TECHNOMETRICS ©, VOL. 24, NO. 4, NOVEMBER 1982

This paper was presented at the TECHNOMETRICS Session of the 26th Annual Fall Technical Conference of the Chemical Division of the American Society for Quality Control and the Section on Physical and Engineering Sciences of the American Statistical Association in Atlantic City, New Jersey, October 21–22, 1982.

Multiparameter Hypothesis Testing and Acceptance Sampling

Roger L. Berger

The Florida State University
Department of Statistics
Tallahassee, FL 32306

The quality of a product might be determined by several parameters, each of which must meet certain standards before the product is acceptable. In this article, a method of determining whether all the parameters meet their respective standards is proposed. The method consists of testing each parameter individually and deciding that the product is acceptable only if each parameter passes its test. This simple method has some optimal properties including attaining exactly a prespecified consumer's risk and uniformly minimizing the producer's risk. These results are obtained from more general hypothesis-testing results concerning null hypotheses consisting of the unions of sets.

KEY WORDS: Consumer's risk; Producer's risk; Multiple inference.

1. INTRODUCTION

In many situations the quality of a product is determined by several parameters. The product is of acceptable quality to the consumer only if each of the parameters meets certain standards. For example, an upholstery fabric must meet standards for strength, colorfastness, and fire resistance. Based on some measurements on the product, the consumer must decide whether the product is acceptable, that is, all of the parameters meet the standards, or unacceptable, that is, one or more of the parameters do not meet the standards. In making this decision the consumer wishes to use a rule that controls the consumer's risk at a small level.

If there is only one parameter and only one kind of measurement, then a standard quality control text such as Burr (1976) or Duncan (1974) gives methods for making this decision. Different methods are given depending on whether the parameter is a mean, variance, or proportion of defectives and on whether the measurements are counts of defective units (sampling by attributes) or measurements on a continuous variable (sampling by measurements). But no text that the author has found deals with the situation in which there are multiple parameters of interest.

This problem will be formulated as a hypothesis-testing problem in which the null hypothesis states that one or more of the parameters do not meet their standards and the alternative hypothesis states that all of the parameters do meet their standards. Note that rejection of the null hypothesis does not correspond to rejection of the product. Rather, rejection of the null hypothesis corresponds to the decision that the product is acceptable. With this formulation the probability of a Type I error will be the consumer's risk. An α-level test will be one for which the consumer's risk is less than or equal to α.

The test proposed herein is so simple it must not be new. But the author has not been able to find the test described in hypothesis-testing or quality control literature. The test is the following. A hypothesis test is done on each parameter individually at level α. The overall test rejects the null hypothesis and decides that all of the parameters meet their standards if and only if each individual test decides that the individual parameter meets its standard.

This test has several interesting properties. First, the individual tests are performed at level α and yet the overall test has level α. Usually when doing simultaneous inference about many parameters (see, e.g., Miller 1966), inferences about individual parameters must be done with an error rate of less than α to achieve an overall error rate of α. This, for example, is the basis of the Bonferroni method of simultaneous confidence intervals. Second, under very mild conditions, the size of this test is exactly α. So the test is not being too conservative by requiring each of the individual tests to decide that the individual parameter meets its standard. Third, under more restrictive conditions a result of Lehmann (1952) can be used to

prove that this test is uniformly most powerful in a reasonable class of tests. In terms of risks this says that this test uniformly minimizes the producer's risk. These properties indicate that not only is the test extremely easy to implement, since it deals with only one parameter at a time, but it also seems to be a reasonably good test.

2. BASIC RESULTS

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector of observations whose distribution is determined by a vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_\ell')$. Let Θ denote the parameter space. Let Θ_i, $i = 1, \ldots, k$, be subsets of Θ. Let $\Theta_0 = \bigcup_{i=1}^k \Theta_i$. Let A' denote the complement of the set A. Note that $\Theta_0' = \bigcap_{i=1}^k \Theta_i'$. The problem to be considered is that of testing H_0: $\theta \in \Theta_0$ versus H_1: $\theta \in \Theta_0'$. In the example in the introduction Θ_i is the hypothesis that θ_i does not meet its standard. If θ_i must be greater than c_i to meet its standard, then $\Theta_i = \{\boldsymbol{\theta}: \theta_i \le c_i\}$. If Θ_i must be between c_i and d_i to meet its standard, then $\Theta_i = \{\boldsymbol{\theta}: \theta_i \ge d_i \text{ or } \theta_i \le c_i\}$. With this formulation, H_0 is the hypothesis that at least one parameter does not meet its standard and H_1 is the hypothesis that every parameter meets its standard. Note that k, the number of subsets, may be less than ℓ, the number of parameters. This will be the case if some of the parameters are nuisance parameters and do not have standards associated with them.

Let α, $0 \le \alpha \le 1$, be fixed. For $i = 1, \ldots, k$, let $\psi_i(\mathbf{x})$ be an α-level test of H_{0i}: $\boldsymbol{\theta} \in \Theta_i$ versus H_{1i}: $\theta \in \Theta_i'$, that is, $E_{\boldsymbol{\theta}}\psi_i(\mathbf{X}) \le \alpha$ for all $\boldsymbol{\theta} \in \Theta_i$. Let Ψ be the test of H_0 versus H_1 that rejects H_0 if and only if every ψ_i rejects H_{0i}.

Other authors such as Birnbaum (1954, 1955), Lehmann (1952), and Spjøtvoll (1972) have considered testing hypotheses H_0 and H_1. But in all these papers, except the one result of Lehmann to be discussed in Section 3.2, the null hypothesis is of the form H_1.

Tsutakawa and Hewett (1978) propose the test Ψ for a problem comparing regression lines. Tsutakawa and Hewett's test is discussed in Example 4.1. Wilkinson (1951) has proposed a test like Ψ but in a very different situation. He assumes that the individual tests are α-level tests for all of H_0, not just H_{0i}. Wilkinson also assumes the individual tests are independent, which typically will not be the case in the problems considered herein.

The facts that Ψ is always an α-level test and under mild conditions has size exactly α are stated in Theorems 1 and 2. Theorem 2 is proved in the Appendix.

Theorem 1. Ψ is an α-level test of H_0 versus H_1, that is, $E_{\boldsymbol{\theta}}\Psi(\mathbf{X}) \le \alpha$ for all $\boldsymbol{\theta} \in \Theta_0$.

Proof of Theorem 1. Let R_i be the event that ψ_i rejects H_{0i}. Then $R = \bigcap_{i=1}^k R_i$ is the event that Ψ rejects H_0. Fix $\boldsymbol{\theta} \in \Theta_0$. Then $\boldsymbol{\theta} \in \Theta_i$ for some i and

$$E_{\boldsymbol{\theta}}\Psi(\mathbf{X}) = P_{\boldsymbol{\theta}}(R) \le P_{\boldsymbol{\theta}}(R_i) = E_{\boldsymbol{\theta}}\psi_i(\mathbf{X}) \le \alpha$$

since $\boldsymbol{\theta} \in \Theta_i$ and ψ_i is an α-level test of H_{0i}.

Theorem 2. Suppose Θ_i and ψ_i, $i = 1, \ldots, k-1$, satisfy (i), (ii), and (iii).

(i) The standard for θ_i is one-sided, that is, $\Theta_i = \{\boldsymbol{\theta}: \theta_i \le c_i\}$.

(ii) The power of ψ_i depends only on θ_i and $\theta_{k+1}, \ldots, \theta_\ell$.

(iii) There exist numbers b_i (possibly infinite) such that, for any fixed values of $\theta_{k+1}, \ldots, \theta_\ell$, $\lim_{\theta_i \to b_i} E_{\boldsymbol{\theta}}\psi_i(\mathbf{X}) = 1$.

Suppose Θ_k and ψ_k satisfy (iv).

(iv) There exist values $\theta_k^*, \theta_{k+1}^*, \ldots, \theta_\ell^*$ such that, for any values of $\theta_1, \ldots, \theta_{k-1}$, the vector $\boldsymbol{\theta}^* = (\theta_1, \ldots, \theta_{k-1}, \theta_k^*, \ldots, \theta_\ell^*)$ satisfies $\boldsymbol{\theta}^* \in \Theta_k$ and $E_{\boldsymbol{\theta}^*}\psi_k(\mathbf{X}) = \alpha$.

Then Ψ has size exactly α, that is, $\sup_{\boldsymbol{\theta} \in \Theta_0} E_{\boldsymbol{\theta}}\Psi(\mathbf{X}) = \alpha$.

Theorem 1 puts no restrictions on the form of the hypotheses Θ_i. They may be one-sided, two-sided, or any other form. The test Ψ is easily constructed and implemented since the individual tests ψ_i test only one hypothesis Θ_i at a time. This simplicity and the generality of the result in Theorem 1 may be useful in some problems. But the test Ψ might be quite conservative in that the maximum consumer's risk (size) might be much less than α. This might put a heavy burden on the producer.

Theorem 2, on the other hand, gives conditions under which the test Ψ is not too conservative in that the maximum consumer's risk (size) is exactly α. Theorem 2 requires $k-1$ of the standards to be one-sided, that is, θ_i must be greater than c_i to meet its standard (recall Θ_i is the set of parameters for which θ_i does *not* meet its standard). If some of the standards are one-sided but in the opposite direction, that is, θ_i must be less than c_i to meet its standard, then Condition (i) is still satisfied. The model may be reparameterized in terms of $\phi_i = -\theta_i$ and Θ_i may be written as $\{\boldsymbol{\theta}: \phi_i \le -c_i\}$, which is the form for Condition (i). Many standard one-sided tests satisfy Conditions (ii) and (iii) of Theorem 2. These include:

1. one-tailed z test if θ_i is a normal mean and the variance is known,
2. one-tailed t test if θ_i is a normal mean and the variance is a nuisance parameter,
3. one-tailed binomial test if θ_i is the proportion of nondefective items,
4. one-tailed χ^2 test if θ_i is a normal variance and the mean is a nuisance parameter,
5. nonparametric one-tailed tests such as

the Wilcoxon Signed Rank Test or the Sign Test if θ_i is a location parameter (see Hollander and Wolfe 1973 for explicit models.) (2.1)

The other standard Θ_k and test ψ_k may be one-sided or two-sided. If Θ_k is one-sided and ψ_k is one of the tests in the list (2.1), then Condition (iv) is satisfied with $\theta_k^* = c_k$. If Θ_k is two-sided, that is, $\Theta_k = \{\boldsymbol{\theta}: \theta_k \geq d_k \text{ or } \theta_k \leq c_k\}$, then some care must be taken to construct a test ψ_k that satisfies Condition (iv). This is not the usual two-sided test since here the alternative hypothesis, not the null hypothesis, is the interval (c_k, d_k). A description of how a t test for a normal mean can be constructed, which satisfies Condition (iv) in the two-sided case, will be given in Example 4.2.

3. THE PRODUCER'S RISK

The test Ψ controls the maximum consumer's risk at α. In this section a further examination of the producer's risk is made. This is done in two ways. First, an example that, although simple, displays typical behavior for the producer's and consumer's risks is examined. Then a theoretical result of Lehmann's (1952), which implies that in certain cases Ψ has the best possible producer's risk, is discussed.

3.1 A Typical Power Function

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent, identically distributed bivariate normal observations with mean vector (θ_1, θ_2) (unknown), variances σ_1^2 and σ_2^2 (known), and correlations $\rho = .7$. Let $\Theta_1 = \{\boldsymbol{\theta}: \theta_1 \leq c_1\}$ and $\Theta_2 = \{\boldsymbol{\theta}: \theta_2 \leq c_2\}$. This means θ_1 must exceed c_1 to meet its standard and θ_2 must exceed c_2 to meet its standard. Suppose the consumer wishes to have a maximum consumer's risk of $\alpha = .05$. The standard normal theory z tests for testing Θ_1 and Θ_2 are these. ψ_1 rejects Θ_1 if $\bar{X} > c_1 + 1.645\, \sigma_1/\sqrt{n}$ and ψ_2 rejects Θ_2 if $\bar{Y} > c_2 + 1.645\, \sigma_2/\sqrt{n}$.

Table 1. *Power Function of Ψ for Example 3.1*

$\frac{(\theta_2-c_2)}{\sigma_2/\sqrt{n}}$ \ $\frac{(\theta_1-c_1)}{\sigma_1/\sqrt{n}}$	$-\infty$	-1.0	-.5	0	.5	1	2	3	4	5	∞
∞	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
-1.0	.000	.001	.002	.003	.004	.004	.004	.004	.004	.004	.004
-.5	.000	.002	.005	.009	.013	.015	.016	.016	.016	.016	.016
0.0	.000	.003	.009	.020	.033	.043	.050	.050	.050	.050	.050
.5	.000	.004	.013	.033	.063	.093	.123	.126	.126	.126	.126
1.0	.000	.004	.015	.043	.093	.159	.245	.258	.259	.259	.259
2.0	.000	.004	.016	.050	.123	.245	.520	.629	.635	.639	.639
3.0	.000	.004	.016	.050	.126	.258	.629	.864	.909	.912	.912
4.0	.000	.004	.016	.050	.126	.259	.635	.909	.983	.991	.991
5.0	.000	.004	.016	.050	.126	.259	.639	.912	.991	.999	1.000
∞	.000	.004	.016	.050	.126	.259	.639	.912	.991	1.000	1.000

Some values of the power function of the test Ψ, which decides both parameters meet their standards if and only if ψ_1 and ψ_2 both reject, are listed in Table 1. The values above or left of the dotted line are values of the consumer's risk for various parameter values. The values of the power function below and to the right of the dotted line are equal to one minus the producer's risk for various parameter values. Large values of the power function correspond to small producer's risks and vice versa.

The following characteristics of the risks, which can be noted in Table 1, are typical of the more general situation in which all of the tests $\psi_1, \ldots, \psi_k$ are chosen from the list (2.1). (In the following remarks, terms like "much better" or "much worse" must be interpreted relative to the associated standard deviations.)

Remark 3.1. Only the marginal distributions of $\bar{X}$ and $\bar{Y}$, not the joint distribution of $(\bar{X}, \bar{Y})$, were considered in constructing ψ_1 and ψ_2. In general, only the marginal distribution of ψ_i need be considered to insure that ψ_i is an α-level test of Θ_i. The joint distribution of the test statistics need not be considered, or even known, to construct the test Ψ. This contributes to the simplicity of Ψ.

Remark 3.2. The maximum consumer's risk of α occurs when θ_1 is much better than its standard and θ_2 just fails to meet its standard or vice versa. In general, as the proof of Theorem 2 shows (see Appendix), the maximum consumer's risk occurs when all but one of the parameters are much better than their standards and the one remaining parameter just fails to meet its standard.

Remark 3.3. If either θ_1 or θ_2 is much worse than its standard, then the consumer's risk is near zero. In general, if at least one of the parameters is much worse than its standard, then the consumer's risk is near zero.

Remark 3.4. The maximum producer's risk occurs when both θ_1 and θ_2 are just slightly better than their standards. In general, the maximum producer's risk occurs when all of the parameters just slightly exceed their standards. If each θ_i is only slightly better than its standard, the probability that ψ_i rejects Θ_i is at most slightly larger than α. The probability that all the ψ_i reject is at most slightly larger than α and is likely to be much less than α. If the tests ψ_i are all independent, then the maximum producer's risk is $1 - \alpha^k$.

Remark 3.5. The producer's risk is near zero only if θ_1 and θ_2 are both much better than their standards. In general, the producer's risk will be near zero only if all of the parameters are much better than their standards.

These observations indicate that the test Ψ provides very good protection for the consumer and presents the producer with a burden of proof. Remark 3.3 indicates that even if only one parameter is far below its standard, the consumer's risk will be near zero. On the other hand, all the parameters must far exceed their standards for producer's risk to be near zero (Remark 3.5). But this problem can be overcome by the use of large samples. As the sample size increases, the amount by which θ_i must exceed its standard for the producer's risk to be acceptably small decreases. Remark 3.4 indicates that the producer's risk will be large if each parameter barely exceeds its standard. Yet the result discussed in Section 3.2 indicates that the producer's risk is as small as can be expected.

A test that is based on the Bonferroni Inequality and that controls the *producer's* risk at α can be constructed in this way. Let H_0': $\theta_1 > c_1$ and $\theta_2 > c_2$. Each of the hypotheses, H_{0i}': $\theta_i > c_i$, could be tested at level $\alpha/2$. H_0' would be rejected if either H_{0i}' were rejected. Then the producer's risk is

$$P(\text{reject } H_0' \mid H_0') \le P(\text{reject } H_{01}' \mid H_0') + P(\text{reject } H_{02}' \mid H_0') \le \alpha.$$

In this normal example, for $\alpha = .05$, H_{01}' would be rejected if $\bar{X} < c_1 - 1.96\sigma_1/\sqrt{n}$ and similarly for H_{02}'. This test is related to the test Ψ and, in fact, Theorem 2 shows that the maximum consumer's risk is exactly $1 - \alpha/2$. The test that decides that $\theta_1 \le c_1$ if $\bar{X} < c_1 - 1.96\sigma_1/\sqrt{n}$ and that $\theta_1 > c_1$ if $\bar{X} > c_1 - 1.96\sigma_1/\sqrt{n}$ is a test of H_{01}: $\theta_1 \le c_1$ versus H_{11}: $\theta_1 > c_1$ with size $1 - \alpha/2$. The same holds true for the test based on $\bar{Y}$. Thus, by Theorem 2, the maximum consumer's risk is $1 - \alpha/2$. In general, if the k individual tests in the Bonferroni test satisfy the conditions of Theorem 2, then the maximum consumer's risk is exactly $1 - \alpha/k$. One would choose between the Bonferroni test and Ψ depending on whether one wished to control the producer's risk or the consumer's risk.

3.2. An Optimality Result

Despite facts like Remarks 3.4 and 3.5, which indicate that the producer's risk may be high even if all the parameters exceed their standards, Theorem 4.2 of Lehmann (1952) indicates that the test Ψ uniformly minimizes the producer's risk in a reasonable class of tests. Lehmann's result will now be informally described.

Suppose all the standards are one-sided, that is, $\Theta_i = \{\boldsymbol{\theta}: \theta_i \le c_i\}$. Suppose each test ψ_i is of the form "reject Θ_i if $Y_i > d_i$" for an appropriate test statistic Y_i. Suppose each ψ_i has size exactly α and satisfies (ii) and (iii) of Theorem 2. The joint distribution of $(Y_1, \ldots, Y_k)$ must satisfy a condition that roughly requires that as all the parameters $\theta_1, \ldots, \theta_k$ increase, all the test statistics tend to increase. A rejection region is monotone if $(y_1, \ldots, y_k)$ is in the rejection region and all the coordinates of $(z_1, \ldots, z_k)$ satisfy $z_i \ge y_i$ imply $(z_1, \ldots, z_k)$ is in the rejection region. Then Lehmann's result asserts that among all level-α tests with monotone rejection regions that are based on these test statistics, Ψ uniformly minimizes the producer's risk.

The conditions of Lehmann's theorem will be satisfied, for example, if all the tests ψ_i are from the list (2.1) and are independent. But the conditions will also be satisfied in many cases when the test statistics are correlated. For example, if for each test either θ_i is a location parameter for Y_i or θ_i is a scale parameter for Y_i and Y_i is positive, then the conditions of Lehmann's theorem will be met, regardless of the correlations between the test statistics. Most dispersion tests fall into the scale parameter framework.

4. EXAMPLES

Two examples of the use of the test Ψ are presented in this section. In the first example, a regression problem examined by Tsutakawa and Hewett (1978) is considered. The test they propose is the test Ψ. The second example considers an acceptance sampling problem from the textile industry.

Example 4.1. This problem was considered by Tsutakawa and Hewett (1978). Let $Y_{ij} = \alpha_i + \beta_i X_{ij} + \varepsilon_{ij}$, $i = 1, 2; j = 1, \ldots, n_i$. $\alpha_1, \beta_1, \alpha_2$, and β_2 are unknown parameters; the X_{ij} are known constants; and the ε_{ij} are independent identically distributed normal random variables with mean zero and variance σ^2. It is desired to compare the two regression lines $\alpha_1 + \beta_1 X$ and $\alpha_2 + \beta_2 X$ on the finite interval $I = [X_*, X^*]$. Specifically, we desire a test for H_0: $\alpha_1 + \beta_1 X \le \alpha_2 + \beta_2 X$ for some $X \in I$ versus H_1: $\alpha_1 + \beta_1 X > \alpha_2 + \beta_2 X$ for all $X \in I$. Letting $\theta_1 = \alpha_1 + \beta_1 X_* - \alpha_2 - \beta_2 X_*$ and $\theta_2 = \alpha_1 + \beta_1 X^* - \alpha_2 - \beta_2 X^*$, H_0 and H_1 can be written as H_0: $\theta_1 \le 0$ or $\theta_2 \le 0$ versus H_1: $\theta_1 > 0$ and $\theta_2 > 0$. Letting $\Theta_1 = \{\boldsymbol{\theta}: \theta_1 \le 0\}$ and $\Theta_2 = \{\boldsymbol{\theta}: \theta_2 \le 0\}$, H_0 and H_1 are seen to be of the form considered in Theorems 1 and 2. To derive the test Ψ, it is necessary to derive the tests ψ_1 for Θ_1 and ψ_2 for Θ_2. Let a_1, b_1, a_2, and b_2 denote the least squares estimates of $\alpha_1, \beta_1, \alpha_2$, and β_2. Then the test statistic $T_1 = (a_1 + b_1 X_* - a_2 - b_2 X_*)/S\sqrt{\sigma_{11}}$ has a t distribution with $\nu = n_1 + n_2 - 4$ degrees of freedom if $\theta_1 = 0$, where

$$S^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - a_i - b_i X_{ij})^2/(n_1 + n_2 - 4)$$

and

$$\sigma_{11} = \sum_{i=1}^{2} \frac{\sum_{j=1}^{n_i} X_{ij}^2/n_i - 2X_* \bar{X}_i + X_*^2}{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}.$$

(See Tsutakawa and Hewett 1978 for details of the derivation.) Thus a size α test of Θ_1 versus Θ_1' rejects Θ_1 if $T_1 > t_{1-\alpha}(v)$, the $1-\alpha$ percentile of the t distribution with $v = n_1 + n_2 - 4$ degrees of freedom. This test is ψ_1. The test ψ_2, a size α test of Θ_2 versus Θ_2', rejects Θ_2 if $T_2 > t_{1-\alpha}(v)$, where $T_2 = (a_1 + b_1 X^* - a_2 - b_2 X^*)/S\sqrt{\sigma_{22}}$ and σ_{22} is the same as σ_{11} except X^* replaces X_*. The test Ψ then rejects H_0: $\theta_1 \le 0$ or $\theta_2 \le 0$ and accepts H_1: $\theta_1 > 0$ and $\theta_2 > 0$ only if $T_1 > t_{1-\alpha}(v)$ and $T_2 > t_{1-\alpha}(v)$. This is the test proposed by Tsutakawa and Hewett. The fact that Ψ has size exactly α follows from Theorem 2 since both ψ_1 and ψ_2 are of the form of 2 in (2.1). The Tsutakawa and Hewett test is an example of how the results of Theorems 1 and 2 yield a size α test when H_0 can be conveniently written as the union of two or more sets.

Example 4.2. An example of the use of the test Ψ in an acceptance sampling situation will be taken from the textile industry. Table 2 lists specifications for upholstery fabric from the American Society for Testing and Materials. The specifications give standards for nine parameters related to strength, colorfastness, flammability, and dimensional stability.

The first three parameters might be assumed to be normal means; mean breaking strength, mean tear strength and mean abrasion resistance. Each standard says the mean must be greater than some value. Each of these standards could be tested with a one-sided t test. For example, ψ_1 would be the size α t test of H_{01}: $\theta_1 \le 50$ versus H_{11}: $\theta_1 > 50$.

The next five parameters, colorfastness and flammability, might be measured by binomial variables, each variable counting the number of units in a sample that pass the corresponding test. Each parameter would then be the proportion of units in the population (a particular manufacturer's output) that achieve one of the standards. The usual one-sided binomial test could be used to test each of these five hypotheses. For example, suppose θ_4 is the proportion of the manufacturer's output that will attain a grade of class four or above in the water colorfastness test. Suppose that at least 90 percent of the output must attain this grade for the upholstery to be acceptable. Then X_4 would be the binomial variable that counts the number of units in a sample that attain a class four grade or better. ψ_4 would be the size α binomial test of H_{04}: $\theta_4 \le .90$ versus H_{04}: $\theta_4 \le .90$ versus H_{14}: $\theta_4 > .90$ based on X_4.

Table 2. *Standard Specification for Woven Upholstery Fabric—Plain, Tufted, or Flocked*

Test	Minimum Standard
Breaking Strength	50 pounds
Tong Tear Strength	6 pounds
Surface Abrasion (heavy duty)	15,000 cycles
Colorfastness to:	
Water	class 4
Crocking	
Dry	class 4
Wet	class 3
Light-40 AATCCF Fading Units	class 4
Flammability	Pass
Dimensional Change	5% shrinkage, 2% gain

Source: 1978 Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, Part 32, page 717.

The last parameter θ_9 might be assumed to be a normal mean, mean percentage change. The standard is two-sided. Thus ψ_9 must be a size α test of H_{09}: $\theta_9 \ge .02$ or $\theta_9 \le -.05$ versus H_{19}: $-.05 < \theta_9 < .02$. If an upper bound for the variance can be assumed, say $\sigma^2 \le \sigma_0^2$, then ψ_9 can be constructed in this way. Let $X_1, \ldots, X_m$ be independent identically distributed normal random variables with mean θ, and variance σ^2. Assume $\sigma^2 \le \sigma_0^2$. Let $\bar{X}$ and S^2 denote the sample mean and variance. Consider testing H_0: $\theta \ge d$ or $\theta \le c$ versus H_1: $c < \theta < d$. Let $T = (\bar{X} - \theta_0)/(S/\sqrt{m})$, where $\theta_0 = (c+d)/2$. T has a noncentral t distribution with noncentrality parameter $\delta = \sqrt{m}(\theta - \theta_0)/\sigma$ and $m-1$ degrees of freedom. Let $\delta_0 = \sqrt{m}(d-c)/2\sigma_0$. Let a be the number that satisfies $P(-a < T_{\delta_0} < a) = \alpha$, where T_{δ_0} has a noncentral t distribution with noncentrality parameter δ_0 and $m-1$ degrees of freedom. Then the test that rejects H_0 if $-a < T < a$ is a size α test of H_0 versus H_1. This is true since $P(-a < T_\delta < a)$ is a decreasing function of $|\delta|$ and δ_0 is the smallest possible value of $|\delta|$ under H_0. In our example, $d = .02$, $c = -.05$. Assume $m = 9$ and $\sigma \le .05$. Then $\delta_0 = 2.1$. If $\alpha = .10$ is used, then the value of a satisfying $P(-a < T_{2.1} < a) = .10$ is $a = .8300$. Thus, the test ψ_9 rejects H_{09} if $-.8300 < T < .8300$. If $\alpha = .05$ is used, then $a = .5055$ and ψ_9 rejects H_{09} if $-.5055 < T < .5055$. These values of a were computed using the program MDTN of the IMSL library to compute noncentral t probabilities. The values of a could be approximated using tables of noncentral t probabilities such as those of Locks, Alexander, and Byars (1963).

Since the first eight tests are from the list (2.1) and the ninth test has the probability of a Type I error, exactly α when $\theta_9 = .02$ and $\sigma = .05$, Theorem 2 guarantees that the test Ψ obtained by combining these nine tests has size exactly α. It would be very difficult to posit a realistic multivariate model for the observations on these variables. Some are discrete and some are continuous. Some are likely to be correlated. Yet it is relatively easy to construct a size α test for

each parameter individually. These nine tests can be combined into the overall test Ψ to test the hypothesis H_0 with a maximum consumer's risk of exactly α. The upholstery will be deemed satisfactory only if each of the nine tests rejects its hypothesis at level α.

5. CONCLUSIONS

Acceptance sampling procedures for individual parameters are well known. This article proposes a way of combining these procedures in the situation in which the quality of a product is measured by standards on several parameters. Not only is the method easy to implement, but it controls the consumer's risk at exactly a preassigned level in typical situations when the standards are one-sided (either upper or lower bounds). Under slightly more restrictive conditions, this method also uniformly minimizes the producer's risk. This method can be used in hypothesis-testing problems other than acceptance sampling if the null hypothesis is a union of sets.

ACKNOWLEDGMENT

This research was supported by the U.S. Army Research Office Grant No. DAAG 29-79-C-0158.

The values in Table 1 were computed using a computer program written by Professor Charles E. McCulloch, Florida State University. The author gratefully acknowledges this assistance. The author also wishes to thank the referees and editor for comments that improved the presentation of this article.

APPENDIX

Proof of Theorem 2. Let R_i and R be defined as in the proof of Theorem 1. Let $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{k-1i}, \theta_k^*, \theta_{k+1}^*, \ldots, \theta_\ell^*)$, $i = 1, 2, \ldots$ be a sequence of parameter points satisfying $\theta_{ji} \to b_j$, $j = 1, \ldots, k-1$, as $i \to \infty$. Then $\boldsymbol{\theta}_i \in \Theta_k$ for all i and $P_{\boldsymbol{\theta}_i}(R_k') = 1 - E_{\boldsymbol{\theta}_i}\psi_k(\mathbf{X}) = 1 - \alpha$ for all i. Also, for $j = 1, \ldots, k-1$, $\lim_{i\to\infty} P_{\boldsymbol{\theta}_i}(R_j') = 1 - \lim_{i\to\infty} P_{\boldsymbol{\theta}_i}(R_j) = 1 - 1 = 0$. Therefore,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} E_{\boldsymbol{\theta}}\Psi(\mathbf{X}) &\geq \lim_{i\to\infty} E_{\boldsymbol{\theta}_i}\Psi(\mathbf{X}) \\
&= \lim_{i\to\infty} P_{\boldsymbol{\theta}_i}\left(\bigcap_{j=1}^{k} R_j\right) \\
&= \lim_{i\to\infty}\left(1 - P_{\boldsymbol{\theta}_i}\left(\bigcup_{j=1}^{k} R_j'\right)\right) \\
&\geq 1 - \lim_{i\to\infty}\sum_{j=1}^{k} P_{\boldsymbol{\theta}_i}(R_j') \\
&= 1 - \lim_{i\to\infty}\left((1-\alpha) + \sum_{j=1}^{k-1} P_{\boldsymbol{\theta}_i}(R_j')\right) \\
&= 1 - (1-\alpha) - 0 = \alpha.
\end{aligned}$$

From Theorem 1, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} E_{\boldsymbol{\theta}}\Psi(\mathbf{X}) \leq \alpha$. Thus the size of Ψ is exactly α.

[*Received October 1980. Revised February 1982.*]

REFERENCES

BIRNBAUM, ALLEN (1954) "Combining Independent Tests of Significance," *Journal of the American Statistical Association*, 49, 559–574.

——— (1955) "Characterizations of Complete Classes of Some Multiparameter Hypotheses, With Applications to Likelihood Ratio Tests," *Annals of Mathematical Statistics*, 26, 21–36.

BURR, IRVING W. (1976) *Statistical Quality Control Methods*, New York: Marcel Dekker.

DUNCAN, ACHESON J. (1974) *Quality Control and Industrial Statistics* (4th ed.), Richard D. Irwin, Inc., Homewood, Illinois.

HOLLANDER, MYLES, and WOLFE, DOUGLAS A. (1973) *Nonparametric Statistical Methods*, New York: John Wiley.

LEHMANN, E. L. (1952) "Testing Multiparameter Hypotheses," *Annals of Mathematical Statistics*, 23, 541–552.

LOCKS, M. O., ALEXANDER, M. J., and BYARS, B. J. (1963), *New Tables of the Noncentral t Distribution*, Canoga Park, Calif.: Rocketdyne, A Division of North American Aviation.

MILLER, RUPERT G. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill.

SPJØTVOLL, EMIL (1972), "On the Optimality of Some Multiple Comparison Procedures," *Annals of Mathematical Statistics*, 43, 398–411.

TSUTAKAWA, R. K., and HEWETT, J. E. (1978), "Comparison of Two Regression Lines Over a Finite Interval," *Biometrics*, 34, 391–398.

WILKINSON, B. (1951), "A Statistical Consideration in Psychological Research," *Psychological Bulletin*, 48, 156–157.