2 + z_3^2 + \cdots z_n^2$$

However, if $x_1, x_2, \cdots x_n$ are n independently and indentically distributed random variables each of which is normally distributed with mean μ and variance σ^2, then

$$\chi_n^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$$

is a chi-square(χ^2) distribution with n degree of freedom.

1.d: What are the properties of the chi-square distribution? List at least five.

Ans: Properties of χ^2 distribution:-

(i) χ^2 is a continuous type of distribution and its range is 0 to ∞. i.e $0 < \chi^2 < \infty$.

(ii) This distribution contains only one parameter which is the degree of freedom of the distribution.

(iii) The mean and variance of χ^2 distribution for n d.f. is n and 2n respectively.

(iv) The mode of χ^2 distribution for n d.f is (n-2).

(v) The moment generating function of χ^2 distribution for n d.f. is $\frac{1}{(1-2t)^{n/2}}$ or $(1-2t)^{-n/2}$.

(vi) χ^2 distribution tends to normal distribution for large degree of freedom.

(vii) It is positively skewed distribution for smaller values of n.

(viii) The distribution becomes symmetrical as n tends to infinity ($n \to \infty$)

* Application Of Chi-square distribution

(i) To test if the hypothetical value of the population variance is $\sigma^2 = \sigma_0^2$.

(ii) To test the goodness of fit.

(iii) To test the independence of attributes.

(iv) To test the homogenity of independent estimates of the population variance.

(v) To test the homogenity of independent estimates of the population correlation cofficient.

1.e: How does the chi-square distribution behave as the degree od freedom increase?

Ans: As the degrees of freedom(k) increase, the chi-square distribution;

1. Becomes more Symmetric → Initially right skewed, but approches a normal distribution for large k.

2. Mean Increases → The mean is k, So it Shifts right as k grows.

3. Variance Increases → The variance is 2k, making the distribution spread out more.

4. Peak shift Right → The mode moves from near zero to k-2.

5. Approximates Normal Distribution → For k>30, it behaves like $N(k, 2k)$.

1.f: How is the chi-square distribution used to test the goodness of fit?

Ans: The chi-square goodness of-fit-test checks if observed data matches an expected distribution.

Steps:-

1. Set hypothesis:-

H_0: Observed data follows the expected distribution.

H_1: Observed data does not follow the expected distribution.

2. Calculate Expected Frequencies:-

$E_i = N \times P_i$ (where, N = total observations, P_i = expected proportion)

3. Compute Chi-square Statistics:-

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

4. Compare with critical value/P-value:-

Use chi-square table with K-1 degree of freedom.

if P-value $< \alpha$, Reject H_0

1.9 Derive the moment generating function (MGF) of the Chi-square distribution.

Ans:

Derivation:- Let $x_1, x_2 \cdots x_n$ be n independent random variable from $N(\mu,\sigma^2)$ i.e $x_i \sim N(\mu,\sigma^2)$; $i=1,2,3 \cdots n$, x_i are independent.

Now, we want to find the distribution of $\chi_n^2 = \Sigma y_i = \Sigma \left(\frac{x_i-\mu}{\sigma}\right)^2$ by mgf technique.

Hence, the mgf of χ^2 is given by

$$M_{\chi^2}(t) = M_{\Sigma y_i}(t) = \prod_{i=1}^{n} M_{y_i}(t) \quad [y_i \text{ are independent}]$$

$$M_{\chi^2}(t) = \prod_{i=1}^{n} \left[M_{\left(\frac{x_i-\mu}{\sigma}\right)^2}(t) \right]$$

??

$$= \prod_{i=1}^{n} E\left[e^{t\left(\frac{x_i-\mu}{\sigma}\right)^2} \right]$$

$$= \prod_{i=1}^{n} E\left[e^{tu^2} \right] \quad \left(u = \frac{x_i-\mu}{\sigma} \right)$$

$$= \prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} e^{tu^2} f(u)\, du \right\}$$

$$= \prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} e^{tu^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{tu^2 - \frac{1}{2}u^2}\, du \right\}$$

[Since the integrand is an even function of u]

$$= \prod_{i=1}^{n} \left\{ \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty e^{-(1/2 - t)u^2} du \right\}$$

Let $z = (1/2 - t)u^2$

$\Rightarrow u^2 = \dfrac{z}{\frac{1-2t}{2}}$

$\Rightarrow u = \sqrt{\dfrac{2z}{1-2t}}$

$dz = (\frac{1}{2} - t)u^2$

$= (\frac{1}{2} - t) \cdot 2u \cdot du$

$\Rightarrow 2u \cdot du = \dfrac{dz}{\left(\frac{1-2t}{2}\right)}$

$\Rightarrow du = \dfrac{dz}{\cancel{2}u\frac{(1-2t)}{2}}$

$\therefore du = \dfrac{dz}{(1-2t)u}$

$$\Rightarrow du = \frac{dz}{(1-2t)\sqrt{\frac{2z}{(1-2t)}}} = \frac{dz}{\sqrt{1-2t}\sqrt{2z}}$$

$$\therefore M_{x^2}(t) = \prod_{i=1}^{n} \left\{ \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty e^{-z} \cdot \frac{dz}{\sqrt{1-2t}\sqrt{2z}} \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{\cancel{\sqrt{2}}}{\sqrt{\pi}\sqrt{1-2t} \cdot \cancel{\sqrt{2}}} \int_0^\infty e^{-z} \cdot z^{-1/2} dz \right\}$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \int_0^\infty e^{-z} z^{1/2 - 1} dz \right.$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \sqrt{1/2} \right\} \quad \left[\because \sqrt{n} = \int_0^\infty e^{-x} x^{n-1} dx \right]$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\cancel{\sqrt{\pi}}\sqrt{1-2t}} \cdot \cancel{\sqrt{\pi}} \right\}$$

$$= \prod_{i=1}^{n} \left\{(1-2t)^{-1/2}\right\}$$

$$= (1-2t)^{-n/2}$$

$$\therefore M_{x^2}(t) = (1-2t)^{-n/2}.$$

which is the mgf of gamma distribution with shape parameter $\alpha = n/2$ and scale parameter $\beta = 2$

Therefore, the pdf x^2 distribution is as

$$f(x^2) = \frac{1}{2^{n/2}\sqrt{n/2}} (x^2)^{\frac{n}{2}-1} \cdot e^{-x^2/2}, \quad x^2 > 0, (0 < x^2 < \infty)$$

This is the pdf of x^2-variable with n degree of freedom.

1.h: Using the MGF, Find the mean and variance of the Chi-square distribution.

Ans: The moment Generating function (MGF) of chi-square distribution random variable $X \sim X_n^2$

$$\therefore M_x(t) = (1-2t)^{-n/2} \quad \text{for } (t < \tfrac{1}{2})$$

Compute mean

The mean (first moment) is given by

$$E(X) = M_x'(0) = \frac{d}{dt} M_{(x)}t \Big|_{t=0}$$

Differentiate $M_x(t)$

$$M_x(t) = (1-2t)^{-n/2}$$

$$\Rightarrow \frac{d}{dt} M'_{(x)} t = -\frac{n}{2}(1-2t)^{-n/2-1} \cdot (-2)$$

$$= \frac{n}{(1-2t)^{n/2+1}}$$ $\boxed{\cancel{\frac{n}{2}} \text{ neglect} \quad \cancel{-\frac{n}{2} = 0}}$

$\therefore$ Now simplify $M'_{(x)} t = \frac{n}{(1-2t)^{n/2+1}}$

Evaluating $t = 0$.

$$E(x) = M'_x(0) = \frac{n}{(1-0)^{n/2+1}} = n$$

$$\therefore E[x] = n$$

---

* Compute the variance $Var(x)$

The variance given by:-

$$Var[x] = E[x^2] - (E[x])^2$$

To find $E[x^2]$, we compute the second derivative of $M_x(t)$.

Differentiate $M_x'(t)$ again.

we have $M_x'(t) = \frac{k}{1-2t}$

Differentiate again using the quotient rule

$$M_x''(t) = \frac{k \cdot 2}{(1-2t)^2} = \frac{2k}{(1-2t)^2}$$

Evaluting at $t=0$

$$E[x^2] = M_x''(0) = \frac{2k}{1^2} = 2k$$

∴ Now, we compute the varince:

$$Var(x) = E[x^2] - (E[x])^2$$

$$Var(x) = 2k - k^2$$

∴ Thus the varince of chi-square distribution is $var(X) = 2k$

Now differentiate again

$$M_x'(t) = n(1-2t)^{-n/2-1}$$

Applying chain rule again,

$$M_x''(t) = \cancel{k}.\,n.\left(-\frac{n}{2}-1\right)(1-2t)^{-n/2-2}.(-2)$$

$$= n\left(\frac{n}{2}+1\right)(1-2t)^{-n/2-2}.2$$

$$= n(n/2+1).2.(1-2t)^{-n/2-2}$$

∴ Evaluting $t=0$

$$E[x^2] = M_x''(0) = n\left(\frac{n}{2}+1\right).2(1-0)^{-n/2-2}$$

$$= n\left(\frac{n}{2}+1\right).2$$

$$= \frac{n^2+2n}{2}.2$$

$$= n^2+2n$$

$$\therefore Var[x] = E(x^2) - (E[x])^2$$

$$= n^2+2n-n^2$$

$$= 2n$$

$$\therefore Var[x] = 2n$$

CamScanner

CHAPTER-2

2.a Define the Students t-distribution and explain how it is observed.

Chapter-16

Tests Of Hypothesis

5a: Explain the concept of hypothesis testing and its importance in statistical analysis.

Ans: Concept of Hypothesis Testing:-

Hypothesis testing is a statistical method used to make inferences or decision about a population based on sample data. It involves formulating two competing hypothesis:-

(i) Null hypothesis (H_0): This is the default assumption that there is no effect or no difference

(ii) Alternative Hypothesis (H_a or H_1): This is what the researcher want to prove. It suggests there is significant effects or difference.

Steps in Hypothesis testing: (5-d)

1. State the hypothesis: Define H_0 and H_1
2. Set the Significance Level (α):- commonly 0.05 (5%) indicating the the tolerance for Type-1 error.

3. Choose a Test Statistic:- Select an appropriate test based on data type.

4. Compute the p-value:- Measure the Probability of observing the data if H_0 is true.

5. Make a Decision:-

If p-value $< \alpha$ → Reject H_0

If p-value $\geq \alpha$ → Fail to reject H_0

Importance of hypothesis Testing in Statistical Analysis:-

1. Decision making:- It helps business, Scientists and Policymakers in data driven choices.

2. Research Validation:- Ensure finding are statistically Significant, not random.

3. Quality control:- Detects Defects and maintains Products Standards in industries.

4. Medical Studies:- Evaluates the effectiveness of treatments and drugs.

5. Economical analysis:- (মূল্যায়ন) Assesses Policy Impacts and market trends.

5.b:- Whate are Type-I and Type-II errors in hypothesis testing? How are they related to the level of significance (α) and the power of the test,

Ans: Type I and Tgpe II Ennons in hypothesis Testing:-

1. Type I error (α): Occurs when they null hypothesis (H_0) is reseckd even though it is actually true. This is also called a false Positive.

2. Type II Error (β):- Occurs when the null hypothesis (H_0) is not reject even though it is actually false. This is also called a false negative.

Relation to (α) and Power of the Test:

(1) Lowering α (reducing type I error) increase the risk of type II error (β).

(ii) Increasing the power of the test (reducing type II error) requires increasing the sample size or effect size.

(iii) A balance between α and β is needed to minimize both errors while maintaining Statistical accuracy.

CamScanner

5.C:- List the applications of test statistics.

Ans: Application of test Statistics in hypothesis testing:-

1. Z-Test:- Used for hypothesis testing when the sample size is large ($n \geq 30$) and the population variance is known.

2. T-test: Compares means of two groups when the population variance is unknown.

3. Chi-square test:- Tests categorical data for independent goodness-of-fit or homogenity.

4. Anova (Analysis of variance); Compares mean of three or more groups to detect significant differences.

5. F-test:- Compares variance between two population to check equality.

6. Regression Analysis: Tests relationship between variables and predicts outcomes.

7. Non-Parametric test:- Used when data does not follow a normal distribution.

5.e: What is the difference between one tailed and two-tailled test? When would you use each.

Ans: Difference between one tailled and two tailed tests

One-tailed test	two-tailed Test
1. tests for an effect in one specific direction (greater or less)	1. Tests for an effect in both direction (different but unspecified)
2. Ha: The parameter is greater than or less than a certain value.	2. Ha:- The parameter is not equal to a certain value.
3. Entire rejection region is on one side of the distribution.	3. Rejection regions are split equally on both sides of the distribution.
4. Example:- $H_a: (\mu_1 > \mu_2)$	Example: $H_a: (\mu_1 \neq \mu_2)$

When to use each:

(I) One tailed test: When you expect a specific directional effect.

(II) Two-tailed test; When ~~on~~ you only want to check for any difference, regardless of direction.

5f: What is the P-value and critical value and how are these interpreted in hypothesis testing.

Ans: P-value and Critical value in hypothesis testing:-

① P-value:- The p value is the Probability of obtaining the observed results assuming the null hypothesis (H_0) is true.

Interpretation:

(i) If P-value < significance level (α), reject H_0.

(ii) If P-value $\geq \alpha$, fail to reject H_0.

② Critical Value: The critical value is the threshold that defines the rejection region in hypothesis testing. It is determined based on the significance level (α) and the type of test.

Interpretation:

(i) If the test statistics exceeds the critical value, reject H_0,

(ii) If the test statistics does not exceeds the critical value, fail to reject H_0, 5.Q:- Explain the concept of the critical region and acceptance region in hypothesis testing? How is it determined.

Ans: Critical region and Acceptance region in hypothesis testing:-

1. Critical Region (Rejection Region):-

(i) The critical region consist of values of the test statistic where reject the null hypothesis.

(ii) It is determined based on the significance Level (α) and the type of test. (one or two tailed).

(iii) Larger deviations from the null hypothesis fall into this region.

2. Acceptance Region:

(i) The acceptance region consists of values where we fail to reject the null hypothesis (H_0).

(ii) It includes values of the test statistics that suggest the sample data is consistent with H_0.

How is it determined:

(i) Choose the significance level (0.005 or 5%)

(ii) Find the critical value from statistical table based on α and test type.

(iii) The critical region consist of value beyond the critical value, while the acceptance region contains values within them.

Set 6

6.a Explain the hypothesis testing for sigle population mean?

Ans: The distinction between small and large samples is important when testing a single population mean.

There is no strict cutoff between small and large samples, but typically, if sample size exceeds 29 $(n > 29)$, it is considered large sample.

The methods used for hypothesis testing differ based on the sample size because the assumptions made for large sample may not hold for small samples.

Assumptions for hypothesis testing:-

① For small samples $(n \leq 29)$

(i) The sample is randomly selected from a normally distributed population.

(ii) The variance (σ^2) is assumed to be known.

(iii) The z-test is typically used if the population variance is known, otherwise, a t-test is used.

② For large sample $(n > 29)$

(i) The sample is randomly selected from a normally distributed population.

(ii) The variance is ~~unkn~~ unknown, so ~~so~~ the sample variance is used as an estimate.

(iii) The z-test is ~~due~~ used due to the Central Limit Theorem (CLT), which states that for large samples.

6.b Describe steps of mean test when variance unknown and sample size is small ($n < 30$).

Ans: Steps of Mean test (variance unknown, small sample ($n < 30$)):-

① State the hypothesis

$H_0 : \mu = \mu_0$ (mean equals given value)

$H_1 : \mu \neq \mu_0$, $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$.

2. Select significance level α :- (0.005 or 5%)

3. Calculate t-static:

$$T = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

where
$\bar{x}$ = sample mean
S = SD
n = Sample size.

4. Find the critical value or p-value:-

① Use t-table with $df = n-1$ for criteria t.

(ii) Alternatively, compare p-value with α.

5. Make decision:-

If $|T| >$ critical t, reject H_0.

If ~~for~~ P-value $< \alpha$, reject H_0.

6. Conclusion:-

Reject $H_0 \rightarrow$ Significant different exists.

Fail to Reject $H_0 \rightarrow$ No Sufficient evidence of a difference.

6.c: What do you mean by Z-test? Mention the applications of it.

Ans: A Z-test is a statistical test used to compare a sample mean with a population mean (or two sample mean) when:-

(i) The population variance is known.

(ii) The sample size is large $(n > 30)$

(iii) The data follows a normal distribution.

Formula:-

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

where $\bar{x}$ = Sample mean

μ = Poplulation mean

σ = Population Standard deviation

n = Sample size.

Applications od z-statistic

(i) Test of a single population mean.

(ii) Test od equality of two population mean

(iii) Test od a single population proportion.

(iv) Test for difference between two population proportions.

(v) Test of a specified correlation co-efficient.

(vi) Test of ~~a~~ equality od two population correlation co-efficient.

6.d:- What is test of hypothesis concerning attributes? Discuss test procedure of a hypothesis about proportion test od a population.

Ans: A test of hypothesis concerning attributes is used to assess if the proportion of a specific attributes in a population matches a hypothesized value. it involves.

1. Stating null and alternative hypothesises.

2. Calculating a test statistic.

3. Comparing the statistic to critical values. or using the p-value to make a decision.

Procedure of a hypothesis about proportion test of a population:-

1. State Procedure:-

Null hypothesis $H_0: P = P_0$

Alternative hypothesis $H_1: P \neq P_0$ (two-tailed)

$P > P_0$ (right-tailed)

$P < P_0$ (left-tailed)

2. Choose significance level (α):

commonly $\alpha = 0.05$ or 5%.

3. Calculate the test statistic (z)

$$z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

where
$\hat{p}$ = Sample porportion
P_0 = hypothesized population porportion.

4. Find the critical value or p-value:

Compare the test statistic to the critical value or calculate the p-value.

5. Make a decision:-

if z is in the rejection region or p-value $< \alpha$,

reject H_0.

Otherwise fail to reject H_0.

6. Conclusion: conclude if the proportion differs significantly from the hypothesized value.

6. Q: What is χ^2 test and what are its application in hypothesis testing?

Ans: The sum of square of n independent standard normal variables is called chi-square (χ^2) distribution with n degree of freedom.

$$\cancel{\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma}\right)^2} \quad \chi^2 = \sum \frac{(O_i-E_i)^2}{E_i}$$

Applications of χ^2-statistic

(i) Test of population variance with specific value

(ii) Test of equality of several variances.

(iii) Test of equality of several correlation co-efficient

(iv) Test of equality of several population proportions.

(v) Test of independence of attributes.

(vi) Test of goodness of fit.

6.f: What is the role of the test statistic in hypothesis testing? How it is used to make decisions about null hypothesis.

Ans: In hypothesis testing, the test statistic is a key value used to assess the validity of the null hypothesis (H_0). It quantifies the difference between the observed sample data and what is expected under H_0.

Role of test statistic:

1. Quantifying the difference: It measures how far the sample data deviates from the expected (Fz(t)χ(s)) values under the null hypothesis.

2. Standardizing the difference: It standardizes this difference by ~~dividing~~ dividing it by a measure of variability.

Decision Process: The test statistic is compare against a critical value or p-value:

(i) If the test statistic is extreme, reject the null hypothesis.

(ii) If the test statistic is not extreme, fail to reject null hypothesis.

Q.9: Explain the test of hypothesis about difference between two population proportions.

Ans: