OPINION BASED ANSWERS, FOR EACH QUESTIONS!

KAVERI'S IAS TEST SERIES

UPSC ANSWER BOOKLET: GENERAL STUDIES

UPSC MAINS ZOOLOGY TEST SERIES

RESEARCH BASED Q & A: BIOSTATISTICS

Q. What is Meant by Chi Square Test? Discuss the Procedure that Involved in the Estimation of Chi Square Value with examples. How the Value is having Significant in the Genetics & Breeding Techniques? 20 marks

Ans. Chi-square test – denoted by χ^2 is a statistical tool to test the significant difference between the observed data and the expected data by conducting an experiment or making an observation. It was developed by Karl Pearson (1900).

Formula: $\chi^2 = \frac{\sum(O-E)^2}{E}$

O = observed data

E = expected data

It is the sun of squared difference between the observed and expected data divided by the expected data of all possible categories.

Chi-square test is used to test the "goodness of fit" to see if a particular distribution fits the expected distribution or if a population distribution is valid. Eg: A cross between 2 pea plants give offspring 639 and 241 (green yellow, respectively). A test to check if it corresponds to Mendelian ratio can be done by χ^2 test.

H_0 = The pea population ratio is not same as the Mendelian ratio 3:1

H_1 = The pea population ratio is the same as the Mendelian ratio 3:1

Total = 880

Expected =	Green	:	yellow
	668	:	220
Observed	639	:	241

$$\chi^2 = \frac{\sum(O-E)^2}{E}$$

$$= \frac{(639-660)^2}{660} + \frac{(241-220)^2}{220}$$

$$= \frac{(21)^2}{660} + \frac{(21)^2}{220}$$

$$= \frac{441}{660} + \frac{441}{220}$$

χ^2 = calculated

To find χ^2 tabulated value,

a. Degree of freedom = No. of category – 1

= 2 – 1 (green, yellow)

= 1

b. Level of significance = 0.05

Using a and b, from the χ^2 table,

χ^2 tabulated is found.

Interpretation of results

χ^2 calculated $<\chi^2$ tabulated

Difference is insignificant as Reject Null hypothesis

χ^2 calculated $>\chi^2$ tabulated

Difference is insignificant as Fail to reject Null hypothesis.

The chi-square value is significant in the breeding techniques as it can

- Show whether the breeding populations corresponds to the expected Mendelian ratio (same genotype and phenotype)
- If they breed as Mendelian or Non-Mendelian ratio (same genotype and different phenotypes)
- If the initial assumptions of hypothesis or objectives are incorrect or whether there are experimental design error,

χ^2 –test is also used as a test of homogeneity and test of independence.

Q. What is meant by Correlation & Regression Analysis? How it is having Significant in Breeding Techniques? Discuss the Estimation of -1 to + 1 Correlation Expression with Examples. 20 Marks

Ans. **Correlation:** it is the simultaneous change (increase or decrease) of the two variables. It shows the relationship between 2 variables such that change in one variable results in the corresponding change in the second variable. It tests the degree of association and it shown by correlation coefficient which is a single number that shows the extent and degree of relationship Mathematically. Ex: Rain and humidity.

Regression analysis shows the relationship between 2 variables in such a way that a known variable can help predict the unknown variable. By determining X-variable, the independent variable can be found.

Ex: Different known drug dosage and change in systolic blood pressure.

Correlation and Regression analysis are significant in breeding techniques as it:

- Can predict or forecast the number of progeny in the cross (Monohybrid or dihybrid) along the known ratios.
- It can also indicate the phenotypes that can be obtained along the known phenotypes.
- It can predict the independent variables (a particular trait / enzyme activity or protein product that can be obtained by crossing the parents with the desired gene inserts)

  Thus it has great applications in genetics, antibiotics production, livestock generation (yield and variety), research and plant breeding experiments.

Range: –1 to +1 for correlation expression.

Q. Define Test of Significance? Discuss in detail Test of Significance during Comparison of one way mean and two ways mean. 15 Marks

Ans. Tests of significance

In statistics, "significance" means non-chance difference between the observed and the expected data. If the difference is significantly large enough not to be ignored, it is not by chance and tests of significance have to be conducted. If the difference is small, it is by chance and can be ignored. There are many examples of tests of significance:

- Chi-square test
- Student 't' test
- Fischer 'F' test
- Analysis of variance test etc.

ANOVA or Analysis of variance:

It is a statistical tool that tests the variation between the means of groups or categories. There are 2 types – one way or two-way ANOVA.

One-way ANOVA

In test the difference between the means of at least 3 categories and each category having at least 3 observations.

Eg: the length of pawn Penaeus monodon is 3 ponds (3) separate populations

	Ponds		
No. of observations	1	2	3
	2	3	4
	3	3	3
	4	3	4
Mean of individual ponds	3	3	3.6

Group mean = 3.2

H_0 = The mean length of the prawns in 3 ponds are the same

H_1 = The mean length of the prawns is 3 ponds are not the same.

We calculate the following:

- Degree of freedom (between) = No. of ponds – 1

  = 3 – 1 = **2**

- Degree of freedom (within) = No. of all observations - …

  = 9 – 3 = **6**

  From the above, using ANOVA table we get **F tabulated.**

- Sum of squares SS
- SS between and SS (within)
- Mean square

  MS (between) and MS (within)

- F calculated = $\frac{MS\ (between)}{MS\ (within)}$

  = F calculated

Result interpretation:

$F_{Cal}<F_{tab}\rightarrow$ mean difference is insignificant

$F_{Cal}>F_{tab}\rightarrow$ Fail to Reject Null hypothesis

Difference is significant, Reject null hypothesis.

Two-Way ANOVA:

In this 2 or more independent variable can be tested to see their effect on the dependent or the measured variable individually or whether the interaction between the 2 independent variable have an effect on the dependent variable.

Eg: Whether age and gender has any effect on drug assimilation

H_0 = Age has no effect on drug assimilation

H_1 = Age has effect

H_0 = Gender has no effect on drug assimilation

H_1 = Gender has effect

H_0 = Age and gender interaction has no effect on drug assimilation

H_1 = Age and gender interaction has effect.

In the above experiment. The number of observations are similar in all category.

We calculate:

- ✓ degree of freedom
- ✓ sum of squares between and within
- ✓ mean square between and within

Result:

If $F_{Calculated} < F_{tabulated}$ = as significant difference so Fail to reject Null hypothesis

Of $F_{Calculated} < F_{tabulated}$

Difference is significant reject Null hypothesis.

The result is individually assessed for all 3 Null hypothesis to know the effect of independent variable (age or gender) or both combined (as and gender) on the measured variable (drug assimilation).

Q. What is meant by Design of Experiment? Discuss the Important Tools that involved the DOX? And its Importance. 20 marks

Ans. Design of experiment:

In science, an experiment is a systematic process carried under controlled condition to illustrate a known result or discover an unknown effect or to test a hypothesis.

Design of experiment is a systematic approach to investigate a system or a process by carrying out a series of structure tests in a planned manner where planned changes are introduced and the results (pre-determined) are analyzed, compared and inference are made and verified.

It was developed by R.A. Fisher (1935) in his book "Design of Experiment".

The importance of Design of experiments:

- ✓ It helps to pre-plan the pre-requisites and components to assure a smooth process
- ✓ Cost and benefit analysis
- ✓ Expected results can be examined
- ✓ Desired changes can be introduced
- ✓ Errors can be identified and quantified
- ✓ Correlation and causation can be segregated.

Q. Discuss the important measures that involved in the Measures of central tendency. 10 marks

Ans. Measures of Central tendency:

The set of statistical tool or tests that indicates the centrality of a distribution data are known as measures of central tendency. The various measures are mean, Median, Mode.

Mean:

It is the sum of all observations divided by the total number of observations For ungrouped data,

Mean = $\frac{\sum x}{n}$

x = all observations $x_1, x_2, \ldots\ldots\ldots\ldots x_n$

For grouped data:

a. Discrete series:

Mean = $\frac{\sum x}{\sum f}$

x = value of variables

f = frequency of each variable

Σf = Total no. of observations

b. Class format:

Mean = $\frac{\sum fm}{\sum f}$

m = midpoint of class intervals

f = frequency of each class

Σf = Total observations

Median: it is a positional average. It represents the middle value of the variables.

For ungrouped data:

a. Odd series – 5, 10, **11**, 6, 2

Median = middlemost

b. Even series – 5, 10, **11, 6**, 2, 9

Median is the mean or average of the two middle values = $\frac{11+6}{2}$ = $\frac{17}{2}$ = 8.5

For grouped data:

a. Discrete series: median is that value of the variable (x) corresponding to $\frac{n+1}{2}$ (where n total frequency) or next higher value in the cumulative frequency table.

	X	f	cf (cumulative frequency)
	1	5	5
	3	2	7
Median	**9**	11	**18**
	2	3	**21**

n = 21

$\frac{n+1}{2}$ = $\frac{21+1}{2}$ = 11 [∴Median = 9]

b. Class format:

X	f	cf (cumulative frequency)
5-10	6	6
10-15	**3**	**9**
15-20	5	14

n = 14

$\frac{n}{2}$ = $\frac{14}{2}$ = 7

Median = $LL + \frac{\frac{n}{2} + cf_p}{fm} \times i$

LL = lower limit of Median class i.e. 10

n = total frequency i.e. 7

cf_p = cumulative frequency of preceding class

fm = frequency of median class

i = class interval

Mode: it is that value in the distribution series that occurs most often or in most frequent.

For ungrouped data:

Eg: Unimodal – 10, 10, 2, 4, 6

Mode = 10

Bimodal = 10, 10, 2, 2, 46

Mode = 10, 2

For grouped data:

a. Discrete series – value of the variable with highest frequency

x	f
10	60
11	**70**
12	20

b. Class format

x	f	
10-20	20	
20-30	**30**	**Mode class**
30-40	25	

Mode = $LL + \frac{fm - f_1}{2fm - f_1 - f_2} \times i$

LL = lower limit of modal class

fm = Maximum frequency

f_1 = frequency preceding maximum freq.

f_2 = frequency succeeding maximum freq.

i = class interval of modal class

Q. What is meant By Degree of freedom? Discuss the key facts of Degree of freedom with Examples. 10 marks

Ans. Degree of freedom (df)

It refers to the number of categories or groups or variables take into consideration subtracted by 1. Degree of freedom is key in tests of significance such as 't' test, 'F' test, ANOVA, χ^2test.

Ex: χ^2test of independence to see whether 2 organisms of a species habitat are associated with each other.

H_0 = 2 species A and B have no influence over each other in a particular habitat

H_1 = 2 species A and B have influence.

Degree of freedom = 2 – 1 = 1

With level of significance as 0.05

We can find out tabulated value which is then compared with calculated value to test the significance.

- df is depended on the number of variables taken into consideration
- it's value is based on the standard level of significance.
- The degree of freedom cannot be manipulated.
- It is not prone to error as values are fixed from the table.
- Based on tests of significance applied, df can be one or more but ultimately seeks to find the tabulated value. In ANOVA – Two way test it is used to find calculated value too.

Q. What is meant ANOVA & ANCOVA? Discuss its Significance. 10 marks

Ans. ANOVA – it is known as analysis of variance. It is a statistical tool to test the variation of means between 3 categories (minimum). It is divided into 2 types along with it significance.

One-way – Tests one independent variable on one measured variable with at least 3 categories and at least observations in each category. It has only 1 Null and 1 Alternate hypothesis. It compares the means of 3 categories or variables.

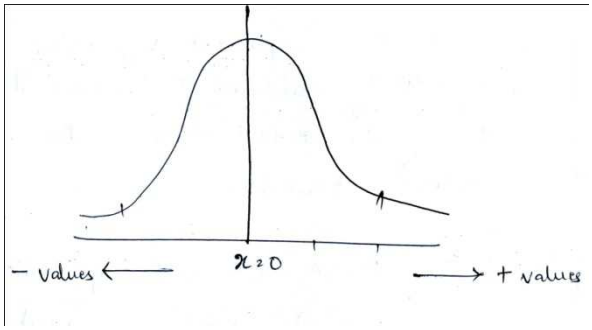
Two-way ANOVA: it tests the effect of the independent variable over the dependent variables.

- It can test two or more independent variables at the same time.
- It has 3 null and 3 alternate hypothesis
- It test the independent variables independently and also the combined effect of the 2 independent variables on the measured variables.
- ANOVA test finds application is new drug formulations research testing the efficacy of drugs under any condition, testing different parameters of drugs, association of species with habitat and other parameters (biotic and abiotic factors).

Q. Discuss the Relationship between mean, median and mode & their Significance. 10 marks

Ans. Mean, median, mode are the measures of Central tendency which are statistical tools to find the centrality of a distribution series.

Mean, Median and Mode coincide at: x = 0



Mean, Median and Mode are always equal at the point of their meeting. They are connected by the equation.

Mode = 3 Median – 2 mean

Significance of Mean, Median and Mode

- It indicates the center of the distribution
- It represents the standard position and the relative difference of other values can be determined.
- Median is usually not affected by extreme values
- Mode is sometimes the only average possible
- Median reduces the absolute standard deviations.
- They are easy to understand and compute.

Q. F Test & T Test.

Ans. F- Test

It is a statistical tool to measure tests of significance. It measures the variance of populations. There are 2 types.

2-Tailed F-Test:

It tests the variance in 2 population. Ex: the length of a fish in 2 separate ponds (2 populations)

	Pond 1	Pond 2
Observations	5	4
Length of fish	6	5
	7	5

I-Tailed F-Test:

It tests the variance of the sample is a single population.

Ex: the length of a fish when it is reared in a single pond for 6 months.

Students 't' test:

First developed by William's Cosset and later improvised by R.A. Fischer, T test is a tests of significance to tests the mean difference in a small population.

It is divided into 2 types:

2-tailed T- test and 1-Tailed T-test

In both student, 't' test and 'F' test the results an interpreted as:

→ Calculated value < tabulated value difference is insignificant

→ Calculated value > tabulated value difference is significant

Q. Types of correlation with Diagram

Ans. Correlation Types

Correlation analysis shows the simultaneous change between 2 variables. It shows correlation but not causation. It is denoted by correlation coefficient 'r' given by Karl Pearson.

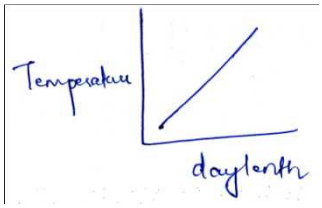
It is divided in 5 types based on degree, extent and distribution. Correlation measures the degree of association.

1. Completely Positive:

   2 variables are directly proportional to each other. Increase in one variable will result in similar change in the 2nd variable.

   R = ±1

   Ex: Day length and Temperature

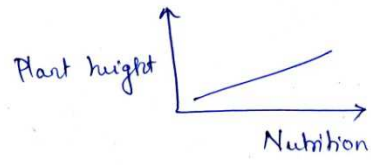


2. Partially Positive

   2 variables are related but not exactly proportional. Increase in 1 variable may not result in exact increase in the second variable.

   r = 0 — +1

   ex: Nutrition and Plant height

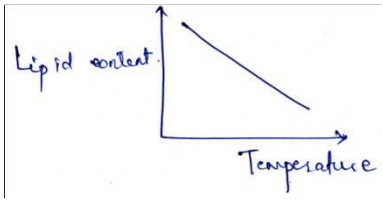


3. Perfectly Negative:

   Decrease in 1 variable will result in increase in another variable. The 2 variables are inversely related.

   R = –1

   ex: Temperature and lipid content in some fish

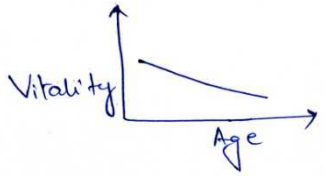


4. Partially negative:

   2 variables are not completely inversely related.

   R = 0 – -1

   Ex: Age and vitality

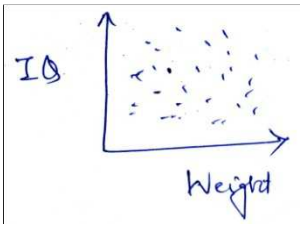


5. No correlation:

   2 variables are not correlated in any way. They are completely independent such that no imaginary line of relationship can be drawn.

   r=0

   ex: Weight and IQ



   Thus, correlation gives a clear picture of association, the degree of association or no association between 2 variable.

Q.

Ans. Mean Deviation and Standard Deviation

Dispersion measures the variability of items. It helps to determine the variation among the items and around the average. The square root of dispersion gives the standard deviation. It shows how the data are distributed, their relative position to itself and around the central mean.

For grouped discrete data:

A. Dispersion of simple S^2 =

$$S^2 = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}$$

n = No. of items

x = value of variable

$\bar{x}$ = mean

Standard deviation of sample S^2 =

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}}$$

B. Dispersion of population: σ^2

$\sigma^2 = \frac{\sum_{i=1}^{n}(x-r)^2}{N}$

r = Mean

N = No. of items

SD = $\sqrt{\frac{\sum_{i=1}^{n}(x-r)^2}{N}}$

JAI HIND

KAVERIS IAS