STATE OF IDAHO DEPARTMENT OF FISH & GAME

Title: Analysis of Variance Calculations As Applied to Creel Census Data

Author: Daniel R. Embody, Consulting Statistician

Date: June 2, 1954

INTRODUCTION

Experimental data when used with the analysis of variance must meet several requirements for the tests of significance to be strictly valid. References (1) and (2) present a thorough statement of these requirements and the consequences to be expected if they are not fulfilled.

To illustrate, consider a situation in which the observations consist of the number of fish removed from a stream. Involved in the situation are several years, several periods within years, and kinds of days within periods. Let Xijs represent the observation in the ith year, the jth periods, and the sth day.

The variability of Xijs can be assigned to 4 sources, namely:

1. Variability associated with years which is equal within years but varies from year to year.
2. Variability associated with periods which is equal within periods but varies from period to period.
3. Variability associated with days which is equal within days but varies from day to day.
4. A random element referred to as error.

The equation for each Xijs is

$$Xijs = M + ai + {}^{b}j + {}^{c}s + {}^{e}ijs$$

where

M = General mean of the population

ai = Population mean for year (i)

bj = Population mean for period (j)

cs = Population mean for day (

eijs =Error

Rearranging the equation, the error for observation X_{ijs} is

$$e_{ijs} = X_{ijs} - M - a_i - b_j - c_s$$

Briefly, the requirements which experimental data must meet are as follows:

1. Errors (e_{ijs}) must be normally distributed
2. Errors (e_{ijs}) must be uncorrelated
3. Errors (e_{ijs}) must form one homogeneous population
4. Main effects under consideration must be additive

Requirement Number 1 is not a serious restriction in that the distribution of errors can depart considerably from the normal distribution before the probabilities evaluated in the test of significance are seriously affected. In biological data, the possibility of correlated errors is very unlikely and can be safely ignored in most cases.

Experience with past creel censuses has shown that data consisting of counts of fish almost always violate requirement Number 3. Fortunately, there are a number of transformations which will equalize the variances of creel data, see references (3,4,5,6,7, and 8). Also, it usually happens that the transformation which equalizes the variances also normalizes the errors and renders main effects additive as indicated in requirement 4.

The tests of significance for interactions in the analysis of variance show whether the treatment effects are additive or not. Additivity of treatment effects may break down if serious changes are made in field methods without compensating for such changes in the analysis of the data.

In addition to providing information as to the importance of the various slain effects and interactions, the creel census data provides a means for estimating the total catch of fish. When the analysis of variance is made with transformed data, it is a somewhat involved procedure to translate the results back into original numbers.

For example, when catch data are transformed to logarithms the treatment means and tests of significance are made in terms of the logarithms. Converting mean of logarithms to antilogarithms gives geometric means of the catch rather than arithmetic means. In evaluating the total catch, it is usually desirable to use the arithmetic means. Techniques for converting means of logarithms tc arithmetic means have been developed (reference 8); but have not been generally available to fishery scientists.

It is the purpose of this report to present the details of a complete statistical analysis of the data obtained in a typical creel census. The proper transformation will be located, the analysis of variance will be made and interpreted, and estimates of the total catch with confidence limits will be made.

SPECIFICATION OF DATA

The census data to be analysed were collected under the direction of Mr. Leon W. Murphy as a part of an investigation of the fisheries of the Little Salmon River. References 9 and 10 describe the data, methods, and purposes o f the investigation.

Briefly, the plan of the census for each year consisted of dividing the 200 day season into 8 periods of 25 days each. For each period one Sunday, one Saturday, and one weekday was selected at random for census operations. The 3 3iys per period for 8 periods gave a total sample of 24 days each year. Identical sampling plans were used for both 1952 and 1953.

Census data consist of counts of the total fish removed from a 24 mile section of the Little Salmon River for each census day. Table 1 shows a summary of actual catch records.

SELECTING THE PROPER TRANSFORMATION

The ranges in Table 1 show that the experimental errors of the data are related to the mean values. Chart I shows that the relationship of ranges plotted against means seems to approximate the straight line. As indicated in the table at the end of reference (8), the logarithm transformation is appropriate for these data.

If the catch data had consisted of numbers of 10100 it is probable that the square root transformation would be indicated. In the case that some of the catch estimates had been zero (not missing data) the number one would ha'- been added to all of the catches and these values would be converted to logarithms. Reference 5 explains the transformations. For information concerning the appropriate transformation to use, the reader is invited to consult references (5), and (8).

Original catch data have been transformed to base 10 logarithms as presented in Table 2. Table 3 was formed by combining the corresponding values for the years. The value for period I Sunday is 1.732 1-2.869 = 4.601, and for period VIII Saturday 2.179 +1.699 = 3.878.

COMPENSATING FOR MISSING VALUES

At this stage it will be seen that each cell in Table 1 contains a number- other than zero. For this problem, therefore, missing values are no problem

Table 1

Fish Reported Caught By Days, Periods, and Years

	1952			1953		
Period	Sunday	Saturday	Weekday	Sunday	Saturday	Week
I	54	11	65	739	188	21
II	496	499	93	324	146	194
III	711	73	68	206	140	68
IV	877	652	127	828	785	327
V	622	490	278	874	597	210
VI	900	215	207	334	493	174
VII	251	64	42	245	133	73
VIII	163	151	71	109	50	86
Mean	509	269	119	457	316	144
Range	846	641	236	765	735	306

Since the problem frequently does arise with census work it seems advisable to discuss it further.

A distinction should be made between values of zero catch and missing values. When the random selection of days happens to pick a day that is cold and rainy and no fishermen. try their luck, then zero is the number of fish caught. Under these circumstances the zero is as good a number as any other and the analysis should proceed using the zero.

When it happens, however, that a census day is missed through a planning error and the omission is not discovered until months after the census has been finished, then the problem of "missing values" is present. If a sizable number of data are missing, then the more complex non—orthogonal methods of analysis must be used.

If only 1 or 2 items are missing, it is possible to estimate numbers which can be inserted into the table such that the analysis of variance can be made with standard procedures. The formula for estimating missing values for a randomized block experiment is described in references (11) and (12).

As an illustration of the missing value technique, consider the transformed data for 1952 as listed in Table 2. Here it is assumed that the value 1.732 period I Sunday was omitted as a result of accidental causes. Table 4 lists the data again. New row and column totals were obtained which of course excluded the missing value.

The formula used to estimate the missing value as described in reference (11) is as follows:

Table 2

Original Catch Records Transformed to Base 10 Logarithms

	1952				1953			
Period	Sun	Sat	Wkdy	Totals	Sun	Sat	Wkdy	Tot al
I	1.732	1.041	1.813	4.586	2.869	2.274	1.322	6.J 5
I I	2.696	2.698	1.968	7.362	2.510	2.164	2.288	6.962
I I I	2.852	o.863	1.832	6.547	2.314	2.146	1.832	5.292
IV	2.943	2.814	2.104	7.861	2.918	2.895	2.514	8.327
V	2.794	2.690	2.444	7.928	2.942	2.776	2.822	8.540
VI	2.954	2.332	2.316	7.602	2.524	2.693	2.240	7.457
VII	2.400	1.806	1.623	5.829	2.389	2.124	1.863	
VIII	2.212	2.179	1.851	6.242	2.037	1.699	1.934	5.670
Totals	20.583	17.423	15.951	53.957	20.503	18.771	16.815	56.089
Means	2.573	2.178	1.994	2.248	2.563	2.346	2.102	2.33-

Table 3

Sums of Corresponding Values for 1952 and *1953* Catch Records

Period	Sunday	Saturda	Weekday	Total	Mean
I	4.601	3.315	3.135	11.051	1.842
II	5.206	4.862	4.256	14.324	2.38'7
III	5.166	4.009	3.664	12.839	2.140
IV	5.861	5.709	4.618	16.188	2.698
V	5.736	5.466	5.266	16.468	2.745
VI	5.478	5.025	4.556	15.059	2.510
VII	4.789	3.930	3.486	12.205	2.031
VIII	4.249	3.878	3.785	11.912	1.985
Total	41.086	36.194	32.766	110.046	2.2926
Mean	2.5679	2.2621	2.0479	2.2926	

$$X = \frac{tT + bB - S}{(t - 1)(b - 1)}$$

where in this situation

X = missing item

t = number of days

b = number of periods

T = sum of items with same day as missing item

B = sum of items with same period as missing item

S = sum of all observed items

The calculations for estimating the value X = 1.940 are illustrated at the bottom of Table 4. The estimated value X = 1.940 appears to agree very well with the actual value of 1.732.

THE ANALYSIS CF IA\T IANCE

Using Tables 2 and 3, the calculations for the analysis of variance were lade. Table 5 summarizes the analysis.

Details of computations are as follows:

(a) Correction factor C.F. $= \frac{(SX)^2}{N} = \frac{(110.046)^2}{48} = 252.294210$

(b) Corrected sums of squares for total

From Table 2

$$SS = (1.732)^2 + (2.696)^2 + \cdots + (1.041)^2 + (2.698)^2 + \ldots + (1.863)^2\ (1.934)^2$$

$$= 262.356878$$

Subtract correction factor

SS - C.F. = 262.356878 - 252.294210 = 10.062668

Table 4

Catch Records Transformed to Logarithms Assuming One

Missing Value

Period	Sunday	Saturday	Weekday	Totals
I	(missing)	1.041	1.813	2.854
II	2.696	2.698	1.968	7.362
III	2.852	1.863	1.832	6.547
IV	2.943	2.814	2.104	7.861
V	2.794	2.690	2.444	7.928
VI	2.954	2.332	2.316	7.602
VII	2.400	1.806	1.623	5.829
VIII	2.212	2.179	1.851	6.242
	18.851	17.423	15.951	52.225

Calculation of Missing Value

$$X = \frac{8(2.854) + 3(18.851) - (52.225)}{(3 - 1)(8 - 1)}$$

$$= \frac{27.160}{14} = 1.940$$

t = 3

b = 8

T = 18.851

B = 2.854

S = 52.225

Table [illegible]

Summary of the Analysis of Variance

Source of Variance	Degrees of Freedom	Sums of Squares	Mean Square	Ratio	Significance
Total	47	10.062668			
Years	1	.094697	.094697	1.21	n.s.
Days	2	2.185526	1.092763	13.9	**
Periods	7	4.875642	.696520	8.87	**
Y x D	2	.065928	.032964		n.s.
Y x P	7	.737765	.105395		n.s.
D x P	14	.530975	.037927		n.s.
Y x D x P	14	1.572133	.112295		
Pooled Interactions	37	2.906803	.078562		

n.s. Not significant (Value of F for P = .95 not exceeded)

** Highly significant (Value of F for P = .99 exceeded)

Standard Deviation

$$s = \sqrt{.078562} = .2828$$

(c) Corrected sums of squares for years

From Table 2

$$SS = \frac{(53.957)^2 + (56.089)^2}{24} = 252.388907$$

Subtracting the correction factor

SS - C.F. = 252.388907 - 252.294210 = .094697

(d) Corrected sums of squares for days

From Table 3

$$SS = \frac{(41.086)^2 + (36.194)^2 + (32.766)^2}{16} = 254.479736$$

Subtracting the correction factor

SS - C.F. = 254.479736 - 252.294210 = 1.092763

(e) Corrected sums of squares for periods

From Table 3

$$SS = \frac{(11.051)^2 + (14.324)^2 + \cdots + (11.912)^2}{6} = 257.169852$$

Subtracting the correction factor

SS - C.F. = 257.169852 - 252.294210 = .696520

(f) Corrected sums of squares for Y x D interaction

From Table 2

$$SS = \frac{(20.583)^2 + (17.423)^2 + (15.951)^2 + (20.503)^2 + (18.771)^2 + (16.[illegible]}{8}$$

$$= 254.640361$$

Subtracting the correction factor

254.640361 - 252.294210 = 2.346151

Subtract corrected sums of squares for years and for days

Corrected SS for Y x D = 2.346151 - .094697 - 2.185526 = .065928

(g) Corrected sums of squares for Y x P interaction

From Table 2

$$SS = \frac{(4.586)^2 + (7.362)^2 + \cdots + (6.242)^2 + (6.465)^2 + (6.962)^2 + \cdots + (5.670)^2}{3}$$

$= 258.002316$

Subtract correction factor

258.002316 - 252.294210 = 5.708106

Subtract corrected sums of squares for years and for periods

Corrected SS for (Y x P) = 5.708106 - .094697 - 4.875642 = .737767

(h) Corrected sums of squares for (D x P) interaction

From Table 3

$$SS = \frac{(4.601)^2 + \cdots + (4.249)^2 + \cdots + (3.315)^2 + \cdots + (3.878)^2 + \cdots + (3.135)^2 + \cdots + (3.755)^2}{2}$$

$= 259.886353$

Subtract correction factor

259.886353 - 252.294210 = 7.592143

Subtract corrected sums of squares for days and for periods

7.592143 - 2.185526 - 4.875642 = .530975

(i) Obtain corrected sums of squares for (Y x D x P) interaction by subtracting all main effects and first order interactions from total

(j) Since none of the first order interactions were significant the sums of squares and degrees of freedom for all interactions were added together to obtain a combined error sum of squares.

(k) Mean squares were obtained by dividing sums of squares by appropriate degrees of freedom.

(1) Ratios obtained by dividing mean squares for treatments by mean square for pooled interactions.

The analysis of variance as summarized in Table 5 provides a number of definite conclusions. The second order interaction (Y x D x P) provides an estimate of experimental error to which the census data are subject. Since the mean squares for the 3 first order interactions (Y x D), (Y x P), (D x P) were numerically less than the mean square for (Y x D x P) it is obvious that none of the first order interactions are significant.

As mentioned earlier these tests of significance for the first order interactions indicate that the main effects for years, periods, and days are additive.

Since the interactions were not significant, their sums of squares and degrees of freedom were pooled to obtain a better estimate of the experimental error variance. Tests of significance for the 3 main effects were made with tie mean square for pooled interactions.

The difference between years was not significant. Differences among days of the week were highly significant, Differences among the 8 periods were highly significant.

The mean catches representing the two years are as follows:

Year	Catch Mean in logarithms
1952	2.248
1953	2.337
Difference	.089

Standard error of difference $s_d = \sqrt{\frac{2 \cdot s^2}{24}} = .0809$

The antilog of .089 is 1.228. Thus although the estimated catch for 1953 shows a 22.8% increase over the estimated catch for 1952, this difference is about what might be expected from chance causes and as such cannot be judged as significant. Conversely we may not say that the difference is not real. Actually it may be a real increase, but our experiment is not sensitive enough to detect it.

Another way of looking at the difference between years is that if it is significant statistically it is not large enough to be considered important in a biological sense.

Mean catches representing the 3 days are as follows:

	Catch mean in logarithms	Standard error
Sunday	2.568	.1000
Saturday	2.2621	.1000
Weekday	2.0479	.1000

Among the 3 means for days there are 2 independent comparisons that can be made. Using the techniques for single degrees of freedom outlined in reference (11) the two comparisons are given in Table 6.

In the setup of the census it was of particular interest to determine if the difference in mean catch between Sundays and Saturdays is real. If the difference were not real changes could be made in the design of future censuses to take advantage of the knowledge. The test of significance (Table 6) of this contrast indicates high significance. It is judged therefore that the difference between Sundays and Saturdays is real.

The second contrast concerns the difference in the mean catch for Sundays

Table 6

Computations for Individual Contrasts Between Days

	Sunday	Saturday	Weekday	Denominator	Products	Mean Square	Ratio
Catch totals in logs	41.086	36.194	32.766				
Contrast							
Sun. vs Sat.	1	- 1		32	4.892	.7479	9.52**
Sun. + Sat. vs Weekdays	1	1	- 2	96	11.748	1.4377	18.3**
						2.1856	

Error mean square = .078562
with 37 degrees of freedom

**Highly significant

and Saturdays vs weekdays. The test of significance (Table 6) shows this difference to be highly significant. It is judged, therefore, that the mean catches representing the 3 kinds of days indicate differences which are judged to be real.

Mean catches in logs for the $ periods are listed in Table 3. The analysis of variance indicated that differences among these means are highly significant. It would be possible to separate the sums of squares representing the 7 degrees of freedom into 7 independent contrasts. Such an exercise, however, would not solve the particular problems at issue and would be quite difficult to interpret.

Actually the fishing problem concerns the possible trends among period means within the season. The interaction between years and periods (Y x P) was not significant indicating that the differences among the period means were consistent over the two years and a trend within years is strongly suggested.

The study of trends may be accomplished by regression methods. A polynomial function in terms of time is fitted to the data. If any of the terms of the polynomial are found to be significant a time trend is indicated.

In particular, the method of orthogonal polynomials will be used in this study. References (11) and (13) describe the methods and supply tables of orthogonal variables. Table 7 outlines the computational procedures.

The tests of significance (column entitled "ratio" in Table 7) show that terms 2 and 5 were highly significant. At this point we judge that the time trend among periods is real. Using the orthogonal variates and the regression coefficients of Table 7, the $ points representing the trend of the catch in time were Calculated. Chart 2 plots the observed and calculated mean catches in logarithms for the 8 periods, The trend shown represents the mean for the two years.

The plotted trend in Chart 2 indicates that the catch increases for each period until period V where the peak is reached. After period V the catch falls off rapidly. A study of hours of effort and numbers of fishermen present on the stream throughout the season would reveal if the catch peak was related to effort or if the fish are merely more catchable. The management of the fishery would depend upon the possible causes for the trend in the catch.

Separate trends could have been fitted to the catches for each year separately In this case the difference between years was not significant and the interaction between years and periods (Y x P) was not significant. Under these condition;; very little could be gained by two charts.

Since the first order interactions were found to be not significant it will not be necessary to illustrate the meaning of interactions. As mentioned earlier if census methods are changed without making compensating changes in the analysis; the result will be a significant interaction. If factors which have important effects on the catch are operating in the situation but are not compensated for in the analysis of the census data, the result again may appear as an interaction.

Examples of a likely cause of interactions between years or other categories of the census are water temperature, air temperature, precipitation, water level, water flow, etc. In the planning of a census it is not possible to anticipate variations among these factors, but if measurements of their magnitude are recorded, then it is possible to eliminate their effects on the analysis by the method of covariance.

CONVERTING FROM LOGARITHMS TO ORIGINAL SCALE

As discussed in the introduction, it is desirable to estimate the total catch in the 8 periods in terms of the arithmetic mean rather than the geometric mean. The method by which the results of the analysis of variance made with data transformed to logarithms can be converted to arithmetic means is derived in reference (IL). Fishery workers in general, however, will find reference 14 somewhat difficult to follow because of the technical language. To assist in its understanding the important formulas and concepts will be reviewed here. z.

Dr. Finney discusses the logarithmic transformation as follows: "For a number of biological and other populations the standard error of an individual observation appears to be approximately proportional to the magnitude of the observation. In such cases the data prove more amenable to statistical treatment if first transformed by taking the logarithm of each observation. The distribution of these logarithms will then be normal".

In Chart 1 it was found that the means and ranges for the data for the 3 days in the 2 years approximated a linear relationship. Under these conditions the census data were transformed to logarithms. Although it was not made in this study, a statistical test of the errors would undoubtedly show them not to depart significantly from the normal distribution.

It should be pointed out as described in reference (8) that when the proper transformation has been located it frequently equalized variances, renders errors normally distributed, and causes treatment effects to be additive. These latter considerations are important when the analysis of variance is being considered as the method for analysing data.

Concerning the problem of converting the results of an analysis with transformed data back to the original scale, Dr. Finney states, "Though the transformation of the data has many advantages from the point of view of the comparision of samples by the anlysis of variance, it is nevertheless important tots able to assess from a sample the mean of the untransformed population. The result of transforming back the mean of the logarithms is to obtain the geometric mean of the original sample, which will tend to underestimate the arithmetic mean of the population. The arithmetic mean of the sample, on the other hand, provides a consistent estimate of the population mean but this is not of full efficiency. Similarly, the variance of the untransformed population will not be efficiently estimated by the variance of the original sample."

In other words, if the data must be transformed to be properly used with the analysis of variance, then the transformation cannot be ignored in makeing estimates of untransformed means, totals and standard deviations without losing information,

To convert means in logarithms to means in the original scale, Dr. Finney derives the following equation:

$$m = e^{\bar{x} + \frac{1}{2}s^2}\left\{1 - \frac{s^2(s^2+2)}{4n} + \frac{s^4\,(3s^4 + 44s^2 + 84)}{96n^2}\right\}$$

where

m = mean in original arithmetic scale

7 = mean of logarithms (base e)

s = standard deviation in logarithms (base e)

n number of observations

e = Naperian logarithm base (2.71828)

The above formula assumes that data are analysed in terms of base e logarithms rather than base 10 logarithms. For most fishery situations the term: in the bracket may be safely ignored.

From reference (15) the terms in the bracket are dropped and constants are changed to base 10 logarithms giving the conversion equation

$\log_{10} m \quad \bar{x} \quad 1.15129\, s^2$

where

m = mean in original arithmetic scale

$\bar{x}$ = mean of logarithms (base 10)

s = standard deviation in logarithms (base 10)

Equations for converting the variance from logarithms to the arithmetic scale as derived by Finney are as follows:

$$v = e^{\bar{x}^2 + s^2}\left[e^{s^2}\left\{1 - \frac{2s^2(2s^2+1)}{n} + \frac{2s^4(12s^4+44s^2+21)}{n}\right\} - \left\{1 - \frac{s^2(s^2-2)}{n} + \frac{s^4(3s^4-28s^2+42)}{6n^2}\right\}\right]$$

where

V = variance in arithmetic scale

$\bar{x}$ = mean in logarithms (base e)

s = standard deviation in logarithms (base e)

n = number of observations

Again from reference (15) the equation is simplified and converted for use with base 10 logarithms as follows:

$$s(o) = \sqrt{V} = m\sqrt{10^{2.302585\, s^2} - 1}$$

where

$s(o)$ = standard deviation in arithmetic scale

V = variance in arithmetic scale

m = mean in arithmetic scale

s = standard deviation in base 10 logarithms

In nearly all analyses with creel census data the simplified formulas given above should be used by fishery workers. For those who may care to read recent work on the subject references (16) and (17) are recommended. Both of these references contain extensive bibliographies.

A numerical example of the use of the conversion equations is given using the results of the analysis of variance described in the previous section. Table 2 summarizes the computational procedures. It will be noted that the standard deviation expressed as a percentage of the mean (coefficient of variation) in original numbers is a constant 70%.

Approximate 95% confidence limits for the means of 16 observations representing the 3 kinds of days over 2 years were estimated in part 3 of Table 2. Since there were 37 degrees of freedom in the estimate of the error term for the analysis of variance at value of t 2.03 was used. The approximate 95% confidence limits were estimated as

$$m \pm \frac{2.03 \ s(o)}{\sqrt{16}}$$

The $\sqrt{16}$ was included since each mean contained 16 observations.

ESTIMATING TOTAL CATCH

The total catch is estimated by multiplying the mean daily catch by the total number of days.

$$C_T = C_D \cdot N_T$$

where

C_T = total catch

C_D = mean daily catch

N_T = total number of days in season sampled

Since differences in catches of the 3 kinds of days were found to be significant, the mean daily catch must be weighted according to the 3 kinds of days. Within the season the relative frequencies are Sundays (1/7), Saturdays (1/7), and weekdays (5/7).

Table 7

Computations for the Fitting of Orthogonal Polynomials

Period	Observed Mean Log Catch	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	Calculated Mean Log Catch
1	1.842	-7	7	-7	7	-7	1.858
II	2.387	-5	1	5	-13	23	1.960
III	2.140	-3	-3	7	-3	-17	2.317
IV	2.698	-1	-5	3	9	-15	2.540
V	2.745	1	-5	-3	9	15	2.765
VI	2.510	3	-3	-7	-3	17	2.579
VII	2.034	5	1	-5	-13	-23	1.985
VIII	1.985	7	7	7	7	7	1.997
	18.341						

$\bar{X}$ = 2.292625

Term	$S(\xi y)$	Sum of Squares	Denom. (x)(6)	Ratio	Regression Coefficient
1	.393	.000919	168		.002399
2	-9.955	.589893	168	45.05**	-.059256
3	.035	.000004	264		.000133
4	4.353	.030760	616		.0070666
5	16.115	.118907	2184	9.08**	.007379
		.740483			

Mean square for error $s^2 = \frac{.078562}{6} = .013904$

Table 8

Computational Procedures for Converting Means in Logarithms

To Arithmetic Means in Original Units

1. Estimating Means

Day	Mean in Base 10 Logarithms $\bar{X}$	Std. Dev. in Base 10 logs. s	s^2	$1.15129s^2$	Log m $[\bar{X}+1.15129 s^2]$	Est. Mean Daily Catch m
Sunday	2.5679	.2829	.07856	.09045	2.6583	455
Saturday	2.2621	.2829	.07856	.09045	2.3525	225
Weekday	2.0479	.2829	.07856	.09045	2.1383	138

2. Estimating Standard Deviations

Day	$2.302585s^2$	$10^{2.302585s^2}-1$	$\sqrt{10^{2.302585s^2}-1}$	s(o)	Coefficient of Variation
Sunday	.1809	.517	.7190	327	72
Saturday	.1809	.517	.7190	162	72
Weekday	.1809	.517	.7190	99	72

3. 95% Confidence Limits for Means

Day	m	Upper Limit $m + .5075\ s(o)$	Lower Limit $m - .5075\ s(o)$
Sunday	455	621	289
Saturday	225	307	143
Weekday	138	188	88

The weighted mean daily catch is

$$C_D = l_1 C_{su} + l_2 C_{sa} + l_3 C_{wk}$$

where

C_D = mean daily catch

C_{su} = mean Sunday catch

C_{sa} = mean Saturday catch

C_{wk} = mean weekday catch

$l_1 = (1/7)$, $l_2 = (1/7)$, $l_3 = (5/7)$

The total estimated catch for the 200 day season, therefore, is

$$C_{200} = 200\left(\frac{C_{su}}{7} + \frac{C_{sa}}{7} + \frac{5C_{wk}}{7}\right)$$

where

C_{200} catch for 200 day season

other symbols defined above.

Table 9 presents the computational procedures for estimating the total catch. Since the difference between years was found to be not significant, one mean catch was estimated representing either year. In future years, the catch will probably change to a significant degree. At that time it may be appropriate to estimate total catches for all years independently. The same equations and methods would be used in any case.

The final statement regarding the catch is that for each year the best estimate of the total catch is

$$C_T = 39100 \text{ fish}$$

and the range 30,200 to 48,000 brackets the true catch with a probability of approximately 95%.

Table 9

Computations for Estimating the Total Catch

Day	Mean Daily Catch *	Weight ($200\ l_i$)	Products
Sunday	455	28.57	13,000
Saturday	225	28.57	6,400
Weekday	138	142.86	19,700

Estimated Total Catch (C_T) = 39,100

*Totals from Table 8

Table 10

Computations for Estimating 95% Confidence Limits For Catch

Day	$200\ l_i$	$(200\ l_i)^2$	s(o)	$(s(o))^2$	$(200\ l_i)^2(s(o))^2$
Sunday	28.57	816.2	327	106,929	87,275,000
Saturday	28.57	816.2	162	26,244	21,420,000
Weekday	142.86	20,409.0	99	9,801	200,029,000
				Total	308,724,000

$$s_{C_T} = \frac{1}{4}\sqrt{308{,}724{,}000} = (.25)(17570) \approx 4400$$

95% limits = 39,1000 ± (2.03)(4,400)

39,100 ± 8,900

Upper limit = 48,000 fish

Lower limit = 30,200 fish

At first glance one might believe that a range of 30.200 to 48,000 fish indicates that the data are too variable to be of much use to the fishery scientist. Actually, however, variation in the catch over these limits would probably not be noticed unless it was consistent over a number of years. If it were consistent over a number of years the census would easily detect it as more data became available.

If it is desired to reduce the variability of the catch records two avenues may be followed. First the standard deviation can be halved by increasing the number of census days by 4. That is instead of one Sunday, one Saturday, and one weekday for each period, four Sundays, four Saturdays and 4 weekdays could be taken. An investigation of the precision of the census relative to the precision and number of the different kinds of days might indicate a better combination of days than was used. Further work is planned along these lines.

A second avenue for increasing the precision of the census would be to test the effect of concomitant factors on the catch. Such factors as weather, temperature, water flows and fish planting schedules might have an important effect on the catch,

By including these factors in the analysis of ee ar°iance, it is possible that the overall variability could be significantly reduced without additional field work. Work along these lines is being studied in connection with other censuses.

Daniel R. Embody
Consulting Statistician

June 2. 1952

-

REFERENCES

(1) Eisenhart, Churchill. The assumptions underlying the analysis of variance. Biometric[s], Vol. 3 No. 1. March 1947.

(2) Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics. Vol. 3. No. 1. March 1947.

(3) Bartlett, M. S. Square-root transformations in the analysis of variance. J. Royal Stat. Soc. Suppl. Vol. 3, No.1. 1936.

(4) Williams, C. B. The use of logarithms in the interpretation of certain entomological problems. Ann. Appl. Biol. Vol. 24. 1937.

(5) Cochran W. G. Some difficulties in the statistical analysis of replicated experiments. The Empire Journal of Experimental Agriculture. Vol. VI, No. 22. April,1935:

(6) Beall, Geoffrey. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika, Vol. 32, 1912.

(7) Curtiss, J. H. On transformation used in the analysis of variance. The Annals of Mathematical Statistics, Vol. XIV, No. 2, June 1943.

(8) Bartlett, M. S. The use of transformations. Biometrics, Vol. 3, No. 1. March 1917.

(9) Murphy, Leon W. Random creel census of the Little Salmon River. Quarterly progress report for Investigations Projects Federal Aid in Fish and Wildlife Restoration Act. Idaho. Project F 7-R-1. Feb. 2, 1953.

(10) Murphy, Leon W. Random creel census of the Little Salmon River. Job completion reports for project No. F 7-R-2. Feb. 22, 1954.

(11) Snedecor, George W. Statistical Methods. The Iowa State College Press, Ames, Iowa. 3rd edition. 1940.

(12) Cochran, W. G. & Gertrude M. Cox. Experimental Designs. John Wiley & Sons Inc. New York. 2nd printing, 1950.

(13) Fisher, R. A. & F. Yates, M.A. Statistical Tables for Bioligical, Agricultural and Medical Research. Oliver & Boyd, London, England, 1950.

(14) Finney, D. J. On the distribution of a varate whose logarithm is normally distributed. Supplement to the Journal of the Royal Statistical Society. Vol. VII No. 2, 1941.

(15) Crtiss, J. H. Transformation from logs. Personal communication. 1944.

(16) Moshman, Jack. Critical values of the log-normal distribution. Jour, Amer. Stat. Assoc. Vol. 48. No. 263, Sept. 1953.

(17) Cohen, A. C. Jr. Estimating parameters of logarithmic-normal distributions by maximum likelihood. Jour. Amer. Stat. Assoc. Vol. 46. No. 254, 1951.

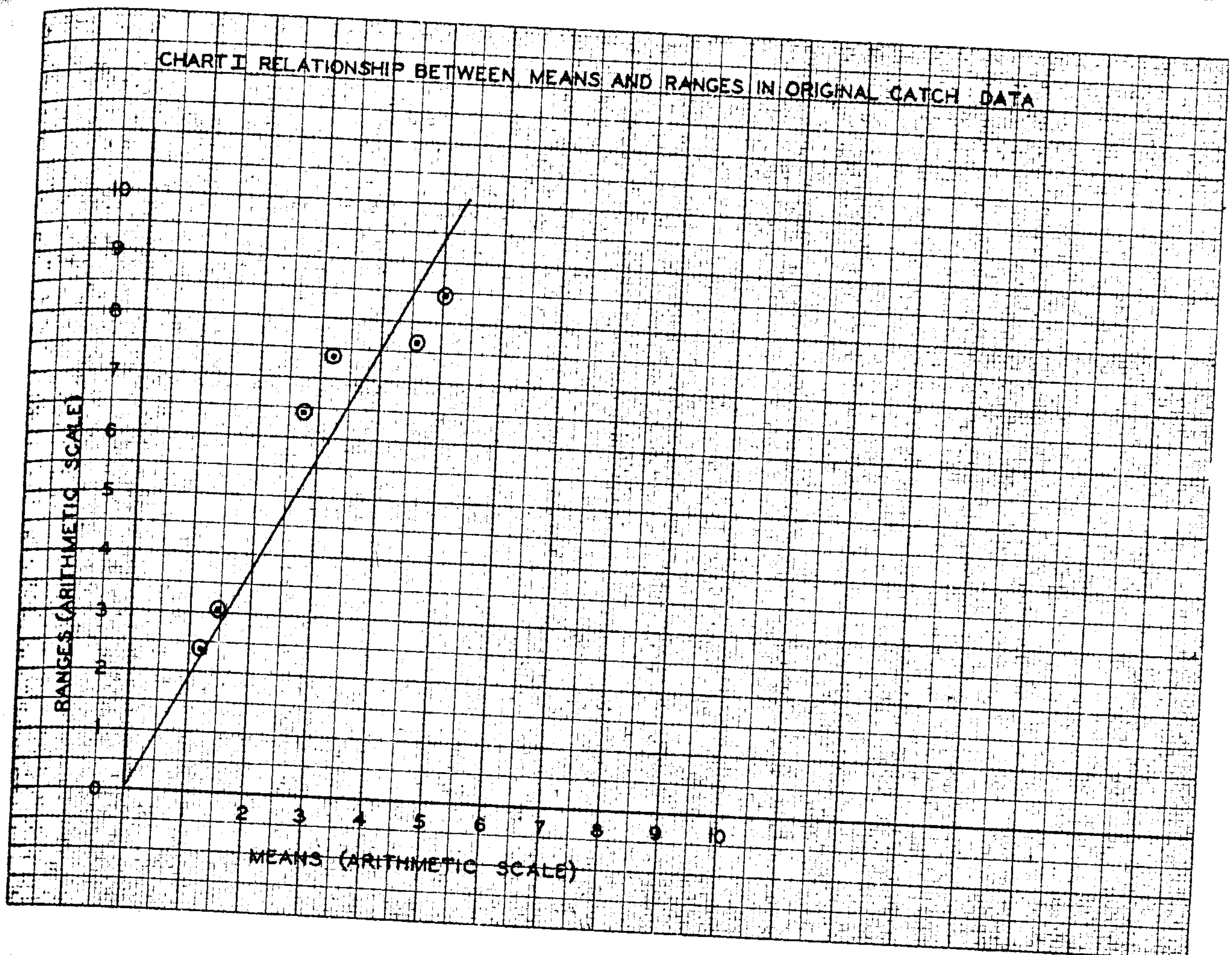

CHART I RELATIONSHIP BETWEEN MEANS AND RANGES IN ORIGINAL CATCH DATA
RANGES (ARITHMETIC SCALE)
MEANS (ARITHMETIC SCALE)

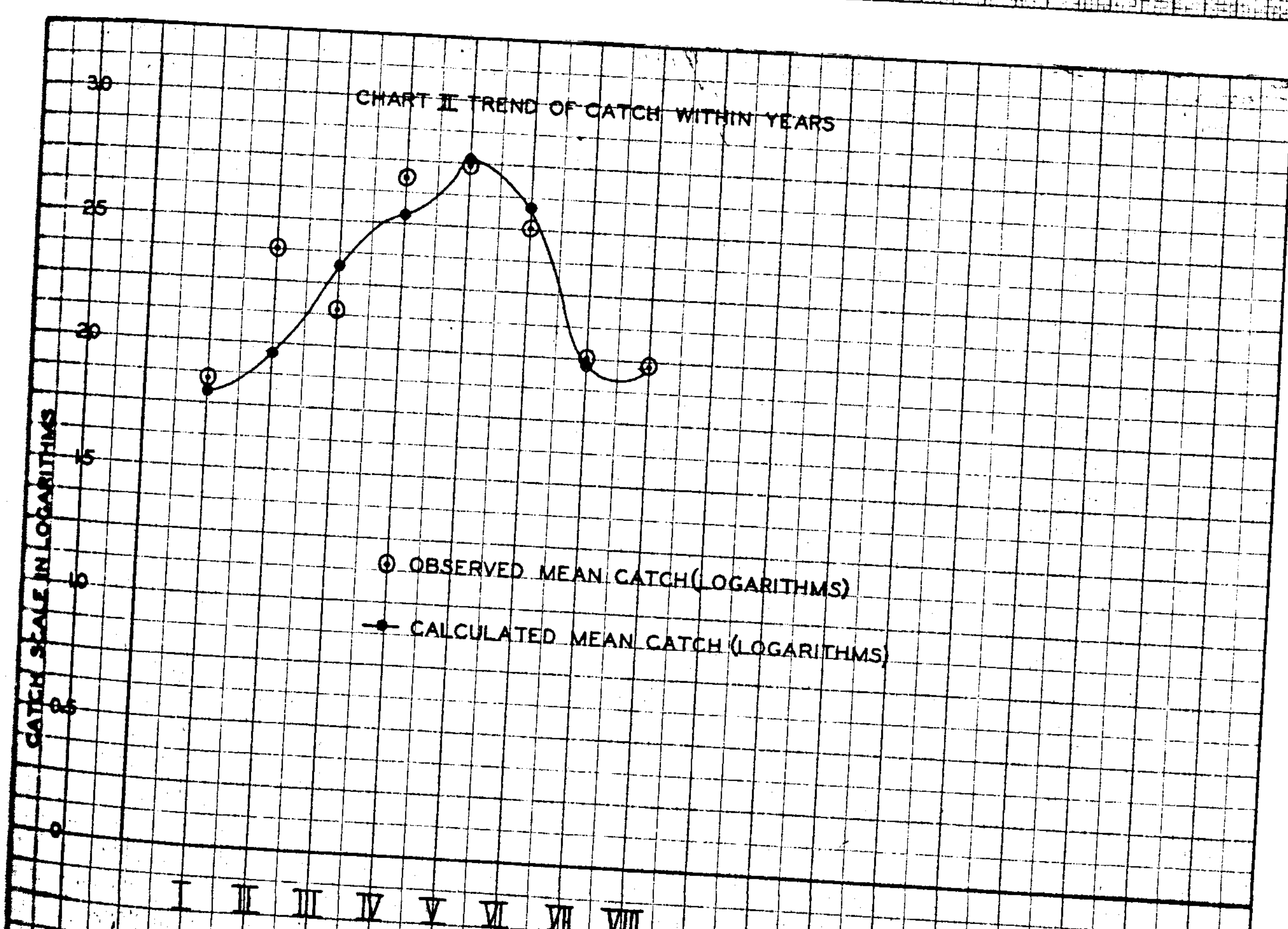

CHART II TREND OF CATCH WITHIN YEARS
CATCH SCALE IN LOGARITHMS
OBSERVED MEAN CATCH (LOGARITHMS)
CALCULATED MEAN CATCH (LOGARITHMS)
I II III IV V VI VII VIII