A NOTE ON REGRESSION ANALYSIS

By Abraham Wald

Columbia University

1. Introduction. In regression analysis a set of variables $y, x_1, \cdots, x_p$ is considered where y is called the dependent variable and $x_1, \cdots, x_p$ are the independent variables. Let y_α denote the αth observation on y and $x_{i\alpha}$ the αth observation on x_i, $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$. The observations $x_{i\alpha}$ are treated as given constants, while the observations $y_1, \cdots, y_N$ are regarded as chance variables. The following two assumptions are usually made concerning the joint distribution of the variates $y_1, \cdots, y_N$:

(a) The variates $y_1, \cdots, y_N$ are normally and independently distributed with a common unknown variance σ^2.

(b) The expected value of y_α is equal to $\beta_1 x_{1\alpha} + \cdots + \beta_p x_{p\alpha}$ where $\beta_1, \cdots, \beta_p$ are unknown constants.

In some problems it seems reasonable to assume that the regression coefficients $\beta_1, \cdots, \beta_p$ are not constants, but chance variables. This leads to a different probability model for regression analysis and the object of this note is to discuss certain aspects of this model. In what follows in this note we shall make the following assumptions concerning the joint distribution of the chance variables $y_1, \cdots, y_N; \beta_1, \cdots, \beta_p$.

Assumption 1. For given values of $\beta_1, \cdots, \beta_p$ the joint conditional probability density function of $y_1, \cdots, y_N$ is given by

$$\frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{\alpha=1}^{N}(y_\alpha - \beta_1 x_{1\alpha} - \cdots - \beta_p x_{p\alpha})^2\right] \tag{1.1}$$

Assumption 2. The regression coefficients $\beta_1, \cdots, \beta_p$ are independently distributed.

Assumption 3. The regression coefficients $\beta_1, \cdots, \beta_r$, $(r \leq p)$, are normally distributed with zero means and a common variance σ'^2.

The purpose of this note is to derive confidence limits for the ratio $\frac{\sigma'^2}{\sigma^2}$. Such confidence limits have been derived by the author [1] for analysis of variance problems assuming that there are only main effects but no interactions. The regression problem treated in the present note is much more general and includes all the analysis of variance problems with or without interactions as special cases.

It should be remarked that Assumptions 2 and 3 do not exclude the case where $\beta_{r+1}, \cdots, \beta_p$ are constants.

2. Derivation of confidence limits for the ratio $\frac{\sigma'^2}{\sigma^2}$. Let $b_1, \cdots, b_p$ be the sample estimates of $\beta_1, \cdots, \beta_p$ obtained by the method of least squares. We