

NEW APPROXIMATE INFERENTIAL METHODS FOR THE RELIABILITY
PARAMETER IN A STRESS-STRENGTH MODEL: THE NORMAL CASE

Huizhen Guo and K. Krishnamoorthy
Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504-1010
krishna@louisiana.edu

Key Words: explanatory variables; generalized p-value; lognormal distribution; noncentral t distribution; parametric bootstrap; quantile; Satterthwaite approximation; size; tolerance limits.

ABSTRACT

This article considers the problem of testing and interval estimation of the reliability parameter $P(Y_1 > Y_2)$, where Y_1 and Y_2 are independent normal random variables with unknown means and variances. A new approximate method is proposed, and is compared with two existing approximate methods and a generalized variable method. Simulation studies indicate that the sizes of the existing approximate tests exceed the nominal level considerably in some situations while the new method controls the sizes satisfactorily in all situations considered. The studies also indicated that the generalized variable test is too conservative for small samples. A confidence limit for the reliability parameter based on the new approximate test is also given. The results are extended to the case where Y_1 and Y_2 depend on some explanatory variables. The methods are illustrated using three examples, and some recommendations are made regarding what method should be used for practical applications.

1. INTRODUCTION

The classical stress-strength reliability problem concerns the proportion of the times the strength Y_1 of a component exceeds the stress Y_2 to which it is subjected. If $Y_1 \leq Y_2$, then either the component fails or the system that uses the component may malfunction. If both Y_1 and Y_2 are random, then the reliability R of the unit can be expressed as $R = P(Y_1 > Y_2)$, where P denotes the probability. To assess the reliability of the component, inference on R is desired. This reliability problem arises in the areas of aeronautical, civil, mechanical and nuclear engineering. Some specific examples are as follows: Hall (1) provided an example of a system application where the breakdown voltage Y_1 of a capacitor must exceed the voltage output Y_2 of a transverter (power supply) in order for a component to work properly. Guttman et. al. (2) presented a rocket-motor experiment data where Y_1 represents the chamber burst strength and Y_2 represents the operating pressure. These two examples along with another example are illustrated in Section 3 of this article.

In this article, we consider some approximate inferential procedures for R when Y_1 and Y_2 are independent normal random variables with unknown means and variances. Relevant references for the present problem can be found in the papers by Reiser and Guttman (3) and Weerahandi and Johnson (4). To formulate the present problem, let $Y_1 \sim N(\mu_1, \sigma_1^2)$ independently of $Y_2 \sim N(\mu_2, \sigma_2^2)$. Then the reliability parameter R can be expressed as

$$R = P(Y_1 > Y_2) = \Phi \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right), \quad (1.1)$$

where Φ is the standard normal distribution function. Let us consider the problem of testing

$$H_0 : R \leq R_0 \text{ vs } H_a : R > R_0, \quad (1.2)$$

where R_0 is a specified probability which is usually close to 1. Let z_{R_0} denote the R_0 th quantile of the standard normal distribution. Then

$$\eta = \mu_1 - \mu_2 - z_{R_0} \sqrt{\sigma_1^2 + \sigma_2^2}. \quad (1.3)$$

is the $(1 - R_0)$ th quantile of the distribution of $Y_1 - Y_2$. Note that $R > R_0$ if and only if $\eta > 0$, and hence the hypotheses in (1.2) are equivalent to

$$H_0 : \eta \leq 0 \text{ vs } H_a : \eta > 0. \quad (1.4)$$

The null hypothesis in (1.4) is rejected at the nominal level α whenever a lower $1 - \alpha$ confidence limit η_L for η is greater than 0. The lower limit η_L is referred to as the R_0 content - $(1 - \alpha)$ coverage one-sided lower tolerance limit for the distribution of $Y_1 - Y_2$. That is, the interval (η_L, ∞) would contain at least R_0 proportion of data from the distribution of $Y_1 - Y_2$ with confidence level $1 - \alpha$. Thus, a test for the reliability parameter R can be obtained using a one-sided tolerance limit for the distribution of $Y_1 - Y_2$. The problem of constructing one-sided tolerance limit for the distribution of $Y_1 - Y_2$ has been addressed in

Hall (1). Specifically, Hall developed an exact one-sided tolerance limit when the variance ratio $q_1 = \sigma_1^2/\sigma_2^2$ is known, and an approximate limit when q_1 is unknown. Although, Hall has argued that the two problems are related, the above description is simple and clearly points out the relation. The theoretical developments in Hall (1) can be used to construct lower confidence limits for R in (1.1). In particular, the results given in Reiser and Guttman (3) and Guttman et. al. (2) can be easily deduced from the results in Hall (1).

The approximate methods suggested in the above articles are essentially based on the well-known Aspin-Welch approximate test (see Casella and Berger (5), p. 409) for the Behrens-Fisher problem. Let \bar{x}_i and s_i^2 denote respectively the mean and variance of a sample of n_i observations from $N(\mu_i, \sigma_i^2)$ population, $i = 1, 2$. Hall's approximate tolerance limit is obtained by first developing a limit which depends on the variance ratio q_1 , and then replacing q_1 by a suitable estimate. Hall recommended using the unbiased estimator $\hat{q}_1 = (n_2 - 3)s_1^2/((n_2 - 1)s_2^2)$ whereas, in the problem of reliability estimation, Reiser and Guttman (3) suggested the usual estimator s_1^2/s_2^2 . Although, the procedures based on these two estimates may not differ significantly for large n_2 , they do differ for small n_2 . See the sizes of the tests based on these two estimates in Figures 1-3 (H and R-G tests). Guttman et. al. (2) extended the results of Reiser and Guttman (3) to the case where Y_1 and Y_2 depend on some covariates. Weerahandi and Johnson (4) proposed inferential procedures for the reliability parameter R based on the generalized p-value approach. As claimed by Weerahandi and Johnson (4), the sizes of the generalized p-value test are very close to the nominal level, and it performs better than the approximate test due to Church and Harris (6) when Y_1 and Y_2 are normal and the distribution of Y_2 is known. However, the generalized p-value test is approximate, often conservative, when the distribution of Y_2 is unknown (see Test W-J in Figures 1 and 3).

It is clear from the above discussion that the Hall's (1984) approach depends on the labeling of the populations. Specifically, if the variance ratio is defined as σ_2^2/σ_1^2 , and is estimated by the unbiased estimator $(n_1 - 3)s_2^2/((n_1 - 1)s_1^2)$, then the resulting test will be different from the one due to Hall. The two tests may be combined to overcome the label dependency. In this article, we propose a new approximate test that is obtained by combining the two tests.

This article is organized as follows. For the sake of completeness and easy reference, we present in Section 2 the exact test for the reliability parameter based on the Hall's approach when the variance ratio is known. In Section 3, we describe the existing approximate tests, the new test mentioned in the preceding paragraph, and the generalized variable test due to Weerahandi and Johnson (4). The tests are compared with respect to the size using the Monte Carlo simulation. The simulation studies indicate that the new test performs better than the existing tests, and behaves like an exact test when the sample sizes are equal and the common sample size is 15 or more. Furthermore, we observed from our simulation studies that the generalized variable test is often very conservative when sample sizes are moderate and close to each other. Interval estimates for R based on the approximate tests are given. Construction of the tests and interval estimates are illustrated using three examples. In Section 4, we extend the results of Section 3 to the reliability problem when Y_1 and Y_2 have

covariates. Some concluding remarks are given in Section 5.

2. AN EXACT TEST FOR R WHEN $q_1 = \sigma_1^2/\sigma_2^2$ IS KNOWN

We shall now develop an exact test for the reliability parameter R using the Hall's approach for constructing tolerance limits for $Y_1 - Y_2$. Letting

$$s_d^2 = (1 + q_1^{-1})((n_1 - 1)s_1^2 + (n_2 - 1)q_1 s_2^2)/(n_1 + n_2 - 2), \quad (2.1)$$

it can be shown that $s_d^2/(\sigma_1^2 + \sigma_2^2) \sim \chi_{n_1+n_2-2}^2/(n_1 + n_2 - 2)$, where χ_b^2 denotes the chi-square random variable with $df = b$. Using this result and standardizing $\bar{y}_d = \bar{y}_1 - \bar{y}_2$, we see that

$$\frac{\bar{y}_d - \eta}{s_d} = \frac{Z + z_{R_0}\sqrt{m_1}}{\sqrt{m_1}\sqrt{\chi_{n_1+n_2-2}^2/(n_1 + n_2 - 2)}} \sim \frac{1}{\sqrt{m_1}}t_{n_1+n_2-2}(z_{R_0}\sqrt{m_1}), \quad (2.2)$$

where Z is the standard normal random variable, η is given in (1.3), $m_1 = \frac{n_1(1+q_1)}{q_1+n_1/n_2}$ and $t_n(\xi)$ denotes the noncentral t variable with $df = n$ and noncentrality parameter ξ . This result in (2.2) leads to an exact $1 - \alpha$ lower limit for η (or R_0 content - $(1 - \alpha)$ coverage lower tolerance limit for the distribution of $Y_1 - Y_2$), and is given by

$$\bar{y}_d - t_{n_1+n_2-2, 1-\alpha}(z_{R_0}\sqrt{m_1}) \frac{s_d}{\sqrt{m_1}}, \quad (2.3)$$

where $t_{n,p}(c)$ denotes the 100 p th percentile of a noncentral t distribution with $df = n$ and noncentrality parameter c . The test based on (2.3) rejects the null hypothesis in (1.2) when the lower tolerance limit in (2.3) is greater than zero or equivalently, when the p-value

$$P\left(t_{n_1+n_2-2}(z_{R_0}\sqrt{m_1}) > \frac{\sqrt{m_1}\bar{y}_d}{s_d}\right) < \alpha. \quad (2.4)$$

3. TESTS FOR R WHEN $q_1 = \sigma_1^2/\sigma_2^2$ IS UNKNOWN

If q_1 is unknown, an approximate test can be obtained by substituting an estimate for q_1 in (2.4). This approximate test, as pointed out by Hall (1) in the problem of constructing tolerance limits for $Y_1 - Y_2$, is too liberal in some situations. Hall also suggested an alternative approach that can be used to develop an approximate test for R . Using the usual moment approximation (popularly known as the Satterthwaite's (7) approximation), it can be shown that $(s_1^2 + s_2^2)/(\sigma_1^2 + \sigma_2^2)$ is approximately distributed as $\chi_{f_1}^2/f_1$, where f_1 is given in (3.1). Using this result and standardizing \bar{y}_d , it can be readily verified that

$$\frac{\bar{y}_d - \eta}{\sqrt{s_1^2 + s_2^2}} \sim \frac{1}{\sqrt{m_1}} \left(\frac{Z + z_{R_0}\sqrt{m_1}}{\sqrt{\chi_{f_1}^2/f_1}} \right) \text{ approximately,}$$

where

$$m_1 = \frac{n_1(1 + q_1)}{q_1 + n_1/n_2} \text{ and } f_1 = \frac{(n_1 - 1)(q_1 + 1)^2}{q_1^2 + (n_1 - 1)/(n_2 - 1)}. \quad (3.1)$$

Noticing that $\left(\frac{Z+z_{R_0}\sqrt{m_1}}{\sqrt{\chi_{f_1}^2/f_1}}\right) \sim t_{f_1}(z_{R_0}\sqrt{m_1})$, we get a $1 - \alpha$ lower confidence limit for η as

$$\bar{y}_d - t_{f_1, 1-\alpha}(z_{R_0}\sqrt{m_1})\sqrt{\frac{s_1^2 + s_2^2}{m_1}}. \quad (3.2)$$

Thus, if q_1 is known, then (3.2) is an approximate $(R_0, 1 - \alpha)$ lower tolerance limit for the distribution of $Y_1 - Y_2$. At the level of significance α , the null hypothesis in (1.2) will be rejected whenever the limit in (3.2) is greater than zero or equivalently, whenever the p -value

$$P\left(t_{f_1}(z_{R_0}\sqrt{m_1}) > \frac{\sqrt{m_1}\bar{y}_d}{\sqrt{s_1^2 + s_2^2}}\right) < \alpha. \quad (3.3)$$

It can be verified that the test based on (3.3) is exact only when $n_1 = n_2$ and $\sigma_1^2 = \sigma_2^2$. Nevertheless, when q_1 is unknown, the following approximate tests based on (3.3) perform better than the corresponding ones based on (2.4). We shall now present three approximate tests, and the generalized p -value test due to Weerahandi and Johnson (4) for the case of unknown variance ratio.

The H Test: If q_1 is unknown, replacing q_1 in (3.1) by its unbiased estimator $\hat{q}_1 = s_1^2(n_2 - 3)/(s_2^2(n_2 - 1))$, we get an approximate test that rejects the H_0 in (1.2) when the p -value

$$P_1 = P\left(t_{\hat{f}_1}(z_{R_0}\sqrt{\hat{m}_1}) > \hat{\delta}\sqrt{\hat{m}_1}\right) < \alpha, \quad (3.4)$$

where $\hat{\delta} = \bar{y}_d/\sqrt{s_1^2 + s_2^2}$, \hat{m}_1 and \hat{f}_1 can be obtained from (3.1) by replacing q_1 with \hat{q}_1 . As we already mentioned, this unbiased estimate of q_1 is used in Hall (1) for constructing tolerance limit for the distribution of $Y_1 - Y_2$. For this reason, we refer to this test as the H test.

The R-G Test: The estimate $\hat{q} = s_1^2/s_2^2$ is used to construct a lower bound for R in the paper by Reiser and Guttman (3). This choice of estimate yields a test that rejects the H_0 in (1.2) when the p -value

$$P\left(t_{\hat{f}}(z_{R_0}\sqrt{\hat{m}}) > \hat{\delta}\sqrt{\hat{m}}\right) < \alpha, \quad (3.5)$$

where \hat{m} and \hat{f} can be obtained from (3.1) by replacing q_1 with $\hat{q} = s_1^2/s_2^2$.

The G-K Test: As we already mentioned, the p -value in (3.4) depends on the definition of the variance ratio. If the variance ratio is defined as $q_2 = \sigma_2^2/\sigma_1^2$, and it is estimated by $\hat{q}_2 = s_2^2(n_1 - 3)/(s_1^2(n_1 - 1))$, then the test based on \hat{q}_2 rejects the H_0 in (1.2) when

$$P_2 = P\left(t_{\hat{f}_2}(z_{R_0}\sqrt{\hat{m}_2}) > \hat{\delta}\sqrt{\hat{m}_2}\right) < \alpha, \quad (3.6)$$

where

$$\hat{m}_2 = \frac{n_2(1 + \hat{q}_2)}{\hat{q}_2 + n_2/n_1}, \quad \text{and} \quad \hat{f}_2 = \frac{(n_2 - 1)(\hat{q}_2 + 1)^2}{\hat{q}_2^2 + (n_2 - 1)/(n_1 - 1)}. \quad (3.7)$$

It is clear that the p-values P_1 in (3.4) and P_2 in (3.6) are different. Furthermore, our preliminary numerical studies indicated that the sizes of the H test based on (3.4) are very close to the nominal level when $q_1 = \sigma_1^2/\sigma_2^2$ is small and $n_1 \geq n_2$, and the sizes of the test based on P_2 in (3.6) are close to the nominal level when $q_2 = \sigma_2^2/\sigma_1^2$ is small and $n_2 \geq n_1$. In particular, we found if one of the tests is too liberal for some sample size and parameter configurations, then the other test performs satisfactorily at the same configurations. These findings suggest that the test that rejects H_0 in (1.2) whenever

$$P_a = \max\{P_1, P_2\} < \alpha \quad (3.8)$$

should be less liberal than the Hall's test and the one based on P_2 in (3.6). Notice that the value of P_a does not depend on labels of the samples. In other words, P_a is a symmetric function of (n_1, s_1^2) and (n_2, s_2^2) . The test based on (3.8) is referred to as the G-K test.

The W-J Test: The generalized variable statistic due to Weerahandi and Johnson (4) is given by

$$T = \frac{\bar{y}_d - Z \sqrt{\frac{(n_1-1)s_1^2}{n_1 U_1} + \frac{(n_2-1)s_2^2}{n_2 U_2}}}{\sqrt{\frac{(n_1-1)s_1^2}{U_1} + \frac{(n_2-1)s_2^2}{U_2}}}, \quad (3.9)$$

where Z , U_1 and U_2 are independent random variables with $Z \sim N(0, 1)$ and $U_i \sim \chi_{n_i-1}^2$, $i = 1, 2$. For given sample sizes, \bar{y}_d , s_1^2 and s_2^2 , the test based on T rejects the null hypothesis in (1.2) whenever the generalized p-value

$$P_{Z,U_1,U_2} (T < \Phi^{-1}(R_0)) < \alpha. \quad (3.10)$$

3.1 Simulation Studies of the Sizes

The sizes of the approximate tests given in the preceding section are estimated using Monte Carlo simulation consisting of 10,000 runs. From the development of the tests, we see that the p-values depend only on the variance ratio q_1 . The sizes are computed when $\eta = 0$, or equivalently when $\mu_1 - \mu_2 = z_{R_0} \sqrt{\sigma_1^2 + \sigma_2^2}$, and $R_0 = 0.95$. To estimate the size of the test, for example, the H test based on (3.4), we computed the probability $P_1 = P\left(t_{\hat{f}_1}(z_{R_0} \sqrt{\hat{m}_1}) > \frac{\sqrt{\hat{m}_1} \bar{y}_d}{\sqrt{s_1^2 + s_2^2}}\right)$ for each simulated statistic $(\bar{y}_d, s_1^2, s_2^2)$, and then found the proportion of the values of P_1 which are less than the nominal level α . For a good test, this proportion should be close to α . To compute the noncentral t probabilities, we used the algorithm due to Benton and Krishnamoorthy (8), which is an enhanced version of the algorithm due to Lenth (9). To estimate the sizes of the generalized p-value test in (3.10), we used 2500 simulated values of $(\bar{x}_1, s_1^2, \bar{x}_2, s_2^2)$, and for each such simulated value, we used 5000 simulated values of (Z, U_1, U_2) to estimate the probability in (3.10). The percentage of the 2500 probabilities which are less than α is an estimate of the size of the generalized p-value test. IMSL subroutines RNCHI and RNNOA are used to generate respectively the chi-square random numbers and the normal random numbers.

The following tests are considered in our simulation studies.

1. The H test based on P_1 in (3.4); $\hat{q}_1 = \frac{(n_2-3)s_1^2}{(n_2-1)s_2^2}$
2. The R-G test based on (3.5); $\hat{q} = \frac{s_1^2}{s_2^2}$
3. The G-K test based on (3.8); the new test
4. The W-J test based on the generalized p-value in (3.10)

The estimated sizes of the tests are plotted in Figure 1 when the sample sizes are equal, $0.05 \leq \sigma_1^2/\sigma_2^2 \leq 20$, and the nominal level $\alpha = 0.05$. When $n_1 = n_2 = 5$, we see that the tests H and R-G are liberal when the ratio $q_1 = \sigma_1^2/\sigma_2^2 \geq 3$ and the new test G-K is either slightly conservative or very close to the nominal level 0.05. When the common sample size is 10 there is only a slight difference between the sizes of tests H and R-G; the new test G-K is almost exact except in the neighborhood of $q_1 = 1$. We observe from Figures 1(a-d) that the test G-K is almost exact when the common sample size greater than or equal to 15. The generalized p-value test is in general conservative; it is too conservative when σ_1^2/σ_2^2 is small.

In Figures 2(a-e), we plotted the sizes of the tests when the sample sizes are unequal. It is clear that the sizes of the tests H and R-G could be as high as 0.11 (see Figure 2(a)) when n_1 is much smaller than n_2 and σ_1^2/σ_2^2 is small. The differences among the sizes are not appreciable when $|n_1 - n_2|$ is not too large; still the test G-K controls the sizes very satisfactorily. The sizes of the test G-K range from 0.03 to 0.07 while the sizes of the tests H and R-G range from 0.04 to 0.11 for all the cases considered. The generalized p-value test is slightly liberal in some cases (see Figure 2(b)), and in other situations its sizes are close to the nominal level 0.05.

Since we noticed that the difference between the sample sizes has impact on the sizes, we plotted the sizes in Figures 3(a-h) as a function of n_2 (ranging from 5 to 140) while $n_1 = 7$ and 15, and $q_1 = 0.2, 0.8, 2$ and 8. We again observe from these graphs that the test G-K performs better than the other two approximate tests. The sizes of the new test ranges from 0.04 to 0.07 where as the sizes of the other two approximate tests could go as high as 0.10. Thus, we see that the test G-K is clearly preferable to other two approximate tests for practical applications. The generalized p-value test is again either conservative or almost exact for all the cases considered in Figure 3.

Our overall conclusions based on Figures 1-3 are as follows. The new test is certainly preferable to the other approximate tests when the sample sizes are close to each other. Our numerical studies (not reported here) showed that if both sample sizes are at least 5, and their ratio is between 0.7 and 1.3, then the sizes of the new test are very close to the nominal level. The generalized p-value test W-J can be recommended for practical use if the ratio of the sample sizes is less than 0.7 or greater than 1.3.

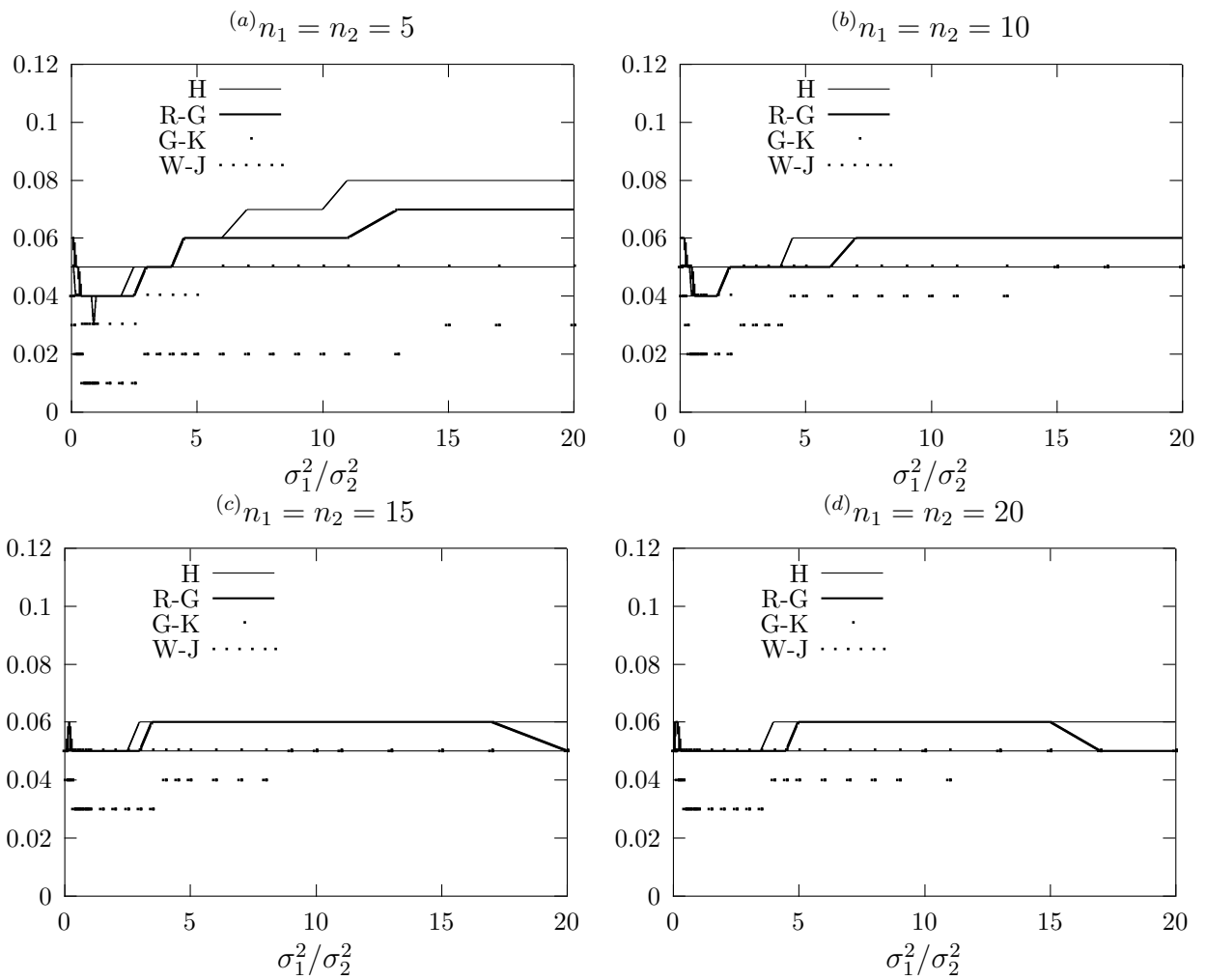


Figure 1. Sizes of the Tests When Sample Sizes are Equal and $\alpha = 0.05$; Test H is based on (3.4), Test R-G is based on (3.5), Test G-K is based on (3.8) and Test W-J in (3.10)

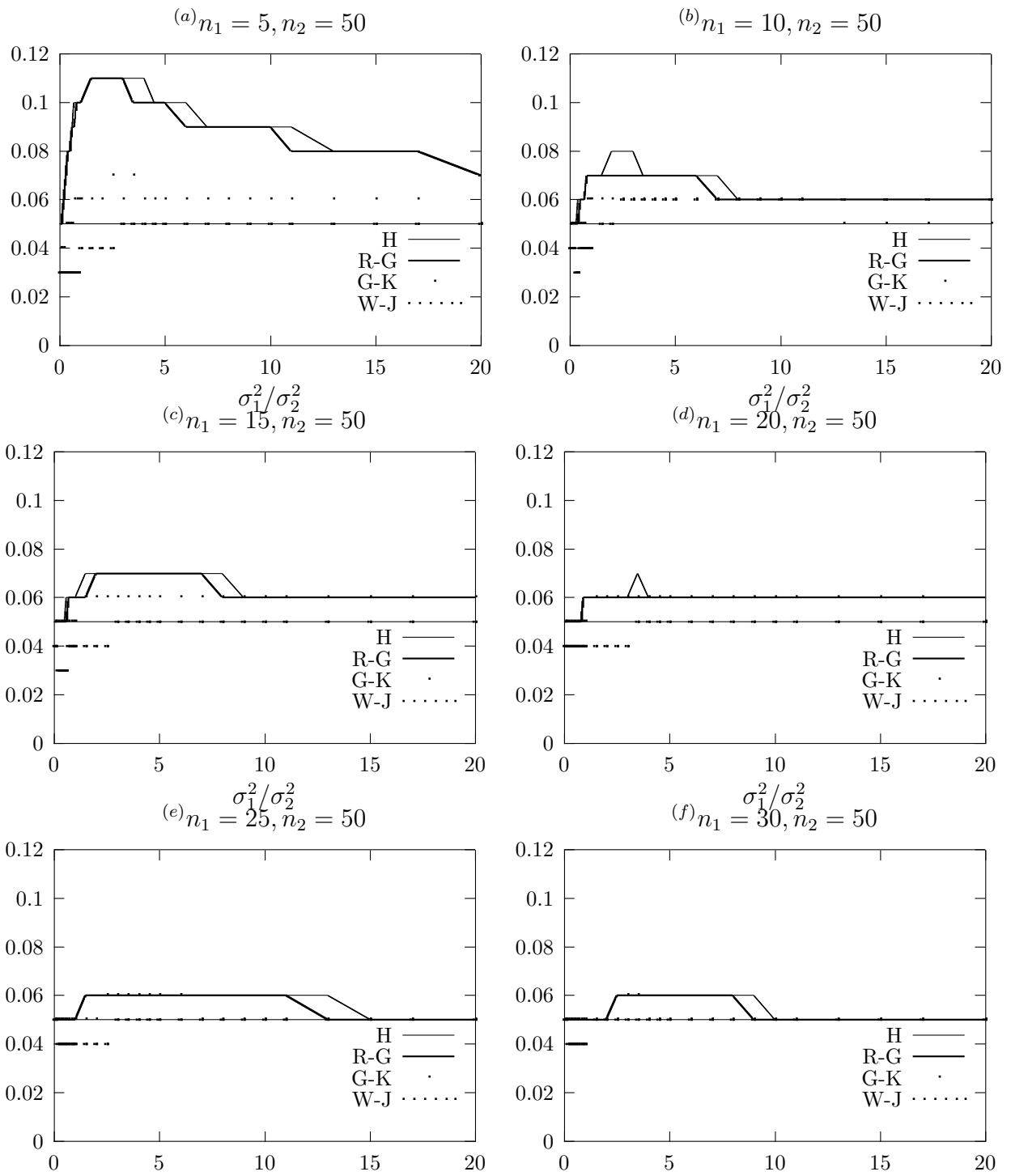


Figure 2. Sizes of the Tests When Sample Sizes are Unequal and $\alpha = 0.05$; Test H is based on (3.4), Test R-G is based on (3.5), Test G-K is based on (3.8) and Test W-J Test in (3.10)

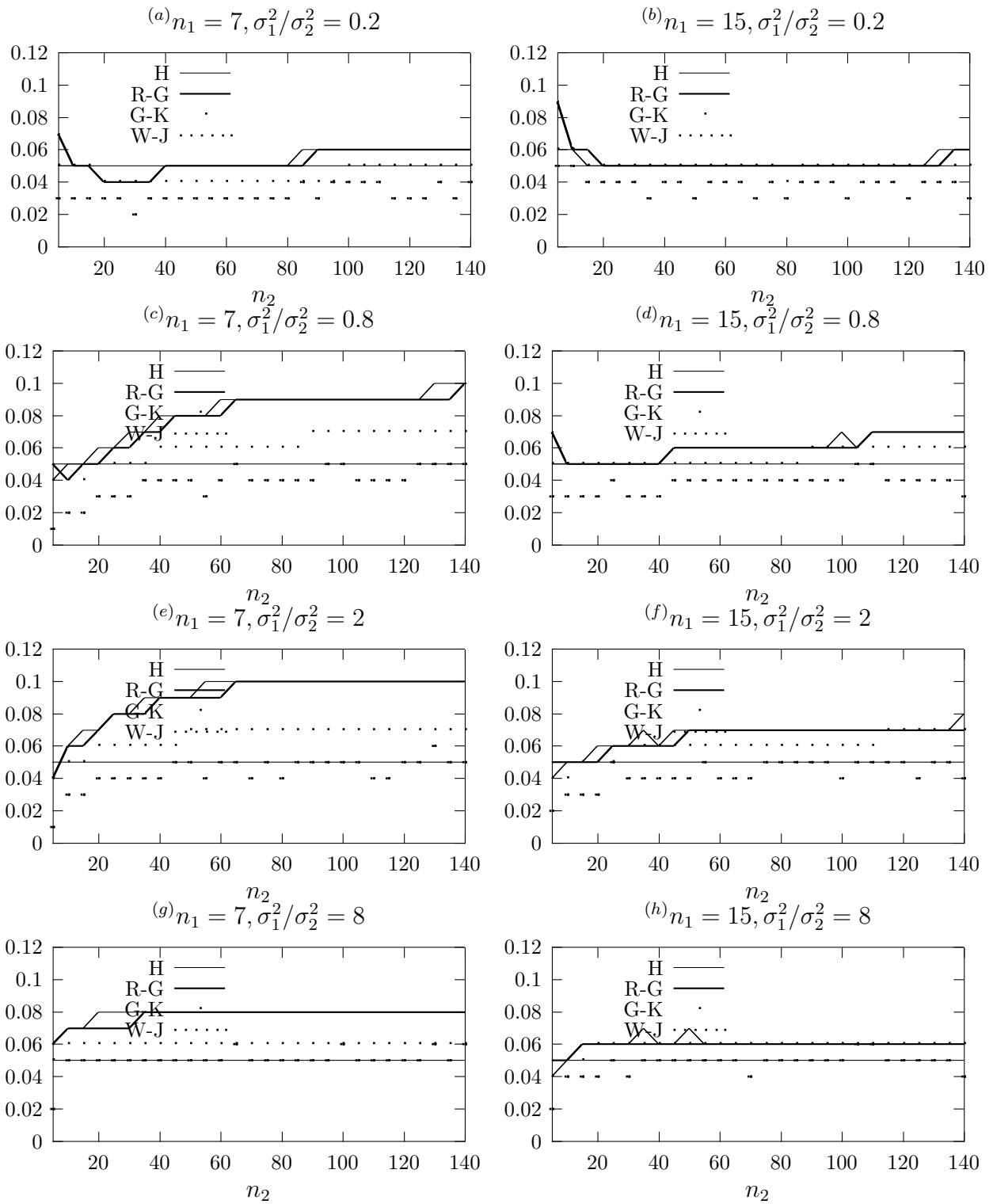


Figure 3. Sizes of the Tests as a Function of n_2 ; Test H is based on (3.4), Test R-G is based on (3.5), Test G-K is based on (3.8) and Test W-J Test in (3.10); $\alpha = 0.05$.

3.2 Lower Confidence Bounds for R

We shall now consider interval estimation of R on the basis of the tests in the previous section. Let $\delta = (\mu_1 - \mu_2)/\sqrt{\sigma_1^2 + \sigma_2^2}$. Since $R = \Phi(\delta)$ is a one-one function of δ , it suffices to find a confidence bound for δ . To get confidence bound for δ , we note that

$$\hat{\delta} = \frac{\bar{y}_d}{\sqrt{s_1^2 + s_2^2}} = \frac{\frac{\bar{y}_d - \mu_d}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \delta\sqrt{m_1}}{\sqrt{m_1}\sqrt{\frac{s_1^2 + s_2^2}{\sigma_1^2 + \sigma_2^2}}} \sim \frac{Z + \sqrt{m_1}\delta}{\sqrt{m_1}\sqrt{\chi_{f_1}^2/f_1}} \text{ approximately,} \quad (3.11)$$

where m_1 and f_1 are given in (3.1). Note that $\hat{\delta} \sim t_{f_1}(\sqrt{m_1}\delta)/\sqrt{m_1}$ approximately. Let $F(x; n, \xi)$ denote the noncentral t cdf with $df = n$, and noncentrality parameter ξ . For any given x and n , the noncentral t cdf is strictly decreasing with respect to its noncentrality parameter, and hence a $1 - \alpha$ lower bound δ_L for δ can be obtained as the solution of

$$P(t_{f_1}(\sqrt{m_1}\delta_L) < \sqrt{m_1}\hat{\delta}) = F(\sqrt{m_1}\hat{\delta}; f_1, \sqrt{m_1}\delta_L) = 1 - \alpha. \quad (3.12)$$

If the variance ratio q_1 is unknown, then replacing m_1 and f_1 by their estimates, we get the lower bound δ_{1L} as the solution of

$$F(\sqrt{\hat{m}_1}\hat{\delta}; \hat{f}_1, \sqrt{\hat{m}_1}\delta_{1L}) = 1 - \alpha, \quad (3.13)$$

where \hat{m}_1 and \hat{f}_1 are defined as in (3.4). The lower bound corresponding to \hat{m}_2 and \hat{f}_2 in (3.7) is the δ_{2L} that satisfies

$$F(\sqrt{\hat{m}_2}\hat{\delta}; \hat{f}_2, \sqrt{\hat{m}_2}\delta_{2L}) = 1 - \alpha. \quad (3.14)$$

Thus, the combined lower bound is given by

$$\delta_L = \min\{\delta_{1L}, \delta_{2L}\}. \quad (3.15)$$

If q_1 is known, then the exact lower bound for δ based on the test (2.4) is given by δ_{EL} which satisfies

$$F(\sqrt{m_1}\hat{\delta}_*; n_1 + n_2 - 2, \sqrt{m_1}\delta_{EL}) = 1 - \alpha, \quad (3.16)$$

where $\hat{\delta}_* = \bar{y}_d/s_d$, s_d^2 is defined in (2.1) and m_1 is given in (3.1).

The generalized $1 - \alpha$ lower limit for δ on the basis of T in (3.9) is given by the α th quantile T_α of T , which leads to the lower limit $\Phi(T_\alpha)$ for R .

3.3 Examples

We shall now illustrate the methods using the examples given in Hall (1), Reiser and Guttman (3) and Weerahandi and Johnson (4). The noncentral t probabilities and critical values are calculated using the software *StatCalc*, and it can be downloaded for free from <http://www.etext.net/catalog/StatCalc/> or using an online calculator at <http://calculators.stat.ucla.edu/>. The generalized p-values and limits are computed using 1,000,000 simulation runs.

Example 1. (Hall 1984) A sample of $n_1 = 50$ capacitors yielded mean breakdown voltage $\bar{y}_1 = 6.75$ kV and $s_1^2 = 0.123$. The voltage output from $n_2 = 20$ transverters (power supplies) produced $\bar{y}_2 = 4.00$ kV and $s_2^2 = 0.53$. We have $\bar{y}_d = \bar{y}_1 - \bar{y}_2 = 2.75$. Using the methods in the preceding sections, we like to compute 95% lower limits for R . To compute δ_{cL} in Section 2.2, we note that

$$\begin{aligned}\hat{q}_1 &= 0.2077, \hat{m}_1 = 22.3007, \hat{f}_1 = 27.2543, \\ \hat{q}_2 &= 4.1331, \hat{m}_2 = 22.6471, \hat{f}_2 = 28.6559 \text{ and } \hat{\delta} = 3.4031\end{aligned}$$

The equation (3.13) is $F(\sqrt{\hat{m}_1\hat{\delta}}; f_1, \sqrt{\hat{m}_1\hat{\delta}_{1L}}) = F(16.0707; 27.2543, 4.7224\delta_{1L}) = 0.95$. Solving this for δ_{1L} , we found $\delta_{1L} = 2.5560$. Similarly, solving the equation (3.14), we found $\delta_{2L} = 2.5745$. Thus, $\delta_L = \min\{\delta_{1L}, \delta_{2L}\} = 2.5560$. The 95% lower limit for R is $\Phi(\delta_L) = \Phi(2.5560) = 0.9947$. The 95% lower limit using $\hat{q} = s_1^2/s_2^2 = 0.2321$ is $\Phi(2.5693) = 0.9949$. This limit, due to Reiser and Guttman (3), can be obtained from (3.12) with (m_1, f_1) replaced by (\hat{m}, \hat{f}) defined in (3.5). The generalized limit given in Section 2.2 is $\Phi(T_\alpha) = \Phi(2.5118) = 0.9940$. All these estimates are very close to each others.

Example 2. The summary statistics of this example are taken from Reiser and Guttman (3). The data are pertaining to mechanical component that yielded $\bar{y}_1 = 170,000$ psi, $s_1 = 5,000$ psi, $\bar{y}_2 = 144,500$ psi, $s_2 = 8,900$ psi and $n_1 = n_2 = 32$. Using the formulas of the preceding sections, we computed

$$\begin{aligned}\hat{q}_1 &= 0.29525, \hat{f}_1 = 47.8377, \\ \hat{q}_2 &= 2.9640, \hat{f}_2 = 49.7800, \\ \hat{m}_2 &= \hat{m}_1 = 32 \text{ and } \hat{\delta} = 2.4980.\end{aligned}$$

We are interested in obtaining a 90% lower bound for R . Using $\hat{m}_1, \hat{f}_1, \hat{\delta}$ and $1 - \alpha = 0.90$ in (3.13), we obtained $\delta_{1L} = 2.08946$. Similarly, using (3.14) we got $\delta_{2L} = 2.0951$. The 90% lower bound based on (3.15) for R is $\Phi(2.08946) = 0.9817$. The 90% lower bound using the approach of Reiser and Guttman (3) is 0.9819. The 90% generalized lower limit for this example is 0.9804.

Suppose we want to test $H_0 : R \leq 0.95$ vs $H_a : R > 0.95$. Then using (3.4), we have $P_1 = P(t_{\hat{f}_1}(z_{R_0}\sqrt{\hat{m}_1}) > \sqrt{\hat{m}_1\hat{\delta}}) = P(t_{47.8377}(1.64485 \times \sqrt{32}) > 14.1306) = 0.0027$. Using (3.6), we have $P_2 = 0.0024$. Thus, the p-value of the G-K test is 0.0027. The p-value of the R-G test in (3.5) is 0.0024. The generalized p-value in (3.10) is 0.0042. The p-values of the approximate tests are very close to each other while the generalized p-value is slightly higher than the others.

Example 3. In this example, we consider the rocket-motor experiment data set given in Guttman et. al. (2) which was later used by Weerahandi and Johnson (4) for illustration purpose. We are interested in making inferences on the reliability of the rocket motor at the

highest operating temperature of 59 degrees centigrade. At this temperature, the operating pressure Y_2 distribution tends to be closest to the chamber burst strength Y_1 distribution. Assumption of normality has been verified in the paper just cited. The hypotheses considered by Weerahandi and Johnson are $H_0 : R \leq 0.999999$ and $H_a : R > 0.999999$. This is equivalent to testing $H_0 : \delta \leq \Phi^{-1}(0.999999) = 4.75059$ vs $H_a : \delta > 4.75059$. A sample of $n_1 = 17$ and a sample of $n_2 = 24$ observations yielded the statistics

$$\bar{y}_1 = 16.485, s_1^2 = 0.3409, \bar{y}_2 = 7.789, s_2^2 = 0.05414,$$

$$\hat{q}_1 = 5.7491, \hat{m}_1 = 17.7678, \hat{f}_1 = 21.5956$$

$$\hat{q}_2 = 0.13896, \hat{m}_2 = 17.6272, \hat{f}_2 = 20.4806 \text{ and } \hat{\delta} = 13.8356.$$

Using $\hat{m}_1, \hat{f}_1, \hat{\delta}$ and $Z_{R_0} = 4.75059$ in (3.4), we computed $P_1 = 0.00000024$. Similarly, using (3.6), we computed $P_2 = 0.00000052$. Thus, the p-value of the G-K test is 0.00000052. The generalized p-value based on our own simulation is 0.000004 (the reported value in Weerahandi and Johnson (4) is 0.0000042). Even though both tests provided strong evidence in favor of the alternative hypothesis H_a , we see that the p-value of the new test is much smaller than the generalized p-value.

4. INFERENCE ON THE RELIABILITY PARAMETER WHEN Y_1 AND Y_2 DEPEND ON COVARIATES

We shall now extend the results of the preceding sections to the case where Y_1 and Y_2 depend on some covariates. Specifically, we consider the models

$$Y_{1i} = \beta_1' X_{1i} + \epsilon_{1i}, \quad i = 1, \dots, n_1,$$

and

$$Y_{2i} = \beta_2' X_{2i} + \epsilon_{2i}, \quad i = 1, \dots, n_2,$$

where ϵ_{1i} 's are independent $N(0, \sigma_1^2)$ random variables, ϵ_{2i} 's are independent $N(0, \sigma_2^2)$ random variables, X_{1i} 's are $p_1 \times 1$ vectors and X_{2i} 's are $p_2 \times 1$ vectors. Let $\mathbf{X}'_i = (X_{i1}, \dots, X_{in_i})$, and $\mathbf{Y}'_i = (Y_{i1}, \dots, Y_{in_i})'$, $i = 1, 2$. The least squares estimator of β_i is $\hat{\beta}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{Y}_i$, $i = 1, 2$, and the residual sums of squares based on the models are $(n_1 - p_1)s_1^2$ and $(n_2 - p_2)s_2^2$. All these variables are independent with

$$\hat{\beta}_i \sim N_{p_i}(\beta_i, \sigma_i^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}), \quad i = 1, 2,$$

$$(n_i - p_i)s_i^2 / \sigma_i^2 \sim \chi_{n_i - p_i}^2, \quad i = 1, 2.$$

The reliability parameter is defined by

$$R(\mathbf{x}_2, \mathbf{x}_2) = P(Y_1 > Y_2) = \Phi \left(\frac{\beta_1' \mathbf{x}_1 - \beta_2' \mathbf{x}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right),$$

where \mathbf{x}_1 and \mathbf{x}_2 are given values for the explanatory variables. Writing $c_i = \mathbf{x}'_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{x}_i$, $i = 1, 2$, we see that $\hat{\beta}_i' \mathbf{x}_i \sim N(\beta_i' \mathbf{x}_i, c_i \sigma_i^2)$, $i = 1, 2$. Because we deal with the same distributions as those in Section 3, the results of Section 3 can be extended to the present problem in

a straightforward manner. For the sake of completeness, we shall present the test equivalent to the new approximate test in (3.8). Let

$$\hat{q}_1 = \frac{(n_2 - p_2 - 2)s_1^2}{(n_2 - p_2)s_2^2}, \quad \hat{m}_1 = \frac{c_1^{-1}(1 + \hat{q}_1)}{\hat{q}_1 + c_2/c_1} \quad \text{and} \quad \hat{f}_1 = \frac{(n_1 - p_1)(1 + \hat{q}_1)^2}{\hat{q}_1^2 + (n_1 - p_1)/(n_2 - p_2)}. \quad (4.1)$$

The terms \hat{q}_2 , \hat{m}_2 and \hat{f}_2 can be defined by interchanging the subscripts 1 and 2 in (4.1). Define $\hat{\delta} = (\hat{\beta}'_1 \mathbf{x}_1 - \hat{\beta}'_2 \mathbf{x}_2) / \sqrt{s_1^2 + s_2^2}$. The p-value P_1 similar to the one in (3.4) is given by $P(t_{\hat{f}_1}(z_{R_0} \sqrt{\hat{m}_1}) > \sqrt{\hat{m}_1} \hat{\delta})$, and the one similar to P_2 in (3.6) is given by $P(t_{\hat{f}_2}(z_{R_0} \sqrt{\hat{m}_2}) > \sqrt{\hat{m}_2} \hat{\delta})$. The null hypothesis $H_0 : R(\mathbf{x}_2, \mathbf{x}_2) \leq R_0$ will be rejected whenever the maximum of these two p-values is less than α .

The lower bounds δ_{1L} and δ_{2L} can be computed using the above (\hat{m}_1, \hat{f}_1) and (\hat{m}_2, \hat{f}_2) respectively in (3.13) and (3.14). The combined lower bound for $R(\mathbf{x}_2, \mathbf{x}_2)$ is then given by $\delta_L = \min\{\delta_{1L}, \delta_{2L}\}$.

5. CONCLUDING REMARKS

In this article, we explored the Hall's (1) approximate method of constructing tolerance limits for the distribution of $Y_1 - Y_2$, and proposed an approximate inferential procedures for the reliability parameter R . The proposed methods require computations of the noncentral t cumulative probabilities, critical points and the noncentrality parameter (given other parameters of the noncentral t) with non-integer degrees of freedom. All these quantities can be computed using freely available PC calculators (see example section for a URL). If the data are from lognormal distributions, then the proposed methods can be applied to the logged data. Even though the proposed methods are accurate enough for practical purpose still they are not exact. Some readers may want to explore the applicability of the parametric bootstrap (PB) method. Nowadays, such numerically intensive methods are commonly used to tackle non-standard problems such as the present one. We indeed numerically investigated the properties of the test based on the PB method. Our extensive simulation studies showed the test based on the PB method is liberal (in some situations, the size of the test exceeds 0.2 when the nominal level is 0.05). For these reasons, we have not discussed the PB test here. Our overall conclusion is that, among the methods considered in this article, the G-K test and the confidence limit based on it (the δ_L in (3.15)) are preferable for practical applications when the sample sizes are close to each other, and the generalized variable method W-J is preferable to others when one of the sample sizes is drastically different from the other.

BIBLIOGRAPHY

- (1) Hall, I. J. Approximate One-Sided Tolerance Limits for the Difference or Sum of Two Independent Normal Variates. *Journal of Quality Technology*, **1984**, *16*, 15-19.
- (2) Guttman, I.; Johnson, R. A.; Bhattacharya, G. K.; Reiser, B. J. Confidence Limits for Stress-Strength Models With Explanatory Variables. *Technometrics*, **1988**, *30*, 161-168.
- (3) Reiser, B. J.; Guttman, I. Statistical inference for $\Pr(Y < X)$: The normal case. *Technometrics*, **1986**, *28*, 253-257.
- (4) Weerahandi, S.; Johnson, R. A. Testing Reliability in a Stress-Strength Model When X and Y are Normally Distributed. *Technometrics*, **1992**, *34*, 83-91.
- (5) Casella, G.; Berger, R. L. *Statistical Inference*, Duxbury, CA: Pacific Grove, **2002**.
- (6) Church, J. D.; Harris, B. The Estimation of Reliability From Stress-Strength Relationships. *Technometrics*, **1970**, *12*, 49-54.
- (7) Satterthwaite, F. E. An Approximate Distribution of Estimates of Variance Components. *Biometric Bulletin*, **1946**, *2*, 110-114.
- (8) Benton, D.; Krishnamoorthy, K. Computing Discrete Mixtures of Continuous Distributions: Noncentral Chisquare. Noncentral t and the Distribution of the Square of the Sample Multiple Correlation Coefficient. *Computational Statistics and Data Analysis*, **2003**, *43*, 249-267
- (9) Lenth, R. V. Cumulative Distribution Function of the Noncentral t Distribution. *Applied Statistics*, **1989**, *38*, 185-189.