APPENDIX X

THE STATISTICAL TREATMENT OF THE DATA

(a) The "t" tests in covariance method:

In the analysis of covariance, the "between" and "within" variances are adjusted, before making the F-test. If the adjusted F-ratio turns out to be statistically significant, the next step, in the traditional procedure, is to compare the different "methods" group-means by "t" tests. But the method to get the "t" ratios of adjusted differences between any two group means is different from the ordinary method of making "t" tests. This method has been fully discussed by Cochran and Cox (16, pp. 79-80). Their discussion is briefly summarized here.

The difference between two adjusted means may be written:

$$(\bar{Y}_2 - \bar{Y}_1)_{adj} = (\bar{Y}_2 - \bar{Y}_1) - C(\bar{X}_1 - \bar{X}_2) \qquad \text{(xv)}$$

where $\bar{Y}_2$ is the mean score on the final measure for one film version, $\bar{Y}_1$ is the mean score on the final measure for another film version, $\bar{X}_2$ is the mean score on the initial measure for the film version which has $\bar{Y}_2$ as its final measure and $\bar{X}_1$ is the mean score on the initial measure for the film version which has $\bar{Y}_1$ as its final measure, and C is an adjustment factor. The value of C can be expressed as follows:

$$C = \frac{\Sigma yx}{\Sigma x^2} \qquad \text{(xvi)}$$

where Σyx is the sum of "within" cross products required for the analysis of covariance (as in Table 21) and Σx^2 is the "within" sum of squares in the analysis of variance of the initial measure (as in Table 19).

In this study, the analysis of covariance of the learning scores of female groups for different film versions by the covariance method A necessitated the use of the special form of "t"-tests described above. For our "t"-tests

$$C = \frac{\sum yx}{\sum x^2} \qquad \text{(xvi)}$$

$$= \frac{63676.7822}{85265.0659} = .7468 \qquad \text{(xvii)}$$

A difficult feature about this method is that the 'X' values enter into the variance so that every comparison necessitates a separate computation of the variance estimate. As a time-saving approximation, Cochran and Cox, following Finney, propose that an average value for the contribution of the term in the X's be used. This amounts to using the formula

$$V_{est} = S_{y.x}{}^2 \left[1 + \frac{t_{xx}}{\sum x^2}\right] \qquad \text{(xviii)}$$

for the effective residual error mean square where V_{est} is the population estimate of variance, $S_{y.x}{}^2$ is the residual error variance, t_{xx} is the "between" mean square for x. Thus, for the variance of the difference between the adjusted means for different film versions, the experimenter used the formula:

$$V_{est(f2 - f1)} = V_{est}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad \text{(xix)}$$

where $V_{est(f2 - f1)}$ is the variance estimate for any two film versions, and n_1 and n_2 are the number of female subjects assigned to the two film versions in question. Then

$$t = \frac{(\bar{Y}_2 - \bar{Y}_1)_{adj}}{\sqrt{V_{est}(f_2 - f_1)}} \qquad \text{(xx)}$$

In the case of the "t" tests summarized in Table 22, the estimate of the population was calculated from the formula:

$$V_{est} = S_{y \cdot x}^2 \left[1 + \frac{t_{xx}}{\Sigma x^2}\right] \qquad \text{(xviii)}$$

Thus

$$V_{est} = 136.1555 \left[1 + \frac{669.8134}{85265.0659}\right] = 137.2311 \qquad \text{(xxi)}$$

(b) <u>Calculation of the deviations from the regressed scores of a group</u>:

The theory of regression and prediction is too well-known to need an exposition here. But a point of special interest in this study was to calculate the deviations from the predicted score of each individual for a particular film repetition from the data on the sums of squares (both "within" and "between") and of cross products of Y (the final measure) and X (the initial measure) already available for each film repetition.

From general regression equation in deviation form was, therefore, derived an equation which is convenient to calculate the value of X, if certain constants such as M_x, M_y, etc., are known.

From the theory of regression, the predicted deviation $\bar{y}$ of a score Y from its mean M_y is written as follows:

$$\bar{y} = bx \qquad \text{(xxii)}$$

where b is the regression coefficient of Y on X, x the deviation of a

X from its mean M_x. Now as shown by Edwards (22, p. 336)

$$b = r \frac{Sy}{Sx} \qquad \text{(xxiii)}$$

where r = the product moment correlation coefficient between X's and Y's, Sy is the standard deviation of Y scores and Sx is the standard deviation for X scores.

or

$$b = \frac{\Sigma xy}{\Sigma x^2} \qquad \text{(xxiv)}$$

In the regression equations needed for our analysis, Σxy is the total sum of cross products for all the film versions for a particular film repetition, and Σx^2 is the sum of squares of X's over all methods for a particular film repetition.

From (xxii), (xxiii), and (xxiv)

$$\bar{y} = \frac{\Sigma xy}{\Sigma x^2} \cdot x \qquad \text{(xxv)}$$

Expressing $\bar{y}$ as $Y - M_y$ and x as $X - M_x$ we get

$$Y - M_y = \frac{\Sigma xy}{\Sigma x^2} (X - M_x) \qquad \text{(xxvi)}$$

or

$$Y = M_y - \frac{\Sigma xy}{\Sigma x^2} \cdot M_x + \frac{\Sigma xy}{\Sigma x^2} \cdot X \qquad \text{(xxvii)}$$

$$= C + \frac{\Sigma xy}{\Sigma x^2} \cdot X \qquad \text{(xxviii)}$$

where $C = M_y - \frac{\Sigma xy}{\Sigma x^2} \cdot M_x$ for a particular film repetition and is a constant.

Thus, getting the predicted score for each subject for a particular film repetition was made relatively easy because the values C and $\frac{\Sigma xy}{\Sigma x^2}$ were the same for all the subjects, so far as a particular film repetition was concerned.

(c) Kelley's method of getting the significance levels of the F-ratio not given in standard tables

It will be remembered that Kelley's method was used to get the significance level of certain F-ratios in the main experiment. The procedure given by Kelley (47, pp. 325-331) is briefly summarized here.

Let V_i be the variance of the sum of i independent standard scores, and V_j another similar variance, independent of the former, having j degrees of freedom.

$$F_{ij} = \frac{V_i/i}{V_j/j} \qquad \text{(xxix)},$$

a variance ratio having i degrees of freedom for the numerator and j degrees of freedom for the denominator.

If the hypothesis is that $V_i/i = V_j/j$ the divergence of F_{ij} from 1.00 (i.e., from its median value) is a measure of improbability of the hypothesis. The probability measure of any such divergence is obtained as follows:

$$\text{Let} \quad f_{ij} = \sqrt[3]{F_{ij}} \qquad \text{(xxx)}$$

$$\text{Let} \quad \theta_i = \frac{3}{\sqrt{2}} - \frac{\sqrt{2}}{3i} \text{ and } \theta_j = \frac{3}{\sqrt{2}} - \frac{\sqrt{2}}{3j} \qquad \text{(xxxi)}$$

But Kelley gives a table which gives θ_i and θ_j values for given degrees of freedom. From this table, it is easy to get the value of 'd' for a given F_{ij} from the following equation:

$$d = \frac{-\theta_i + \theta_j f_{ij}}{\sqrt{\frac{1}{i} + \frac{1}{j} f_{ij}^2}} \qquad \text{(xxxii)}$$

The 'd' is a deviate in a unit normal distribution. From 'd', the proportion to the right of the point 'd' can be obtained. This is the desired probability for F_{ij}, the probability that a variance ratio as great as that observed would arise as a matter of chance under the hypothesis that $V_{i/i} = V_{j/j}$.