

A Note of Modified Two-Sample-Theoretical Estimation on Biological Assay

Asano, Choichiro
Shionogi Pharmaceutical Co. | Kyushu University

<https://doi.org/10.5109/12993>

出版情報：統計数理研究. 9 (2/3), pp.41-56, 1960-10. Research Association of Statistical Sciences

バージョン：

権利関係：



A NOTE OF MODIFIED TWO-SAMPLE-THEORETICAL ESTIMATION ON BIOLOGICAL ASSAY

By

Chooichiro ASANO

(Received February 20, 1960)

§ 1. Summary and introduction

The statistical methods given in this note are based upon the method of two-sample-theoretical estimations developed by various authors, and attempt some extension by the consideration of the validity test of observations supported by certain physical proof or experiential grounds.

In some practical situations an allowable discrepancy between the true value and its estimates of a population in our concern may be merely assigned a priori. There are, however, other situations where the allowable discrepancy should be better determined relatively in view of the difference between the estimate of our population and that of other one or more populations which are chosen for our comparison, as may happen in biological assay. In this note the estimation of biological standard line and those of main effects after preliminary analysis of variances are discussed from this standpoint and are based upon the sample size of sufficiently large number in securing the power of the validity test after testing the validity of the observations.

There have been developed two methods of two-sample formulation. The first one was originated by C. Stein [1] and developed later by Seelbinder [1], while the second one was due to Kitagawa et al [1], [2]. Their methods are different, but are common in assigning a constant discrepancy d , and based on the relation

$$Pr.\{|\bar{x} - \mu| \leq d\} = Pr.\{t_{n-1}(\alpha) \sqrt{\hat{\sigma}^2/n} \leq d\} \geq 1 - \beta ,$$

for determining the sample size n , which becomes consequently a random variable. They are concerned with one unique population. When we are concerned with more than one populations, it may not be a realistic formulation to assign an absolute allowable discrepancy a priori. It seems to us that more realistic formulation may be given in some situations by assigning a relative allowable discrepancy in view of the first sample and consequently the estimation after the significance test in the analysis of variance will more or less naturally follow in such a situation.

The author wishes his hearty thanks to Prof. T. Kitagawa for his kind suggestions and encouragement.

§ 2. Two-sample theory of biological standard line estimation

Consider the response metameter obtained from subjects receiving a dose metameter x and suppose that there exists a linear regression of y on x as a dose-response curve, i.e., the relationship may be written by $E_x(y) = \beta_0 + \beta_1 x$ and that a value of y about $E_x(y)$ is a random variable distributed normally with zero mean and variance σ^2 where x is a fixed variate and β_0 , β_1 and σ^2 are the unknown parameters depending upon the homogeneity of the subject, the control of environment, the history of bred stock of animals, etc, at that time. In case of an unstabilizing assay, the test preparation is tested in comparison with the experimental results of a standard preparation at the same time. In standard line assay, a standard line may be fitted immediately after observing certain n responses on various x_i ($i = 1, 2, \dots, n$), of standard preparation, without the preliminary test of linearity, and then an average response, \bar{y}_T , of test preparation is projected on the x axis by means of the standard line now obtained. This is an ordinary procedure.

In our situation of this assay type, we confirm the assumption that there exists a regression line over certain interval of x . However, there are many actual cases when a linearity is not certain because of the fluctuations of observations, and hence a preliminary test of linearity is first required for our careful statistical procedure based upon sample subjects of sufficiently large number in securing the power of the test, and then we shall be able to estimate a standard line after the above validity check of observations.

In such cases a rule may be prescribed in the following modified two-sample theoretical estimation scheme.

Let us define

$$(2.1) \quad u_1^2 = (\hat{\beta}_1 - \beta_1)^2 \sum_{j=1}^{n_1} (x_j - \bar{x})^2, \quad u_2^2 = \sum_{j=1}^{n_1} (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j)^2 / (n_1 - 2),$$

$$F' = u_1^2 / u_2^2, \quad A = F_{1, n_1-2}(\alpha_1), \quad \lambda = F_{1, n_1-2}(\alpha_2),$$

where n_1 is the size of the first sample, and $F_{1, n_1-2}(\alpha_i)$, $i = 1, 2$, is an α_i -point of F -distribution with the pair of the degrees of freedom $(1, n_1 - 2)$ for an assigned value of α_i in $0 < \alpha_2 \leq \alpha_1 < 1$.

Then our statistical procedure is defined in the following way:

(1°) *If $F' \geq A$, we stop the experimentation. Then the estimates of β_0 , β_1 and hence $\beta_0 + \beta_1 x$ for each assigned x are given by the first sample of n_1 , that is, are defined as follows:*

$$(2.2) \quad \hat{\beta}_1 = \sum (y_i - \bar{y})(x_i - \bar{x}) / \sum (x_i - \bar{x})^2, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

and $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

(2°) If $\lambda \leq F' < A$, then make n_2 additional observations, where $F' > 1$ and n_2 is defined by

$$(2.3) \quad n_2 = \left[\frac{A-1}{F'-1} n_1 \right] + 1 - n_1,$$

where the notation $[]$ denoting the Gauss' symbol. The estimates of β_0 , β_1 and hence $\beta_0 + \beta_1 x$ for each assigned x are given by the amalgamation of the first and the second samples, that is, the estimates are defined by the $(n_1 + n_2)$ observations as follows:

$$(2.4) \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n_1+n_2} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n_1+n_2} (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

and $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

(3°) If $0 < F' < \lambda$, then we make n_3 afresh, where $\lambda > 1$ and n_3 is prescribed, and the estimates of β_0 , β_1 and hence $\beta_0 + \beta_1 x$ for each assigned x are given by n_3 samples merely, that is, are given by (2.2) where $i=1, 2, \dots, n_3$.

Our statistical procedure is based upon a following practical situation. That is to say, in case of (1°) we confirm the linearity of the regression line under a significant level α_1 prescribed and stop the experimentation with the firm belief that the observations are obtained under the well-controlled circumstances, and the regression line is estimated by n_1 observations, and in case of (2°) we are afraid that we may be confronted with an insufficiency of sample size because of the observation containing a rather large fluctuation, and therefore we may and now we shall in this case make the additional samples of n_2 so as to find the trustworthy linearity of the regression line by sample subjects of sufficiently large number $n_1 + n_2$, and then the regression line is estimated by (2.4). And in case of (3°), under our interpretation that the sample of n_1 observations are in contradiction to our trustworthy linearity of the regression line supported by our long experiential or some physical ground, we shall in this case reject the information given by n_1 observations and make the observations of a prescribed size n_3 , and the regression line is estimated finally by these n_3 observations. Sometimes we may put that $n_3=0$, that is, we give up the estimation of the line at that time and may make the experiment under well-controlled circumstance in another occasion. Under some circumstances of bioassay, we may have a good reason to put λ to equal A .

2.1. Investigation of the sample size

The fact that sample size n_2 is a random variable is of the crucial importance in our approach, and our final aim is to decide the first sample size n_1 so as to minimize the expected value of n under the assigned λ and

4. In some biometrical and pharmaceutical researches, especially, such a consideration concerning to the sample size and the expectation of the regression line may be needed because of the individual subjects containing a large fluctuation.

Now let us consider the expectation of the sample size under the assigned n_1 , λ and A . For the sake of convenience, let us put $m=n/2$ and $m_i=n_i/2$, ($i=1, 2, 3$), and let δ be a noncentricity $\beta_1^2 S_x / 2\sigma^2$ of F' defined by (2.1). For the sake of estimating straight line, we know that the observation points had better be fixed at two extreme points, x_{10} and x_{20} . Now let us put

$$(2.5) \quad \bar{x} = (x_{10} + x_{20})/2, \quad S_0 = 2(x_{10} - \bar{x})^2.$$

Hence we have

$$(2.6) \quad S_x = \sum_{i=1}^{2m_1} (x_i - \bar{x})^2 = m_1 S_0.$$

The reason why we take n_2 defined by (2.3) comes rather intuitively from the following consideration.

First, in view of the relation

$$(2.7) \quad \frac{E\{u_1^2\}}{E\{u_2^2\}} - 1 = m_1 \frac{\beta_1^2 S_0}{\sigma^2},$$

the statistic $(F' - 1)/m_1$ will give us an approximate value of $m_1 \beta_1^2 S_0 / \sigma^2$.

Secondly, there exists always a positive value m such that

$$(2.8) \quad m \frac{\beta_1^2 S_0}{\sigma^2} = A - 1.$$

The combination of these two facts gives us,

$$(2.9) \quad \frac{F' - 1}{m_1} = \frac{A - 1}{m}.$$

This leads us to the following relationship which is essentially equivalent to somewhat generalized expression (2.3).

$$(2.10) \quad n_2 = 2 \left[\frac{A-1}{F'-1} \frac{n_1}{2} \right] + 2 - n_1 \quad \text{or} \quad m_2 = \left[\frac{A-1}{F'-1} m_1 \right] + 1 - m_1.$$

Then the expected sample size $E\{n | n_1\}$, namely, $E\{m | m_1\}$ on the basis of this rule may be written as follows:

$$(2.11) \quad E\{m | m_1\} \\ = (m_1 + m_2) \int_0^\lambda f(F'; m_1, \delta) dF' + \int_\lambda^\Lambda \left(\left[\frac{A-1}{F'-1} m_1 \right] + 1 \right) f(F'; m_1, \delta) dF'$$

$$\begin{aligned}
& + m_1 \int_{\Lambda}^{\infty} f(F'; m_1, \delta) dF' \\
& = m_1 + m_3 \int_0^{\Lambda} f(F'; m_1, \delta) dF' + \int_{\lambda}^{\Lambda} \left(\left[\frac{A-1}{F'-1} m_1 \right] + 1 - m_1 \right) f(F'; m_1, \delta) dF',
\end{aligned}$$

where

$$\begin{aligned}
(2.12) \quad f(F'; m_1, \delta) dF' &= \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k! B\left(\frac{1}{2} + k, m_1 - 1\right)} \left(\frac{1}{2(m_1 - 1)} \right)^{\frac{1}{2} + k} F'^{k - \frac{1}{2}} \\
&\quad \cdot \left(1 + \frac{F'}{2(m_1 - 1)} \right)^{-(m_1 + k - \frac{1}{2})} dF'.
\end{aligned}$$

Then

$$(2.13) \quad \int_0^{\Lambda} f(F'; m_1, \delta) dF' = \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right)$$

and

$$(2.14) \quad I(\lambda, A) \equiv \int_{\lambda}^{\Lambda} \frac{1}{F'-1} f(F'; m_1, \delta) dF' = \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k! B\left(\frac{1}{2} + k, m_1 - 1\right)}$$

$$\begin{aligned}
& \cdot \int_{\lambda/2(m_1-1)}^{\Lambda/2(m_1-1)} \frac{T^{k-\frac{1}{2}} (1+T)^{-(m_1+k-\frac{1}{2})}}{2(m_1-1)T-1} dT = \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k! B\left(\frac{1}{2} + k, m_1 - 1\right)} \sum_{i=0}^{m_1-1} \frac{(-1)^i \binom{m_1-1}{i}}{(2m_1-1)^{k+i+\frac{1}{2}}} \\
& \cdot \left\{ 2 \sum_{j=0}^{k+i} \frac{(2m_1-1)^{\frac{2(k+i-j)-1}{2}} \left\{ \frac{\Theta}{2(k+i-j)-1} - \theta \right\}}{2(k+i-j)-1} + \log \left\{ \left| \frac{\sqrt{(2m_1-1)\Theta-1}}{\sqrt{(2m_1-1)\Theta+1}} \right| \cdot \left| \frac{\sqrt{(2m_1-1)\theta+1}}{\sqrt{(2m_1-1)\theta-1}} \right| \right\} \right\},
\end{aligned}$$

where

$$(2.15) \quad I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) = B_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) / B \left(\frac{1}{2} + k, m_1 - 1 \right),$$

$$\theta = \frac{\lambda}{2(m_1 - 1) + \lambda}, \quad \Theta = \frac{A}{2(m_1 - 1) + A}.$$

Furthermore we may evaluate the right side of the function (2.11) by a relation,

$$(2.16) \quad \int_{\lambda}^{\Lambda} \left(\frac{A-1}{F'-1} m_1 - 1 \right) f(F'; m_1, \delta) dF' \leq \int_{\lambda}^{\Lambda} \left[\frac{A-1}{F'-1} m_1 \right] f(F'; m_1, \delta) dF'$$

$$< \int_{\lambda}^{\Lambda} \left(\frac{A-1}{F'-1} \right) m_1 f(F'; m_1, \delta) dF'.$$

Hence we have $E\{m|m_1\}$ as follows:

$$(2.17) \quad (m_1 + m_3) \sum_k \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) + m_1 \left\{ 1 + (A-1) I(\lambda, A) - \sum_k \frac{e^{-\delta} \delta^k}{k!} \right. \\ \cdot I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) \} \leq E\{m|m_1\} < (m_1 + m_3 - 1) \sum_k \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) \\ + m_1 \left\{ 1 + (A-1) I(\lambda, A) - \sum_k \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right) \right\} + \sum_k \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{1}{2} + k, m_1 - 1 \right).$$

The mean square error around $E\{m|m_1\}$ of (2.11) is given by

$$(2.18) \quad E\{(m'_1 - E\{m|m_1\})^2\} = E\{m'^2\} - 2E\{m'\} \cdot E\{m|m_1\} + (E\{m|m_1\})^2,$$

where m' is a value unadjusted to an integer and $E\{m'\}$ is given by the function (2.11) removing Gauss' symbol and 1 from the right side and

$$(2.19) \quad E(m'^2) = (m_1 + m_3)^2 \int_0^{\Lambda} f(F'; m_1, \delta) dF' + \int_{\lambda}^{\Lambda} \left(\frac{A-1}{F'-1} \right)^2 m_1^2 f(F'; m_1, \delta) dF' \\ + m_1^2 \int_{\Lambda}^{\infty} f(F'; m_1, \delta) dF' = m_1^2 + m_3(2m_1 + m_3) \int_0^{\Lambda} f(F'; m_1, \delta) dF' + m_1^2 \\ \cdot \int_{\lambda}^{\Lambda} \left\{ \left(\frac{A-1}{F'-1} \right)^2 - 1 \right\} f(F'; m_1, \delta) dF'.$$

If we put that

$$(2.20) \quad J(\lambda, A) \equiv \int_{\lambda}^{\Lambda} \frac{1}{(F'-1)^2} f(F'; m_1, \delta) dF' \\ = \sum_k \frac{e^{-\delta} \delta^k}{k! B\left(\frac{1}{2} + k, m_1 - 1\right)} \sum_i \frac{(-1)^i \binom{m_1}{i}}{(2m_1 - 1)^{k+i+1/2}} \left[(2m_1 - 1)^{\frac{2(k+i)+1}{2}} \left(\frac{\theta^{\frac{2(k+i)+1}{2}}}{(2m_1 - 1)\theta - 1} - \frac{\theta^{\frac{2(k+i)+1}{2}}}{(2m_1 - 1)\theta - 1} \right) \right. \\ \left. + \frac{2(k+i) - 1}{2} \left\{ 2 \sum_{j=0}^{k+i} \frac{(2m_1 - 1)\theta^{\frac{2(k+i-j)-1}{2}}}{2(k+i-j) - 1} - ((2m_1 - 1)\theta) \right\} \right. \\ \left. + \log \left\{ \left| \frac{\sqrt{(2m_1 - 1)\theta - 1}}{\sqrt{(2m_1 - 1)\theta + 1}} \right| \cdot \left| \frac{\sqrt{(2m_1 - 1)\theta + 1}}{\sqrt{(2m_1 - 1)\theta - 1}} \right| \right\} \right],$$

the value of (2.19) can be evaluated as follows:

$$(2.21) \quad E(m'^2) = m_3(2m_1 + m_3) \sum_k \frac{e^{-\delta} \delta^k}{k!} I_\theta\left(\frac{1}{2} + k, m_1 - 1\right) + m_1^2 \left\{ 1 - \sum_k \frac{e^{-\delta} \delta^k}{k!} \cdot \left(I_\theta\left(\frac{1}{2} + k, m_1 - 1\right) - I_\theta\left(\frac{1}{2} + k, m_1 - 1\right) \right) + (A-1)^2 J(\lambda, A) \right\}.$$

2.2. Precision of estimate

Now let us consider the precision of estimate according to above two-sample theoretical standard line in current case when $n_3=0$.

(1) Estimation of y on each assigned x

Let us consider the simple case in which the precision of $\hat{y} (= \hat{\beta}_0 + \hat{\beta}_1 x_0)$ on an assigned x_0 will be discussed after fitting a standard line. $E\{\hat{y}\}$ and $V\{\hat{y}\}$ for a fixed sample size n are shown by

$$(2.21) \quad E\{\hat{y}\} = \beta_0 + \beta_1 x_0$$

$$(2.22) \quad V\{\hat{y}\} = \frac{2\sigma^2}{n} \left(\frac{1}{2} + \frac{(\bar{x} - x_0)^2}{S_0} \right)$$

because of the facts

$$(2.23) \quad E\{\hat{\beta}_0\} = \beta_0, \quad E\{\hat{\beta}_1\} = \beta_1, \quad E\{\hat{\sigma}^2\} = \sigma^2,$$

$$V\{\hat{\beta}_0\} = \frac{\sigma^2}{n} \frac{(x_{10}^2 + x_{20}^2)}{S_0}, \quad V\{\hat{\beta}_1\} = \frac{2\sigma^2}{nS_0}, \quad Cov\{\hat{\beta}_0, \hat{\beta}_1\} = -\frac{2\sigma^2}{nS_0}.$$

In the consequence under our two-sample-theoretical estimation we will have

$$(2.24) \quad E\{\hat{y}\} = E\{\hat{y}_{(n_1)}\} Pr.(n=n_1) + E\{\hat{y}_{(n_1+n_2)}\} \cdot \{1 - Pr.(n=n_1)\} = \beta_0 + \beta_1 x_0,$$

and

$$(2.25) \quad V\{\hat{y}\} = V\{\hat{y}_{(n_1)}\} Pr.(n=n_1) + V\{\hat{y}_{(n_1+n_2)}\} Pr.(n=n_1+n_2)$$

$$\begin{aligned} &= \frac{A\sigma^2}{m_1} \left\{ \int_{\Lambda} f(F') dF' + \frac{1}{A-1} \int_{\Lambda} (F'-1) f(F') dF' \right\}, \\ &= \frac{A\sigma^2}{m_1} \left[1 + \frac{(m_1-1)}{A-1} \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} \left(\frac{1+2k}{m_2-2} \right) \left\{ I_\theta\left(\frac{3}{2} + k, m_1-2\right) - I_\theta\left(\frac{3}{2} + k, m_1-2\right) \right\} \right. \\ &\quad \left. - \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} \left\{ \frac{A I_\theta\left(\frac{1}{2} + k, m_1-1\right) - I_\theta\left(\frac{1}{2} + k, m_1-1\right)}{A-1} \right\} \right], \end{aligned}$$

where

$$(2.26) \quad \Delta \equiv -\frac{1}{2} + \frac{(x_0 - \bar{x})^2}{S_0}.$$

(2) Estimation of x corresponding to observed \bar{y}_t

In case of current standard line assay, a standard line $y = \beta_0 + \beta_1 x$ is obtained as a regression relationship between a dose of standard preparation (its potency is known a priori) and the correspondent response observed, and then an unknown potency x_0 of test preparation is estimated by using this standard line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ and the observed average response \bar{y}_t of test preparation. The maximum likelihood estimate of the x_0 is given by

$$(2.27) \quad \hat{x}_0 = (\bar{y}_t - \hat{\beta}_0) / \hat{\beta}_1,$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by (2.2). Now in case when the two-sample theoretical standard line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ and \bar{y}_t of test preparation resulting from l observations are obtained, the estimate of x_0 is also expressed by the same (2.27), where $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by (2.2) and (2.4), and further the prediction limits for x_0 are given by using a following statistic z ,

$$(2.28) \quad z = \bar{y}_t - \hat{\beta}_0 - \hat{\beta}_1 x_0,$$

where z follows a normal distribution with

$$(2.29) \quad E\{z\} = 0,$$

$$(2.30) \quad V\{z\} = \sigma^2 \left\{ \frac{1}{l} + \frac{\Delta}{m_1} \frac{1}{r'} \right\}$$

and where $1/r'$ is a coefficient of $\Delta\sigma^2/m_1$ of (2.25), namely, a bracketed expressions.

Furthermore a maximum likelihood estimate of σ^2 is obtained as follows,

$$(2.31) \quad \hat{\sigma}^2 = \left\{ \sum_{i=1}^{m_1 r'} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 + \sum_{j=1}^l (y_j - \bar{y}_{(t)})^2 \right\} / (m_1 r' + l),$$

where $\hat{\sigma}^2(m_1 r' + l)$ is known to a χ^2 -distribution with the $(m_1 r' + l)$ degrees of freedom.

These relations show us that the statistic t defined by

$$(2.32) \quad t = \frac{z}{\sqrt{\frac{1}{l} + \frac{\Delta}{m_1 r'}}} \bigg/ \sqrt{\frac{(l + m_1 r') \hat{\sigma}^2}{(l + m_1 r' - 3)}}$$

is distributed in the t -distribution with the $(l + m_1 r' - 3)$ degrees of freedom, that is,

$$(2.33) \quad Pr. \{ |t| \leq t_{l+m_1 r'-3}(\alpha/2) \} = 1 - \alpha,$$

for each assigned α in $0 < \alpha < 1$.

Therefore when the $\alpha/2$ -point value from the t -table is substituted for $t_{l+m_1r'-3}(\alpha/2)$, the numerator and denominator and the whole expression in the bracket of (2.33) remain a quadratic of x_0 . The solutions of this quadratic equation of x_0 give us two fiducial limits, \hat{x}_0^L and \hat{x}_0^U , in the following way:

$$(2.34) \quad Pr.\{\hat{x}_0^L \leq x_0 \leq \hat{x}_0^U\} = 1 - \alpha,$$

where we put

$$(2.35) \quad (\hat{x}_0^L, \hat{x}_0^U) = \left(\bar{x} + \frac{(\hat{x}_0 - \bar{x}) \pm \left(\frac{\hat{\sigma}t}{\hat{\beta}_1} \right) \sqrt{1 + \frac{3}{l+m_1r'}} \sqrt{\left(\frac{1}{l} + \frac{1}{m_1r'} \right) \left\{ 1 - \frac{2}{m_1r'S_0} \left(\frac{\hat{\sigma}t}{\hat{\beta}_1} \right)^2 \left(1 + \frac{3}{l+m_1r'} \right) \right\} + \frac{2(\hat{x}_0 - \bar{x})^2}{m_1r'S_0}}} \right)}{1 - \frac{2}{m_1r'S_0} \left(\frac{\hat{\sigma}t}{\hat{\beta}_1} \right)^2 \left(1 + \frac{3}{l+m_1r'} \right)} \right).$$

2.3. Power of test

Let us summarize here the power of test for the linearity derived from our two-sample formulation of biological standard line estimation in current case when $n_3 = 0$.

That is to say, the power of our test is defined on the basis of testing a hypothesis $\beta_1 = 0$ against an alternative hypothesis $\beta_1 \neq 0$, and is expressed by the sum of two probabilities $Pr.\{D_1\} + Pr.\{D_2\}$, where

$$(2.36) \quad Pr.\{D_1\} = Pr.\{F'_{(n_1)} \geq F_{1, n_1-2}(\alpha)\},$$

$$(2.37) \quad Pr.\{D_2\} = Pr.\{F'_{(n_1)} < F_{1, n_1-2}(\alpha), F'_{(n_1+n_2)} \geq F_{1, n_1+n_2-2}(\alpha)\},$$

and $F_{1, n-2}(\alpha)$ is the α -point value of F -distribution with the pair of the degrees of freedom $(1, n-2)$.

(2.36) is obtained by

$$(2.38) \quad Pr.\{D_1\} = 1 - \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} I_{\Theta} \left(\frac{1}{2} + k, \frac{n_1-2}{2} \right),$$

and (2.37) is obtained by

$$(2.40) \quad Pr.\{D_2\} = \int_{\lambda}^{\Lambda} \int_{D_1} \left\{ \int_{D(v, w)} f(u, v, w, n_2) dv dw \right\} du dF'$$

in terms of

$$(2.41) \quad f(u, v, w, n_2) = \frac{\Gamma(n_1 + n_2 - 1)}{\pi \Gamma\left(\frac{n_1-2}{2}\right) \Gamma\left(\frac{n_2-2}{2}\right)} u^{-\frac{1}{2}} v^{\frac{n_1-3}{2}} w^{\frac{n_1-2}{2}} (1 + w + uv + wvu)^{-\frac{n_1+n_2-2}{2}},$$

where

(2.42) {

$$\overline{D}_1: 0 < u < \theta_{21} \cdot F_{1, n_1-2}(\alpha) / (n_1-2), \quad \theta_{21} = 1 + \frac{\beta_1^2 S_{x(1)}}{\sigma^2},$$

$$D(v, w): \theta_{13, 24} \cdot F_{1, n_1+n_2-2}(\alpha) / (n_1+n_2-2) < \frac{uv+1}{v+w^{-1}} < \infty \text{ for fixed } u,$$

$$\theta_{13, 24} = \theta_{21} + \frac{\beta_1^2 S_{x(2)}}{\sigma^2}, \quad n_2 = n_1 \left(\frac{A-1}{F'-1} - 1 \right),$$

$$S_{x(1)} = \sum_{j=1}^{n_1} (x_j - \bar{x})^2, \quad S_{x(2)} = \sum_{i=1}^{n_2} (x_i - \bar{x})^2.$$

2.4. Numerical table for the routine standard line assay

In case of a routine work ($n_3=0$) of biological standard line estimation using an arbitrary universe, let us consider the following decision procedure in deciding the first sample size for the application of this scheme.

Let us put

(2.43)

$$\delta^* = \beta_1^2 S_0 / \sigma^2,$$

and prepare the following table of $E\{m|m_1\}$ for various m_1 and α_1 in case when we now put $\lambda=1.01$ for any m_1 . The limit value of $E\{m|m_1\}$ for m_1 are given by $\{F_{1,\infty}(\alpha_1)-1\}/\delta^*$.

Table of $E\{m|m_1\}$

($\alpha_1=0.01$)

$\delta^* \backslash m_1$	0.05	0.1	0.15	0.25	0.4	0.5	0.75	1	2	4
3	—	—	—	—	—	—	—	—	—	—
6	169	86	53	41	26	15	9	6		
12	150	73	48	34	15	12	12			
20	137	61	42	25	22	20				
limit	113	56	38	23	14	11	8	6	3	2

($\alpha_1=0.05$)

$\delta^* \backslash m_1$	0.05	0.1	0.15	0.25	0.4	0.5	0.75	1	2	4
3	—	—	—	—	—	7	5	3		
6	80	41	28	15	7	6				
12	72	36	24	12	12					
20	64	31	21	20						
limit	57	28	19	7	6	4	3			

$(\alpha_1=0.10)$

$\delta^* \backslash m_1$	0.05	0.1	0.15	0.25	0.4	0.5	0.75	1	2	4
3	—	—	—	—	5	3				
6	33	26	9	7	6					
12	30	23	12							
20	26	20	20							
limit	24	17	11	7	4	3				

From a general view of the table, it may be said that if $\delta^* > 2$, the assay in the routine work is most desirable and is completed with a small number of subjects and that if $\delta^* < 0.2$, the assay is not so suitable for the restriction of subjects and the control of assay circumstances.

Now while we do not know the true value δ_{true}^* of the present assay, we may and we shall assume two extreme (lower and upper) values (δ_L^* and δ_U^*) to be considered among possible values δ^* 's in the above table. Then the principle of a choice criterion of the first size may be a minimax procedure that minimizes the maximul number of the difference between $E\{m|m_1\}$ and $\lim_{m_1 \rightarrow \infty} E\{m|m_1\}$ corresponding to each $\delta^* (\delta_L^* \leq \delta^* \leq \delta_U^*)$. That is, if we have the information on δ_{true}^* throughout the trials of sufficiently large $n_1 (= \infty)$, we must have taken $\lim_{m_1 \rightarrow \infty} E\{m|m_1\}$ as a sample size at one time. We, however, cannot know such a δ_{true}^* and therefore, as a second best plan, our decision procedure may have good reason to choose $E\{m|m_1\}$ which gives us a minimul loss in the maximul value of the sample-size loss, $E\{m|m_1\} - \lim_{m_1 \rightarrow \infty} E\{m|m_1\}$, among δ_L^* and δ_U^* . For example, when we consider the first sample size in a range from 0.5 to 2 of δ^* at that table, we calculate the difference as following form,

$\delta^* \backslash m_1$	0.5	0.75	1	2	max.
3	*	3	1	0	* (>4)
6	4	1	0	3	4
12	1	4	6	9	9
20	8	12	14	17	17

and the right column shows the maximul value of each row. Then we may choose m_1 corresponding to the minimul value between them. In this example, $n_1=6$ is chosen.

§ 3. Further discussion of two-sample-theoretical estimation of the effects in analysis of variance test

Our considerations applied earlier at § 2 have also a similar meaning generally in case of estimating certain main effects or interaction-effects

after the preliminary test of analysis of variances.

In our situation of the present estimation problem, we confirm the assumption that there exists some differences among their mean effects from a proof with physical grounds. However, there are many actual cases when a non-centricity among them is not certain by the fluctuations of observations and at that time it seems inadequately to us that the effects are estimated immediately by that observations. Hence a preliminary test of non-centricity is first required for our careful statistical procedure which should be based upon observations of sufficiently large numbers in securing the power of test and secondly we shall estimate the effects after the above validity test of observations.

Now let us consider the randomized block design having l factors and m replications as a fundamental example of analysis of variance test without loss of generality. Let x_{ij} be a random sample drawn from $N(\theta_{..} + \theta_{i.} + \theta_{.j}, \sigma^2)$, ($i=1, 2, \dots, l$ and $j=1, 2, \dots, m$), where $\theta_{i.}$ and $\theta_{.j}$ denote a treatment effect and a block effect respectively and where we put $\sum \theta_{i.} = \sum \theta_{.j} = 0$. Then the treatment effects are tested by the ratio F' defined by

$$(3.1) \quad F' = u_1^2 / u_2^2,$$

where $u_1^2 = m_1 \sum_{i=1}^l (\bar{x}_{i.} - \bar{x}_{..})^2 / (l-1)$ and $u_2^2 = \sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2 / (l-1) \times (m-1)$, and let δ be a noncentricity $m_1 \sum \theta_{i.}^2 / 2\sigma^2$ of F' defined by (3.1) where m_1 denotes the first number of replications and of course

$$(3.2) \quad E\{u_1^2\} = \sigma^2 + m_1 \sum_i \theta_{i.}^2, \quad E\{u_2^2\} = \sigma^2.$$

Then the procedure of our modified two-sample scheme is defined in the following way:

(a) *At first we make $l \times m_1$ observations and test analysis of variance for treatment effects given by (3.1)*

(b-1) *If $F' \geq A$, we stop the experimentation and estimate the treatment effects because of that it shows a significant discrepancy among the treatment effects and therefore the sample size may be already sufficient large.*

(b-2) *If $\lambda \leq F' < A$, then make m_2 additional replications all over the treatments, which are defined by the result of analysis of variance, because of that the first sample size is insufficiently large, and after making the additional observations the treatment effects are estimated by pooling of data.*

(b-3) *If $0 < F' < \lambda$, then we make afresh m_3 replications all over the treatments, where $\lambda > 1$ and m_3 is prescribed a priori, and estimate the treatment effects by $l \times m_3$ observations because of the same reasoning as (3°) in § 2, where we define that $A = F_{(l-1), (l-1)(m_1-1)}(\alpha_1)$, $\lambda = F_{(l-1), (l-1)(m_1-1)}(\alpha_2)$.*

The inference rule proposed by us is practised as a routine work very

often and intuitively and it seems to us that the procedure has some reason to apply. Sometimes we are used to put that $m_3=0$, that is, we give up the estimation of the treatment effects at that time by the reason that the difference of their effects is beneath our notice or that we make the experiment in good circumstances at another day. In some circumstances of bioassay, we may have a good reason to put λ equal to A .

Then our present consideration of the necessary size of additional sample comes after the analysis of variance of the first observations.

First in view of the relation

$$(3.3) \quad \frac{E\{u_1^2\}}{E\{u_2^2\}} - 1 = m_1 \frac{\sum \theta_i^2}{\sigma^2},$$

the statistic $(F'-1)/m_1$ may give us an approximate value of $\sum \theta_i^2/\sigma^2$ as an information drawn from the first observations.

Secondly, we may employ a sufficiently large m and there exists always a positive value m such that

$$(3.4) \quad \frac{m \sum \theta_i^2}{\sigma^2} = A-1.$$

Thus the combination of these two facts gives us

$$(3.5) \quad (F'-1)/m_1 = (A-1)/m,$$

which leads us the following relationship

$$(3.6) \quad m_2 = \left[\frac{A-1}{F'-1} m_1 \right] + 1 - m_1,$$

where the notation $[]$ denoting Gauss' symbol. Hence the crucial importance is that the additional number of replications m_2 is a random variable.

As the ratio F' is distributed in the noncentral F -distribution with the pair of the degrees of freedom $(l-1, (l-1)(m_1-1))$ and with the non-centrality parameter $\delta = m \sum \theta_i^2 / 2\sigma^2$ under our situation $\theta_i \neq 0$ for some i , the whole expected number of replications $E\{m|m_1\}$ is given as follows:

$$(3.7) \quad E\{m|m_1\} = (m_1 + m_2) \int_0^\lambda f(F'; m_1, \delta) dF' + \int_\lambda^\Lambda \left(\left[\frac{A-1}{F'-1} m_1 \right] + 1 \right) \cdot f(F'; m_1, \delta) dF' + m_1 \int_\Lambda^\infty f(F'; m_1, \delta) dF',$$

where

$$f(F'; m_1, \delta) dF' = \sum \frac{e^{-\delta} \delta^k}{k! B\left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2}\right)} \left(\frac{1}{m_1-1} \right)^{\frac{l-1}{2} + k} F'^{\frac{l-3}{2} + k} \left(1 + \frac{F'}{m_1-1} \right)^{-\frac{m_1(l-1)}{2} - k} dF'.$$

Then

$$(3.8) \quad \int_0^\lambda f(F'; m_1, \delta) dF' = \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right),$$

and

$$(3.9) \quad I(\lambda, A) \equiv \int_\lambda^\Lambda \frac{1}{F'-1} f(F'; m_1, \delta) dF'$$

$$= \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k! B \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right)} \sum_{i=0}^{(l-1)(m_1-1)/2} \frac{(-1)^i \binom{(l-1)(m_1-1)}{i}}{m_1^{\frac{l-1}{2} + k + i}}$$

$$\cdot \left[2 \sum_{j=0}^{k+i} \frac{m_1^{\frac{l-3}{2} + k + i - j} \left\{ \theta^{\frac{l-3}{2} + k + i - j} - \theta^{\frac{l-3}{2} + k + i - j} \right\}}{2(k+i-j) + l-3} + 2 \sum_{j=0}^{\frac{l-4}{2}} \frac{m_1^{\frac{l-3}{2} - j} \left\{ \theta^{\frac{l-3}{2} - j} - \theta^{\frac{l-3}{2} - j} \right\}}{l-3-2j} \right.$$

$$\left. + \log \left\{ \left| \frac{\sqrt{m_1 \theta - 1}}{\sqrt{m_1 \theta + 1}} \right| \cdot \left| \frac{\sqrt{m_1 \theta + 1}}{\sqrt{m_1 \theta - 1}} \right| \right\} \right], \quad (l: \text{even}),$$

$$= \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k! B \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right)} \sum_i \frac{(-1)^i \binom{(l-1)(m_1-1)}{i}}{m_1^{\frac{l-1}{2} + k + i}}$$

$$\cdot \left[2 \sum_{j=0}^{k+i} \frac{m_1^{\frac{l-3}{2} + k + i - j} \left\{ \theta^{\frac{l-3}{2} + k + i - j} - \theta^{\frac{l-3}{2} + k + i - j} \right\}}{2(k+i-j) + l-3} + 2 \sum_{j=0}^{\frac{l-5}{2}} \frac{m_1^{\frac{l-3}{2} - j} \left\{ \theta^{\frac{l-3}{2} - j} - \theta^{\frac{l-3}{2} - j} \right\}}{l-3-2j} \right.$$

$$\left. + \log \left| \frac{m_1 \theta - 1}{m_1 \theta + 1} \right| \right], \quad (l: \text{odd}),$$

where $\theta = \lambda / (m_1 - 1 + \lambda)$ and $\Theta = A / (m_1 - 1 + A)$.

Furthermore we may evaluate the right side of the function (3.7) by the same relation as (2.17) formally. Hence we have $E\{m m_1\}$ as follows:

$$(3.10) \quad (m_3 + m_1 - 1) \sum_k \frac{e^{-\delta} \delta^k}{k!} I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) + m_1 \left\{ 1 + (A-1) I(\lambda, A) \right.$$

$$\left. - \sum_k \frac{e^{-\delta} \delta^k}{k!} I_\Theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) \right\} \leq E\{m m_1\} < (m_3 + m_1 - 1)$$

$$\cdot \sum_k \frac{e^{-\delta \delta^k}}{k!} I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) + m_1 \left\{ 1 + (A-1)I(\lambda, A) - \sum_k \frac{e^{-\delta \delta^k}}{k!} \right. \\ \left. \cdot I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) \right\} + \sum_k \frac{e^{-\delta \delta^k}}{k!} I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right).$$

The mean square error around $E\{m|m_1\}$ of (3.7) is given by the same expression as (2.18), where $E\{m'^2\}$ is also that of (2.19).

Then if we put that

$$(3.11) \quad J(\lambda, A) \equiv \int_{\lambda}^{\Lambda} \frac{1}{(F'-1)^2} f(F'; m_1, \delta) dF' \\ = \sum_k \frac{e^{-\delta \delta^k}}{k! B \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right)} \sum_{i=0}^{\frac{(l-1)(m_1-1)}{2} + 1} \frac{(-1)^i \binom{\frac{(l-1)(m_1-1)}{2} + 1}{i}}{m_1^{\frac{l-1}{2} + k + i}} \left[\left\{ \frac{(m_1 \theta)}{m_1 \theta - 1} - \right. \right. \\ \left. \frac{2(k+i)+l-1}{2} \frac{(m_1 \theta)}{m_1 \theta - 1} \right\} + \frac{2(k+i)+l-3}{2} \left\{ \sum_{j=0}^{k+i} \frac{2}{2(k+i-j)+l-3} \left((m_1 \theta)^{\frac{2(k+i-j)+l-3}{2}} - (m_1 \theta)^{\frac{2(k+i-j)+l-3}{2}} \right) \right. \\ \left. + \sum_{j=0}^{\frac{l-4}{2}} \frac{2}{l-3-2j} \left((m_1 \theta)^{\frac{l-3-2j}{2}} - (m_1 \theta)^{\frac{l-3-2j}{2}} \right) + \log \left[\frac{\sqrt{m_1 \theta} - 1}{\sqrt{m_1 \theta} + 1} \cdot \frac{\sqrt{m_1 \theta} + 1}{\sqrt{m_1 \theta} - 1} \right] \right\}, (l: \text{even}), \\ = \sum_k \frac{e^{-\delta \delta^k}}{k! B \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right)} \sum_{i=0}^{\frac{(l-1)(m_1-1)}{2} + 1} \frac{(-1)^i \binom{\frac{(l-1)(m_1-1)}{2} + 1}{i}}{m_1^{\frac{l-1}{2} + k + i}} \left[\left\{ \frac{(m_1 \theta)}{m_1 \theta - 1} \right. \right. \\ \left. - \frac{2(k+i)+l-1}{2} \frac{(m_1 \theta)}{m_1 \theta - 1} \right\} + \frac{2(k+i)+l-3}{2} \left\{ \sum_{j=0}^{k+i} \frac{2}{2(k+i-j)+l-3} \left((m_1 \theta)^{\frac{2(k+i-j)+l-3}{2}} - (m_1 \theta)^{\frac{2(k+i-j)+l-3}{2}} \right) \right. \\ \left. + \sum_{j=0}^{\frac{l-5}{2}} \frac{2}{l-3-2j} \left((m_1 \theta)^{\frac{l-3-2j}{2}} - (m_1 \theta)^{\frac{l-3-2j}{2}} \right) + \log \left[\frac{m_1 \theta - 1}{m_1 \theta + 1} \right] \right\}, (l: \text{odd}),$$

the value of $E(m'^2)$ can be evaluated as follows:

$$(3.12) \quad E(m'^2) = m_3(2m_1 + m_3) \sum_k \frac{e^{-\delta \delta^k}}{k!} I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) \\ + m_1^2 \left[1 - \sum_k \frac{e^{-\delta \delta^k}}{k!} \left\{ I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) - I_\theta \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right) \right\} \right. \\ \left. + (A-1)^2 J(\lambda, A) \right].$$

Furthermore the power of test is defined on the basis of testing the hypothesis $\delta=0$ against an alternative hypothesis $\delta \neq 0$ and according to our estimation procedure in the current case of $m_3=0$ is given by

$$(3.13) \quad Pr.\{F' \geq F_{l-1, (l-1)(m_1-1)}(\alpha_1)\} \\ + \int_{\lambda}^{\Lambda} Pr.\{F' < F_{l-1, (l-1)(m_1-1)}(\alpha_1), F' \geq F_{l-1, (l-1)(m_1+m_2-1)}(\alpha_1)\} g(F') dF',$$

where the first term is equal to

$$(3.14) \quad 1 - \sum_{k=0}^{\infty} \frac{e^{-\delta} \delta^k}{k!} I_{\theta} \left(\frac{l-1}{2} + k, \frac{(l-1)(m_1-1)}{2} \right),$$

and the second term is

$$(3.15) \quad \int_{\lambda}^{\Lambda} \int_{D_1^c} \left\{ \iint_{\hat{D}(v, w)} f(u, v, w, F') dv dw \right\} du dF',$$

in which we put

$$(3.16) \quad \left\{ \begin{aligned} & f(u, v, w, F') \\ &= \frac{\Gamma(l+m_{12}-1)}{\Gamma\left(\frac{m_1-1}{2}\right) \Gamma\left(\frac{m_{12}-m_1-1}{2}\right) \left\{ \Gamma\left(\frac{l-1}{2}\right) \right\}^2} u^{\frac{l-3}{2}} v^{\frac{l+m_1}{2}-2} w^{\frac{m_1-5}{2}} \\ & \quad \cdot (1+w+uv+vwu)^{-(l+m_{12}-1)}, \\ & m_{12} = m_1(A-1)/(F'-1), \quad \theta_{21} = \sum_{i=1}^l \theta_i^2 / \sigma^2, \\ & D_1^c: 0 < u < F_{(l-1), (l-1)(m_1-1)}(\alpha_1) \cdot \theta_{21} / (m_1-1), \\ & D(v, w): F_{(l-1), (l-1)m_{12}}(\alpha_1) / m_{12} < \frac{uv+1}{v+w^{-1}} < \infty \text{ for fixed } u. \end{aligned} \right.$$

KYUSHU UNIVERSITY and SHIONOGI PHARMACEUTICAL CO.

References

- KITAGAWA, T. [1]: *Successive process of statistical inferences*, (4), Bull. Math. Stat., 5 (1952), 35-50.
- KITAGAWA, T., KITAHARA, T., NOMACHI, Y. and WATANABE, N. [2]: *On the determination of sample size from the two sample theoretical formulation*, Bull. Math. Stat., 5 (1953).
- SEELBINDER B. M. [1]: *On Stein's two-stage sampling scheme*, Ann. Math. Stat., 24 (1953), 640-647.
- STEIN, C. [1]: *A two-sample test for a linear hypothesis whose power is independent of the variance*, Ann. Math. Stat., 16 (1945), 243-258.