

## On the Determination of Sample Size from the Two Sample Theoretical Formulation

Kitagawa, Toshio  
Kyushu University

Kitahara, Teisuke

Nomachi, Yukio  
Kyushu University

Watanabe, Nobuo

<https://doi.org/10.5109/12960>

---

出版情報 : 統計数理研究. 5 (3/4), pp.35-45, 1953-06. Research Association of Statistical Sciences

バージョン :

権利関係 :



# ON THE DETERMINATION OF SAMPLE SIZE FROM THE TWO SAMPLE THEORETICAL FORMULATION

By

Tosio KITAGAWA, Teisuke KITAHARA,  
Yukio NOMACHI and Nobuo WATANABE

## § 1. Two sample theoretical formulation.

Let us consider a normal population  $N(m, \sigma^2)$  with unknown mean  $m$  and unknown variance  $\sigma^2$ , and let  $O_n: (x_1, x_2, \dots, x_n)$  be a random sample of size  $n$  from this population. Then, as well known, the confidence interval for the mean  $m$  with confidence coefficient  $1 - \alpha$  is given by

$$(1.01) \quad (\bar{x} - t_{n-1}(\alpha) s n^{-1/2}, \quad \bar{x} + t_{n-1}(\alpha) s n^{-1/2}),$$

where  $\bar{x}$  and  $s^2$  are the sample mean and unbiased variance estimate from obtained from the sample respectively, and  $t_{n-1}(\alpha)$  is the  $\alpha$  significant level of  $t$ -distribution with the  $n - 1$  degrees of freedom. The length of confidence interval obtained in this way is  $2t_{n-1}(\alpha) s n^{-1/2}$ , and the expectation of this length is equal to  $2t_{n-1}(\alpha) E\{s\} n^{-1/2} = 2t_{n-1}(\alpha) \sigma n^{-1/2} + O(n^{-3/2})$ , which tends to zero as  $n$  increases infinitely and moreover we have  $4t_{n-1}^2(\alpha) n^{-1} \sigma^2 \{s\} = 2t_{n-1}^2(\alpha) \sigma^2 n^{-2} + O(n^{-3})$ . In view of these properties of length of the confidence interval, it will be suggested that there may be some possibilities that a length of confidence interval with any assigned confidence coefficient  $1 - \alpha$  can be smaller than any assigned length  $2d$ , by choosing the sample size sufficiently large. The real situations, however, should be more carefully formulated. There are two essential points which should be remarked. The first point is that the relation  $t_{n-1}(\alpha) s n^{-1/2} \leq d$  cannot be established with the probability one, for any fixed  $n$ , while the second one is that since the population variance  $\sigma^2$  is unknown to us, the equation such as  $t_{n-1}(\alpha) s n^{-1/2} \leq d$  cannot give us any definite procedure for determining the sample size  $n$ . In view of these two facts, let us now appeal to the two sample theoretical approach, which will suggest following type of two sample procedure: (i) Let us at first draw a sample of size  $n_1$   $O_{n_1}: (x_{1,1}, x_{1,2}, \dots, x_{1,n_1})$  from our population, and let us make the unbiased estimate  $s_1^2$  for  $\sigma^2$ :

$$(1.02) \quad s_1^2 = (n_1 - 1)^{-1} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2.$$

(ii) In view of this  $s_1$ , let us determine a size of second sample by some procedure, and draw a second sample of this size  $O_{n_2}: (x_{21}, x_{22}, \dots, x_{2n_2})$  from our population. Then our situation is to expect to satisfy the following relation:

$$(1.03) \quad \text{Pr. } \{t_{n_2-1}(\alpha) s_2 n_2^{-1/2} \leq d\} \geq 1 - \beta,$$

where

$$(1.04) \quad s_2^2 = (n_2 - 1)^{-1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2.$$

It is to be noted that the sample size  $n_2$  is itself a random variable as a function of the stochastic variables. In what follows let us denote the stochastic variable which represents the sample size  $n_2$  by a bold type  $\mathbf{n}_2$  in order to distinguish from a constant number  $n_2$ . In order that we may realize this type of two sample procedure, what remains to us to determine is how to select the size of the second sample  $n_2$ . Here let us appeal to the  $F$ -distribution which gives us, for two independent unbiased variance estimates  $s_1^2$  and  $s_2^2$  with the degrees of freedom  $n_1 - 1$  and  $n_2 - 1$  respectively.

$$(1.05) \quad \text{Pr. } \{s_2/s_1 < (F_{n_1-1}^{n_2-1}(\beta))^{1/2}\} = 1 - \beta.$$

In combination of (1.03) and (1.05) we have the relation

$$(1.06) \quad \text{Pr. } \left\{s_2 \leq \frac{dn_2^{1/2}}{t_{n_2-1}(\alpha)}\right\} \geq 1 - \beta = \text{Pr. } \{s_2 \leq s_1 \sqrt{F_{n_1-1}^{n_2-1}(\beta)}\}.$$

Therefore we propose to determine the size of second sample  $n_2$  such as the following relations shall hold simultaneously:

$$(1.07) \quad \frac{dn_2^{1/2}}{t_{n_2-1}(\alpha)} \geq s_1 \sqrt{F_{n_1-1}^{n_2-1}(\beta)},$$

and

$$(1.08) \quad \frac{d(n_2 - 1)^{1/2}}{t_{n_2-2}(\alpha)} < s_1 \sqrt{F_{n_1-1}^{n_2-2}(\beta)},$$

that is,

$$(1.09) \quad \frac{d(n_2 - 1)^{1/2}}{t_{n_2-2}(\alpha) \sqrt{F_{n_1-1}^{n_2-2}(\beta)}} < s_1 \leq \frac{dn_2^{1/2}}{t_{n_2-1}(\alpha) \sqrt{F_{n_1-1}^{n_2-1}(\beta)}}.$$

Consequently we have following relations:

$$\begin{aligned} (1.10) \quad & \text{Pr. } \left\{ \frac{t_{n_2-1}(\alpha)}{\sqrt{n_2}} s_2 < d \right\} \\ &= \sum_{n_2=3}^{\infty} \text{Pr. } \left\{ \frac{t_{n_2-1}(\alpha)}{\sqrt{n_2}} s_2 < d/n_2 \right\} \text{Pr. } \{n_2 = n_2\} \\ &= \sum_{n_2=3}^{\infty} \text{Pr. } \left\{ \frac{d(n_2 - 1)^{1/2}}{t_{n_2-2}(\alpha) \sqrt{F_{n_1-1}^{n_2-2}(\beta)}} < s_1 \leq \frac{dn_2^{1/2}}{t_{n_2-1}(\alpha) \sqrt{F_{n_1-1}^{n_2-1}(\beta)}} \right\} \\ & \quad \cdot \text{Pr. } \left\{ \frac{t_{n_2-1}(\alpha)}{\sqrt{n}} s_2 < d/s_1 \right\} \end{aligned}$$

$$= \sum_{n_2=3}^{\infty} \int_{\frac{t_{n_2-2}(\alpha) \sqrt{F_{n_1-1}^{n_2-2}(\beta)}}{d(n_2-1)^{1/2}}}^{\frac{dn_2^{2/1}}{t_{n_2-1}(\alpha) \sqrt{F_{n_1-1}^{n_2-1}(\beta)}}} \varphi_{n_1}(s_1) ds_1 \int_0^{\frac{dn_2^{1/2}}{t_{n_2-1}(\alpha)}} \varphi_{n_2}(s_2) ds_2.$$

## § 2. Numerical considerations.

Each term of the sums of the integrals in the last right hand side of (1.10) may be calculated by means of the tables of the incomplete  $I'$ -function [1]. We have constructed the distribution functions of the stochastic variables  $n_2$  with respect to the set of the values of  $\alpha, \beta, d^2\sigma^{-2}$  and  $n_1$ , where  $\alpha = 0.01, 0.05$ ;  $\beta = 0.01, 0.05$ ;  $d^2\sigma^{-2} = 0.75, 0.25$ ;  $n_1 = 5, 10, 15, 21, 25, 31$ . The results are given in Tables I, II, III and IV. These tables suggest us the way how to choose a sample size  $n_2$  in our two sample theoretical formulations; indeed this choice will be seen to depend upon the values of  $\alpha, \beta$  and  $d^2\sigma^{-2}$ .

## § 3. General remarks.

There remains the problem of two dependent samples. That is to say, in stead of taking consideration a new independent second sample  $O_2$ , we may consider a pooled sample of size  $n_1 + n_2$ , adding a new sample of size  $n_2$  to that of size  $n_1$ . This problem was considered by MOOD [1], but, as he himself pointed out in MOOD [2], there was an error because he had been in failing to take account of the fact that  $n$  is a random variable.

In these problems there may be various procedures which will give us two sample formulations. The method due to STEIN [1] is specially worth while to mention, and KITAGAWA [1] gave some stochastic considerations about similar problems in Part III, and KITAGAWA [4] applies STEIN-BARNARD method to BEHRENS-FISHER's test. The two dependent sample formulations will be reconsidered in another occasion.

## References

- Marilyn HARRIS, D. G. WORVITZ and A. M. MOOD. [1]: *On the determination of sample sizes in designing experiments*, Jour. Ame. Stat. Associ., sep. 1948.
- A. M. MOOD; Erratum [2]: *On the determination of sample size in designing experiments*, Jour. Ameri. Stat. Associ., December 1951, vol. 46, No. 256.
- Karl PEARSON, F.R.S. [1]: *Tables of the incomplete  $\Gamma$ -function*, 1922, London.
- Charles STEIN. [1]: *A Two Sample Test for a linear Hypothesis whose Power is Independent of the Variance*, Ann. Math. Stat., vol. 15, 1945, p. 234.
- KITAGAWA, T. [1]: *Successive process of statistical inferences*, Mem. Fac. Sci., Ser. A, 5 (1950), 139-180.
- KITAGAWA, T. [2]: *Successive process of statistical inferences*, Mem. Fac. Sci., Ser. A, 6 (1951), 55-95.
- KITAGAWA, T. [3]: *Successive process of statistical inferences*, Man. Fac. Sci., Ser. A, 7 (1952), 131-155.
- KITAGAWA, T. [4]: *Successive process of statistical inferences*, Bull. Math. Stat. Resrech Association of Stat. Sci., vol. 5, No. 1-2.



**Table I — (2)** ( $\alpha = \beta = 1\%$ ;  $d^2\sigma^{-2} = 0.5$ )

$n_1 \backslash n_2$	11	12	13	14	15	16	17	18	19	20	C.S.
25	0	0	0	0	0	0.0002	0.0004	0.0025	0.0015	0.0028	0.0074
31	0	0	0	0	0	0.0001	0.0002	0.0006	0.0011	0.0026	0.0046
$n_1 \backslash n_2$	21	22	23	24	25	26	27	28	29	30	C.S.
25	0.0047	0.0085	0.0091	0.0133	0.0171	0.0208	0.0244	0.0299	0.0339	0.0384	0.2075
31	0.0045	0.0067	0.0110	0.0154	0.0219	0.0262	0.0319	0.0379	0.0442	0.0499	0.2562
$n_1 \backslash n_2$	31	32	33	34	35	36	37	38	39	40	C.S.
25	0.0407	0.0410	0.0432	0.0451	0.0463	0.0466	0.0464	0.0459	0.0443	0.0427	0.6497
31	0.0526	0.0524	0.0543	0.0551	0.0550	0.0537	0.0521	0.0492	0.0454	0.0418	0.7658
$n_1 \backslash n_2$	41	42	43	44	45	46	47	48	49	50	C.S.
25	0.0406	0.0316	0.0338	0.0290	0.0267	0.0241	0.0219	0.0196	0.0173	0.0154	0.9097
31	0.0375	0.0296	0.0263	0.0228	0.0197	0.0170	0.0142	0.0120	0.0099	0.0072	0.9620
$n_1 \backslash n_2$	51	52	53	54	55	56	57	58	59	60	T.S.
25	0.0133	0.0113	0.0102	0.0087	0.0073	0.0063	0.0052	0.0043	0.0036	0.0031	0.9840
31	0.0065	0.0051	0.0044	0.0034	0.0026	0.0021	0.0016	0.0012	0.0009	0.0007	0.9905

[illegible]





**Table III—(1)** ( $\alpha = 5\%$ ,  $\beta = 1\%$ ;  $d^2\sigma^{-2} = 0.75$ )

[illegible]

Table III — (2) ( $\alpha = 5\%$ ,  $\beta = 1\%$ ;  $d^2\sigma^{-2} = 0.5$ )

$n_1 \backslash n_2$	9	10	11	12	13	14	15	16	17	18	C.S.
21	0	0.0003	0.0010	0.0024	0.0055	0.0091	0.0163	0.0242	0.0322	0.0405	0.1315
25	0	0.0002	0.0007	0.0019	0.0050	0.0095	0.0175	0.0282	0.0382	0.0502	0.1514
31	0	0.0001	0.0004	0.0031	0.0041	0.0091	0.0188	0.0334	0.0469	0.0631	0.1772
$n_1 \backslash n_2$	19	20	21	22	23	24	25	26	27	28	C.S.
21	0.0498	0.0576	0.0641	0.0651	0.0667	0.0670	0.0652	0.0596	0.0551	0.0502	0.7319
25	0.0615	0.0715	0.0786	0.0777	0.0776	0.0751	0.0701	0.0608	0.0534	0.0458	0.8235
31	0.0774	0.0905	0.0954	0.0924	0.0875	0.0800	0.0697	0.0549	0.0445	0.0348	0.9043
$n_1 \backslash n_2$	29	30	31	32	33	34	35	36	37	38	C.S.
21	0.0450	0.0392	0.0339	0.0272	0.0229	0.0193	0.0157	0.0128	0.0104	0.0081	0.9694
25	0.0382	0.0314	0.0251	0.0185	0.0145	0.0112	0.0084	0.0064	0.0047	0.0033	0.9852
31	0.0260	0.0192	0.0136	0.0088	0.0062	0.0041	0.0028	0.0018	0.0011	0.0007	0.9886
$n_1 \backslash n_2$	39	40	41	42	43	44	45	46	47	48	C.S.
21	0.0065	0.0040	0.0039	0.0026	0.0021	0.0016	0.0012	0.0009	0.0007	0.0005	0.9904
25	0.0024	0.0019	0.0009	0.0007	0.0005	0.0004	0.0002	0.0002	0.0001	0.0001	0.9926
31	0.0004	0.0003	0.0001	0.0001	0	0	0	0	0	0	0.9895
$n_1 \backslash n_2$	49	50	51	52	53	54	55	56	57	59	T.S.
21	0.0003	0.0003	0.0002	0.0002	0.0001	0					0.9915
25	0	0									0.9926
31	0	0									0.9895

Table III — (3) ( $\alpha = 5\%$ ,  $\beta = 1\%$ ;  $d^2/\sigma^2 = 0.25$ )

$n_1 \backslash n_2$	16	17	18	19	20	21	22	23	24	25	C.S.
31	0	0.0001	0.0003	0.0006	0.0012	0.0023	0.0036	0.0055	0.0083	0.0114	0.0333
$n_1 \backslash n_2$	26	27	28	29	30	31	32	33	34	35	C.S.
31	0.0154	0.0194	0.0239	0.0285	0.0337	0.0384	0.0415	0.0440	0.0462	0.0478	0.3721
$n_1 \backslash n_2$	36	37	38	39	40	41	42	43	44	45	C.S.
31	0.0491	0.0498	0.0496	0.0489	0.0475	0.0447	0.0384	0.0359	0.0332	0.0305	0.7997
$n_1 \backslash n_2$	46	47	48	49	50	51	52	53	54	55	C.S.
31	0.0272	0.0245	0.0218	0.0193	0.0170	0.0144	0.0124	0.0107	0.0090	0.0088	0.9648
$n_1 \backslash n_2$	56	57	58	59	60	61	62	63	64	65	T.S.
31	0.0050	0.0051	0.0044	0.0037	0.0028	0.9858					

Table IV — (2)      ( $\alpha = \beta = 5\%$ ;  $d^2\sigma^{-2} = 0.5$ )

$n_1 \backslash n_2$	3	4	5	6	7	8	9	10	11	12	C.S.
10			0	0.0002	0.0008	0.0016	0.0034	0.0063	0.0110	0.0159	0.0392
15				0	0.0002	0.0006	0.0021	0.0051	0.0104	0.0176	0.0360
21					0	0.0002	0.0011	0.0034	0.0087	0.0153	0.0287
$n_1 \backslash n_2$	13	14	15	16	17	18	19	20	21	22	C.S.
10	0.0207	0.0253	0.0315	0.0362	0.0390	0.0413	0.0433	0.0443	0.0450	0.0448	0.4106
15	0.0274	0.0355	0.0469	0.0557	0.0602	0.0642	0.0660	0.0657	0.0646	0.0588	0.5810
21	0.0362	0.0390	0.0615	0.0730	0.0823	0.0865	0.0862	0.0820	0.0752	0.0633	0.7139
$n_1 \backslash n_2$	23	24	25	26	27	28	29	30	31	32	C.S.
10	0.0439	0.0418	0.0397	0.0377	0.0358	0.0338	0.0315	0.0295	0.0271	0.0243	0.7557
15	0.0546	0.0496	0.0449	0.0383	0.0335	0.0288	0.0244	0.0207	0.0170	0.0140	0.9068
21	0.0538	0.0445	0.0352	0.0282	0.0213	0.0161	0.0119	0.0087	0.0062	0.0043	0.9439
$n_1 \backslash n_2$	33	34	35	36	37	38	39	40	41		T.S.
10	0.0246	0.0205	0.0186	0.0169	0.0153	0.0139	0.0126	0.0117			0.8898
15	0.0111	0.0092	0.0074	0.0059	0.0048	0.0038	0.0028	0.0025	0.0017		0.9560
21	0.0030	0.0021	0.0014	0.0010	0.0007	0.0004	0.0003	0.0002	0.0001		0.9531

