

Bootstrap Testing for Reliability of Stress-Strength Model with Explanatory Variables

Jin Pyo Park¹ · Sang Gil Kang² · Jang Sik Cho³

Abstract

In this paper, we consider some approximate testings for the reliability of the stress-strength model when the stress X and strength Y each depends linearly on some explanatory variables \mathbf{z} and \mathbf{w} , respectively. We construct a bootstrap procedure for testing for various values of the reliability and compare the power of the bootstrap test with the test based on Mann-Whitney type estimator by Park et.al.(1996) for small and moderate sample size.

Key Words and Phrases: Bootstrap test, Stress-strength model, Actual power.

1. Introduction

An important extension of the stress-strength model allows the strength X and the stress Y to depend on some explanatory variables. In many cases, an experimenter has access to the measurements of some explanatory variables that affect the strength or influence the stress. The additional information can play an important role in the analysis by extending the classical stress-strength model to include explanatory variables.

Duncan(1986) gave some specific examples of the strength-stress model with explanatory variables. Guttman, Johnson, Bhattacharyya and Reisser (1988) obtained an approximate confidence interval for reliability, $R = P(X < Y|\mathbf{z}, \mathbf{w})$. Park, Kim and Park(1996) considered Mann-Whitney type statistic to estimate the reliability.

Since the true distribution of the estimator for R is often skewed and biased for a small sample and/or large value of R , the power of testing based on Park et. al.

¹Professor, Department of Computer Engineering, Kyungnam University, Masan, 631-701, Korea.

²Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

³Full-time Lecturer, Department of Computer Science, Kyungsoo University, Pusan, 608-736, Korea.

may deteriorate the accuracy. So we will use the bootstrap method to rectify these problems.

Efron(1979, 1981) initially introduced the bootstrap method to assign the accuracy for an estimator. General theory for bootstrap hypothesis testing is discussed briefly by Hinkley(1988) during a survey of bootstrap methods, and at greater length by Hinkley(1989). Beran(1988) discussed pivoting in the context of bootstrap hypothesis testing. Hall and Wilson(1991) and Becher(1993) illustrated the two guidelines of pivoting and sampling under null hypothesis by applying bootstrap tests to specific data sets.

In this paper, we construct the bootstrap procedure for testing reliability using some bootstrap methods and compare the powers with the test based on the Mann-Whitney type statistic via Monte Carlo simulation in small and moderate samples.

2. Preliminaries

Suppose that X is related to p explanatory variables \mathbf{z} and \mathbf{Y} is related to q explanatory variables \mathbf{w} according to the linear relationships,

$$X = \mu + \underline{\beta}'(\mathbf{z} - \bar{\mathbf{z}}) + \delta \quad (1)$$

and

$$Y = \nu + \underline{\gamma}'(\mathbf{w} - \bar{\mathbf{w}}) + \epsilon, \quad (2)$$

where $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ and $\underline{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)'$ are regression coefficients and the errors δ and ϵ are independent random variables with distribution F and G , respectively, such that $E(\delta) = E(\epsilon) = 0$, $Var(\delta) = \sigma^2 < \infty$ and $Var(\epsilon) = \tau^2 < \infty$. Here, the errors δ and ϵ are not necessarily normal. Suppose that (X_i, \mathbf{z}_i) and (Y_j, \mathbf{w}_j) , $i = 1, \dots, m$, $j = 1, \dots, n$ be samples from the models in (1), and let $\bar{\mathbf{z}} = m^{-1} \sum_{i=1}^m \mathbf{z}_i$, $\bar{\mathbf{w}} = n^{-1} \sum_{j=1}^n \mathbf{w}_j$. Also let $P(\theta) = P(X < Y | \mathbf{z}, \mathbf{w})$, where $\underline{\theta} = (\mu, \nu, \underline{\beta}', \underline{\gamma}')'$. Then by Park, et.al.(1996), $P(\theta)$ can be estimated by

$$\widehat{U}(\widehat{\underline{\theta}}) = (mn)^{-1} \sum_{i=1}^m \sum_{j=1}^n I(\widehat{\delta}_i - \widehat{\epsilon}_j < \widehat{x}) = (mn)^{-1} \sum_{i=1}^m \sum_{j=1}^n \widehat{U}_{ij}, \quad (3)$$

where I denotes the indicator function,

$$\begin{aligned} \widehat{\delta}_i &= X_i - \widehat{\mu} - \widehat{\underline{\beta}}'(\mathbf{z}_i - \bar{\mathbf{z}}), \\ \widehat{\epsilon}_j &= Y_j - \widehat{\nu} - \widehat{\underline{\gamma}}'(\mathbf{w}_j - \bar{\mathbf{w}}), \end{aligned}$$

$$\hat{x} = -\hat{\mu} + \hat{\nu} - \hat{\beta}'(\mathbf{z} - \bar{\mathbf{z}}) + \hat{\gamma}'(\mathbf{w} - \bar{\mathbf{w}}) \quad \text{and} \quad \hat{U}_{ij} = I(\hat{\delta}_i - \hat{\epsilon}_j \leq \hat{x}).$$

Under mild conditions Park(1995) established the asymptotic normality of $\sqrt{N} [\hat{U}(\hat{\theta}) - P(\theta)]$. That is,

$$\sqrt{N} [\hat{U}(\hat{\theta}) - P(\theta)] \rightarrow^d N(0, \Sigma). \tag{4}$$

The consistent estimator of the asymptotic variance for $\sqrt{N} [\hat{U}(\hat{\theta}) - P(\theta)]$ is given

$$\begin{aligned} \hat{\Sigma} = & \frac{N}{(m-1)(n-1)} [(m-1)(\hat{p}_1 - \hat{U}(\hat{\theta}))^2 + (n-1)(\hat{p}_2 - \hat{U}(\hat{\theta}))^2 + \hat{U}(\hat{\theta}) - \hat{U}(\hat{\theta})^2]^2 \\ & + N(\mathbf{z} - \bar{\mathbf{z}})'[(\mathbf{Z} - \bar{\mathbf{Z}})'(\mathbf{Z} - \bar{\mathbf{Z}})]^{-1}(\mathbf{z} - \bar{\mathbf{z}})\hat{p}_3\hat{\sigma}^2 \\ & + N(\mathbf{w} - \bar{\mathbf{w}})'[(\mathbf{W} - \bar{\mathbf{W}})'(\mathbf{W} - \bar{\mathbf{W}})]^{-1}(\mathbf{w} - \bar{\mathbf{w}})\hat{p}_3\hat{\tau}^2, \end{aligned} \tag{5}$$

where

$$\begin{aligned} \hat{p}_1 &= \frac{\sum_{i=1}^m \sum_{j=1}^n \sum_{k \neq i}^m \hat{U}_{ij} \hat{U}_{kj}}{mn(m-1)}, \\ \hat{p}_2 &= \frac{\sum_{i=1}^m \sum_{j=1}^n \sum_{k \neq i}^n w U_{ij} \hat{U}_{ik}}{mn(n-1)}, \\ \hat{p}_3 &= \frac{\hat{H}(\hat{x} + h) - \hat{H}(\hat{x} - h)}{2h}, \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^m (X_i - \hat{\mu} - \hat{\beta}'(\mathbf{z}_i - \bar{\mathbf{z}}))^2}{(m-p-1)}, \\ \hat{\tau}^2 &= \frac{\sum_{j=1}^n (Y_j - \hat{\nu} - \hat{\gamma}'(\mathbf{w}_j - \bar{\mathbf{w}}))^2}{(n-q-1)}, \end{aligned}$$

here $\hat{H}(x) = \sum_{i=1}^m \sum_{j=1}^n I(\hat{\delta}_i - \hat{\epsilon}_j \leq x)/mn$ and h is a bandwidth of N^b , $-1/2 \leq b \leq 0$. Thus we construct an approximate test for $H_0 : P(\theta) = R_0$ v.s. $H_1 : P(\theta) \leq R_0$ based on $\hat{U}(\hat{\theta})$ as follows.

$$\phi_{MW}(x, y) = \begin{cases} 1, & \sqrt{N} [\hat{U}(\hat{\theta}) - R_0]/\sqrt{\hat{\Sigma}} \leq -z_\alpha, \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where z_α is the upper $100\alpha\%$ quantile of the standard normal distribution.

3. The Bootstrap Procedure

In this section we consider the bootstrap test for $H_0 : P(\theta) = R_0$ v.s. $H_1 : P(\theta) \leq R_0$ based on some bootstrap methods. Related to the Park et.al.'s procedure is the bootstrap procedure which is a resampling scheme that one attempts to learn the sampling properties of a statistic by recomputing its value on the basis of a new sample realized from the original one. The bootstrap procedure provides an approximate testing by using the plug-in principle for $P(\theta)$.

The bootstrap procedure can be described as follows:

- (1) Select B independent bootstrap samples $(\mathbf{X}, \mathbf{z})^{*1}, (\mathbf{X}, \mathbf{z})^{*2}, \dots, (\mathbf{X}, \mathbf{z})^{*B}$, each consisting of m data values drawn with replacement from (\mathbf{X}, \mathbf{z}) . And select B independent bootstrap samples $(\mathbf{Y}, \mathbf{w})^{*1}, (\mathbf{Y}, \mathbf{w})^{*2}, \dots, (\mathbf{Y}, \mathbf{w})^{*B}$, each consisting of n data values drawn with replacement from (\mathbf{Y}, \mathbf{w}) , respectively.
- (2) Evaluate the bootstrap replication corresponding to each bootstrap samples,

$$\widehat{U}(\widehat{\theta}^{*b}) = (mn)^{-1} \sum_{i=1}^m \sum_{j=1}^n \widehat{U}_{ij}^{*b}, \quad b = 1, 2, \dots, B, \tag{7}$$

where \widehat{U}_{ij}^{*} is the bootstrap version of \widehat{U}_{ij} .

Therefore we propose the bootstrap test by using two methods, that is, percentile, percentile-t methods.

3.1. Percentile method

The test by the bootstrap percentile method (percentile test) is obtained by percentiles of the empirical bootstrap distribution of $\widehat{U}(\theta^*)$. Let \widehat{H}^* be the empirical cumulative distribution function of $\widehat{U}(\theta^*)$.

Then it is constructed by $\widehat{H}^*(s) = \frac{1}{B} \sum_{b=1}^B I(\widehat{U}(\theta^{*b}) \leq s)$, where s is arbitrary real value and $I(\cdot)$ is an indicator function.

Then we construct the percentile test function for $H_0 : P(\theta) = R_0$ v.s. $H_1 : P(\theta) \leq R_0$ as follows:

$$\phi_{PER}(x, y) = \begin{cases} 1, & \widehat{U}(\widehat{\theta}) \leq c_1^* \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where c_1^* is calculated as a number such that $\widehat{H}^*(s) = \alpha$, that is,

$$c_1^* = \widehat{H}^{*-1}(\alpha) = \inf\{s : \widehat{H}^*(s) \leq \alpha\}. \tag{9}$$

In other word, $\widehat{H}^{*-1}(\alpha)$ is the $(B \cdot \alpha)$ th value in the ordered list of the B replications of $\widehat{U}(\widehat{\theta}^*)$.

3.2 Percentile-*t* method

The test by the bootstrap percentile-*t* method(percentile-*t* test) is constructed by using the bootstrap distribution of an approximate pivotal quantity for $\hat{U}(\hat{\theta})$ instead of the bootstrap distribution of $\hat{U}(\hat{\theta})$. We define an approximate bootstrap pivotal quantity for $\hat{U}(\hat{\theta})$ by

$$\hat{U}(\hat{\theta}^*)_{STUD} = \frac{\sqrt{N} [\hat{U}(\hat{\theta}^*) - \hat{U}(\hat{\theta})]}{\sqrt{\hat{\Sigma}^*}}, \tag{10}$$

where $\hat{\Sigma}^*$ is the bootstrap version of $\hat{\Sigma}$.

We compute the empirical distribution function \hat{H}^*_{STUD} of $\hat{U}(\hat{\theta}^*)_{STUD}$ by

$$\hat{H}^*_{STUD}(s) = \frac{1}{B} \sum_{b=1}^B I(\hat{U}(\hat{\theta}^{*b})_{STUD} \leq s), \tag{11}$$

for all real value *s*.

Then we construct the percentile-*t* test function for $H_0 : P(\theta) = R_0$ v.s. $H_1 : P(\theta) \leq R_0$ as follows:

$$\phi_{per-t}(x, y) = \begin{cases} 1, & \sqrt{N} [\hat{U}(\hat{\theta}) - R_0]/\sqrt{\hat{\Sigma}} \leq c_2^* , \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $c_2^* = \hat{H}^{*-1}_{STUD}(\alpha) = \inf\{s : \hat{H}^*_{STUD}(s) \leq \alpha\}$.

4. Comparisons

In this Section, we compare the power of the bootstrap test presented in Section 3 with the test based on Mann-Whitney type estimator by Park et.al.(1996) for small and moderate sample size. The distributions of errors under the consideration are as follows:

Case I : $\delta, \epsilon \sim N(0, 1)$

Case II : $\delta \sim 0.95N(0, 1) + 0.05N(0, 3^2)$ and $\epsilon \sim 0.95N(0, 1) + 0.05N(0, 10^2)$

Case III : $\delta, \epsilon \sim 0.9N(0, 1) + 0.1N(0, 10^2)$.

As one can see, we consider the standard normal distributions in Case I and the variance-contaminated normal distributions in Case II and III. For all such cases, both δ and ϵ have symmetric and unimodal distributions. Thus, every $\delta - \epsilon$ have symmetric distributions which are unimodal.

In this Section, we compare the power performance of the test $H_0 : P(\theta) = 0.9$ v.s. $H_1 : P(\theta) \leq 0.9$. To the actual powers against H_0 , we take the values of $P(\underline{\theta})$ under H_1 (say $P_0(\underline{\theta})$) as $P_0(\underline{\theta}) = 0.85, 0.80, 0.75, \dots, 0.60$. So the regression parameters $\mu, \nu, \underline{\beta}, \underline{\gamma}$ are chosen so that $P(\underline{\theta})$ takes the values of 0.85, 0.80, 0.75, 0.70, 0.65, 0.60. For the sake of convenience, we only consider the simple linear regression models for X and Y . We set both \mathbf{z}_i and \mathbf{w}_i as $\pm(i-1)/n$, $i = 1, 2, \dots, n/2$, symmetrically around the point zero. For each case, we try simulation when $\mathbf{z} = \bar{\mathbf{z}} = 0$ and $\mathbf{w} = \bar{\mathbf{w}} = 0$.

The equally chosen sample sizes n and m are 10, 20, 30, and bandwidth h is selected as $(m+n)^{-0.2}$. The number of pairs of samples generated for each combination of $P(\underline{\theta})$ and $n(=m)$ is 1000. For each independent random samples, the approximated bootstrap tests were constructed by each method with bootstrap replications $B = 1000$ times. Also the used significance levels α is 0.05 and 0.10. Tables 4.1-4.3 give the actual powers of the approximated tests, respectively. The graphs for some cases of tables 4.1-4.3 are given Figures 4.1-4.2.

Figure 4.1 represents the plot of the actual powers against reliabilities when $n = m = 10$ in case II. Figure 4.1 illustrates that the test function $\phi_{per-t}(\cdot)$ are nearly always better than test function $\phi_{MW}(\cdot)$ regardless of the reliabilities.

Figure 4.2 represents the plot of the actual powers against sample sizes when $P(\underline{\theta}) = 0.8$ in Case II. From Figure 4.2, we show that the actual powers of $\phi_{per-t}(\cdot)$ is better than that of $\phi_{MW}(\cdot)$. Since simulation results for other values of reliabilities are similar, we don't report here.

Consequently, it is observed the test based on $\phi_{per-t}(\cdot)$ performs better.

Bootstrap methods can require even more computing than Park et.al.'s method, and up to hundreds to thousands of times more computing time than using Park et.al.'s method. However, with high speed computers, even this may not be a severe problem, and the improvement may often be worth the extra cost.

Table 4.1 The Actual Powers of the Tests for $H_0 : P(\underline{\theta}) = 0.90$
 v.s. $H_1 : P(\underline{\theta}) = P_0(\underline{\theta})$ in Case I.

$n(=m)$	$P_0(\underline{\theta})$	$\phi_{MW}(\cdot)$	$\phi_{PER}(\cdot)$	$\phi_{Per-t}(\cdot)$
10	0.85	0.2650(0.1200)	0.2830(0.1140)	0.4120(0.1980)
	0.80	0.4070(0.2380)	0.4120(0.2270)	0.5600(0.3070)
	0.75	0.5890(0.4200)	0.5760(0.3960)	0.6980(0.4480)
	0.70	0.7130(0.5400)	0.6950(0.5270)	0.7940(0.5370)
	0.65	0.8310(0.7010)	0.7910(0.6930)	0.8540(0.6280)
	0.60	0.8680(0.7680)	0.8570(0.7470)	0.8910(0.6920)
20	0.85	0.4030(0.2240)	0.3980(0.1910)	0.5170(0.3450)
	0.80	0.6310(0.4490)	0.6360(0.4370)	0.7210(0.5740)
	0.75	0.8560(0.7240)	0.8420(0.6920)	0.9020(0.8050)
	0.70	0.9220(0.8360)	0.9080(0.7970)	0.9840(0.8750)
	0.65	0.9810(0.9340)	0.9730(0.9260)	0.9910(0.9600)
	0.60	0.9950(0.9770)	0.9910(0.9610)	0.9970(0.9800)
30	0.85	0.4820(0.2810)	0.4880(0.2710)	0.5840(0.4200)
	0.80	0.7840(0.6200)	0.7790(0.5940)	0.8440(0.7310)
	0.75	0.9560(0.8960)	0.9500(0.8710)	0.9700(0.9270)
	0.70	0.9860(0.9610)	0.9790(0.9650)	0.9930(0.9710)
	0.65	0.9970(0.9900)	0.9960(0.9860)	1.0000(0.9950)
	0.60	0.9990(0.9980)	0.9990(0.9960)	1.0000(0.9990)

Table 4.2 The Actual Powers of the Tests for $H_0 : P(\theta) = 0.90$
v.s. $H_1 : P(\theta) = P_0(\theta)$ in Case II.

$n(=m)$	$P_0(\theta)$	$\phi_{MW}(\cdot)$	$\phi_{PER}(\cdot)$	$\phi_{Per-t}(\cdot)$
10	0.85	0.2440(0.1290)	0.2850(0.1210)	0.3980(0.1960)
	0.80	0.4300(0.2380)	0.4290(0.2490)	0.5570(0.3080)
	0.75	0.5900(0.3850)	0.5800(0.3630)	0.6830(0.4160)
	0.70	0.7220(0.5320)	0.7230(0.5190)	0.7790(0.5340)
	0.65	0.8080(0.6260)	0.7950(0.5900)	0.8440(0.5920)
	0.60	0.8520(0.7410)	0.8410(0.7280)	0.8920(0.6880)
20	0.85	0.3840(0.2130)	0.3810(0.2270)	0.4960(0.3150)
	0.80	0.6690(0.4810)	0.6640(0.4700)	0.7640(0.6040)
	0.75	0.8290(0.6860)	0.8150(0.6720)	0.8880(0.7720)
	0.70	0.9500(0.8590)	0.9480(0.8310)	0.9660(0.9190)
	0.65	0.9820(0.9400)	0.9770(0.9300)	0.9890(0.9650)
	0.60	0.9960(0.9800)	0.9930(0.9790)	0.9980(0.9870)
30	0.85	0.4610(0.2720)	0.4570(0.2540)	0.5700(0.3780)
	0.80	0.7840(0.6020)	0.7820(0.5910)	0.8520(0.7200)
	0.75	0.9300(0.8590)	0.9220(0.8440)	0.9620(0.8940)
	0.70	0.9860(0.9610)	0.9850(0.9570)	0.9940(0.9830)
	0.65	0.9980(0.9950)	0.9960(0.9880)	0.9980(0.9950)
	0.60	0.9990(0.9980)	0.9990(0.9960)	1.0000(0.9990)

Table 4.3 The Actual Powers of the Tests for $H_0 : P(\theta) = 0.90$ v.s. $H_1 : P(\theta) = P_0(\theta)$ in Case III.

$n(=m)$	$P_0(\theta)$	$\phi_{MW}(\cdot)$	$\phi_{PER}(\cdot)$	$\phi_{Per-t}(\cdot)$
10	0.85	0.2710(0.1360)	0.2790(0.1270)	0.3870(0.1930)
	0.80	0.4530(0.2650)	0.4610(0.2590)	0.5820(0.3410)
	0.75	0.5980(0.3980)	0.5450(0.3800)	0.6940(0.4150)
	0.70	0.6980(0.5250)	0.6790(0.5140)	0.7710(0.5250)
	0.65	0.8350(0.6540)	0.7980(0.6350)	0.8700(0.6120)
	0.60	0.8640(0.7480)	0.8580(0.7320)	0.8700(0.6710)
20	0.85	0.4170(0.2220)	0.4050(0.1920)	0.5300(0.3440)
	0.80	0.6580(0.4620)	0.6380(0.4570)	0.7520(0.5690)
	0.75	0.8280(0.6920)	0.8180(0.6810)	0.8850(0.7780)
	0.70	0.9290(0.8570)	0.9270(0.8480)	0.9560(0.8970)
	0.65	0.9810(0.9410)	0.9780(0.9360)	0.9910(0.9660)
	0.60	0.9930(0.9800)	0.9880(0.9740)	0.9960(0.9840)
30	0.85	0.4640(0.2930)	0.4590(0.2810)	0.5750(0.4030)
	0.80	0.8080(0.6340)	0.7880(0.5810)	0.8740(0.7410)
	0.75	0.9270(0.8660)	0.9260(0.8410)	0.9500(0.9060)
	0.70	0.9830(0.9610)	0.9780(0.9500)	0.9870(0.9740)
	0.65	0.9990(0.9950)	0.9980(0.9870)	0.9990(0.9970)
	0.60	0.9990(0.9980)	0.9990(0.9990)	0.9990(0.9990)

References

1. Becher, H.(1993). Reader reaction : Bootstrap hypothesis testing procedures. *Biometrics*, 49, 1268-1272.
2. Beran, R. (1988). Prepivoting test statistics : A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83, 687-697.
3. Duncan, A. G. (1986). *Quality and Industrial Statistics(5th ed.)*, Homewood, IL:Richard D. Irwin.
4. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7, 1-26.
5. Efron, B. (1981). Nonparametric Standard Errors and Confidence Intervals (With Discussion), *Canadian Journal of Statistics*, 9, 139-172.
6. Guttman, I., Johnson, R., Bhattacharyya, G.K., and Reisser, B. (1988). Confidence Limits for Stress-Strength Models with explanatory variables, *Technometrics*, 30, 161-168.

7. Hall, P. and Wilson, S.R.(1991). Two guidelines for bootstrap hypothesis testing, *Biometrics*, 47, 757-762.
8. Hinkely, D.V. (1988). Bootstrap methods.(With discussion), *Journal of the Royal Statistical Society, Series B*, 50, 321-337.
9. Hinkely, D. V.(1989). Bootstrap Significance tests, *Proceeding 47th Session of the International Statistical Institute*, Paris, 29 August-6 September 1989, 3, 65-74.
10. Park, E. S., Kim, J. J., Park, S. H.(1996). A Comparison of Confidence Intervals for the Reliability of the Stress-Strength Model with Explanatory Variables, *The Korean Communications in Statistics*, 3(1), 73-85.

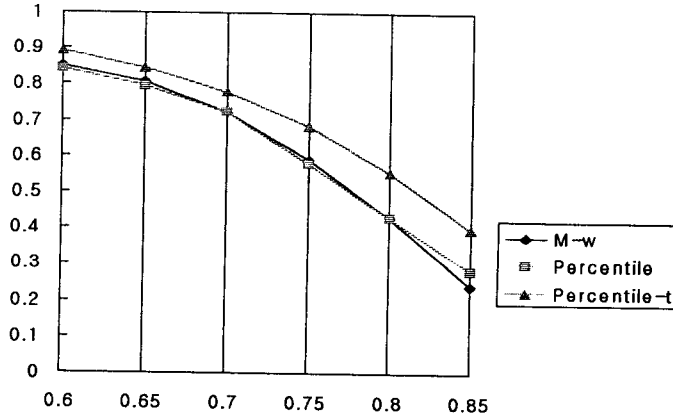


Figure 4.1 Plot of the actual powers against reliabilities when $n = m = 10$ in case II.

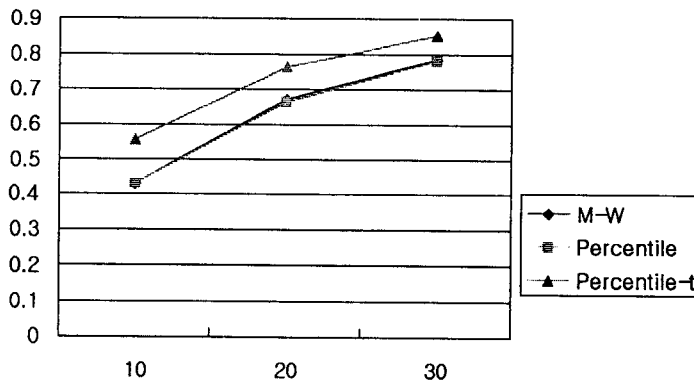


Figure 4.2 Plot of the actual powers against sample sizes when $P_0(\theta) = 0.8$ in Case II.