Measurement Techniques, Vol. 51, No. 9, 2008

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

POWER AND ROBUSTNESS OF CRITERIA USED TO VERIFY THE HOMOGENEITY OF MEANS

B. Yu. Lemeshko and S. B. Lemeshko

UDC 519.24

It is demonstrated that parametric tests of the homogeneity of means is robust with respect to disturbances in the assumption of normality of observations of random variables. The power of parametric and nonparametric tests is investigated.

Key words: *power and robustness of Student's, Mann–Whitney, and Crustal–Wallis tests, F-test.*

Tests to verify the hypothesis of homogeneity of means (homogeneity of mathematical expectations) are resorted to when monitoring measuring instruments and in the statistical analysis of the results of experiments and quality control for checking whether disturbances are present in the course of a process.

In the general case, a given hypothesis of equality of mathematical expectations corresponding to k samples will have the form

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

under the competitive hypothesis

$$H_1 : \mu_{i_1} \neq \mu_{i_2},$$

for at least some pair of indices i_1, i_2.

There are a number of parametric tests that may be used to compare two sample means to check some hypothesis H_0: with known variances; with unknown, but equal variances (Student's test); with unknown and unequal variances; and with the F-test. There also exists a number of nonparametric tests that may be used for this purpose, e.g., the Wilcoxon, Mann–Whitney, and Kruskal–Wallis tests. Membership of the particular sample being analyzed to a normal law is the basic assumption determining whether parametric tests should be used. Nonparametric tests are free of this requirement.

Despite what would appear to be the utter clarity of all the nuances associated with the application of these tests, there are least two points that have not been sufficiently elucidated in the literature. First, it is not clear how important it is to verify whether the samples being analyzed belong to a normal law when parametric tests are used to verify the homogeneity of means. Because of a number of objective and subjective factors, researchers now often resort to verification of the normality of observations, as a consequence of which the potential error of conclusions are also subject to valid criticism. Such a situation in the analysis of biomedical measurements, where samples are encountered that are in good agreement with a normal law is quite typical, though highly questionable. On the other hand, we may note the use of expert judgements of the preferability of nonparametric tests or the lack of any need for verifying normality, for example, when Student's test is used in the case of samples of large volume. The latter point is due to the rather vague data on the power of these tests.

Novosibirsk State Technical University; e-mail: Lemeshko@fpm.ami.nstu.ru. Translated from Izmeritel'naya Tekhnika, No. 9, pp. 23–28, September, 2008. Original article submitted May 19, 2008.

 0543-1972/08/5109-0950

The objective of the present work is to investigate the influence of disturbances in the assumption of normality on the distribution of the statistics of parametric tests and a comparative analysis of the power of the most well-known tests to verify the homogeneity of means.

Tests to Compare Two Sample Means with Known Variances. The use of this comparison test (comparison with respect to two samples) with known and equal variances involves the calculation of the statistic

$$z = (\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2) \sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}\,, \tag{1}$$

where

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

and n_i is the size of the ith sample, $i = 1, 2$.

In the case where the observations (measurement errors) belong to the normal laws, the statistic $z(1)$ obeys a standard normal law.

Student's Test for Comparison of Two Sample Means with Unknown but Equal Variances. With the use of this comparison test, the statistic t is calculated from the expression [1]

$$t = (\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2) \Bigg/ \sqrt{\left[\frac{n_1 + n_2}{n_1 n_2}\right]\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]}, \tag{2}$$

where

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 .$$

If hypothesis H_0 is valid and if the samples belong to a normal law, the (2) statistic will obey a t_ν Student's distribution with number of degrees of freedom $\nu = n_1 + n_2 - 2$.

Test to Compare Two Sample Means with Unknown and Unequal Variances. With unequal volumes of the samples, $n_1 \neq n_2$, the statistic of the test has the form [2, 3]

$$t = (\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2) \sqrt{[s_1^2 / n_1 + s_2^2 / n_2]}\,. \tag{3}$$

In the case of a normal law and where hypothesis H_0 is valid, the (3) statistic will obey a t_ν Student's distribution with number of degrees of freedom

$$\nu = (s_1^2 / n_1 + s_2^2 / n_2)^2 \Bigg/ \left[\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}\right].$$

If the unknown variances are equal, statistics (3) and (2) will be equivalent, while if they are not equal, we will always have the number of degrees of freedom $\nu < n_1 + n_2 - 2$. The greater is the difference between the two variances corresponding to the samples, the stronger will the distribution of the two statistics (3) and (2) differ.

Note that with $n_1 + n_2 > 200$, the difference between tests with statistics (1)–(3) practically vanishes, since with increasing number of degrees of freedom the Student's distribution reduces to a standard normal distribution and the corresponding Student's distributions are practically identical to a standard normal distribution. If a standard normal law is used in such situations to calculate the attainable levels of significance in place of the corresponding Student's distribution, the errors in the calculated probability will not exceed 0.001.

Fisher's Test (*F*-test). In the case where the hypothesis of constancy (equality) of the variances is valid, the hypothesis of the homogeneity of the mathematical expectations over k samples may be verified by means of this test [4].

Suppose there are k samples of volume n. The total sum of squares of the deviations over all samples

$$Q_{kn} = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{kn})^2 ,$$

where $\bar{x}_{kn} = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij} = \bar{\bar{x}}_k$, is decomposed into two components $Q_{kn} = Q_1 + Q_2$, with

$$Q_1 = n \sum_{i=1}^{k} (\bar{x}_{in} - \bar{\bar{x}}_k)^2 = n \sum_{i=1}^{k} (\bar{x}_{in}^2 - k\bar{\bar{x}}_k)^2 ;$$

$$Q_2 = n \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{in})^2 = (n-1) \sum_{i=1}^{k} s_{in}^2 .$$

The component Q_1 is a measure of the difference in the levels of identification between the k samples, whereas Q_2 determines the difference in the levels of identification within these samples. To verify the hypothesis, we use a test with the statistic

$$F = \frac{Q_1 /(k-1)}{Q_2 /[k(n-1)]}. \tag{4}$$

If all the samples are extracted from a normal general population, then if hypothesis H_0 is valid, the (4) statistic will obey a F_{ν_1,ν_2} Fisher distribution with degrees of freedom $\nu_1 = k - 1$, $\nu_2 = k(n - 1)$ [4].

The membership of the samples to a normal distribution is the basic assumption for the tests that have been enumerated here and that are used in the construction of distributions of the statistics in the case where the hypothesis H_0 is to be verified.

Mann and Whitney Test. This ranking test [5–8] is based on the Wilcoxon test [9] for independent samples. It is a nonparametric analog of the t test for comparison of two mean values of continuous distributions. To calculate the statistic, $n_1 + n_2$ values of the combined sample are ordered and the sum of the ranks R_1 corresponding to the elements of the first sample and the sum of the ranks R_2 corresponding to the elements of the second sample are determined, and the following are calculated:

$$U_1 = n_1 n_2 + n_1(n_1 - 1)/2 - R_1;$$

$$U_2 = n_1 n_2 + n_2(n_2 - 1)/2 - R_2.$$

The statistic of the test has the form

$$U = \min\{U_1, U_2\}.$$

Instead of the U statistic, it is more convenient to use

$$\tilde{z} = \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}. \tag{5}$$

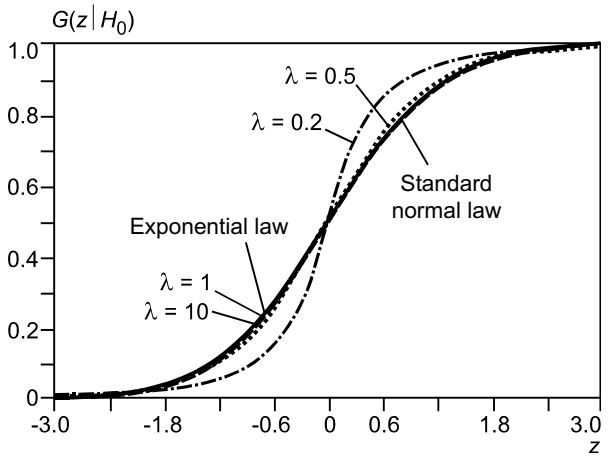

Fig. 1. Empirical distributions of the (1) statistic for different laws of distribution of the observed values, validity of hypothesis H_0, and sample sizes $n_1 = n_2 = 10$.

In the case where hypothesis H_0 is valid, the discrete distribution of the statistic (5) with $n_1 + n_2 > 60$ is approximated quite well by a standard normal law when the size of each of the samples is not too small: $n_1, n_2 \geq 8$. With samples of smaller size, it must be kept in mind that the attainable level of significance (p value) calculated from the value of the statistic in accordance with the distribution function of a standard normal law may differ markedly from the true value.

Kruskal–Wallis Test. This test is a development of the U test for verifying a hypothesis of equality of the means of k samples [10, 11]. A combined sample of size $n = \sum_{i=1}^{k} n_i$ is ordered and the sums of the ranks R_i for the ith sample, $i = \overline{1, k}$ calculated. The statistic for verifying hypothesis H_0 is determined by the expression

$$H = \left[\frac{12}{n(n+1)}\right]\left[\sum_{i=1}^{k} R_i^2 / n_i\right] - 3(n+1), \tag{6}$$

which constitutes the variance of the rank sums. For large n_i and k, this statistic obeys a χ^2_{k-1} distribution in the case where hypothesis H_0 is in fact valid [11]. In these tests, it may be recalled that the χ^2_{k-1} distribution may be used under ordinary circumstances with $n_i \geq 5$, $k \geq 4$.

In fact, with $k = 2$ discreteness is negligible with $n_i \geq 30$. The influence of discreteness rapidly decreases with increasing sample size. If $k = 3$, the distribution of the statistic is approximated quite well by a χ^2_{k-1} distribution, beginning with $n_i \geq 20$, while with $n_i \geq 30$ the agreement of the distribution of a statistic with a χ^2_{k-1} distribution does not deviate with respect to any of the goodness-of-fit tests used [12, 13]. If $k \geq 5$, the agreement with a χ^2_{k-1} distribution does not deviate with $n_i \geq 20$.

Study of the Robustness of Parametric Tests Relative to Disturbance in an Assumed Normality. In conducting these studies, the same technique of computer simulation and analysis of statistical laws was used as in other studies by the present author [12, 13].

The distribution of statistics (1)–(3) under the assumption that hypothesis H_0 is valid was studied for different distribution laws, in particular, in the case where the observations belong to a family with density

$$f(x) = \frac{\lambda}{2\theta_1 \Gamma(1/\lambda)} \exp\left\{-\left(\frac{|x - \theta_0|}{\theta_1}\right)^{\lambda}\right\} \tag{7}$$

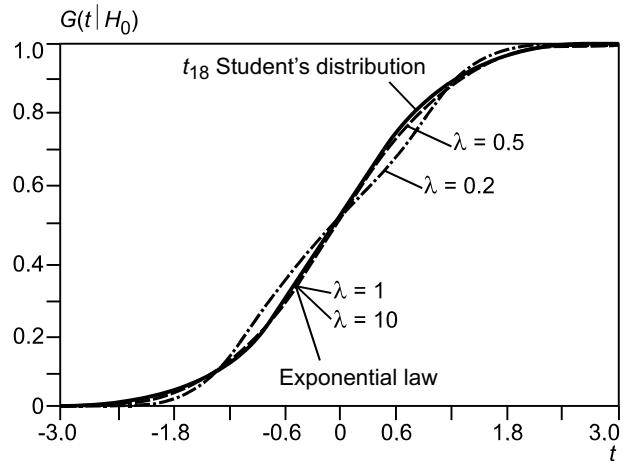

Fig. 2. Empirical distributions of the (2) statistic for different laws of distribution of the observed values, validity of hypothesis H_0, and sample sizes $n_1 = n_2 = 10$.

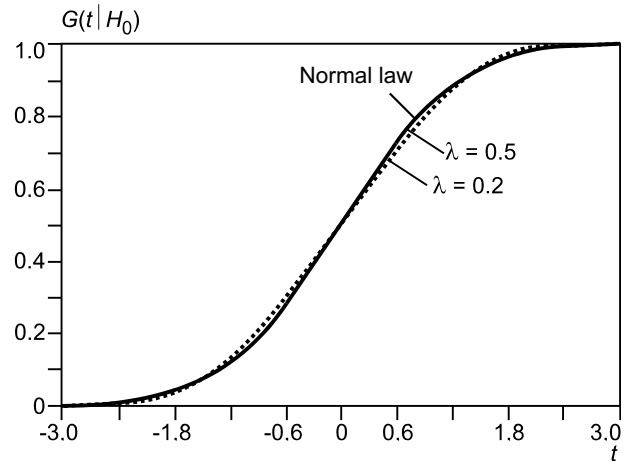

Fig. 3. The same as in Fig. 2, but with sample sizes $n_1 + n_2 = 100$.

with different values of the form parameter λ. With λ = 2, expression (7) yields the density of a normal distribution law. With greater values of λ, the (7) distribution tends to a uniform distribution, while with low λ we obtain symmetric laws with "heavy tails."

Distributions of the (1) statistics obtained by means of simulation, in the case where the observed values belong to the distribution laws of the (7) family for different values of the form parameter and exponential law with density $f(x) = (1/\theta)\exp(-x/\theta)$, are presented in Fig. 1.

The following conclusions may be made on the basis of the results of the studies. Of course, the distribution of the (1) statistic depends on laws to which the samples that are being analyzed belong. Because of the asymmetry of the observed laws, there is a difference between the distribution of the statistic and the standard normal distribution, though this difference is not so great as to lead to major errors in the use of the test. In the case of symmetric laws, it is found that the distribution of the statistic is robust with respect to significant deviations in the observed laws from the normal (right up through a uniform distribution). The distributions of the statistics differ substantially from the standard normal law only for laws with heavy tails, for example, in a Cauchy distribution with λ = 0.5 or λ = 0.2.

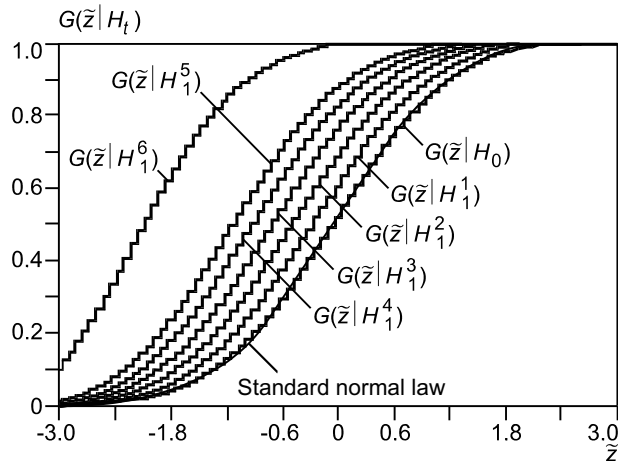

Fig. 4. Empirical distributions of Mann–Whitney statistic (5) where different competing hypotheses are valid and with sample sizes $n_1 = n_2 = 10$.

TABLE 1. Power of Tests Relative to the Alternative H_1^1: $\mu_2 = \mu_1 + 0.1\sigma$

α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
			z test with known variances		
0.10	0.145	0.167	0.186	0.217	0.283
0.05	0.078	0.091	0.105	0.126	0.175
0.01	0.018	0.022	0.026	0.034	0.053
			Student's t test with unknown and equal variances		
0.10	0.144	0.166	0.185	0.216	0.283
0.05	0.077	0.091	0.104	0.125	0.174
0.01	0.017	0.021	0.026	0.034	0.053
			$\tilde{z}$ Mann–Whitney test		
0.10	0.153	0.165	0.184	0.214	0.277
0.05	0.079	0.091	0.101	0.123	0.170
0.01	0.016	0.021	0.024	0.032	0.051
			Fisher F test		
0.10	0.109	0.116	0.125	0.141	0.183
0.05	0.055	0.061	0.067	0.078	0.108
0.01	0.012	0.013	0.015	0.020	0.031
			Kruskal–Wallis H test		
0.10	0.113	0.118	0.123	0.141	0.178
0.05	0.057	0.059	0.066	0.078	0.104
0.01	0.008	0.013	0.015	0.019	0.030

A similar pattern of the dependence of distributions of the (2) statistic on the laws of distribution of the observed values is reflected in Fig. 2 and this again enables us to arrive at identical conclusions regarding the robustness of the Student's test. The distributions of the (3) statistics, which are used in a test with unequal and unknown variances, depend similarly on the observed laws. With increasing size of the samples, the test become even more robust with respect to deviations in the

TABLE 2. Power of Tests Relative to the Alternative H_1^2: $\mu_2 = \mu_1 + 0.2\sigma$

α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
z test with known variances					
0.10	0.202	0.258	0.306	0.389	0.552
0.05	0.115	0.155	0.192	0.259	0.409
0.01	0.030	0.045	0.060	0.092	0.182
Student's t test with unknown and equal variances					
0.10	0.199	0.256	0.304	0.387	0.551
0.05	0.112	0.153	0.190	0.257	0.407
0.01	0.028	0.043	0.059	0.090	0.179
$\tilde{z}$ Mann–Whitney test					
0.10	0.209	0.251	0.299	0.379	0.538
0.05	0.115	0.151	0.184	0.250	0.395
0.01	0.026	0.041	0.054	0.085	0.170
Fisher F test					
0.10	0.131	0.165	0.198	0.261	0.408
0.05	0.071	0.094	0.119	0.168	0.290
0.01	0.017	0.025	0.034	0.056	0.121
Kruskal–Wallis H test					
0.10	0.136	0.164	0.192	0.261	0.394
0.05	0.073	0.091	0.115	0.168	0.278
0.01	0.011	0.023	0.032	0.054	0.113

observed laws from the normal. In Fig. 3, distributions of the (2) statistics with sizes of the samples $n_1 = n_2 = 100$ and where the samples belong to a normal law and to distributions of the (7) family with $\lambda = 0.5$ and 0.2 are shown in Fig. 3.

These results confirm a general law. That is, parametric tests associated with testing a hypothesis of mathematical expectations are very robust with respect to deviations of the observed laws from the normal. This is valid even in the case of multidimensional random variables [14].

The distributions of the (4) statistic of the F test are also robust with respect to deviations in the laws corresponding to the analyzed samples from the normal. However, it should be emphasized that the application of a given test to verify the homogeneity of means presupposes approximate equality of the variances of the samples that are being analyzed. If this condition is not satisfied, the distributions of the statistic $G(F \mid H_0)$ become different from the corresponding F_{ν_1,ν_2} distribution. If the ratio of the maximum variance to the minimum variance of the variances being analyzed does not exceed 4, the deviation in the distribution of this statistic from the Fisher F_{ν_1,ν_2} distribution will not exceed 0.01.

Power of Tests. We considered verification of the homogeneity of the means of two samples. For the tests being studied, the power was analyzed for the case of identical variances of the samples relative to the following alternatives: H_1^1: $\mu_2 = \mu_1 + 0.1\sigma$; H_1^2: $\mu_2 = \mu_1 + 0.2\sigma$; H_1^3: $\mu_2 = \mu_1 + 0.3\sigma$; H_1^4: $\mu_2 = \mu_1 + 0.4\sigma$; H_1^5: $\mu_2 = \mu_1 + 0.5\sigma$; H_1^6: $\mu_2 = \mu_1 + \sigma$. The distributions of the $\tilde{z}$ statistic (5) of the Mann–Whitney test assuming that the verified $G(\tilde{z} \mid H_0)$ and competing $G(\tilde{z} \mid H_1^i)$ hypotheses are valid in the case of sample sizes $n_1 = n_2 = 10$, are presented in Fig. 4. On the other hand, this enables us to judge the power of a test relative to the alternatives that are being considered and, on the other hand, demonstrates the discreteness of the distributions of the statistics, which must be taken into account when comparing the power of different tests.

Estimators of the power $1 - \beta$ of tests calculated on the basis of the resultsof simulation, where β is the probability of an error of the second kind are presented in Tables 1–4 for H_1^1, H_1^2, H_1^5, and H_1^6, respectively, for different values of the level of significance α (probability of an error of the first kind) and different sample sizes. The tests in the tables are ordered

TABLE 3. Power of Tests Relative to the Alternative H_1^5: $\mu_2 = \mu_1 + 0.5\sigma$

α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
z test with known variances					
0.10	0.434	0.618	0.743	0.888	0.988
0.05	0.299	0.475	0.614	0.804	0.971
0.01	0.113	0.228	0.348	0.568	0.887
Student's t test with unknown and equal variances					
0.10	0.424	0.611	0.738	0.886	0.988
0.05	0.285	0.463	0.607	0.799	0.970
0.01	0.099	0.211	0.335	0.556	0.882
$\tilde{z}$ Mann–Whitney test					
0.10	0.430	0.596	0.725	0.875	0.985
0.05	0.283	0.451	0.586	0.781	0.964
0.01	0.090	0.199	0.310	0.529	0.865
Fisher F test					
0.10	0.288	0.464	0.606	0.799	0.963
0.05	0.185	0.338	0.478	0.697	0.929
0.01	0.060	0.144	0.244	0.453	0.801
Kruskal–Wallis H test					
0.10	0.286	0.452	0.586	0.799	0.970
0.05	0.184	0.321	0.457	0.697	0.940
0.01	0.043	0.132	0.227	0.434	0.824

in terms of degree of power. Samples of distributions of the statistics are modeled with sample size $N = 10^6$, which made it possible to estimate the power with error to within $\pm 10^{-3}$. In the situation being considered here, a test with the (3) statistic is equivalent to a test with statistic (2) and has the same power, and therefore is not shown.

The values of the power for multi-sample tests are substantially lower than for tests with statistics (1)–(3) and the Mann–Whitney test (cf. Tables 1–4). In fact, because of the form of their structures, the F test and the Kruskal–Wallis test cannot distinguish between the two alternatives $\mu_2 > \mu_1 + \Delta\mu$ and $\mu_2 < \mu_1 - \Delta\mu$. We obtain an analogous situation if in expressions (1)–(3), (5) the numerator is taken with respect to a modulus. Then the values of the power of these tests for a comparison with multi-sample tests must be taken with significance levels $\alpha/2$.

Let us present some conclusions that may be arrived at on the basis of the results presented in Tables 1–4. First, it is obvious that parametric tests possess greater efficiency than do nonparametric tests. Second, it may be stated that nonparametric tests are absolutely slightly inferior in terms of power to parametric tests, thus, the Mann–Whitney test is inferior to the Student's test, and the Kruskal–Wallis test to the Fisher test, respectively. The apparent advantage of the $\tilde{z}$ test with $n = 10$ as reflected in the tables is explained by the fact that as a consequence of the discrete nature of the distribution of its statistic, the true levels of significance differ from the values of α and slightly exceeds it. This also explains the "advantage" of the H test relative to the F test in certain cases.

And, thirdly, as a rule, only the probability α of an error of the first kind is generally specified in actual practice when these tests are used to verify a hypothesis of homogeneity of the mathematical expectations. The control procedures most often assume small sample sizes. It is not always the case that the concern is whether or not the specification of the probability β of an error of the second kind will alter the the hypothesis being tested when a competing hypothesis proves to be valid. At the same time, it is desirable to ensure that the condition $\beta \le \alpha$ is observed in the test procedure when specifying α. In this case, with a competing hypothesis H_1^1 for $\alpha = 0.1$ and sample sizes $n = 100$, the probability of an error of the

TABLE 4. Power of Tests Relative to the Alternative H_1^6: $\mu_2 = \mu_1 + \sigma$

α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
z test with known variances					
0.10	0.830	0.970	0.995	1.000	1.000
0.05	0.723	0.935	0.987	1.000	1.000
0.01	0.463	0.798	0.939	0.996	1.000
Student's t test with unknown and equal variances					
0.10	0.816	0.967	0.995	1.000	1.000
0.05	0.693	0.927	0.985	1.000	1.000
0.01	0.398	0.764	0.928	0.995	1.000
$\tilde{z}$ Mann–Whitney test					
0.10	0.811	0.961	0.993	1.000	1.000
0.05	0.681	0.917	0.981	0.999	1.000
0.01	0.368	0.739	0.911	0.993	1.000
Fisher F test					
0.10	0.693	0.927	0.985	1.000	1.000
0.05	0.562	0.868	0.967	0.999	1.000
0.01	0.294	0.673	0.882	0.990	1.000
Kruskal–Wallis H test					
0.10	0.680	0.917	0.981	1.000	1.000
0.05	0.548	0.849	0.981	0.999	1.000
0.01	0.231	0.640	0.861	0.987	1.000

second kind is given as $\beta = 1 - 0.283 = 0.717$ for the z test with the statistic (1). If $\alpha = 0.1$, with sample sizes $n = 100$ this test assures a value $\beta = 0.061 < 0.1$ only for the more remote alternative H_1^4, and in order to distinguish with specified quality between the two hypotheses H_0 and H_1^1, samples of size $n \approx 1350$ are needed.

If it is asked which alternatives with the same quality ($\alpha, \beta \le 0.1$) may be distinguished with a sample size $n = 10$, it turns out these are the alternatives in which μ_2 differs from μ_1 by a quantity of at least 1.15σ. Thus, with $n = 20$, if μ_2 differs from μ_1 by a quantity on the order of 0.82σ; with $n = 30$, by a quantity on the order of 0.67σ; with $n = 50$, by a quantity on the order of 0.51σ; and with $n = 100$, by a quantity on the order of 0.364σ.

Thus, the studies have confirmed the robustness of parametric tests for verifying the homogeneity of the mathematical expectations. This means that if the law (laws) of distribution of the samples that are being analyzed differ from the normal, but there is no basis for supposing that the observed quantities belong to laws with heavy tails, it is still correct to apply parametric tests with statistics (1)–(3), or, at least, the use of parametric tests will not lead to serious errors.

If the variances of the samples being analyzed are not known and, possibly, different, it is better to use a test with statistic (3), since in the case of small sample sizes the distribution of the statistic (2) will differ substantially from a $t_{n_1+n_2-2}$ Student's distribution.

With $n_1 + n_2 > 200$, the standard normal law may be used as the distributions for all tests with statistics (1)–(3). The nonparametric analog of tests with these statistics, i.e., the $\tilde{z}$ Mann–Whitney test, is absolutely slightly inferior to the latter in terms of efficiency.

It is best to apply the F test to verify the homogeneity of the mathematical expectations of a series of samples if there is reason for assuming that the variances corresponding to the samples are roughly identical. Otherwise, the F test should be avoided and the Kruskal–Wallis test, which is slightly inferior to the latter in terms of efficiency, should be used.

It should be recalled that besides errors of the first kind there are also errors of the second kind. If the hypothesis which is to be verified for given α has not deviated, this will still not mean that it is valid. In constructing a verification procedure and conjecturing which alternatives must differ, it is necessary to select sample sizes for which β will not less than α.

The present study was completed with the support of the Russian Foundation for Basic Research (Grant No. 06-01-00059).

REFERENCES

1. L. M. Zaks, *Statistical Estimation* [in Russian], Statistika, Moscow (1976).
2. B. L. Welch, *Biometrica*, **34**, 29 (1947).
3. K. Mardia and P. Zemroch, *Tables of F Distributions and Associated Distributions* [Russian translation], Nauka, Moscow (1984).
4. H.-J. Mittag and H. Rinne, *Statistical Methods of Quality Control* [Russian translation], Mashinostroenie, Moscow (1995).
5. N. V. Mann and D. R. Whitney, *Ann. Math. Statist.*, **18**, 50 (1947).
6. R. C. Milton, *J. Amer. Statist. Assoc.*, **59**, 925 (1964).
7. M. Hollander and D. A. Wolfe, *Non-Parametric Statistical Methods*, 2nd ed., Wiley, New York (1999).
8. W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed., Wiley, New York (1999).
9. F. Wilcoxon, *Biometrics Bulletin*, No. 1, 80 (1945).
10. W. H. Kruskal and W. A. Wallis, *J. Amer. Statist. Assoc.*, **47**, 583 (1952).
11. W. H. Kruskal and W. A. Wallis, *J. Amer. Statist. Assoc.*, **48**, 907 (1953).
12. R 50.1.033-2001, *Recommendations on Standardization. Applied Statistics. Rules for Verifying the Agreement of a Test Distribution with a Theoretical Distribution. Part I. Chi-Square Type Tests*.
13. R 50.1.037-2002, *Recommendations on Standardization. Applied Statistics. Rules for Verifying the Agreement of a Test Distribution with a Theoretical Distribution. Part II. Nonparametric Tests*.
14. B. Yu. Lemashko and S. S. Pomadin, *Sibir. Zh. Industr. Matem.*, **5**, No. 3, 115 (2002).