Josef Machek

On a two-sample procedure for testing Student's hypothesis using mean range

Aplikace matematiky, Vol. 4 (1959), No. 3, 211–223

ON A TWO-SAMPLE PROCEDURE FOR TESTING STUDENT'S HYPOTHESIS USING MEAN RANGE

JOSEF MACHEK

(Received September 9th 1958) DT: 519.271

Stein's two-sample procedure for testing a hypothesis concerning the mean of a normal distribution is modified for the case where the usual mean-square estimate of the standard deviation is replaced by an estimate based on sample range or on mean range of several samples. Expressions for the mean number of necessary observations are derived and some illustrative tables given. Tables of 5% and 10% critical points of the corresponding test statistic are given.

1. INTRODUCTION AND STATEMENT OF PROBLEM

Throughout the paper we shall be concerned with one-sided tests of a hypothesis about the mean of a normal distribution where the value of the population standard deviation is not known and it is nevertheless desired that the power-function of the test pass through two prescribed points. Thus we shall consider tests which have to provide a given protection not only against an unjust rejection of the hypothesis when the mean takes the value specified by it, but also against accepting it when the mean has a given value different from that specified by the hypothesis. Moreover, we shall try to make the test rather simple in application.

The need for such tests is most likely to occur in problems of sampling inspection. An example is provided by acceptance sampling of materials; suppose, e. g., that a lot of material, say a chemical, is to be sampled. A quantitative characteristic such as concentration or percentage content of some ingredient is relevant for the quality of the material. Lots with values of this characteristic smaller than a given standard are perfectly suitable and should be accepted on the average at least in a prescribed proportion of cases, say 95 per cent. With increasing values of the characteristic the quality of lots decreases so that lots in which the relevant characteristic has a value as high as another given constant or higher are unacceptable (we may well imagine that the use of them is associated with serious difficulties) and the consumer is prepared to run only a risk of no more than one in ten, say, of accepting such lots. If the standard deviation of the results of corresponding analyses were known and some further assumptions (concerning the homogeneity of the lots and the method of sampling) valid, the construction of a sampling plan would present no difficulties of principle. If the standard deviation cannot be regarded as known, the solution is more complicated.

The problem may be described as follows: Let us have a population whose distribution is normal, with mean μ and standard deviation σ. Both μ and σ are unknown. Two numbers are given, μ_0 and Δ ($\Delta > 0$). It is desired to test the hypothesis $\mu \leqq \mu_0$ against the alternative $\mu > \mu_0$ and the test has to satisfy the following conditions ($\boldsymbol{P}(\mu)$ denotes the power of the test, i. e. the probability that the hypothesis will be rejected when μ is the true value of the mean):

$$\left.\begin{aligned} \boldsymbol{P}(\mu_0) &= \alpha\,, \\ \boldsymbol{P}(\mu_0 + \Delta) &= 1 - \beta\,. \end{aligned}\right\} \tag{1}$$

Here α and β are previously fixed numbers from the interval (0, 1), usually equal to 0·10, 0·01 or the like. They represent, respectively, the highest probability of rejecting the hypothesis if it is in fact true and the least probability of rejecting the hypothesis if μ is at the distance Δ or larger from the largest value in the hypothesis. Thus α is what is called producer's risk in quality control problems and β the consumer's risk.

It has been proved [1], that no single-sample test satisfying (1) exists. It is possible, however, to devise a test employing two samples (the size of the first sample being fixed and the size of the second directed by the observations in the first) that insures the fulfillment of requirements (1) whatever be the value of σ. The idea is due to Ch. Stein [2]. In statistical quality control — as far as is known — Stein's method is not widely used, although with attribute sampling two-sample inspection plans (and indeed more complicated plans) are used occasionally. At the same time with the problem just stated the two-phase procedure is not merely a means for reducing the number of observations, but the only way to ensure the fulfillment of (1). The reason for which Stein's procedure has not so far been used more extensively seems to be that its application requires the computation of sample standard deviation which is not suitable for routine work. Therefore we propose a modification of Stein's procedure in which the usual estimate of standard deviation is replaced by mean range — an estimate that is common in quality control problems. In section 2 the modified procedure is described, in section 3 the theoretical background is explained and the construction of tables described. In section 4 the formulae for average sample size are developed and the average sample size tabled for a particular case.

2. THE TEST PROCEDURE

The test procedure satisfying requirements (1) involves the following steps:

1. A random sample of an arbitrary size $n_1 = kn$ is drawn (k and n are integers). Suitable choice of n_1 will be dicussed in section 4. This sample is composed of k mutually independent subsamples of size n. From this sample we compute: the mean, to be denoted by $\bar{x}_1$, the range of each subsample, $R_1, R_2, \ldots, R_k$ and the mean range of all subsamples,

$$\bar{R} = \frac{1}{k}\sum_{i=1}^{k} R_i \, .$$

The sample is decomposed in subsamples for two reasons; first, for larger values of kn (say, $kn \geqq 12$) the mean range of k groups of n is — when k and n have been suitably chosen — a more efficient estimate of σ than the range of the total sample of kn; second, it is easier to pick out the extreme observations in a smaller group than in the total sample.

2. Another sample is drawn the size of which, n_2, is given by the formula

$$n_2 = \min\left\{s : s > \frac{\bar{R}}{\Delta^2 Z^2} - n_1 \, , \; s \text{ a positive integer}\right\}. \qquad (2)$$

In words, n_2 is the least positive integer greater than $\frac{\bar{R}^2}{\Delta^2 Z^2} - n_1$. Thus if $\bar{R}^2/(\Delta^2 Z^2) \leq n_1$, we have $n_2 = 1$; if

$$n_1 + s - 1 \leqq \frac{\bar{R}^2}{\Delta^2 Z^2} < n_1 + s \, ,$$

we have $n_2 = s$.

In formula (2) Z denotes a number depending on the chosen risks α and β, $Z = \frac{1}{z_\alpha + z_\beta}$. The numbers $z_{0.05}$ and $z_{0.10}$ are for certain combinations of k and n given in tables 1 and 2, respectively. From the data in the second sample the mean $\bar{x}_2$ is computed.

3. The root a of the equation

$$\frac{a^2}{n_1} + \frac{(1-a)^2}{n_2} = \frac{\Delta^2 Z^2}{\bar{R}^2} \qquad (3)$$

is determined.

The solution of (3) and the determination of n_2 from (2) are the only more complicated operations in the application of the procedure. However, if the test is to be used on repeated occasions with data regarding the same quality characteristic, suitable charts may be devised for these operations so that the computational labour is reduced to a minimum.

Table 1.

Values $z_{0\cdot05}$ defined by the relation $P\left\{\frac{\xi}{w} > z_{0\cdot05}\right\} = 0\cdot05$ (ξ has a normal distribution with zero mean and unit variance, w is distributed independently of ξ as the mean range of k mutually independent samples of size n from a normal distribution with unit variance).

k \ n	5	6	7	8	9	10	11	12	15	20
1	0·869	0·765	0·698	0·648	0·616	0·588	0·566	0·544	0·510	0·464
2	0·780	0·703	0·654	0·612	0·583	0·560	0·541	0·525	0·490	0·452
3	0·754	0·684	0·635	0·600	0·573	0·551	0·533	0·518	0·484	0·448
4	0·742	0·675	0·629	0·594	0·568	0·547	0·529	0·515	0·482	0·446
5	0·735	0·669	0·624	0·591	0·565	0·544	0·527	0·513	0·480	0·445
6	0·730	0·666	0·622	0·589	0·563	0·543	0·526	0·511	0·479	0·444
7	0·726	0·663	0·620	0·587	0·562	0·542	0·525	0·510	0·478	0·444
8	0·724	0·662	0·618	0·586	0·561	0·541	0·524	0·510	0·478	0·443
9	0·722	0·660	0·617	0·585	0·560	0·540	0·523	0·509	0·477	0·443
10	0·721	0·659	0·616	0·584	0·559	0·539	0·523	0·509	0·477	0·443
12	0·718	0·657	0·615	0·583	0·558	0·539	0·522	0·508	0·476	0·442
15	0·716	0·656	0·613	0·582	0·558	0·538	0·521	0·507	0·476	0·442
20	0·714	0·654	0·612	0·581	0·557	0·537	0·521	0·507	0·475	0·441
30	0·712	0·652	0·611	0·580	0·556	0·536	0·520	0·506	0·475	0·441
60	0·709	0·651	0·610	0·579	0·555	0·535	0·519	0·505	0·474	0·441
∞	0·707	0·649	0·608	0·578	0·554	0·534	0·518	0·505	0·474	0·440

Table 2.

Values $z_{0\cdot10}$ defined by the relation $P\left\{\frac{\xi}{w} > z_{0\cdot10}\right\} = 0\cdot10$ (ξ has a normal distribution with zero mean and unit variance, w is distributed independently of ξ as the mean range of k mutually independent samples of size n from a normal distribution with unit variance).

k \ n	5	6	7	8	9	10	11	12	15	20
1	0·623	0·558	0·514	0·484	0·460	0·441	0·426	0·413	0·384	0·354
2	0·585	0·530	0·493	0·466	0·445	0·428	0·414	0·403	0·376	0·349
3	0·573	0·522	0·487	0·461	0·440	0·424	0·411	0·400	0·374	0·347
4	0·567	0·517	0·483	0·458	0·438	0·422	0·409	0·398	0·373	0·346
5	0·564	0·515	0·481	0·456	0·437	0·421	0·408	0·397	0·372	0·345
6	0·561	0·514	0·480	0·455	0·436	0·420	0·407	0·396	0·372	0·345
7	0·560	0·512	0·479	0·455	0·435	0·420	0·407	0·396	0·371	0·344
8	0·559	0·512	0·479	0·454	0·435	0·419	0·407	0·396	0·371	0·344
9	0·558	0·511	0·478	0·454	0·434	0·419	0·406	0·395	0·371	0·344
10	0·557	0·510	0·478	0·453	0·434	0·419	0·406	0·395	0·371	0·344
12	0·556	0·510	0·477	0·453	0·434	0·418	0·406	0·395	0·370	0·344
15	0·555	0·509	0·476	0·452	0·433	0·418	0·405	0·395	0·370	0·344
20	0·554	0·508	0·476	0·452	0·433	0·418	0·405	0·394	0·370	0·344
30	0·553	0·507	0·475	0·451	0·432	0·417	0·405	0·394	0·370	0·344
60	0·552	0·507	0·475	0·451	0·432	0·417	0·404	0·394	0·370	0·343
∞	0·551	0·506	0·474	0·450	0·432	0·417	0·404	0·393	0·369	0·343

4. The hypothesis $\mu \leq \mu_0$ is accepted if the test statistic

$$T = \frac{a\bar{x}_1 + (1 - a)\,\bar{x}_2 - \mu_0}{\Delta\,.\,Z} \qquad (4)$$

does not reach the critical value z_α from table 1 or 2 (according to the chosen risk α) corresponding to the pair k and n that has been employed in subdividing the size of the first sample, and rejected otherwise.

By a modification of the inequalities we obtain a test satisfying the conditions

$$\mathsf{P}(\mu_0) = \alpha ,$$

$$\mathsf{P}(\mu_0 - \Delta) = \beta ,$$

i. e. a test of $\mu \geq \mu_0$ directed against the alternatives

$$\mu < \mu_0 .$$

3. THEORETICAL BACKGROUND

Let us examine the distribution of the random variable

$$\zeta = \frac{a\bar{x}_1 + (1 - a)\,\bar{x}_2 - \mu}{\Delta Z} ,$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of two random samples from a normal population with mean μ and standard deviation σ taken in accordance with the principles enumerated in section 2, a is a function of the first sample given by equation (3), Δ and Z have the same meaning as in section 2. First we consider the conditional distribution of ζ, given $\bar{R}$. This distribution is normal (since it is a linear combination of means of two normal samples) and has the expectation

$$\frac{1}{\Delta Z} E\,\{a\bar{x}_1 + (1 - a)\,\bar{x}_2 - \mu\} = 0$$

and variance

$$\frac{\sigma^2}{\Delta^2 Z^2}\left\{\frac{a^2}{n_1} + \frac{(1 - a)^2}{n_2}\right\} = \frac{\sigma^2}{\bar{R}^2}$$

(the last equality follows from (3)).

Thus the conditional distribution function of ζ given $\bar{R}$ is

$$P_R(z) = \mathsf{P}\{\zeta \leq z \mid \bar{R}\} = \Phi\left(\frac{z\bar{R}}{\sigma}\right),$$

where Φ denotes the distribution function of the normal distribution with zero mean and unit variance. And the unconditional distribution function of the random variable ζ is

$$P(z) = \mathsf{P}\{\zeta \leq z\} = \int_0^\infty \Phi(zv)\, p_{k,n}(v)\, \mathrm{d}v , \tag{5}$$

where $p_{k,n}(v)$ denotes the probability density of the distribution of mean range of k mutually independent samples of size n from a normal population with unit variance. It may be easily verified that (5) is the distribution function of the ratio of two random variables, say X/Y, where X has a normal distribution with zero mean and unit variance and Y is distributed as mean range of k independent samples of size n from a normal population with unit variance. Also, this is the distribution of the test statistic T under the hypothesis $\mu = \mu_0$.

The numbers z_q in tables 1 and 2 are the solutions of the equation

$$1 - P(z_q) = q \tag{6}$$

for $q = 0{\cdot}05$ and $q = 0{\cdot}10$, respectively.

It is seen that the procedure described in section 2 satisfies the requirements (1) of section 1 since

$$\boldsymbol{P}(\mu_0) = \mathsf{P}\{T \geqq z_\alpha \mid \mu = \mu_0\} = 1 - P(z_\alpha) = \alpha$$

and

$$\boldsymbol{P}(\mu_0 + \Delta) = \mathsf{P}\{T \geqq z_\alpha \mid \mu = \mu_0 + \Delta\} = 1 - P\left(z_\alpha - \frac{1}{Z}\right) = 1 - P(-z_\beta) = \\ = 1 - \beta\,.$$

The last equality follows from the symmetry of the distribution of ζ.

The values in tables 1 and 2 have been computed by means of an approximation auggested by P. B. Patnaik [3]. Numerical comparisons which have been made so far show that the approximation is very close. We shall describe it briefly, since we shall have occasion to use it in section 4 in the calculation of the average number of necessary observations.

Patnaik's approximation consists in the replacement of the distribution of mean range by the distribution of the random variable $c\sqrt{\chi^2/\nu}$, where χ^2 has a χ^2 distribution with ν "degrees of freedom" and c is a positive constant. The constants c and ν are determined so that the first two moments of $c\sqrt{\chi^2_\nu/\nu}$ coincide with the first two moments of mean range. Thus the "number of degrees of freedom" may assume non-integral values, too. The numbers c and ν depend, of course, on the number of subsamples, k, and on their size, n. A table of c and ν values for some combinations of k and n is given in [4]. For the purposes of the present paper a more detailed table had to be prepared.

The critical values z_q were then computed as follows. Let ξ have a normal distribution with zero mean and unit variance, further let w be distributed as mean range of k independent samples of size n from a normal population with unit variance; assume that w is independent of ξ. Since w is distributed approximately as $c\sqrt{\chi^2_\nu/\nu}$, the ratio ξ/w is distributed approximately as the ratio $\xi/[c\sqrt{\chi^2_\nu/\nu}]$, i. e.

$$\mathsf{P}\left\{\frac{\xi}{w} \leqq z\right\} \approx \mathsf{P}\left\{\frac{\xi}{c\sqrt{\chi^2_\nu/\nu}} \leqq z\right\} = G_\nu(cz)\,, \tag{7}$$

where G_ν denotes the distribution function of Student's distribution with ν "degrees of freedom". The values z_q defined by (6) are thus approximately given by

$$z_q = \frac{t_{1-q}}{c}, \tag{8}$$

where t_{1-q} is the $100(1-q)$ percent fractile of Student's distribution with ν "degrees of freedom".

For some particular values of q the critical values z_q may be computed from the values tabled by LORD in [5]. Lord's tables are constructed for a two-sided test of Student's hypothesis when an unbiased estimate of σ based on mean range is used. Accordingly they give, for selected values of q and for a number of combinations of k and n (i. e. the number of subsamples and their size), the values u_q defined by the relation

$$\mathsf{P}\left\{\frac{|\xi|}{w/d_n} > u_q\right\} = q\,,$$

where ξ is a random variable with a normal distribution with zero mean and unit standard deviation and w is a random variable that is distributed as mean range of k mutually independent samples of size n from a normal population with unit variance; d_n is the expectation of the range of a sample of n elements from a unit normal population, tabled e. g. in [4] or [6]. The critical values z_q are connected with u_q through the relation

$$z_q = \frac{u_{2q}}{d_n}\,. \tag{9}$$

4. THE AVERAGE SAMPLE SIZE

In the procedure described in section 2 only the first sample has a fixed size n_1; the size of the second sample, n_2, depends on the mean range of the first sample. The total number of observations necessary to reach the decision, $N = n_1 + n_2$, is thus, in fact, a random variable. Its mean value is

$$\mathsf{E}\{N\} = (n_1 + 1)\,\mathsf{P}\{\bar{R} < \Delta Z \sqrt{n_1 + 1}\} + \sum_{k=n_1+2}^{\infty} k\,\mathsf{P}\left\{k - 1 \leq \frac{\bar{R}^2}{\Delta^2 Z^2} < k\right\} =$$

$$= (n_1 + 1)\,\mathsf{P}\left\{\frac{\bar{R}^2}{\sigma^2} < \frac{\Delta^2 Z^2 (n_1 + 1)}{\sigma^2}\right\} + \sum_{k=n_1+2}^{\infty} k\,\mathsf{P}\left\{\frac{\Delta^2 Z^2 (k-1)}{\sigma^2} \leq \frac{\bar{R}^2}{\sigma^2} < \frac{\Delta^2 Z^2 k}{\sigma^2}\right\}.$$

The meaning of the symbols $\bar{R}$, Δ, Z and σ^2 is the same here as in sections 1 and 2. Denote by δ the "distance between the hypothesis and the alternative", expressed in terms of the population standard deviation, $\delta = \Delta/\sigma$ and apply Patnaik's approximation described in section 3 to the expression for $\mathsf{E}\{N\}$.

We have

$$\mathsf{E}\{N\} \approx (n_1 + 1)\,\mathsf{P}\left\{\chi_\nu^2 < \frac{\delta^2 Z^2 (n_1 + 1)\,\nu}{c^2}\right\} +$$

$$+ \sum_{k=n_1+2}^{\infty} k\,\mathsf{P}\left\{\frac{\delta^2 Z^2 (k-1)\,\nu}{c^2} \leqq \chi_\nu^2 < \frac{\delta^2 Z^2 k\nu}{c^2}\right\},$$

where χ_ν^2, ν and c have the same meaning as in section 3. If we denote by F_ν the distribution function of the χ^2-distribution with ν "degrees of freedom", we may write the expectation $E\{N\}$ as

$$\mathsf{E}\{N\} \approx (n_1 + 1)\, F_\nu(A_{n_1+1}) + \sum_{k=n_1+2}^{\infty} k \int_{A_{k-1}}^{A_k} \mathrm{d}F_\nu(x)\,,$$

where A_k stands for $(\delta^2 Z^2 k\nu)/c^2$. In the second term on the right-hand side we now replace the summands by $\frac{c^2}{\delta^2 Z^2 \nu} \int_{A_{k-1}}^{A_k} x \mathrm{d}F_\nu(x)$, thus introducing a further approximation. We obtain

$$\mathsf{E}\{N\} \approx (n_1 + 1)\, F_\nu(A_{n_1+1}) + \frac{c^2}{\delta^2 Z^2 \nu} \int_{A n_1+1}^{\infty} x \,\mathrm{d}F_\nu(x)\,.$$

Integrating by parts we get finally

$$\mathsf{E}\{N\} \approx (n_1 + 1)\, F_\nu(A_{n1+1}) + \frac{c^2}{\delta^2 Z^2} [1 - F_{\nu+2}(A_{n1+1})]\,. \qquad (10)$$

If the first sample is large, the number ν will usually be large, too, and the expression (10) may be replaced by

$$\mathsf{E}\{N\} \approx (n_1 + 1)\, \Phi(\sqrt{2A_{n_1+1}} - \sqrt{2\nu - 1}) + \frac{c^2}{\delta^2 Z^2} [1 - \Phi(\sqrt{2A_{n_1+1}} - \sqrt{2\nu + 3}),$$
(11)

where Φ denotes the distribution function of the normal distribution with zero mean and unit variance.

Thus the mean number of observations depends on the population standard deviation σ, on the size of the first sample and on the way of its subdivision into subsamples, i. e. on k and n. It is natural to choose the size n_1 and its subdivision so as to make the mean number of observations as small as possible. To this end it is necessary to have some idea of the value of the population standard deviation, to fix the size of subsamples into which the first sample will be divided, to compute $\mathsf{E}\{N\}$ using (10) or (11) for several values of n_1 and to minimize it by trial and error. Numerical comparison shows that the dependence of the expected number of observations upon the method of the subdivision of the first sample is not too strong. Table 3 gives values $\mathsf{E}\{N\}$ corresponding to various choices of k and n for $n_1 = kn = 100$ and to various values of δ, when the risks are $\alpha = 0{\cdot}05$ and $\beta = 0{\cdot}10$. For $\delta = 0{\cdot}4$ the expected number of observations is approximately 101, so that for δ as high as 0·4 the second sample will usually be of size 1, it will consist of a single observation. This means that for $\delta \geq 0{\cdot}4$ the first sample of size 100 is unnecessarily large (in this connection see Table 4 and the corresponding comments). Further it is seen from Table 3 that the expected number of observations is at its minimum value when the subsamples have sizes somewhere between 5 and 15. This is in agreement with the well known fact that the mean range as an estimate of the standard deviation is most efficient when the size of subsamples is about 8.

Table 3.

Expected number of observations for the two-sample procedure with $n_1 = 100$, $\alpha = 0{\cdot}05$, $\beta = 0{\cdot}10$.

Number of subsamples k	Size of subsample n	Difference between hypothesis and alternative in terms of population s. d. σ:					
		$\delta = 0{\cdot}02$	$\delta = 0{\cdot}05$	$\delta = 0{\cdot}10$	$\delta = 0{\cdot}20$	$\delta = 0{\cdot}30$	$\delta = 0{\cdot}40$
5	20	21 954	3 513	878·2	219·5	106·2	101·0
10	10	21 863	3 498	874·5	218·6	102·5	101·0
20	5	21 897	3 503	875·9	219·0	102·7	101·0
50	2	22 234	3 557	889·3	222·3	108·2	101·0
Size of single sample for σ known		21 418	3 427	857	214	95	54

Thus it is sufficient for the determination of the optimal n_1 to compute $\mathsf{E}\{N\}$ for n_1 growing by steps of 10, beginning with n_1 equal to about one half the size of sample necessary for the attainment of the same power with the single-sample test with known standard deviation, and to use subsamples of size 10. In table 3 the sizes for the single-sample test with the same power (the same risks) when the standard deviation is known are shown in the last row.

For illustration which n_1 is approximately optimum we give another table, Table 4, which gives the values $\mathsf{E}\{N\}$ corresponding to various sizes of the first sample, n_1, and various values of δ, when $\alpha = 0{\cdot}05$ and $\beta = 0{\cdot}10$ as before. It is assumed that the first sample is divided into subsamples of size 10. Tables 4 and 3 have been computed by means of formulas (10) and (11).

The last two rows of Table 4 give, respectively, the size of the sample necessary for the single-sample test having the same power $1 - \beta$ against the alternatives specified by δ for the case when σ is known, and for the case when σ is estimated by means of the common mean-square estimate (these values are quoted from [7]). It is seen that the minimal expected number of observations does not exceed much the size necessary for the single-sample test with

Table 4.

Expected number of observations for the two-sample procedure with various sizes of the first sample.

Size of first sample n_1	Difference between hypothesis and alternative in terms of population s. d. σ: δ						
	0·25	0·30	0·35	0·40	0·45	0·50	0·60
10	171·2	118·9	87·4	66·9	52·9	42·9	29·9
20	152·9	106·2	78·0	59·8	47·3	38·5	27·9
30	147·3	102·3	75·2	57·7	46·1	38·6	32·2
40	144·7	100·4	73·8	57·1	47·4	43·0	41·1
50	143·0	99·3	73·3	58·7	52·7	51·2	51·0
60	142·1	98·7	74·1	63·6	61·3	61·0	61·0
70	141·6	98·6	77·2	71·6	71·0	71·0	71·0
80	140·8	99·1	83·1	81·1	81·0	81·0	81·0
90	140·4	101·2	91·3	91·0	91·0	91·0	91·0
100	140·2	105·7	101·1	101·0	101·0	101·0	101·0
120	141·2	121·5	121·0	121·0	121·0	121·0	121·0
150	154·2	151·0	151·0	151·0	151·0	151·0	151·0
Size of single sample for σ known:	137	95	70	54	42	34	24
Size of single sample, σ estimated by mean-square estimate:	139	97	72	55	44	36	26

known σ. Further it is at first sight surprising that no reduction in the number of observations results from the decomposition of the test into two phases in the case of unknown standard deviation. It must be remembered, however, that the described two-sample procedure is devised to satisfy one condition more, viz. the given power against Δ irrespective of σ. If this condition were replaced and the requirement on the power stated only in terms of δ (that is in terms of σ), then, it is believed, it would be possible to find a more economical procedure. This, however, constitutes another problem.

References

[1] *Dantzig, G. B.:* On the Non-Existence of Tests of Student's Hypothesis Having Power Functions Independent of σ, Annals of Mathematical Statistics 11 (1940), p. 186.

[2] *Stein, C.:* A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance, Annals of Mathematical Statistics 16 (1945), p. 243.

[3] *Patnaik, P. B.:* The use of Mean Range as an Estimator of Variance in Statistical Tests, Biometrika 37 (1950), p. 78.

[4] *Pearson, E. S., Hartley, H. O.:* Biometrika Tables for Statisticians, Cambridge University Press 1956.

[5] *Lord, E.:* The Use of Range in Place of Standard Deviation in the t-test, Biometrika 34 (1947), p. 41.

[6] *Janko, J.:* Statistické tabulky, Nakladatelství ČSAV 1958.

[7] *Davies, O. L.:* Design and Analysis of Industrial Experiments, Oliver and Boyd, London 1956.

Souhrn

TEST STUDENTOVY HYPOTHESY DVOJÍM VÝBĚREM PŘI UŽITÍ PRŮMĚRNÉHO ROZPĚTÍ

JOSEF MACHEK

(Došlo dne 9. září 1958.)

V článku je popsána modifikace Steinova dvojfázového testu hypothesy o průměru normálního rozdělení, při které se na místě obvyklého odhadu směrodatné odchylky užívá výběrového rozpětí, resp. průměrného rozpětí několika výběrů.

Podstata úlohy je tato: Je dán základní soubor, jenž má normální rozdělení s neznámou střední hodnotou μ a s neznámou směrodatnou odchylkou σ. Jest ověřiti hypothesu $\mu \leqq \mu_0$ proti alternativě $\mu > \mu_0$. Test má splňovat požadavky (1), kde $P(\mu)$ značí pravděpodobnost, že testovaná hypothesa $\mu \leqq \mu_0$ bude zamítnuta, když skutečná hodnota průměru je μ; Δ je dané číslo (předpokládáme $\Delta > 0$) a α a β jsou zvolená risika chybných rozhodnutí. Požadavky (1) mají být splněny nezávisle na hodnotě směrodatné odchylky σ.

Je dokázáno, že neexistuje test, založený na jediném výběru, který by měl požadované vlastnosti ([1]). Lze však sestrojit test, založený na dvou výběrech, z nichž druhý má rozsah závislý na pozorováních z prvního výběru, který splňuje (1) při jakémkoliv σ ([2]).

Při užití rozpětí na místě obvyklého odhadu směrodatné odchylky zahrnuje test tyto kroky:

1. Vezme se výběr libovolného rozsahu $n_1 = kn$ (kde k a n jsou celá čísla). Výběr je složen z k navzájem nezávislých dílčích výběrů rozsahu n. Vypočte se aritmetický průměr celého prvního výběru, rozpětí každého z dílčích výběrů a průměrné rozpětí $\bar{R}$ všech dílčích výběrů.

2. V závislosti na zvolených hodnotách α a β se určí číslo $Z = \dfrac{1}{z_\alpha + z_\beta}$, kde z_q jsou pro $q = 0{,}05$ a $q = 0{,}10$ uvedena v tabulkách 1 a 2. Provede se

druhý výběr, jehož rozsah n_2, je dán výrazem (2), a vypočte se jeho aritmetický průměr $\bar{x}_2$.

3. Určí se konstanta a řešením rovnice (3).

4. Vypočte se hodnota testové charakteristiky (4) a testovaná hypothesa se zamítne, jestliže $T \geqq z_\alpha$. Kritické hodnoty testové charakteristiky jsou uvedeny v tabulkách 1 a 2.

V § 3 je odvozeno rozdělení testové charakteristiky; charakteristika T má rozdělení jako podíl dvou navzájem nezávislých náhodných veličin, X/Y, kde X má rozdělení normální s parametry *0* a *1* a Y má rozdělení jako průměrné rozpětí k navzájem nezávislých výběrů rozsahu n ze základního souboru s normálním rozdělením s jednotkovým rozptylem.

Při popsaném postupu je rozsah druhého výběru, a tedy i celkový počet pozorování, N, nutných k dosažení rozhodnutí, náhodnou veličinou. Její přibližná střední hodnota je dána výrazy (10) a (11). Závisí na směrodatné odchylce σ základního souboru. Pomocí formulí (10) a (11) byla vypočtena tabulka 4, udávající střední hodnoty počtu pozorování při různých volbách n_1 a při různých hodnotách $\delta = \Delta/\sigma$. Formulí (10) a (11) lze užít při volbě optimálního n_1, tj. takového, při kterém střední hodnota celkového rozsahu výběru je minimální.

Резюме

КРИТЕРИЙ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ СТЬЮДЕНТА ДВОЙНОЙ ВЫБОРКОЙ

ИОСЕФ МАХЕК (Josef Machek)

(Поступило в редакцию 9/IX 1958 г.)

В статье рассматривается следующая задача: Распределение генеральной совокупности нормально с неизвестными параметрами μ (среднее значение) и σ (стандартное отклонение). Требуется проверить гипотезу $\mu \leqq \mu_0$ против альтернативной гипотезы $\mu > \mu_0$. Обозначим через $\mathbf{P}(\mu)$ вероятность непринятия проверяемой гипотезы, если истинное значение среднего равно μ. Критерий для проверки должен удовлетворять требованиям (1), где Δ — заданное число и α и β — заранее выбранные вероятности ошибочных решений, независимо от неизвестного значения параметра σ.

Известно, что при помощи одной выборки эта задача не может быть решена. Для ее решения необходимо воспользоваться двумя выборками, причем объем второй из них является функцией наблюдений, полученных в первой выборке. Этот метод принадлежит Х. Штейну [2]. В настоящей статье метод Штейна видоизменен так, что в качестве оценки неизвестного параметра σ использован размах выборки или средний размах нескольких выборок.

Проверка включает следующие шаги:

1. Из совокупности производят выборку произвольного объема $n_1 = kn$, где k и n — целые числа. Эту выборку разобьют на k частичных выборок объема n и затем подсчитают среднее всей выборки $\bar{x}_1$, размах каждой из частичных выборок и средний размах всех k частичных выборок $\bar{R}$.

2. Возьмут вторую выборку, объем которой n_2 задан формулой (2), где $Z = (z_\alpha + z_\beta)^{-1}$; значения $z_{0,05}$ и $z_{0,10}$ для различных сочетаний k и n приведены в таблицах 1 и 2. По данным второй выборки подсчитают среднее $\bar{x}_2$.

3. Подсчитают значение критерия T, заданного формулой (4), где a есть решение уравнения (3). Гипотеза $\mu \leqq \mu_0$ отвергнута, если $T \geqq z_\alpha$, где α — заранее фиксированная уровень значимости.

В разделе 3 приведено распределение критерия T при гипотезе и при альтернативах.

При только что описанном критерии число наблюдений, необходимых для решения, является случайной величиной. Ее математическое ожидание приблизительно задано формулами (10) или (11), где F_ν обозначает функцию распределения величины χ^2 с ν степенями свободы, и c — постоянная, зависимая от k и n (c и ν табелированы, например, в [4]), $\delta = \Delta/\sigma$, Φ — функция нормального распределения с параметрами 0 и 1. По формулам (10) и (11) была составлена таблица 4, в которой приведены математические ожидания числа наблюдений для некоторых значений δ и n_1. Формулы (10) и (11) могут быть использованы для оптимального выбора объема первой выборки, дающего минимальное среднее число наблюдений.