# METHODS OF STATISTICAL ANALYSIS FOR INTERACTION AND MAIN EFFECTS CONTRIBUTING TO AN ALL OR NOTHING TRAIT

by

CAROLYN FYON

A PROJECT

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

MCMASTER UNIVERSITY

(OCTOBER) 1974

MASTER OF SCIENCE (1974) MCMASTER UNIVERSITY
(MATHEMATICS) HAMILTON, ONTARIO.

TITLE: METHODS OF STATISTICAL ANALYSIS FOR INTERACTION AND MAIN EFFECTS CONTRIBUTING TO AN ALL OR NOTHING TRAIT.

AUTHOR: CAROLYN FYON, B.Sc. (LOYOLA COLLEGE OF MONTREAL)

SUPERVISOR: DR. M. L. TIKU

NUMBER OF PAGES: iii, 53

SCOPE AND CONTENTS: An analysis of the presence or absence of black melanin in broiler chickens as affected by the presence of different traits is studied in the following project. The purpose of this analysis is to show that the simple partitioning of chi-square method is as good as any other method.

This project also shows the equivalence of different statistical methods.

## Acknowledgement

I would like to gratefully acknowledge the help and guidance of Dr. M.L. Tiku, my supervisor.

## Table of Contents

## Introduction

Huntsman, Jerome and Snyder (1960) presented data concerning the incidence of black melanin in the abdominal tissue of broiler chickens. This pigment, when present, is located in the umbilical region of the abdomen and infiltrates the facial tissue there. The character is pleiotropic in nature and its presence or absence appears, for the most part, to be under the control of the plumage color phenotype and sex.

From previous work, it had been noted that three pairs of allelic plumage genes involving three independent loci appear to have distinct influence on the incidence of melanin deposition. The allelic traits concerned were: dominant white (I) and absence of dominant white (i); extended black (E) and restricted black (e); barred (B) and nonbarred (b). The first two pairs on autosomes, the last pair is located on the sex chromosome.

The population of broiler chickens from which the data were obtained was produced by using male parents known to be heterozygous for plumage color alleles at these three loci in a cross with females homozygous for the recessive alleles at each of the three loci. Therefore the males were of the genotype IiEeBb and the females were iieeb-. As a result of these parental genotypes, the eight plumage color phenotypes expected in the population were: IEB, IEb, IeB, Ieb, iEB, iEb, ieB and ieb. (see Fig. 1 on p. 51)

The data is presented in Table I.

Huntsman et al (1960) did no formal analyses on their data but pointed out that certain genetic interactions between genes at different loci appeared to be present.

The purpose of this project is to develop a logical analyses in order to determine the significance of these interlocus interactions and of any main effects produced by alleles within a locus. This will be done by first using a chi-square test and then by analyzing the data using a linear model. A test developed by Woolf will be done for second order interaction.

Chapter I consists of a summary of a number of papers dealing with the chi-square distribution:
Bartlett (1935), Berkson (1955), Berkson (1968), Bhapkar (1966), Bhapkar and Koch (1968), Cochran (1950), Goodman (1963), Kastenbaum and Lamphiear (1959), Plackett (1962), Woolf (1955), Roy and Kastenbaum (1956). There is also a summary of a paper by Grizzle, Starmer and Koch (1969) on linear models and a summary of a paper by Patil (1974) on the analysis of a three dimensional contingency table.

Chapter II consists of a more detailed account of some of the methods used to analyze the data:

(a) partitioning of chi-square
(b) method by Woolf and Plackett to test zero second-order interaction and
(c) methods related to linear models.

Table I

The incidence of melanin pigment deposition in the male offspring produced by a cross of Ii Ee Bb ♂♂ X ii ee b- ♀♀

| Genotype * | Total no. of birds | No. with melanin present | No. with melanin absent | % with melanin present |
|---|---|---|---|---|
| iiEeBb | 76 | 2 | 74 | 2.6 |
| iiEebb | 115 | 102 | 13 | 88.7 |
| iieeBb | 80 | 29 | 51 | 36.3 |
| iieebb | 90 | 17 | 73 | 18.9 |
| IiEeBb | 62 | 1 | 61 | 1.6 |
| IiEebb | 76 | 23 | 53 | 30.3 |
| IieeBb | 73 | 19 | 54 | 26.0 |
| Iieebb | 44 | 6 | 38 | 13.6 |

* the genotypes of the parents were such that the genotypes of the offspring can be deduced from the phenotypes.

## Chapter I

Bartlett (1935) considered the problem of a 2 x 2 x 2 = $2^3$ contingency table. The difference between this table and the ordinary 2 x 2 table is that in the $2^3$ table, the second-order interaction must be taken into account. The $2^3$ table looks as follows:

| | A | | | B | | |
|---|---|---|---|---|---|---|
| | x | y | Total | x | y | Total |
| u | $n_1$ | $n_2$ | $n_1 + n_2$ | $n_5$ | $n_6$ | $n_5 + n_6$ |
| v | $n_3$ | $n_4$ | $n_3 + n_4$ | $n_7$ | $n_8$ | $n_7 + n_8$ |
| Total | $n_1+n_3$ | $n_2+n_4$ | $n_1+n_2+n_3+n_4$ | $n_5+n_7$ | $n_6+n_8$ | $n_5+n_6+n_7+n_8$ |

The standard deviation denoted by x is to be found. This is done by solving: $(n_1+x)(n_4+x)(n_6+x)(n_7+x) = (n_2-x)(n_3-x)(n_5-x)(n_8-x)$. The expected value $a_i$ $(i = 1,2,\ldots,8)$ is then found. Therefore the sum of squares is given by $x^2 \sum_{r=1}^{8} \frac{1}{a_r}$ which is distributed in large sample theory as chi-square with one degree of freedom.

Cochran (1950) wrote a paper describing methods to use when the ordinary chi-square test cannot be used because of matching which may cause correlation between the results in different samples.

If there are only two samples, McNemar's test is used. The 2 x 2 table is of the form:

| | | after | | |
|---|---|---|---|---|
| | | less | more | Total |
| before | less | a | b | a + b |
| | more | c | d | c + d |
| | Total | a + c | b + d | a + b + c + d |

"Matching" means that each sample contains exactly the same subjects. The numbers b and c are tested to see whether they are binomial successes and failures out of n = (b+c) trials with probability ½.

Therefore $\chi^2 = \frac{(b - \frac{n}{2})^2}{\frac{n}{2}} + \frac{(c - \frac{n}{2})^2}{\frac{n}{2}} = \frac{(b - c)^2}{(b + c)}$

with one degree of freedom.

Cochran wanted to extend this test to the situation where there are more than two samples. Suppose we have a table of the form:

| | A | B | C | D | Number |
|---|---|---|---|---|---|
| | $a_1$ | $b_1$ | $c_1$ | $d_1$ | E |
| | $a_2$ | $b_2$ | $c_2$ | $d_2$ | F |
| | $a_3$ | $b_3$ | $c_3$ | $d_3$ | G |
| | $a_4$ | $b_4$ | $c_4$ | $d_4$ | H |
| | $a_5$ | $b_5$ | $c_5$ | $d_5$ | I |
| Total ($T_j$) | K | L | M | N | |

where the "a's" and "b's" are 0's or 1's.

By number, it is meant the number of say cases with that specific combination of the a's and b's. Note K+L+M+N $\neq$ E+F+G+H+I. The values K,L,M and N are obtained by adding up the total number of 1's for each column, e.g. $a_1 = 1$, $a_2 = 1$ and $a_3 = a_4 = a_5 = 0$ e.g., $a_1 = 1$, $a_2 = 1$ and $a_3 = a_4 = a_5 = 0$ Total $T_j = E + F$.

The data is considered as having E+F+G+H+I rows and 4 columns. The test criterion used is $\Sigma(T_j - \bar{T})^2$ where $T_j$ is the total number of successes (1's) in the jth column. This is distributed as $\chi^2 \sigma^2(1 - \rho)$ with (c-1) degrees of freedom where c is the number of variates. Here $\sigma^2 = \sum_i \frac{u_i}{c}(1 - \frac{u_i}{c})$ where $u_i$ represents the successes. The common covariance is given by

$$\rho\sigma^2 = \frac{-\Sigma \frac{u_1}{c}(1 - \frac{u_i}{c})}{(c-1)} = \frac{-\sigma^2}{(c-1)}$$

Therefore $\sigma^2(1-\rho) = \sigma^2 - \rho\sigma^2$

$$= \Sigma \frac{u_i}{c}(1 - \frac{u_i}{c}) + \frac{\sigma^2}{(c-1)}$$

$$= \frac{(c-1)[\Sigma(\frac{u_i}{c})(1 - \frac{u_i}{c})] + \Sigma \frac{u_i}{c}(1 - \frac{u_i}{c})}{(c - 1)}$$

$$= \frac{1}{(c-1)} \Sigma u_i (1 - \frac{u_i}{c}) .$$

Therefore the required test is given by:

$$Q = \frac{(c-1) \sum_j (T_j - \bar{T})^2}{\sum_i u_i(1 - \frac{u_i}{c})} = \frac{c(c-1) \sum_j (T_j - \bar{T})^2}{c(\sum_i u_i) - (\sum_i u_i^2)}$$

which is distributed as chi-square with (c-1) degrees of freedom. Notice that with two samples (i.e. c = 2)

$$Q = \frac{(b-c)^2}{(b+c)}$$

which is the same as that obtained above for the two sample case.

When an example is being worked out: $\sum_i u_i = \sum_j T_j = \sum_i$ (value of $u_i$) frequency and $\sum_i u_i^2 = \sum_i$ (frequency) (value of $u_i^2$). The frequencies are E,F,G,H,I depending on which row is being used. The $u_i$ value is obtained by using the number of 1's in each row that is being considered.

Berkson (1955) wanted to show which is the better: the minimum chi-suare or the maximum likelihood estimate for finite samples where the estimates may differ in their distributions.

We have $P_i = 1 - Q_i = \dfrac{1}{1 - e^{-(\alpha + \beta x_i)}}$

$$p_i = 1 - q_i = \frac{1}{1 - e^{-(a + bx_i)}}$$

The straight line transform of this function called the logit is given by

$$\text{logit } P_i = \ln\left(\frac{P_i}{Q_i}\right) = \alpha + \beta x_i$$

For the maximum likelihood estimates of $\alpha$ and $\beta$, the following equations must be solved:

$$\sum_i n_i \, (p_i - \hat{p}_i) = 0$$

and

$$\sum_i n_i \, x_i \, (p_i - \hat{p}_i) = 0$$

where $n_i$ is the number at $x_i$ and $p_i = 1 - q_i$ is the proportion of $n_i$ observed to respond and $p_i$ is the estimate of $P_i$.

The minimum chi-square is obtained by solving:

$$\sum_i \; n_i \frac{(\hat{p}_i q_i + \hat{q}_i p_i)(p_i - \hat{p}_i)}{\hat{p}_i \hat{q}_i} = 0$$

$$\sum_i \; n_i \frac{(\hat{p}_i q_i + \hat{q}_i q_i) x_i (p_i - \hat{p}_i)}{p_i q_i} = 0$$

To this day, it is still not known which is the better. However, the minimum chi-square has the same asymptotic properties as the maximum likelihood estimate.

Plackett's work (1962) involves interactions in contingency tables. Suppose we have a 2 x 2 x 2 table of the form:

| Combination of Classes | Probability |
|---|---|
| $ABC$ | $p_1$ |
| $A\bar{B}C$ | $p_2$ |
| $\bar{A}BC$ | $p_3$ |
| $\bar{A}\bar{B}C$ | $p_4$ |
| $AB\bar{C}$ | $p_5$ |
| $A\bar{B}\bar{C}$ | $p_6$ |
| $\bar{A}B\bar{C}$ | $p_7$ |
| $\bar{A}\bar{B}\bar{C}$ | $p_8$ |

where $p_1, p_2, \ldots, p_8$ are the probabilities and A denotes the presence of the attribute A and $\bar{A}$ denotes the absence of the attribute A etc. A function $\Psi$ is introduced such that $\Psi(p_1,p_2,p_3,p_4)$ measures the degree of association between A and B in class C. The condition for zero second order interaction is:

$$\Psi(p_1,p_2,p_3,p_4) = \Psi(p_5,p_6,p_7,p_8).$$

Also, $\Psi(p_1,p_3,p_5,p_7) = \Psi(p_2,p_4,p_6,p_8)$ for A and C in class B. Similarly $\Psi(p_1,p_2,p_5,p_6) = \Psi(p_3,p_4,p_7,p_8)$ for B and C in class A.

Bartlett (1935) used for a 2 x 2 table:

$$\Psi(p_1,p_2,p_3,p_4) = \frac{p_1p_4}{p_2p_3} .$$

In a 2 x 2 x 2 table, the condition for zero second-order interaction is

$$p_1p_4p_6p_7 = p_2p_3p_5p_8$$

because $\Psi(p_1,p_2,p_3,p_4) = \Psi(p_5,p_6,p_7,p_8)$

$$\frac{p_1p_4}{p_2p_3} = \frac{p_5p_8}{p_6p_7}$$

$$p_1p_4p_6p_7 = p_2p_3p_5p_8$$

Similarly $\Psi(p_1,p_3,p_5,p_7) = \Psi(p_2,p_4,p_6,p_8)$

$$\frac{p_1p_7}{p_3p_5} = \frac{p_2p_8}{p_4p_6}$$

$$p_1p_2p_4p_6 = p_2p_8p_3p_5$$

Similarly $\Psi(p_1,p_2,p_5,p_6) = \Psi(p_3,p_4,p_7,p_8)$

$$\frac{p_1p_6}{p_2p_5} = \frac{p_3p_8}{p_4p_7}$$

$$p_1p_6p_4p_7 = p_2p_5p_3p_8 \ .$$

In all cases (and therefore consistent)

$$p_1p_4p_6p_7 = p_2p_3p_5p_8$$

There is another way of analyzing a 2 x 2 x t table given by Woolf (1955). Let the frequencies in the kth 2 x 2 table be denoted by $n_{1k}$, $n_{2k}$, $n_{3k}$, $n_{4k}$ where $n_{1k}$, $n_{2k}$ occupy the first row and $n_{1k}$, $n_{3k}$ the first column.

Compute:

$$z_k = \ell n\ n_{1k} - \ell n\ n_{2k} - \ell n\ n_{3k} + \ell n\ n_{4k}$$

and $u_k$ from

$$\frac{1}{u_k} = \frac{1}{n_{1k}} + \frac{1}{n_{2k}} + \frac{1}{n_{3k}} + \frac{1}{n_{4k}}$$

If there is zero second-order interaction

$$X^2 = \sum_k u_k z_k^2 = \frac{(\sum_k u_k z_k)^2}{u_k}$$

is asymptotically distributed as $\chi^2$ with (t-1) degrees of freedom.

Roy and Kastenbaum (1956) also discussed the hypothesis of no interaction: Suppose we have a three way table: let $n_{ijk}$ denote the observed frequency and $p_{ijk}$ the probability in the (ijk)th cell where i = 1,2,...,r; j = 1,2,...,s; k = 1,2,...,t.

Let the marginals be denoted in the usual manner i.e.,

$\sum_i n_{ijk} = n_{\cdot jk}$ and $\sum_{i,j,k} n_{ijk} = n.$ Similarly define

$\sum_i p_{ijk} = p_{\cdot jk}$ etc. The likelihood function is

$$\phi(n_{ijk}\text{'s}) = \phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}}$$

As $n! \sim \prod_{i,j,k} n_{ijk}!$

$$\phi \sim \prod_{i,j,k} p_{ijk}^{n_{ijk}}$$

The hypothesis of independence between (i,j) and k is to be tested. The best way of doing this is by using the following set of conditions:

$$H_o : p_{ijk} = \frac{q_{ij.}\; q_{i.k}\; q_{.jk}}{q_{i..}\; q_{.j.}\; q_{..k}}$$

The role of the q's is to yield certain constraints on the p's.

For an r × s × t table, the following "no interaction" constraints are present:

$$\left.\frac{p_{rst}\, p_{ijt}}{p_{ist}\, p_{rjt}} = \frac{p_{rsk}\, p_{ijk}}{p_{isk}\, p_{rjk}}\right\} \quad \begin{array}{l} i = 1,2,\ldots\ (r-1) \\ j = 1,2,\ldots\ (s-1) \\ k = 1,2,\ldots\ (t-1) \end{array}$$

Therefore there are (t-1)(s-1)(r-1) constraints on the $p_{ijk}$'s.

Maximize $\phi$ subject to the "no interaction" constraints and all $\Sigma p_{ijk} = 1$. To do this, use the Lagrangian multiplier $\lambda_{ijk}$ for $\frac{p_{rst}\, p_{rjt}}{p_{ist}\, p_{rjt}} = \frac{p_{rsk}\, p_{ijk}}{p_{isk}\, p_{rjk}}$ and the Lagrangian multiplier $\mu$ for $\Sigma p_{ijk} = 1$. Solving these equations, the following is obtained:

$$\sum_{i,j,k} \frac{\mu^2_{ijk}}{(n_{ijk} + \eta_{ijk}\ \mu_{ijk})}$$

which is distributed as $\chi^2$ with the degrees of freedom equal to the number of "no interaction" constraints on the p's which is equal to (r-1)(s-1)(t-1) where $i = 1,2,\ldots,r$; $j = 1,2,\ldots,s$; $k = 1,2,\ldots,t$ and

$\eta_{ijk} = 1$ if ijk = rst or if any two subscripts differ

$\eta_{ijk} = 1$ if any one subscript differs from the corresponding pivotal or if all the subscripts differ.

Berkson (1968) discussed logit analysis. Linear formulas of the estimates are used in this kind of analysis. Therefore iterative methods are not necessary. The logit $\chi^2$ is computed directly from the observations.

Consider the following table:

| birth order k | i | No. of mothers with Losses 1 | No losses 2 | Total |
|---|---|---|---|---|
| 1 | 1 | $a_1$ | $b_1$ | $n_{11}$ |
| | 2 | $c_1$ | $d_1$ | $n_{21}$ |
| 2 | 1 | $a_2$ | $b_2$ | $n_{12}$ |
| | 2 | $c_2$ | $d_2$ | $n_{22}$ |
| 3 | 1 | $a_3$ | $b_3$ | $n_{13}$ |
| | 2 | $c_3$ | $d_3$ | $n_{23}$ |

with $p_{1k} = 1 - q_{1k} = a_k/n_{1k}$

$p_{2k} = 1 - q_{2k} = c_k/n_{2k}$

$P_{1k} = 1 - Q_{1k} =$ probability corresponding to $p_{1k}$

$P_{2k} = 1 - Q_{2k} =$ probability corresponding to $p_{2k}$

$\ell_{1k} = \text{logit } p_{1k} = \ell n\ (a_k/b_k)$

$\ell_{2k} = \text{logit } p_{2k} = \ell n\ (c_k/d_k)$

$L_{1k} = \text{logit } P_{1k} = \ell n\ (P_{1k}/Q_{1k})$

$L_{2k} = \text{logit } P_{2k} = \ell n\ (P_{2k}/Q_{2k})$

$\beta_k = L_{1k} - L_{2k}$

$B_k = \ell_{1k} - \ell_{2k}$

$\Psi_k = \text{odds ratio} = \frac{P_{1k}}{Q_{1K}} \quad \frac{P_{2k}}{Q_{2k}} = e^{\beta_k}$

$$C_{1k} = 1/a_k + 1/b_k$$

$$C_{2k} = 1/c_k + 1/d_k$$

$$\tilde{C}_t = \sum_{k=1}^{3} C_{1k} + \sum_{k=1}^{3} C_{2k}$$

$$w_{1k} = \frac{1}{C_{1k}}$$

$$w_{2k} = \frac{1}{C_{2k}}$$

$$\tilde{w}_k = \frac{1}{(C_{1k} + C_{2k})}$$

The analysis is broken into four cases as follows:

Case I : the hypothesis of no interaction is tested. The logit $\chi^2$ is given by

$$\chi^2_\ell = \sum_{k=1}^{3} n_{1k}\, p_{1k}\, q_{1k}\, (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} n_{2k}\, p_{2k}\, q_{2k}\, (\ell_{2k} - L_{2k})^2$$

$$= \sum_{k=1}^{3} w_{1k}\, (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} w_{2k}\, (\ell_{2k} - L_{2k})^2$$

the constraint is

$$L_{11} - L_{21} = L_{12} - L_{22} = L_{13} - L_{23} = \beta$$

which can be written as

$$F_1\ (L) = F_1 = L_{11} - L_{21} - L_{12} + L_{22} = 0$$

$$F_2\ (L) = F_2 = L_{11} - L_{21} - L_{13} + L_{23} = 0.$$

Using Lagrangian multipliers, the following is obtained

$$\chi_{\ell}^2 = \sum_{k=1}^{3} w_{1k} (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} w_{2k} (\ell_{2k} - L_{2k})^2 + \lambda_1 F_1 + \lambda_2 F_2.$$

Setting the differentials with respect to the L's equal to zero and substituting the solutions for L and estimating $\lambda_1$, $\lambda_2$.

$$\hat{\beta} = \frac{\tilde{w}_1 B_1 + \tilde{w}_2 B_2 + \tilde{w}_3 B_3}{\tilde{w}_1 + \tilde{w}_2 + \tilde{w}_3}$$

and

$$\chi_{\ell}^2 = \sum_{k=1}^{3} w_{1k} (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} w_{2k} (\ell_{2k} - L_{1k} + \hat{\beta})^2. \quad (1)$$

Differentiating $\chi_{\ell}^2$ with respect to the $L_{1k}$ and equaling to zero, the estimate of $L_{1k}$ is obtained

$$\hat{L}_{1k} = \frac{w_{1k} \ell_{1k} + w_{2k} \ell_{2k} + w_{2k} \hat{\beta}}{w_{1k} + w_{2k}}$$

(2)

and

$$\hat{L}_{2k} = \hat{L}_{1k} - \hat{\beta}$$

Substituting (2) in (1)

$$\chi_{\ell} = \sum_{k=1}^{3} \tilde{w}_k (B_k - \hat{\beta})^2 = \sum_{k=1}^{3} \tilde{w}_k B_k^2 - \hat{\beta}^2 \sum_{k=1}^{3} \tilde{w}_k .$$

Case II: Is there a difference between problems and controls.

The constraint is

$$F(L) = F = L_{11} + L_{12} + L_{13} - L_{21} - L_{22} - L_{23} = 0.$$

Using Lagrangian multipliers

$$\chi_\ell^2 = \sum_{k=1}^{3} w_{1k} (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} w_{2k} (\ell_{2k} - L_{2k})^2 + \lambda (F). \quad (3)$$

Differentiating with respect to the L's and solving

$$L_{1k} = \ell_{1k} - \frac{\lambda}{2w_{1k}} - \ell_{1k} - \frac{\lambda}{2} C_{1k}$$

(4)

$$L_{2k} = \ell_{2k} + \frac{\lambda}{2w_{2k}} - \ell_{2k} + \frac{\lambda}{2} C_{2k}$$

Substituting (4) in (3)

$$\lambda = \frac{2(\sum_{k=1}^{3} \ell_{1k} - \sum_{k=1}^{3} \ell_{2k})}{\widetilde{C}_t} \quad (5)$$

Substituting (5) in (4)

$$\left.\begin{aligned} \hat{L}_{1k} &= \ell_{1k} - \tfrac{1}{2}\lambda C_{1k} \\ \hat{L}_{2k} &= \ell_{2k} - \tfrac{1}{2}\lambda C_{2k} \end{aligned}\right\} \quad (6)$$

Substituting (6) into

$$\chi_\ell^2 = \sum_{k=1}^{3} w_{1k} (\ell_{1k} - L_{1k})^2 + \sum_{k=1}^{3} w_{2k} (\ell_{2k} - L_{2k})^2$$

the following is obtained:

$$\chi_\ell^2 = \tfrac{1}{4} \widetilde{C}_t \lambda^2.$$

Case III: The hypothesis of equality of birth order effects is to be tested. The constraint is

$$L_{11} + L_{21} = L_{12} + L_{22} = L_{13} + L_{23}.$$

This case is similar to Case I and

$$\hat{\beta}' = \frac{(\widetilde{w}_1 B_1' + \widetilde{w}_2 B_2' + \widetilde{w}_3 B_3')}{\widetilde{w}_1 + \widetilde{w}_2 + \widetilde{w}_3}$$

where $B_1' = \ell_{11} + \ell_{21}$

and $B_2' = \ell_{12} + \ell_{22}$

and $B_3' = \ell_{13} + \ell_{23}$

$$\widetilde{L}_{1k} = \frac{(w_{1k}\,\ell_{1k} - w_{2k}\,\ell_{2k} + w_{2k}\,\hat{B}')}{w_{1k} + w_{2k}}$$

$$\widetilde{L}_{2k} = \hat{B}' - \widetilde{L}_{1k}$$

and $\chi^2_\ell = \sum_{k=1}^{3} \widetilde{w}_k \, (B_k' - B')^2 = \sum_{k=1}^{3} \widetilde{w}_k \, B'^2_k - \hat{B}'^2 \sum_{k=1}^{3} \tilde{w}_k.$

<u>Case IV</u>: The following two hypotheses are tested: is the effect of birth order linear or does it require a second degree polynomial to describe the effect? The restrictions are:

$$L_{11} + L_{21} = L_{13} + L_{23}$$

and $L_{11} + L_{21} - 2L_{12} - 2L_{22} + L_{13} + L_{23} = 0$ and the estimate of B is

$$\hat{B}' = \frac{(\widetilde{w}_1 \ B_1' + \widetilde{w}_3 \ B_3')}{\widetilde{w}_1 + \widetilde{w}_3}$$

where $B' = L_{11} + L_{21} = L_{13} + L_{23}$

and $B_1' = \ell_{11} + \ell_{21}$

$B_3' = \ell_{13} + \ell_{23}$

$$\tilde{L}_{11} = \frac{(w_{11} \ \ell_{11} - w_{21} \ \ell_{21} + w_{21} \ \hat{B}')}{w_{11} + w_{21}}$$

$\tilde{L}_{21} = \hat{B}' - \tilde{L}_{11}$

$\tilde{L}_{12} = \ell_{12}$

$\tilde{L}_{22} = \ell_{22}$

$$\tilde{L}_{13} = \frac{w_{13} \ \ell_{13} - w_{23} \ \ell_{23} + w_{23} \ \hat{B}'}{w_{13} + w_{23}}$$

$\tilde{L}_{23} = \hat{B}' - L_{13}.$

Therefore

$$\chi_{\ell}^2 = \tilde{w}_1 \ (B_1' - \hat{B})^2 + \tilde{w}_3 (B_3' - \hat{B})^2$$

$$= \tilde{w}_1 \ B_1'^2 + w_3 \ B_3'^2 - \hat{B}'^2 (\tilde{w}_1 + \tilde{w}_3).$$

Grizzle, Starmen and Koch (1969) fitted a linear model to analyze categorical data. Various models were used.

Suppose the data is arranged as follows:

Frequency Distribution

| Populations | Categories of response 1 | 2 | ....... | r | Total |
|---|---|---|---|---|---|
| 1 | $n_{11}$ | $n_{12}$ | | $n_{1r}$ | $n_{1\cdot}$ |
| 2 | $n_{21}$ | $n_{22}$ | | $n_{2r}$ | $n_{2\cdot}$ |
| . | . | . | | . | . |
| . | . | . | | . | . |
| . | . | . | | . | . |
| s | $n_{s1}$ | $n_{s2}$ | | $n_{sr}$ | $n_{s\cdot}$ |

Expected Cell Probabilities

| Populations | Categories of response 1 | 2 | ...... | r | Total |
|---|---|---|---|---|---|
| 1 | $\Pi_{11}$ | $\Pi_{12}$ | ...... | $\Pi_{1r}$ | 1 |
| 2 | $\Pi_{21}$ | $\Pi_{22}$ | | $\Pi_{2r}$ | 1 |
| . | . | . | | . | . |
| . | . | . | | . | . |
| . | . | . | | . | . |
| s | $\Pi_{s1}$ | $\Pi_{s2}$ | | $\Pi_{sr}$ | 1 |

Define

$$p_{ij} = n_{ij}/n_{i.}$$

$$\underset{\sim}{p}_i' = [p_{i1}, p_{i2}, \ldots, p_{ir}]$$

$$\text{var} \quad (\underset{\sim}{p}_i) = v(\underset{\sim}{\Pi}_i) = \frac{1}{n_i}\begin{bmatrix} \Pi_{i1}(1-\Pi_{i1}) & -\Pi_{i1}\Pi_{i2} & \cdots & -\Pi_{i1}\Pi_{ii} \\ -\Pi_{i1}\Pi_{i2} & \Pi_{i2}(1-\Pi_{i2}) & \cdots & -\Pi_{i2}\Pi_{ii} \\ \cdot & \cdot & \cdots & \cdot \\ -\Pi_{i1}\Pi_{i2} & \cdots\cdots & & -\Pi_{ii}(1-\Pi_{ii}) \end{bmatrix}$$

$V(p_i)$ = sample estimate of $v(\Pi_i)$ $(p_{ij} = \Pi_{ij})$

$V(p)$ = block diagonal with $V(p_i)$ on the main diagonal

$f_m(\Pi)$ = any function of the elements of p that has partial derivatives up to second order with respect to the $\Pi_{ij}$

$f_m(\Pi) = f_m(p)$ evaluated at $\Pi = p$

$$H = \left[\frac{\partial fm\ (\Pi)}{\partial \Pi_{ij}} \,\bigg|\, \Pi_{ij} = p_{ij}\right]$$

and $S = HV(p)H'$.

Assume $\underset{u \times 1}{F}(\Pi) = \underset{u \times v}{X}\ \underset{v \times 1}{\beta}$

where X is a known design matrix and $\beta$ is a vector of unknown parameters. The test statistic used to see if the data fits a particular model is $SS\ [F(\Pi) = x\beta] = F'\ S^{-1}\ F - b'(X'S^{-1}X)b$ where $b = (X'S^{-1}X)^{-1}\ X'S^{-1}F$ which is distributed as chi-square with (u - v) degrees of freedom. If the value obtained is less than the known value, the data fits that model and therefore row and column effects are tested. This is done by using the following statistic and by choosing the appropriate C matrix (depends on whether row or column effects are being tested) which is a $(d \times v)$ matrix. The test statistic is

$$SS[C\beta = 0] = b'C'[C(X's^{-1}X)^{-1}C']^{-1}\ Cb$$

which is distributed as chi-square with d degrees of freedom.

Grizzle et al (1969) then described other models based on this general model which will be mentioned in the next chapter.

Patil (1974) described another method for analyzing an $r \times s \times t$ contingency table. Suppose our data is arranged as follows:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_{111}$ | $n_{121}$ | ... | $n_{1s1}$ | $n_{211}$ | $n_{221}$ | ... | $n_{2s1}$ | ....... | $n_{r11}$ | ... | $n_{rs1}$ |
| $n_{112}$ | $n_{122}$ | ... | $n_{1s2}$ | $n_{212}$ | $n_{222}$ | ... | $n_{2s2}$ | ....... | $n_{r12}$ | ... | $n_{rs2}$ |
| $n_{11t}$ | $n_{12t}$ | ... | $n_{1st}$ | $n_{21t}$ | $n_{22t}$ | ... | $n_{2st}$ | ....... | $n_{r1t}$ | ... | $n_{rst}$ |

Let $i = 1,r$

$j = 1,s$

$k = 1,t$

Form the matrix $Y_k' = (n_{11k},\ n_{12k}, \ldots,\ n_{1(s-1)k}, \ldots\ n_{(r-1)(s-1)k})$

Let $n_{i.k} = \sum_j n_{ijk}$

$n_{.jk} = \sum_i n_{ijk}$

$n_{..k} = \sum_{ij}\sum n_{ijk}$

Calculate the matrix $\mu_{ijk}$ (mean vector) and $\Sigma_{ijk}$ (covariance matrix) for $k = 1,\ldots,t$ when

$$\mu_{ijk} = E(n_{ijk}) = \frac{n_{i.k}\, n_{.jk}}{n_{..k}}$$

and

$$\Sigma_{ijk} = \frac{n_{i.k}\, n_{.jk}(n_{..k} - n_{i.k})(n_{..k} - n_{.jk})}{n^2_{..k}\,(n_{..k}-1)}$$

Now calculate $\chi^2_k$ for $k = 1,\ldots,t$.

This is done as follows

$$\chi^2_k = (Y_k - \mu_k)' \; \Sigma^{-1}_k (Y_k - \mu_k)$$

and

$$\chi^2_0 = (Y - \mu)' \; \Sigma^{-1}(Y - \mu)$$

where $Y' = \sum_k Y'_k$

and $\Sigma = \sum_k \Sigma_k$

and $\mu' = \sum_k \mu_k'$ .

Therefore the required statistic to test the null hypothesis which tests for zero second-order interaction is

$$\chi^2 = \sum_{k=1}^{t} \chi^2_k - \chi^2_0$$ with $(r-1)(s-1)(t-1)$ degrees of freedom.

## CHAPTER II

### 2.1. Orthogonal partitioning of chi-squares

We now analyze the data by adopting a straightforward linear model.

Consider the following mathematical model:

$$E(Y_{ijk\ell}) = \mu + E_i + B_j + I_k + (EB)_{ij} + (EI)_{ik} + (BI)_{jk} \qquad (1)$$

$$i = 1,2 \qquad j = 1,2 \qquad k = 1,2 \qquad \ell = 1,2,\ldots,n_{ijk}$$

where single letters represent main effects and double letters represent first order interactions. We assume (see Scheffé, page 92).

$$\sum_{i=1}^{2} n_{i..}\, E_i = 0 \qquad \sum_{j=1}^{2} n_{.j.}\, B_j = 0 \qquad \sum_{k=1}^{2} n_{..k}\, I_k = 0$$

$$\sum_{ij} n_{ij.}\, (EB)_{ij} = 0 \qquad \sum_{i}\sum_{k} n_{i.k}\, (EI)_{ik} = 0 \qquad \sum_{k}\sum_{j} n_{.jk} (BI)_{jk} = 0 \qquad (2)$$

The objective is to test the null hypothesis that the main effects and interactions are all zero.

Since the distribution of $Y_{ijk\ell}$ is binomial, i.e., $\theta^Y(1-\theta)^Y \quad Y = 0,1 \quad (3)$ we cannot adopt the conventional analysis-of-variance technique to test the null hypothesis. However, we still can estimate parameters through "least square estimation".

We first want to find the least square estimators of the main effects and interactions. To do this, we must minimize the following expression:

$$A = \sum_i \sum_j \sum_k \sum_\ell \{Y_{ijk\ell} - \mu - E_i - B_j - I_k - (EB)_{ij} - (EI)_{ik} - (BI)_{jk}\}^2$$

$$= \sum_i \sum_j \sum_k \sum_\ell \{Y^2_{ijk\ell} - 2\mu Y_{ijk} - 2E_i Y_{ijk\ell} - 2B_j Y_{ijk\ell} - 2I_k Y_{ijk\ell}$$

$$- 2E_i B_j Y_{ijk\ell} - 2I_k E_i Y_{ijk\ell} - 2B_j I_k Y_{ijk\ell} + \mu^2 + 2\mu E_i + 2\mu B_j$$

$$+ 2\mu I_k + 2\mu E_i B_j + 2\mu E_i I_k + 2\mu B_j I_k + E_i^2 + 2B_j E_i + 2E_i I_b$$

$$+ 2E_i{}^2 B_j + 2E_i{}^2 I_k + 6E_i B_j I_k + B_j{}^2 + 2B_j I_k + 2E_i B_j{}^2 + 2B_j{}^2 I_k + I_k{}^2$$

$$+ 2E_i I_k{}^2 + 2B_j I_k{}^2 + E_i{}^2 B_j{}^2 + 2E_i{}^2 I_k B_j + 2E_i I_k B_j{}^2 + E_i{}^2 I_k{}^2 +$$

$$+ 2E_i B_j I_k{}^2 + B_j{}^2 I_k{}^2\}.$$

Differentiating A with respect to the parameters and solving the resulting equations, the following estimators are obtained:

$$\hat{\mu} = \bar{Y}_{....}$$

$$\hat{E}_i = \bar{Y}_{i...} - \bar{Y}_{....}$$

$$\hat{B}_j = \bar{Y}_{.j..} - \bar{Y}_{....}$$

$$\hat{I}_k = \bar{Y}_{..k.} - \bar{Y}_{....}$$

$$(\hat{EB})_{ij} = \bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....}$$

$$(\hat{EI})_{ik} = \bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....}$$

$$(\hat{BI})_{jk} = \bar{Y}_{.jk.} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y}_{....}$$

where $\bar{Y}_{....} = \frac{1}{n} \sum_i \sum_j \sum_k \sum_\ell Y_{ijk\ell} = \frac{1}{N} Y_{....}$

where $N = \sum_i \sum_j \sum_k n_{ijk}$

and $\bar{Y}_{i..} = \frac{1}{n_{i..}} \sum_j \sum_k \sum_\ell Y_{ijk\ell} = \frac{1}{n_{i..}} Y_{i...}$

where $n_{i..} = \sum_j \sum_k n_{ijk}$ etc.

(These values are seen in Tables II and III). These estimators are asymptotically independently distributed.

As an example, it will be shown how the estimator $\hat{\mu}$ was obtained upon differentiating. When we differentiate A with respect to $\mu$ , the following terms are obtained: (the other terms do not contain $\mu$ and therefore are equal to zero).

$$B = -2 \sum_i \sum_j \sum_k \sum_\ell Y_{ijk\ell} + 2 \sum_i \sum_j \sum_k n_{ijk}\mu + 2 \sum_i n_{i..} E_i$$

$$+ 2 \sum_j n_{.j.} B_j + 2 \sum_k n_{..k} I_k + 2 \sum_i \sum_j n_{ij.} E_i B_j + 2 \sum_i \sum_k n_{i.k} E_i I_k$$

$$+ 2 \sum_j \sum_k n_{.jk} B_j I_k .$$

To minimize this, set it equal to zero. Note also from (2) that $\sum_i n_{i...} E_i = 0$ etc.

Therefore

$$-\sum_i \sum_j \sum_k \sum_\ell Y_{ijk\ell} + \sum_i \sum_j \sum_k n_{ijk}\mu = 0$$

$$\mu = \frac{\sum_i \sum_j \sum_k \sum_\ell Y_{ijk\ell}}{\sum_i \sum_j \sum_k n_{ijk}} = \frac{Y_{....}}{N}$$

TABLE II

Values of $Y_{ijk\ell}$

| | k = 1 | | k = 2 | | |
|---|---|---|---|---|---|
| | j = 1 | j = 2 | j = 1 | j = 2 | $Y_{i...}$ |
| i = 1 | 2 | 102 | 1 | 23 | 128 |
| i = 2 | 29 | 17 | 19 | 6 | 71 |
| $Y_{.j..}$ | 150 | | 49 | | $Y_{....}$ = 199 |

TABLE III

Values of $n_{ijk}$

| | k = 1 | | k = 2 | | |
|---|---|---|---|---|---|
| | j = 1 | j = 2 | j = 1 | j = 2 | $n_{i..}$ |
| i = 1 | 76 | 115 | 62 | 76 | 229 |
| i = 2 | 80 | 90 | 73 | 44 | 287 |
| $n_{.j.}$ | 361 | | 255 | | $n_{...}$ = 616 |

$$\hat{\mu} = \bar{Y}_{....} = \frac{1}{N} Y_{....} = \frac{199}{616} = .3230519$$

standard error $= \sqrt{s^2/N} = \sqrt{.000229} = \pm .0151$

$$\hat{E}_1 = Y_{1....} - Y_{....} = \frac{128}{329} - \frac{199}{616} = .0660058$$

$$\text{s.e.} = \sqrt{s^2(\frac{1}{N_{1...}} - \frac{1}{N})} = \sqrt{.0001998} = \pm .0141$$

$$\hat{E}_2 = Y_{2...} - Y_{....} = \frac{71}{287} - \frac{199}{616} = -.0756652$$

$$\text{s.e.} = \sqrt{s^2(\frac{1}{N_{2....}} - \frac{1}{N})} = \sqrt{.0002625} = \pm .0162$$

$$\hat{I}_1 = Y_{..1.} - Y_{....} = \frac{150}{361} - \frac{199}{616} = .0924605$$

$$\text{s.e.} = \sqrt{.0001616} = \pm .0127$$

$$\hat{I}_2 = Y_{..2.} - Y_{....} = \frac{49}{255} - \frac{199}{616} = -.1308951$$

$$\text{s.e.} = .003242 = \pm .018$$

$$\hat{B}_1 = Y_{.1..} - Y_{....} = \frac{51}{291} - \frac{199}{616} = -.1477942$$

$$\text{s.e.} = \sqrt{.0002558} = \pm .01599$$

$$\hat{B}_2 = Y_{.2..} - Y_{....} = \frac{148}{325} - \frac{199}{616} = .1323327$$

$$\text{s.e.} = \sqrt{.0002051} = \pm .01432$$

$$(\widehat{EB})_{11} = Y_{11..} - Y_{1...} - Y_{.1..} + Y_{....} = \frac{3}{138} - \frac{128}{329} - \frac{51}{291} + \frac{199}{616} = -.2195244$$

$$\text{s.e.} = \sqrt{\left(\frac{1}{138} - \frac{1}{329} - \frac{1}{291} + \frac{1}{616}\right)s^2} = \pm.01838$$

$$(\hat{EB})_{12} = Y_{12..} - Y_{1...} - Y_{.2..} + Y_{....} = \frac{175}{191} - .3890577 - \frac{148}{325}$$

$$+ .3234519 = .1330958$$

$$\text{s.e.} = \sqrt{.0001047} = \pm.0102$$

$$(\hat{EB})_{21} = Y_{21..} - Y_{2...} - Y_{.1..} + Y_{....} = .2141329$$

s.e. = $\pm.0132$

$$(\hat{EB})_{22} = Y_{22..} - Y_{2...} - Y_{12..} + Y_{....} = -.208077$$

s.e. = $\pm.01887$

$$(\hat{EI})_{11} = Y_{1.1.} - Y_{1...} - Y_{..1.} + Y_{....} = .0629844$$

s.e. = $\pm.012165$

$$(\hat{EI})_{12} = Y_{1.2.} - Y_{1...} - Y_{112.} + Y_{....} = -.0842496$$

s.e. = $\pm.0164$

$$(\hat{EI})_{21} = Y_{2.1.} - Y_{2...} - Y_{..1.} + Y_{....} = -.069259$$

s.e. = $\pm.0197$

$$(\hat{BI})_{11} = Y_{.111} - Y_{.1..} - Y_{..1.} + Y_{....} = -.0690003$$

s.e. = $\pm.01605$

$$(\hat{BI})_{12} = Y_{.12.} - Y_{.1..} - Y_{..2.} + Y_{....} = .1037855$$

s.e. = ±.01536

$$(\hat{BI})_{21} = Y_{.21.} - Y_{.2..} - Y_{..1.} + Y_{....} = .0326427$$

s.e. = ±.0096

$$(\hat{BI})_{22} = Y_{.22.} - Y_{.211} - Y_{112.} + Y_{....} = -.0828229$$

s.e. = ±.0204

The standard errors were obtained using the following formulas.

$$v(\hat{\mu}) = \frac{\sigma^2}{N} \qquad N = 616$$

$$v(\hat{E}_i) = \sigma^2 \left(\frac{1}{N_{i..}} - \frac{1}{N}\right)$$

$$v(\hat{B}_j) = \sigma^2 \left(\frac{1}{N_{.j.}} - \frac{1}{N}\right)$$

etc.

where $\sigma^2$ is replaced by its least square estimate.

From above, it is seen that all the estimates of the parameters deviate from zero more than four times their standard errors AND this is an indication that the main effects and interactions are non-zero. The null hypothesis that the main effects and interactions are zero can be tested by using orthogonal partitioning of chi-square as follows. An excellent description of the technique of partitioning chi-square is given by Mather (1957).

The main effects can be tested by calculating the values of

$$\chi^2 = \left\{\frac{(\text{observed-expected})^2}{\text{expected}}\right\} + \ldots\ldots$$

In the tables that follow (1) stands for the observed values and (2) stands for the expected values when the null hypothesis about that specific effect is true.

In Table IV, the null hypothesis that the main effect E is zero is being tested. The values for the table are obtained by using the following formulas:

(1) $\sum_{j}\sum_{k} n_{ijk}\ (\mu + E_i)$

(2) $\sum_{j}\sum_{k} n_{ijk}\ (\mu)$ $\qquad H_o : E_i = 0$

The specific numbers is Table IV were obtained as follows:

for i = 1 (1) valve was obtained from $\sum_{i}\sum_{k} n_{ijk}(\mu + E_1)$
$\cong 329(.323 + .066) = 128$

for i = 2 (1) value was obtained from $\sum_{j}\sum_{k} n_{ijk}\ (\mu + E_2)$
$\cong 287(.323 - .076) = 71$

for i = 1 (2) value was obtained from $\sum_{j}\sum_{k} n_{ijk}\ (\mu) \cong 329(.3230519)$
$= 106.284$

for i = 2 (2) value was obtained from $\sum_{j}\sum_{k} n_{ijk}\ (\mu) \cong 287(.3230519)$
$= 92.715909$

Table IV

Observed and expected values required for the determination of $\chi^2 E$

| | i = 1 | i = 2 | Sum |
|---|---|---|---|
| (1) | 128 | 71 | 199 |
| (2) | 106.284 | 92.715909 | 199 |

Therefore $\chi_E^2 = \frac{(128 - 106.284)^2}{106.284} + \frac{(71 - 92.715909)^2}{92.715909} = 9.523359.$

In a similar manner the values in Table V were obtained using:

(1) $\sum_i \sum_k n_{ijk} (\mu + B_j)$

(2) $\sum_i \sum_j n_{ijk} (\mu)$ $\qquad H_o : B_j = 0$

Table V

Observed and expected values required for the determination of $\chi_B^2$

| | j = 1 | j = 2 | Sum |
|---|---|---|---|
| (1) | 51 | 148 | 199 |
| (2) | 94.008 | 104.992 | 199 |

$$\chi_B^2 = \frac{(51 - 94.008)^2}{94.008} + \frac{(148 - 104.992)^2}{104.992} = 37.293275.$$

The values in Table VI were obtained by using:

(1) $\sum_i \sum_j n_{ijk} (\mu + I_k)$

(2) $\sum_i \sum_j n_{ijk} (\mu)$ $\qquad H_o : I_k = 0$

Table VI

Observed and expected values required for the determination of $\chi^2_I$

| | j = 1 | j = 2 | Sum |
|---|---|---|---|
| (1) | 150 | 49 | 199 |
| (2) | 116.622 | 83.378 | 199 |

$$\chi^2_I = \frac{(150 - 116.622)^2}{116.622} + \frac{(49 - 83.378)^2}{83.378} = 23.077137$$

The interaction chi-squares are now calculated. For the determination of $\chi^2_{(EB)}$, the following are used:

(1) $\sum_k n_{ijk} \{\mu + E_i + B_j + (EB)_{ij}\}$

(2) $\sum_k n_{ijk} (\mu + E_i + B_j)$ $\quad H_o : (EB)_{ij} = 0$

Table VII

Observed and expected values required for the determination of $\chi^2_{(EB)}$

| | | j = 1 | j = 2 | Sum |
|---|---|---|---|---|
| i=1 | (1) | 3 | 125 | 128 |
| | (2) | 33.294 | 99.586 | 132.88 |
| i=2 | (1) | 48 | 23 | 71 |
| | (2) | 152.238 | 50.882 | 66.12 |
| Sum | (1) | 51 | 148 | 199 |
| | (2) | 48.532 | 150.468 | 199 |

Sample calculation:

for i = 1, j = 1 the (1) value was obtained by $\sum_k \{\mu + E_1 + B_1 + (EB)_{11}$

$= 135\ (.323 + .066 - .148 - .2195) = 3.$

Using Table VII, the chi-square values can now be calculated. The total chi-square subclasses are divided into three components: due to differences in row totals, due to differences in column totals and due to interaction between E and B.

$$\chi^2_{\text{subclasses}} = \frac{(3-33.294)^2}{33.294} + \frac{(48-15.238)^2}{15.238} + \frac{(125-99.586)}{99.586} + \frac{(23-50.882)^2}{50.882} = 119.767$$

with three degrees of freedom. It is now divided into its three components: (a) due to differences in row totals:

$$\chi^2 = \frac{(128-132.88)^2}{132.88} + \frac{(71-66.12)^2}{66.12} = .5393866$$

with one degree of freedom

(b) due to differences in column totals:

$$\chi^2 = \frac{(51-48.533)^2}{48.532} + \frac{(148-150.468)^2}{150.468} = .1660116$$

with one degree of freedom

(c) due to interaction between E and B:

$$\chi^2_{(EB)} = 119.76742 - .5393866 - .1660116 = 119.06203$$

with one degree of freedom. This $\chi^2_{(EB)}$ is the one that is used in analyzing the null hypothesis.

The values in Table VIII are obtained from:

(1) $\sum_j n_{ijk} \{\mu + E_i + I_k + (EI)_{ik}\}$

(2) $\sum_j n_{ijk} (\mu + E_i + I_k)$ $\qquad H_o : (EI)_{ik} = 0$

Table VIII

Observed and expected values required for the determination of $\chi^2_{(EI)}$

| | | k = 1 | k = 2 | Sum |
|---|---|---|---|---|
| i=1 | (1) | 104 | 24 | 128 |
| | (2) | 91.970 | 35.676 | 127.596 |
| i=2 | (1) | 46 | 25 | 71 |
| | (2) | 57.774 | 13.63 | 71.404 |
| Sum | (1) | 150 | 49 | 199 |
| | (2) | 149.744 | 49.256 | 199 |

$$\chi^2_{\text{subclasses}} = \frac{(104-91.97)^2}{91.97} + \frac{(24-35.676)^2}{35.676} + \frac{(46-57.774)^2}{57.774} + \frac{(49-49.256)^2}{49.256} = 17.251416$$

with three degrees of freedom. Its three components are:

(a) due to differences in row totals:

$\chi^2 = \frac{(128-127.596)^2}{127.596} + \frac{(71-71.404)^2}{71.404} = .0035649$ with one degree of freedom.

(b) due to differences in column totals:

$$\chi^2 = \frac{(150-149.744)^2}{149.744} + \frac{(49-49.256)^2}{49.256} = .0017681$$ with one degree of

of freedom.

(c) due to interaction between E and I:

$$\chi^2_{(EI)} = 17.251416 - .0035649 - .0017681 = 17.246083$$ with one degree

of freedom.

The values in Table IX are obtained from:

(1) $\sum_i n_{ijk} \{\mu + B_j + I_k + (BI)_{jk}\}$

(2) $\sum_i n_{ijk} \{\mu + B_j + I_k\}$ $\qquad H_o : (BI)_{jk} = 0$

Table IX

Observed and expected values required for the determination of $\chi^2_{(BI)}$

| | | k = 1 | k = 2 | Sum |
|---|---|---|---|---|
| j=1 | (1) | 31 | 20 | 51 |
| | (2) | 41.764 | 5.989 | 47.753 |
| j=2 | (1) | 119 | 29 | 148 |
| | (2) | 112.308 | 38.939 | 151.247 |
| Sum | (1) | 150 | 49 | 199 |
| | (2) | 154.072 | 44.928 | 199 |

$$\chi^2_{\text{subclasses}} = \frac{(31-41.764)^2}{41.764} + \frac{(20-5.989)^2}{5.989} + \frac{(119-112.308)^2}{112.308} + \frac{(29-38.939)^2}{38.939} = 38.487994$$

with three degrees of freedom. Its three components are:

(a) due to the difference in row totals:

$$\chi^2 = \frac{(51-47.753)^2}{47.753} + \frac{(148-151.247)^2}{151.247} = .2904893$$

with one degree of freedom.

(b) due to the difference in column totals:

$$\chi^2 = \frac{(150-154.072)^2}{154.072} + \frac{(49-49.928)^2}{49.928} = .4766809$$

with one degree of freedom.

(c) due to interaction between B and I:

$$\chi^2_{(BI)} = 38.487994 - .2904893 - .4766809 = 37.720824$$

with one degree of freedom.

Note that in Tables VII, VIII and IX, the following equations (Scheffe, p. 92) are very nearly satisfied:

$$\frac{1}{N} \sum_{i=1}^{2} n_{ij.} (EB)_{ij} = \frac{1}{N} \sum_{j=1}^{2} nn_{ij.} (EB)_{ij} = 0 \text{ etc.}$$

Summary of the above:

$$\chi^2_E = 9.52$$

$$\chi^2_I = 23.08$$

$$\chi^2_B = 37.29$$

$$\chi^2_{(EB)} = 119.06$$

$$\chi^2_{(EI)} = 17.25$$

$$\chi^2_{(IB)} = 37.72$$

The above chi-squares are approximately distributed as $\chi^2$ each having one degree of freedom. It is difficult to evaluate how accurate these approximations are but since the calculations are based on a large number of data points, these chi-square approximations are expected to be reasonably accurate. Therefore one could expect the true 5% significance value to be off by only a few units from the tabulated value of 3.841. The above calculated values are greater than this 3.841 value and therefore the null hypothesis that the main effects and interactions are zero can be rejected. The following tentative conclusions can be drawn:

(1) the main effects B and I are important and the main effect E is relatively unimportant.

(2) the interaction (EB) is very important while the interactions (EI) and (BI) are relatively much less important.

## 2.2. Woolf's method as described by Plackett

This method tests for zero second-order interaction.

Arrange the data as follows:

| | E | | I | | B | | EI | | EB | | IB | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | prs. | abs. | prs. | abs. | prs. | abs. | prs. | abs. | prs. | abs. | prs. | abs |
| melanin present | 128 | 71 | 49 | 150 | 51 | 148 | 24 | 175 | 3 | 196 | 20 | 179 |
| melanin absent | 201 | 216 | 206 | 211 | 240 | 177 | 114 | 303 | 135 | 282 | 115 | 302 |
| Total | 329 | 287 | 255 | 361 | 291 | 325 | 138 | 478 | 138 | 478 | 135 | 481 |

This can be considered as six 2x2 tables, i.e. t = 6.

Therefore

| | | | |
|---|---|---|---|
| $n_{11} = 128$ | $n_{21} = 71$ | $n_{31} = 201$ | $n_{41} = 216$ |
| $n_{12} = 49$ | $n_{22} = 150$ | $n_{32} = 206$ | $n_{42} = 211$ |
| $n_{13} = 51$ | $n_{23} = 148$ | $n_{33} = 240$ | $n_{43} = 177$ |
| $n_{14} = 24$ | $n_{24} = 175$ | $n_{34} = 114$ | $n_{44} = 303$ |
| $n_{15} = 3$ | $n_{25} = 196$ | $n_{35} = 135$ | $n_{45} = 282$ |
| $n_{16} = 20$ | $n_{26} = 179$ | $n_{36} = 115$ | $n_{46} = 302$ |

Therefore $z_1 = \ln n_{11} - \ln n_{21} - \ln n_{31} + \ln n_{41} = .6613239$

$z_2 = \ln n_{12} - \ln n_{22} - \ln n_{32} + \ln n_{42} = -1.0948331$

$z_3 = \ln n_{13} - \ln n_{23} - \ln n_{33} + \ln n_{43} = -1.3698758$

$z_4 = \ln n_{14} - \ln n_{24} - \ln n_{34} + \ln n_{44} = -1.0091932$

$z_5 = \ln n_{15} - \ln n_{25} - \ln n_{35} + \ln n_{45} = -3.4428702$

$z_6 = \ln n_{16} - \ln n_{26} - \ln n_{36} + \ln n_{46} = -1.2261585$

$$\frac{1}{u_1} = \frac{1}{n_{11}} + \frac{1}{n_{21}} + \frac{1}{n_{31}} + \frac{1}{n_{41}} = .0315017$$

$u_1 = 31.744318$

$$\frac{1}{u_2} = \frac{1}{n_{12}} + \frac{1}{n_{22}} + \frac{1}{n_{32}} + \frac{1}{n_{42}} = .0366683$$

$u_2 = 27.271512$

$$\frac{1}{u_3} = \frac{1}{n_{13}} + \frac{1}{n_{23}} + \frac{1}{n_{33}} + \frac{1}{n_{43}} = .0361808$$

$u_3 = 27.638968$

$$\frac{1}{u_4} = \frac{1}{n_{14}} + \frac{1}{n_{24}} + \frac{1}{n_{34}} + \frac{1}{n_{44}} = .059453$$

$u_4 = 16.820009$

$$\frac{1}{u_5} = \frac{1}{n_{15}} + \frac{1}{n_{25}} + \frac{1}{n_{35}} + \frac{1}{n_{45}} = .3493884$$

$$u_5 = 2.8621442$$

$$\frac{1}{u_6} = \frac{1}{n_{16}} + \frac{1}{n_{26}} + \frac{1}{n_{36}} + \frac{1}{n_{46}} = .0675933$$

$$u_6 = 14.794365$$

If there is zero second-order interaction, then

$$X^2 = \sum_k u_k z_k^2 - (\sum_k u_k z_k)^2 / \sum_k u_k$$

is asymptotically distributed as chi-square with (t-1) degrees of freedom.

$u_1 z_1 = 20.99326$ $\quad$ $u_1 z_1^2 = 13.883355$

$u_2 z_2 = -29.857754$ $\quad$ $u_2 z_2^2 = 32.689256$

$u_3 z_3 = -37.861953$ $\quad$ $u_3 z_3^2 = 51.866173$

$u_4 z_4 = -16.974638$ $\quad$ $u_4 z_4^2 = 17.130689$

$u_5 z_5 = -\ 9.8539909$ $\quad$ $u_5 z_5^2 = 33.926011$

$u_6 z_6 = -18.140236$ $\quad$ $u_6 z_6^2 = 22.141804$

$\sum_k u_k = 121.1323$ $\quad$ $\sum_k u_k z_k^2 = 171.73827$

$$(\sum_k u_k z_k)^2 = 8408.0271$$

$$(\sum_k u_k z_k)^2 = 69.411933 \quad .$$

Therefore $X^2 = 171.73827 - 69.411933 = 102.33$ with 5 degrees of freedom.

In order to accept the null hypothesis which is that of zero second-order interaction, the $X^2$ value should be less than 11.1. As 102.23 is much greater than this value, we reject the null hypothesis of zero second-order interaction.

## 2.3. Linear model

The data was analyzed using two different models: the first model involved analyzing the data separately (presence or absence of melanin) and the second method was using a logarithmic model of the form: $F(\Pi) = K \log A\Pi$ .

In analyzing the data separately, the data was considered as follows:

| E | I | B | Number with melanin present | probability |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | .0050251 |
| 1 | 1 | 0 | 23 | .1155778 |
| 1 | 0 | 1 | 19 | .0954773 |
| 0 | 1 | 1 | 2 | .0100502 |
| 1 | 0 | 0 | 6 | .0301507 |
| 0 | 1 | 0 | 102 | .5125628 |
| 0 | 0 | 1 | 29 | .1457286 |
| 0 | 0 | 0 | 17 | .0854271 |
| | | Total | 199 | |

where a "1" denotes that the chicken has that trait and a "0" denotes that the trait is not present e.g. $E_1\ I_1\ B_1$ means the chicken has all three traits.

The following null hypothesis is tested: do the three traits have an equal effect on melanin being present i.e. does $E(E) = E(I) = E(B)$. Let $\Pi_1 = \frac{1}{199} = .0050251$, $\Pi_2 = \frac{23}{199} = .1155778$ etc. denote the cell probabilities. Therefore $E(E) = E(I) = E(B)$

$$\Pi_1 + \Pi_2 + \Pi_3 + \Pi_5 = \Pi_1 + \Pi_2 + \Pi_4 + \Pi_6 = \Pi_1 + \Pi_3 + \Pi_4 + \Pi_7$$

$$\Pi_1 + \Pi_2 + \Pi_3 + \Pi_5 = \Pi_1 + \Pi_3 + \Pi_4 + \Pi_7$$

$$\Pi_2 + \Pi_5 = \Pi_4 + \Pi_7$$

$$\Pi_2 - \Pi_7 = \Pi_4 - \Pi_5$$

$$\Pi_2 - \Pi_7 - \Pi_4 + \Pi_5 = 0$$

Also $\Pi_1 + \Pi_3 + \Pi_4 + \Pi_7 = \Pi_1 + \Pi_2 + \Pi_4 + \Pi_6$

$$\Pi_3 + \Pi_7 = \Pi_2 + \Pi_6$$

$$\Pi_2 - \Pi_7 = \Pi_3 - \Pi_6$$

$$\Pi_2 - \Pi_7 - \Pi_3 + \Pi_6 = 0.$$

Therefore, choose $f_1(\Pi) = \Pi_2 - \Pi_7 - \Pi_4 + \Pi_5 = 0$

and $f_2(\Pi) = \Pi_2 - \Pi_7 - \Pi_3 + \Pi_6 = 0$.

Using $f_1(\Pi)$ and $f_2(\Pi)$, A is obtained

$$\begin{array}{cccccccc} \Pi_1 & \Pi_2 & \Pi_3 & \Pi_4 & \Pi_5 & \Pi_6 & \Pi_7 & \Pi_8 \end{array}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & -1 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}$$

$$p = \begin{bmatrix} \Pi_1 = .0050251 \\ \Pi_2 = .1155778 \\ \Pi_3 = .0954773 \\ \Pi_4 = .0100502 \\ \Pi_5 = .0301507 \\ \Pi_6 = .5125628 \\ \Pi_7 = .1457286 \\ \Pi_8 = .0854271 \end{bmatrix}$$

$$V(p) = \begin{bmatrix} .000251 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .0005136 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .0004339 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .0000439 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .0001469 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0\ 0 & .0012554 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .0006255 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .0003926 \end{bmatrix}$$

$a_1p = -.01$ $\qquad\qquad$ $a_2p = .3869347$

$$X^2 = [a_1p \quad a_2p][AV(p)A']^{-1} \begin{bmatrix} a_1p \\ a_2p \end{bmatrix} = 84.286$$

which is distributed as chi-square with two degrees of freedom. As our value is greater than the known value at the 5% level of significance, the null hypothesis is rejected i.e. not all the traits are equally effective on melanin being present, i.e., one trait may be present while another trait being present may not necessarily imply melanin will be present.

Another null hypothesis was tested: is the effect of any two traits independent of the third. This method is based on a previous method by Plackett (1962). For this analysis

$f(\Pi) = \ell n\ \Pi_1 - \ell n\ \Pi_2 - \ell n\ \Pi_3 - \ell n\ \Pi_4 + \ell n\ \Pi_5 + \ell n\ \Pi_6 + \ell n\ \Pi_7 - \ell n\ \Pi_8 =$

The logarithmic model $F(\Pi) = K\ \ell n\ A\Pi$ is used. A is the identity matrix and $K = [1\ -1\ -1\ -1\ 1\ 1\ 1\ -1]$ and

$$D = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_8 \end{bmatrix}$$

Therefore

$$KD^{-1}A = \begin{bmatrix} \frac{1}{p_1} & \frac{-1}{p_2} & \frac{-1}{p_3} & \frac{-1}{p_4} & \frac{1}{p_5} & \frac{1}{p_6} & \frac{1}{p_7} & \frac{1}{p_8} \end{bmatrix}$$

$$f(p) = \ell np_1 - \ell np_2 - \ell np_3 - \ell np_4 + \ell np_5 + \ell np_6 + \ell np_7 - \ell np_8\ .$$

Now $X^2 = f(p)\ KD^{-1}AV(p)A'D^{-1}K'f(p) = .017 = .02$ which is distributed as chi-square with one degree of freedom. As our value is less than the known chi-square value at the 5% level of significance, the null hypothesis is not rejected.

The data was then analyzed using the number of chickens with melanin absent. The data was arranged as follows:

| E | I | B | Number with melanin absent | probability |
|---|---|---|---|---|
| 1 | 1 | 1 | 61 | .1462829 |
| 1 | 1 | 0 | 53 | .1270983 |
| 1 | 0 | 1 | 54 | .1294964 |
| 0 | 1 | 1 | 74 | .177458 |
| 1 | 0 | 0 | 38 | .091127 |
| 0 | 1 | 0 | 13 | .031175 |
| 0 | 0 | 1 | 51 | .1223021 |
| 0 | 0 | 0 | 73 | .1750599 |

The method is the same as above, the only difference is that the $\Pi$ and p values are different. (note $\Pi = p$). When the following null hypothesis was tested: do the three traits have an equal effect on the absence of melanin,

$$X^2 = 10.927.$$

As this is greater than the value at the 5% level of significance, the null hypothesis is rejected.

The null hypothesis: is the effect of any two traits independent of the third trait was then tested. The $X^2$ value obtained was

$$X^2 = 28.87$$

As this is greater than the value at the 5% level of significance, the null hypothesis is rejected.

The data was then analyzed using a linear model of the form $F(\Pi) = X\beta$ but it was found that the SS $[F(\Pi) = X\beta]$ value was too great which meant that the data did not fit this model.

A logarithmic model of the form $F(\Pi) = K \ln A$ was then fitted.

$$K = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix}$$

Notice that a generalized inverse must be used as the variance matrix is singular.

$$SS\,[F(\Pi) = X\beta] = 102.33$$

which is distributed as chi-square with five degrees of freedom. This result is interpreted as a test for no interaction. As this value is greater than the value at the 5% level of significance, the null hypothesis of no interaction was rejected.

Another method introduced by Berkson (1968) using minimum logit chi-square was then used.

Only Cases I and III were considered.

Using Case I, we test for no interaction.

Let $\ell_{1k} = \ell n\,(\frac{a_k}{b_k})$

$$\ell_{2k} = \ell n\,(\frac{c_k}{d_k})$$

$$B_k = \ell_{1k} - \quad_{2k}$$

$$C_{1k} = \frac{1}{a_k} + \frac{1}{b_k}$$

$$C_{2k} = \frac{1}{c_k} + \frac{1}{d_k}$$

$$w_{1k} = \frac{1}{C_{1k}}$$

$$w_{2k} = \frac{1}{C_{2k}}$$

$$\widetilde{w}_k = \frac{1}{C_{1k} + C_{2k}}$$

The formula for the minimum logit chi-square for Case I is

$$\chi^2_\ell = \sum_{k=1}^{6} \widetilde{w}_k B_k^2 - \hat{\beta}^2 \sum_{k=1}^{6} \widetilde{w}_k$$

$$\begin{cases} \tilde{w}_1 = 31.74425809 \\ \tilde{w}_2 = 27.27134362 \\ \tilde{w}_3 = 27.6388297 \\ \tilde{w}_4 = 16.81995136 \\ \tilde{w}_5 = \ \ 2.862140325 \\ \tilde{w}_6 = 14.7943215 \end{cases}$$

$$\sum_{k=1}^{6} \tilde{w}_k = 121.1308445$$

$$\hat{\beta} = \frac{\tilde{w}_1 B_1 + \tilde{w}_2 B_2 + \tilde{w}_3 B_3 + \tilde{w}_4 B_4 + \tilde{w}_6 B_6}{\tilde{w}_1 + \tilde{w}_2 + \tilde{w}_3 + \tilde{w}_4 + \tilde{w}_5 + \tilde{w}_6} = .7545479$$

$$= .569342606$$

$$\sum_{k=1}^{6} \tilde{w}_k \ B_k^2 = 171.01810189$$

Therefore $\chi_\ell^2 = 171.01810189 - (.569342606)(121.1308445)$

$= 102.05316$

Grizzle et al (1969) state that the value obtained by using the logarithmic model for no interaction, one obtains the same result. Looking back, one sees that the value obtained was 102.33 which is close enough. Note also that Woolf's (1955) method yields 102.23 which is close enough.

Case III was then used to test the equality of the traits. The formula used was

$$\chi_\ell^2 \doteq \sum_{k=1}^{6} \tilde{w}_k \ B_k'^{\,2} - \hat{\beta}'^{\,2} \sum_{k=1}^{6} \tilde{w}_k$$

where $\beta_k' = \ell_{1k} + \ell_{2k}$

and $$\hat{\beta}' = \frac{\tilde{w}_1B_1' + \tilde{w}_2B_2' + \tilde{w}_3B_3' + \tilde{w}_4B_4' + \tilde{w}_5B_5' + \tilde{w}_6B_6'}{\tilde{w}_1 + \tilde{w}_2 + \tilde{w}_3 + \tilde{w}_4 + \tilde{w}_5 + \tilde{w}_6}$$

The value obtained was 22.25. As this is greater than the $\chi^2$ value at the 5% level of significance, the null hypothesis was rejected i.e. not all the traits are equal.

The reults obtained by using the various methods are summarized as follows:

## Results

Method I, i.e. systematic partitioning of $\chi^2$

$\chi^2_E = 9.52$ E is relatively unimportant

$\chi^2_B = 37.29$ B is relatively important

$\chi^2_I = 23.07$ I is relatively important

$\chi^2_{(EB)} = 119.06$ interaction EB is very important

$\chi^2_{(EI)} = 17.25$ interaction EI is relatively unimportant

$\chi^2_{(BI)} = 37.72$ interaction BI is relatively unimportant

General conclusion is that the hypothesis that the main effects and interactions are zero is to be rejected.

---

Method II, i.e. Woolf's test for testing second-order interaction

$\chi^2 = 102.33$ reject the hypothesis of zero second-order interaction which agrees with what we have above.

---

| | |
|---|---|
| Method III, i.e. simple linear model | not all the traits have an equal effect which agrees with the above result |
| Method IV, i.e. logarithmic model | $SS[F(\Pi) = X\beta] = 102.33$ which shows that the null hypothesis of no interaction is rejected. This agrees with Method II. |
| Method V, i.e., minimum logit chi-square | $\chi_\ell^2 = \sum_{k=1}^{6} \tilde{w}_k B_k^2 - \hat{\beta}^2 \sum_{k=1}^{6} \tilde{w}_k = 102.05$ which shows that the null hypothesis of no interaction is rejected. It also agrees with Method II. |
| | $\chi_\ell^2 = \sum_{k=1}^{6} \tilde{w}_k B_k'^2 - \hat{\beta}'^2 \sum_{k=1}^{6} \tilde{w}_k = 22.25$ which shows that the null hypothesis that all traits are equal is rejected. This agrees with Method I. |

This completes the analysis of the broiler chicken data. It can be seen that the traits do play a part in determining whether melanin is present or absent, and the interesting thing is that the simple method of systematic partitioning of chi-square developed in this project is more comprehensive and agrees with the published methods of analyzing categorical data.

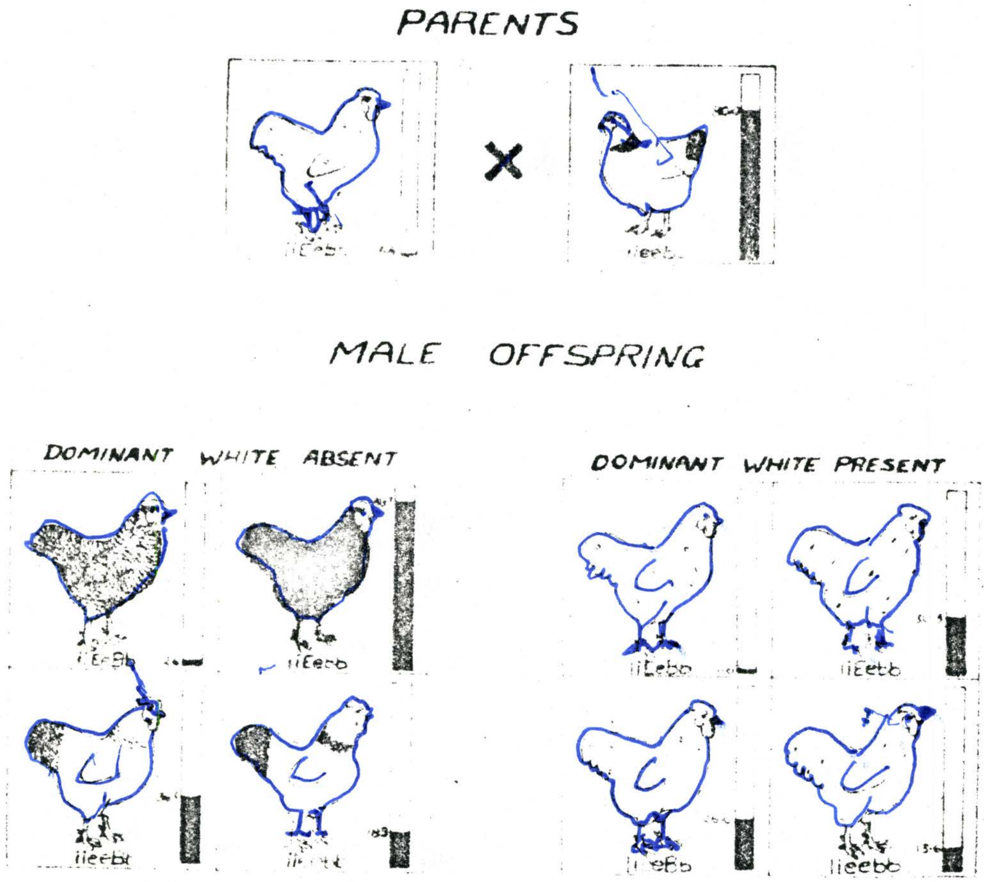


Figure 1. The plumage color genotypes of the parents and male offspring are shown. At the right of the birds, the percent deposition of abdominal melanin for each genotype is represented by the black area in the column.

# Bibliography

Bartlett, M.S. (1935). Contingency table interactions. J.R. Statist. Soc. Suppl. 2, 248-52.

Berkson, J. (1955). Maximum likelihood and minimum chi-square estimates of the logistic function. J. Amer. Stat. Ass. 50. 130-62.

Berkson, J. (1968). Application of minimum logit chi-square estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics 24, 75-95.

Bhapkar, V.P. (1966). A note on the equivalence of two test criteria for hypothesis in categorical data. J.A.S.A. 61, 228-235.

Bhapkar and Kock, G.G. (1968). On the hypothesis of " no interaction " in contingency tables. Biometrics 24, 567-94.

Cochran, W.G. (1950). The comparison of percentages in matched samples. Biometrics 37, 256-66.

Goodman. J.A. (1963). On Plackett's test for contingency table interactions. J.R. Stat.Soc. B25, 179-88.

Goodman, J.A. (1964). Simultaneous confidence intervals for contrasts among multinomial populations. Ann. Math. Stat. 35, 716-25.

Grizzle, J.E.,C.F.Starmer and G.G. Kock. (1969). Analysis of Categorical dats by linear models. Biometrics, 489-504.

Huntsman, C.M., F.N. Jerome and E.S. Snyder, (1960). The relationship between color phenotypes and the presence of black melanin in the abdomen of broiler chickens. Study 2. Poultry Sc. 39, 882-86.

Kastenbaum, M.A.and Lamphiear, P.E. (1959). Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics 15, 107-15.

Mather,K. (1957). Stat. Anal. in Biology. John Wiley and Sons.

Patil, K.D. (1974). Interaction test for three dimensional contingency tables. J.A.S.A. 69.164-68.

Plackett, R.L. (1962). A note on interactions in contingency tables. J.R. Stat. Soc. B24,162-66.

Roy,S.N. and Kastenbaum M.A. (1956). On the hypothesis of " no interaction " in a multi-way contingency table. Ann. Math. Stat. 27, 749-57.

Scheffé,H. (1959). The Analysis of Variance. John Wiley and Sons.

Tiku and Jerome, University of Guelph Report.

Woolf, B (1955). On estimating the relation between blood group and disease, Ann. Hum. Genet. Lond., 19, 251-53.