SEQUENTIAL PROCEDURES FOR ONE–WAY MEASUREMENT ERROR STUDIES

by

Richard W. Andrews
University of Michigan

Andrew J. Barnett
Rain Bird Sprinkler Manufacturing Corporation

David A. Andrews
University of Dallas

KEYWORDS:

Random Effects Model, Gage R & R, Confidence Intervals, Bayesian Procedures

ABSTRACT

Measurement error studies estimate the measurement variance of gages used to check the quality of manufactured products. These studies are often executed as designed experiments with a fixed sample size. We show that sequential procedures can be used in measurement error studies. In a one–way study the levels are different productions of the same part and each part is independently measured m times. The sequential procedure takes as its incremental step either an additional part with m measurements or an additional measurement on each of the existing parts. Sequential probability ratio tests, confidence intervals, and Bayesian sequential procedures are used to determine the acceptability of the measurement error variance as a proportion of the part to part variance. Simulation analyses compare these sequential procedures.

Presented
at
American Statistical Association
Joint Statistical Meetings
August 11, 1999

I. INTRODUCTION

In manufacturing it is important to produce high quality parts and assembled products. To determine if a product meets a standard of high quality it is necessary to take one or more measurements. These measurements are taken according to some measurement system and all measurement systems have measurement errors. These measurement errors are controlled by calibration of the measurement devices and by measurement error studies. Quality certifications, such as ISO 9000, have focused greater attention on calibration and measurement studies.

In most measurement error studies a fixed sample size procedure is executed; however, there can be a substantial savings in time and effort by employing sequential sampling procedures. The application of sequential sampling procedures to measurement error studies has been reported in Barnett and Andrews [1] and Andrews, Barnett, and Andrews [2]. The purpose of this paper is to continue the investigation of sequential sampling procedures as applied to measurement studies. Throughout this study we assume that the measurement device has been calibrated to zero bias and therefore we focus on the measurement variance. In [2] we introduced the concept of a *data engine* for a measurement error study. The term *data engine* is a way to refer to the measurement error study procedures and the statistical model for the data and subsequent analysis. The investigation in [2] considered sequential sampling procedures applied to two of the most basic *data engines*. The first one had independent unbiased observations taken on a known standard, which meant the only unknown quantity was the measurement error standard deviation (σ_m). The second *data engine* relaxed the assumption of a known standard and therefore added another unknown parameter, (μ), the true measured value of the part. In both these cases a single item was measured repeatedly. The results of that study showed that a sequential confidence interval procedure introduced in [1] was as good as the more sophisticated sequential ratio tests and Bayes procedures.

The *data engine* we use throughout this paper is a balanced one way random effects analysis of variance model. This has been suggested in Montgomery[3] and AIAG [4] and is used in practice. For this *data engine* multiple parts from the same production system are repeatedly measured. We assume no bias in the measurement device and that a single operator is taking all the measurements. Therefore, the three unknown quantities are $(\mu, \sigma_p, \sigma_m)$ as defined by the following one–way random effects model.

$$Y_{i,j} = \mu + \delta_i + \epsilon_{i,j}$$

$$i = 1, 2, \ldots, k = \text{ number of parts}$$

$$j = 1, 2, \ldots, m = \text{ number of measurements per part}$$

$$Y_{i,j} = \text{ the } j^{th} \text{ measurement on the } i^{th} \text{ part}$$

$$\mu = \text{ the mean measurement of the production}$$

$$\delta_i \sim i.i.d.\ N(0, \sigma_p^2)$$

$$\epsilon_{i,j} \sim i.i.d.\ N(0, \sigma_m^2)$$

So the two unknown quantities of interest are the part to part standard deviation, σ_p, and the measurement error standard deviation, σ_m. As part of a measurement error study a statement needs to made about the ratio, $\frac{\sigma_m}{\sigma_p}$. That is, we need to know the size of the measurement standard deviation as a proportion of the part to part standard deviation. We will refer to this ratio as the measurement error ratio (MER). If the MER is too large the measurement on the part may not reflect its true quality because of contamination by the measurement error.

There are no specific numerical standards for MER which is unlike the P/T ratio given by

$$P/T = \frac{6\sigma_m}{\text{Tolerance}}.$$

The AIAG[4] gives guidelines for P/T which state that a P/T less than .10 indicates an O.K. gauge and that any P/T less than .30 may be acceptable. We will apply these same standards to the MER by assuming that tolerance is set at $6\sigma_p$. This allows us to use .30 as the cut-off value for MER. Any MER less than .30 is acceptable and any value greater is not.

Given this structure we investigate four different sequential procedures. They are: (1) a sequential probability ratio test, (2) an adjusted sequential probability ratio test, (3) a confidence interval procedure, and (4) a Bayesian procedure. Simulation is used to compare these procedures for settings of MER from 0.05 to 0.40. In all cases we note the proportion of correct decisions where it is correct to accept the measurement system if the MER is less than .30 and it is a correct decision to reject the measurement system if the MER is greater than or equal to .30. In addition we report the sample size required to come to a conclusion. The sample size is reported by noting three numbers from the simulation. The average number of parts needed, $\bar{k}$, the average number of repeated observations per part, $\bar{m}$, and the average total sample size $\bar{t}$ ($t = km$).

Sequential sampling on a random effects one–way layout adds the complexity as to how additional observations are selected. In all cases we start with three parts, ($k = 3$), and eight observations on each part, ($m = 8$). If a decision is not reached at any stage the sequential step can be in either of two directions. One direction is that we can select a new part and take m observations on that part. The other direction is that we can increase the number of observations by one on all the parts in our study. The first procedure increments on k and the second procedure increments on m. If a decision is not made by the time we reach either $k = 10$ or $m = 30$, we stop sampling and use the usual estimates to make a decision. The next four sections describe each class of procedures and compares their results.

II. SEQUENTIAL PROBABILITY RATIO TEST

In Johnson[5], Johnson[6], and Ghosh[7] sequential probability ratio tests are given for testing hypotheses on the ratio of variances from random effects models. The procedures are the same even though they were derived in different ways. The incrementing at a sequential step is either in the k (addition part) direction or in the m (additional repeated measure) direction but not both.

In this paper we use their test procedure but add an additional step which allows us to increment in either direction. This adjustment to the reference procedure destroys the theoretical properties of convergence and creates a situation in which the size of the two types of error are not automatically controlled. However, we are not concerned with convergence since we truncate the procedure and we will use simulation to investigate the error probabilities.

Our procedure uses a hypothesis structure that contrasts an unacceptable MER with an acceptable one.

$$H_0 : \text{MER} = \frac{\sigma_m}{\sigma_p} \geq 0.3 = h_0$$

$$H_a : \text{MER} = \frac{\sigma_m}{\sigma_p} \leq 0.1 = h_1$$

The test statistic is

$$l(k,m) = \frac{m \sum_{i=1}^{k} (\bar{Y}_i - \bar{\bar{Y}})^2}{\sum_{i=1}^{k} \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_i)^2} = \frac{SST}{SSE}$$

Where,

$$\bar{Y}_i = \frac{1}{m} \sum_{j=1}^{m} Y_{ij}$$

$$\bar{\bar{Y}} = \frac{1}{km} \sum_{i=1}^{k} \sum_{j=1}^{m} Y_{ij}$$

The sequential procedure will be to start with $k = 3$ and $m = 8$ and use the following decision rule.

$$\text{If } l(k,m) \leq l_l(k,m), \text{ Accept } H_0$$

$$\text{If } l(k,m) \geq l_u(k,m), \text{ Accept } H_a$$

$$\text{Otherwise, take a sequential step.}$$

Where,

$$l_l(k,m) = \frac{\phi_1 B^{2c}(\frac{\phi_0}{\phi_1})^{p_2 c} - \phi_0}{1 - B^{2c}(\frac{\phi_0}{\phi_1})^{p_2 c}}$$

$$l_u(k,m) = \frac{\phi_1 A^{2c}(\frac{\phi_0}{\phi_1})^{p_2 c} - \phi_0}{1 - A^{2c}(\frac{\phi_0}{\phi_1})^{p_2 c}}$$

$$\phi_0 = 1 + m(\frac{1}{h_0^2})$$

$$\phi_1 = 1 + m(\frac{1}{h_1^2})$$

$$p_1 = k - 1$$

$$p_2 = k(m-1)$$

$$c = (p_1 + p_2)^{-1}$$

$$A = \frac{1-\beta}{\alpha}$$

$$B = \frac{\beta}{1-\alpha}$$

The sequential step can be in either of two directions. We can either sample a new part and take m measurements on that part or we can take an additional observation on each of the k parts. The first step will be referred to as incrementing on k and the alternative step will be incrementing on m. Therefore, to run a test we must set the values of α and β and state the rule for choosing the increment.

An extensive simulation study indicated that reasonable values for the size of the two errors are $\alpha = .05$ and $\beta = .20$. This is reasonable since rejecting a true null hypothesis means we are approving a faulty measurement system. This is a serious error and therefore we hold the size of this error to less than 0.05. Whereas accepting a false null hypothesis results in disapproving a good measurement system and this is a less serious error and therefore $\beta = 0.20$ is reasonable.

As the criterion for determining which direction to increment we will calculate the usual components of variance estimates of σ_p^2 and σ_m^2.

$$\hat{\sigma}_m^2 = \frac{SSE}{mk - k}$$

$$\hat{\sigma}_p^2 = max\{0, \frac{\frac{SST}{k-1} - \hat{\sigma}_m^2}{m}\}$$

Then if the ratio $\frac{\hat{\sigma}_m}{\hat{\sigma}_p}$ is greater than $(h_0 + h_1)/2$ we increment on m, and if $\frac{\hat{\sigma}_m}{\hat{\sigma}_p}$ is less than $(h_0 + h_1)/2$ we increment on k.

The simulation study will set $\mu = 10.0$ which will have no effect on the results. The value of σ_p will be set at 1.0 and σ_m will be set at values to give ratios from 0.05 to 0.40. The increments on m and k will proceed until either a decision is reached or k reaches 10 or m reaches 30. If either k or m reaches these limits the decision will be made by accepting that hypothesis which is closest to the ratio determined by the usual estimates.

Each setting of the ratio will be simulated with 1000 replications and the reported values are given in Table 1. For each MER the table gives the proportion correct, the average number of parts used (K–BAR), the average sample size per part (M–BAR) and the average total sample size (T–BAR).

In addition, the average proportion correct is calculated by equaling weighting each setting of MER. Likewise the average total sample size is found and then a criterion is calculated by dividing the average total sample size by the average proportion correct. From Table 1 we see that for this procedure and for this setting of (α, β) this criterion is approximately 83. This criterion, which trades off correct decisions with sample size requirements, has the interpretation of the sample size needed to make 100% correct decisions over the values of MER chosen. This criterion was used to choose the (α, β) values and will provide a comparison with other procedures.

Figure 1 gives the correct proportions at the assumed MER values. The proportion correct at MER = .10 is approximately .80 which is what was expected with a β equal to .20. However, the percent correct at MER = .30 is slightly larger than the .95 expected based on the α used. However, this procedure is not acceptable for the decisions we need to make because the proportion correct for values close to but less than .30 are extremely low. This is consistent with the findings in [2] and it results because we state that any MER less than .30 is acceptable. In the next section adjustments to this sequential test slightly improve this situation.

III. ALTERNATIVE TESTS

In an attempt to improve on the results of the last section we investigate three alternative ways of approaching this hypothesis structure. The purpose is to see if we can increase the probability of accepting the measurement system when the MER is close but smaller than the cut-off value of 0.30; however, we do not want to greatly increase the sample size. We again use the criterion value as a guide to balance off sample size with proportion of correct decisions.

Figure 2 shows the of values $l_l(k,m)$ and $l_u(k,m)$ as you increase m and k. Notice the erratic jumps are from m = 30 at k to m = 8 at k+1. In an attempt to increase the probability of accepting H_a when the true MER is 0.29, we arbitrary reduce the upper limit to be half way between the lower limit and the old upper limit. This is shown in Figure 3. The resulting test has some favorable results over the results given in the last section.

Compare Table 2 with Table 1 and Figure 4 with Figure 1. We see that using this reduced upper limit, the criterion value has decreased from approximately 83 to 75. This is due to both smaller sample sizes and a slightly increased average proportion correct. The proportion correct at MER = 0.29 has increased from .021 to .077; still not very high. The very critical proportion correct at MER = 0.30 has decreased from .984 to .947. This is still very close to the 5% alpha that was set.

As a second alternative we employed a multiple hypothesis structure as suggested by Armitage[8]. This procedure helped greatly in the basic *data engines* as reported in [2] but as you will read we had much less success with this *data engine.*

The structure we used was to consider three pairs of hypotheses. They were:

STRUCTURE 1:

$$H_0 : \text{MER} = \frac{\sigma_m}{\sigma_p} \geq 0.3 = h_0$$

$$H_a : \text{MER} = \frac{\sigma_m}{\sigma_p} \leq 0.1 = h_1$$

STRUCTURE 2:

$$H_m : \text{MER} = \frac{\sigma_m}{\sigma_p} \geq 0.2 = h_m$$

$$H_a : \text{MER} = \frac{\sigma_m}{\sigma_p} \leq 0.1 = h_1$$

STRUCTURE 3:

$$H_0 : \text{MER} = \frac{\sigma_m}{\sigma_p} \geq 0.3 = h_0$$

$$H_m : \text{MER} = \frac{\sigma_m}{\sigma_p} \leq 0.2 = h_m$$

The decision rule is:

Accept that MER = .3, if H_0 is accepted in Structure 1 and H_0 is accepted in Structure 3

Accept that MER = .1, if H_a is accepted in Structure 1 and H_a is accepted in Structure 2

Accept that MER = .2, if H_m is accepted in Structure 2 and H_m is accepted in Structure 3

Otherwise, increment your sample in the usual way.

In using this procedure anytime we accept MER = .2, we consider that acceptance in the same manner as when we accept MER = .1; that is, we conclude the measurement system is properly functioning. The results of this test are given in Table 3 which shows that the value of the criterion is high at 147 and that the proportion correct at MER = 0.29 has not increased. Therefore, this method has not proved effective for this *data engine.*

As a last alternative to the sequential test we make the simple adjustment of using the alternative hypothesis of

$$H_a : \text{MER} = \frac{\sigma_m}{\sigma_p} \leq 0.2 = h_1$$

The results of this test are given in Table 4. This adjustment increases the proportion correct at 0.29 to 0.105 but at the expense of larger sample sizes as indicated by a criterion value of 134. In the next section we leave ratio testing to investigate a confidence interval procedure.

IV. CONFIDENCE INTERVAL PROCEDURE

Similar to the presentation in [1] and [2], we introduce a sequential confidence interval procedure to determine if the MER is acceptable, where acceptable means any value less that 0.30. Using confidence intervals in a sequential setting was first considered in Knudsen[9].

To develop a confidence interval for MER we use the following procedure. See Burdick and Graybill [10] for details.

Using the model given by (1) a $(1-2\alpha) \times 100\%$ confidence interval for MER $= \frac{\sigma_m}{\sigma_p}$ is

$$[\text{Lower, Upper}] =$$

$$[\sqrt{\frac{m}{U^* - 1}}, \sqrt{\frac{m}{L^* - 1}}]$$

where,

$$L^* = \frac{\text{MST}}{\text{MSE} \times F_{\alpha,n_1,n_2}}$$

$$U^* = \frac{\text{MST}}{\text{MSE} \times F_{1-\alpha,n_1,n_2}}$$

$$\text{MST} = \frac{\text{SST}}{k-1}$$

$$\text{MSE} = \frac{\text{SSE}}{mk-k}$$

$$n_1 = k - 1$$

$$n_2 = mk - k$$

$$F_{\alpha,n_1,n_2} = \text{the F value with } \alpha \text{ in upper tail}$$

For each value of MER reported in Tables 5, 6, and 7, 1000 replications were simulated. At each stage in the sequential procedure, if the upper confidence limit (Upper) is less than 0.30, the MER is considered acceptable. If the lower confidence limit (Lower) is greater than 0.30, the MER is considered unacceptable. If the confidence interval contains 0.30, then a sequential step is executed.

In the sequential step, if 0.30 fell in the upper half of the interval the sample size was incremented by taking an additional part; that is incrementing on k. If 0.30 fell in the lower half of the interval one additional reading was taken on each of the existing parts in the study. That is, there is an increment on m.

From Tables 5, 6, and 7 and Figure 5 we see that there has been a substantial increase in the proportion correct at MER = 0.29. However, there has been some decrease in the proportion correct at the high and low values of MER. Overall, these confidence interval procedures work better than any of the sequential ratio tests attempted in the previous two sections.

Comparing the three confidence levels (80%, 90%, 95%) shows very little difference and therefore to be consistent with [2] we recommend using the 90% level.

V. BAYESIAN PROCEDURE

As a starting point for developing a Bayesian procedure we use a result from Hill [11] which gives the posterior distribution of $\phi = \frac{\sigma_p^2}{\sigma_m^2}$.

$$f(\phi \mid \text{data}) \propto \frac{\phi^{-1}(\phi + m^{-1})^{\frac{1-k}{2}}}{(\text{SSE} + \frac{\text{SST}}{1+m\phi})^{\frac{km-1}{2}}}$$

This assumes use of the noninformative prior, σ_p^{-2}. This prior yields an improper posterior distribution. However, we can use this improper distribution since we employ a discrete approximation.

By using the transformation MER $= \theta = \phi^{-.5}$, we obtain the posterior on MER (θ) to be

$$f(\theta \mid \text{data}) \propto \frac{\theta^{-1}(\theta^{-2} + m^{-1})^{\frac{1-k}{2}}}{(\text{SSE} + \frac{\text{SST}}{1+m\theta^{-2}})^{\frac{km-1}{2}}}.$$

Since we do not know the constant of proportionality we use the following procedure. Setting $k = 3$ and $m = 8$ we evaluate $f(\theta \mid \text{data})$ for 100 values of $\theta = .01, .02, \ldots, 1.00$. For values of SSE and SST we equate these sum of squares to their expected values and set $\sigma_p = 1.0$ and $\sigma_m = .05, .10, .15, .20, .25, .275, .29, .30, .31, .35, .40$.

For example, with $\sigma_p = 1.0$ and $\sigma_m = .30$,

$$\text{E[SST]} = (k-1)(\sigma_m^2 + m\sigma_p^2) = 16.18,$$

$$\text{E[SSE]} = k(m-1)\sigma_m^2 = 1.89.$$

Using this procedure we can evaluate $f(\theta \mid \text{data})$ for the array of MER values. By standardizing the 100 values of $f(\theta \mid \text{data})$ we get approximate posterior probabilities for these 100 points. For each MER setting, we calculated the $P[\theta \geq 0.30 \mid \text{data}]$. Table 8 gives these values. Based on this table we decided to use 0.15 and 0.40 as the cut-off values. That is, if $P[\theta \geq 0.30 \mid \text{data}] \geq 0.40$ at any stage we will reject the measurement system. If $P[\theta \geq 0.30 \mid \text{data}] \leq 0.15$ at any stage we will accept the measurement system. If $0.15 < P[\theta \geq 0.30 \mid \text{data}] < 0.40$ we increment the sample on either k or m. The entire decision rule is:

$$P[\theta \geq 0.30 \mid \text{data}] \begin{cases} \geq 0.40 & \text{reject system;} \\ \leq 0.15 & \text{accept system;} \\ < 0.40 \text{ and } > 0.275 & \text{increment on k;} \\ > 0.15 \text{ and } \leq 0.275 & \text{increment on m.} \end{cases}$$

If either $k = 10$ or $m = 30$ are reached the system will be rejected if $P[\theta \geq 0.30 \mid \text{data}]$ is closer to 0.40 than to 0.15; otherwise it will be accepted.

The simulation results of this Bayesian approach are given in Table 9 and Figure 6. Of all the procedures described in this paper the Bayesian procedure performs the best. Note that the criterion is very small at approximately 43. Also note that the proportion correct at MER = 0.29 is 0.261.

In comparing the Bayesian procedure with the confidence interval procedure at 90% we see that the confidence interval procedure has a higher proportion correct at every value of MER but the required sample size is more than double. Therefore, based on this investigation the Bayes procedure is recommended.

REFERENCES

[1] Barnett, A.J. and Andrews, R.W., "Measurement error study procedure featuring variable sample sizes", *Quality Engineering*, 9(2), pp. 259–267, 1996–1997.

[2] Andrews, R.W., Barnett, A.J, and Andrews, D.A., "Measurement error studies using sequential sampling ", *American Statistical Association Proceeding, Section on Quality & Productivity*, 1997.

[3] Montgomery, D. C., *Introduction to Statistical Quality Control*, 3^{rd} ed., Wiley& Sons, New York, 1996.

[4] AIAG, *Measurement Systems Analysis—Reference Manual*, 2^{nd} ed., Automotive Industry Action Group, Southfield, MI, 1995.

[5] Johnson, N. L., "Some notes on the application of sequential methods in the analysis of variance", *Annals of Mathematical Statistics*, 24, 1953, pp. 614–623.

[6] Johnson, N. L., "Sequential procedures in certain component of variance problems", *Annals of Mathematical Statistics*, 25, 1954, pp. 357–366.

[7] Ghosh, B. K., "Sequential analysis of variance under random and mixed models", *Journal of American Statistical Association*, 62, 1967, pp. 1401–1417.

[8] Armitage, P., "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis", *Journal of the Royal Statistical Society*, B, 12, 1950, pp. 137–144.

[9] Knudsen, L.F.,"A method for determining the significance of a shortage", *Journal of the American Statistical Association*, 38, 1943, pp. 466–70.

[10] Burdick, R. K. and Graybill, F. A., *Confidence Intervals On Variance Components*, Marcel Dekker Inc., New York, 1992.

[11] Hill, B. M., "Inference about variance components in the one–way model", *Journal of the American Statistical Association*, 60, 1965, pp. 806–825.

TABLE 1: ALPHA = .05, BETA = .20

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.946	3.170	8.813	28.027
0.100	0.791	3.770	10.689	39.859
0.150	0.490	4.399	14.086	58.092
0.200	0.220	4.310	16.686	67.171
0.250	0.076	3.915	17.063	63.914
0.275	0.029	3.629	16.574	59.602
0.290	0.021	3.483	17.203	59.907
0.300	0.984	3.386	16.096	54.978
0.310	0.988	3.325	15.842	52.852
0.325	0.990	3.265	15.327	51.061
0.350	0.997	3.186	14.218	45.810
0.400	1.000	3.096	13.268	41.610
AVERAGE->	0.628		AVERAGE->	51.907
			CRITERION->	82.698

TABLE 2: REDUCED UPPER LIMIT

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.947	3.056	8.771	26.875
0.100	0.794	3.250	10.980	36.303
0.150	0.533	3.526	13.554	48.157
0.200	0.310	3.484	16.669	58.865
0.250	0.149	3.414	16.263	56.409
0.275	0.092	3.298	16.442	55.024
0.290	0.077	3.212	16.925	54.985
0.300	0.947	3.226	16.019	52.548
0.310	0.955	3.175	16.250	52.476
0.325	0.963	3.165	15.106	48.630
0.350	0.979	3.121	14.939	47.226
0.400	0.992	3.065	12.854	39.884
AVERAGE->	0.645		AVERAGE->	48.115
			CRITERION->	74.616

TABLE 3: USING H(MIDDLE)

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.941	3.653	9.372	33.690
0.100	0.792	5.646	12.640	61.556
0.150	0.457	6.040	20.046	95.538
0.200	0.216	5.209	24.874	107.074
0.250	0.054	3.985	28.747	107.174
0.275	0.031	3.719	29.344	105.164
0.290	0.019	3.526	29.597	101.750
0.300	0.985	3.413	29.664	99.030
0.310	0.992	3.464	29.803	101.950
0.325	0.996	3.282	29.912	97.580
0.350	0.999	3.213	29.978	96.170
0.400	1.000	3.079	30.000	92.370
AVERAGE->	0.624		AVERAGE->	91.587
			CRITERION->	146.892

TABLE 4: H(a): MER <= 0.2

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.959	3.323	8.965	29.760
0.100	0.863	4.351	11.156	45.868
0.150	0.670	5.455	15.458	71.133
0.200	0.447	5.670	20.537	94.028
0.250	0.229	5.074	25.244	109.172
0.275	0.137	4.558	27.154	109.042
0.290	0.105	4.348	27.888	110.134
0.300	0.923	4.122	28.457	108.538
0.310	0.937	4.074	28.785	110.510
0.325	0.946	3.893	28.899	105.982
0.350	0.973	3.642	29.461	104.002
0.400	0.994	3.343	29.875	99.040
AVERAGE->	0.682		AVERAGE->	91.434
			CRITERION->	134.084

TABLE 5: 95% CONFIDENCE INTERVAL

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.975	3.138	8.623	27.150
0.100	0.893	3.679	10.491	37.882
0.150	0.780	4.462	13.127	54.473
0.200	0.561	5.013	17.908	76.673
0.250	0.367	5.050	22.156	95.472
0.275	0.261	4.894	24.134	101.332
0.290	0.231	4.809	24.701	103.852
0.300	0.780	4.776	24.801	103.301
0.310	0.830	4.508	25.840	103.614
0.325	0.865	4.402	26.482	106.124
0.350	0.913	4.060	27.150	102.527
0.400	0.957	3.610	27.675	96.185
AVERAGE->	0.701		AVERAGE->	84.049
			CRITERION->	119.884

TABLE 6: 90% CONFIDENCE INTERVAL

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.971	3.107	8.609	26.827
0.100	0.900	3.393	10.201	34.577
0.150	0.759	4.133	13.248	52.153
0.200	0.585	4.689	16.698	70.415
0.250	0.366	4.856	21.355	91.021
0.275	0.286	4.807	22.869	97.489
0.290	0.277	4.814	22.950	98.194
0.300	0.759	4.679	23.766	99.224
0.310	0.815	4.481	24.490	100.944
0.325	0.835	4.514	25.128	105.199
0.350	0.881	4.250	26.072	103.963
0.400	0.965	3.746	26.798	99.256
AVERAGE->	0.700		AVERAGE->	81.605
			CRITERION->	116.593

TABLE 7: 80% CONFIDENCE INTERVAL

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.981	3.071	8.401	25.928
0.100	0.914	3.311	9.786	32.431
0.150	0.795	3.693	12.294	44.964
0.200	0.630	4.242	15.454	63.723
0.250	0.416	4.453	19.286	81.448
0.275	0.356	4.381	20.303	83.904
0.290	0.298	4.455	21.353	89.618
0.300	0.756	4.290	22.070	90.191
0.310	0.761	4.325	21.716	90.987
0.325	0.795	4.297	22.596	93.210
0.350	0.876	4.081	23.326	93.848
0.400	0.948	3.797	24.716	94.427
AVERAGE->	0.711		AVERAGE->	73.723
			CRITERION->	103.762

TABLE 8: POSTERIOR PROBABILITIES FOR EQUATED SUM OF SQUARES

SPECIFIED MER	PROBABILITY POSTERIOR GREATER THAN 0.30
0.050	0.000000235
0.100	0.001867
0.150	0.038428
0.200	0.141594
0.250	0.274737
0.275	0.339932
0.290	0.377020
0.300	0.400717
0.310	0.423532
0.350	0.505587
0.400	0.588097

TABLE 9: BAYES PROCEDURE

MER	PROP CORRECT	K-BAR	M-BAR	T-BAR
0.050	0.960	3.017	8.074	24.373
0.100	0.877	3.089	8.632	26.784
0.150	0.695	3.126	8.910	28.001
0.200	0.540	3.197	9.272	30.095
0.250	0.366	3.265	9.241	30.686
0.275	0.306	3.233	9.298	30.722
0.290	0.261	3.204	9.262	30.212
0.300	0.746	3.263	9.480	31.380
0.310	0.776	3.247	9.100	30.008
0.325	0.810	3.258	8.984	29.845
0.350	0.861	3.227	8.804	28.766
0.400	0.912	3.174	8.555	27.401
AVERAGE->	0.676		AVERAGE->	29.023
			CRITERION->	42.944

FIGURE 1: ALPHA = .05, BETA = .20

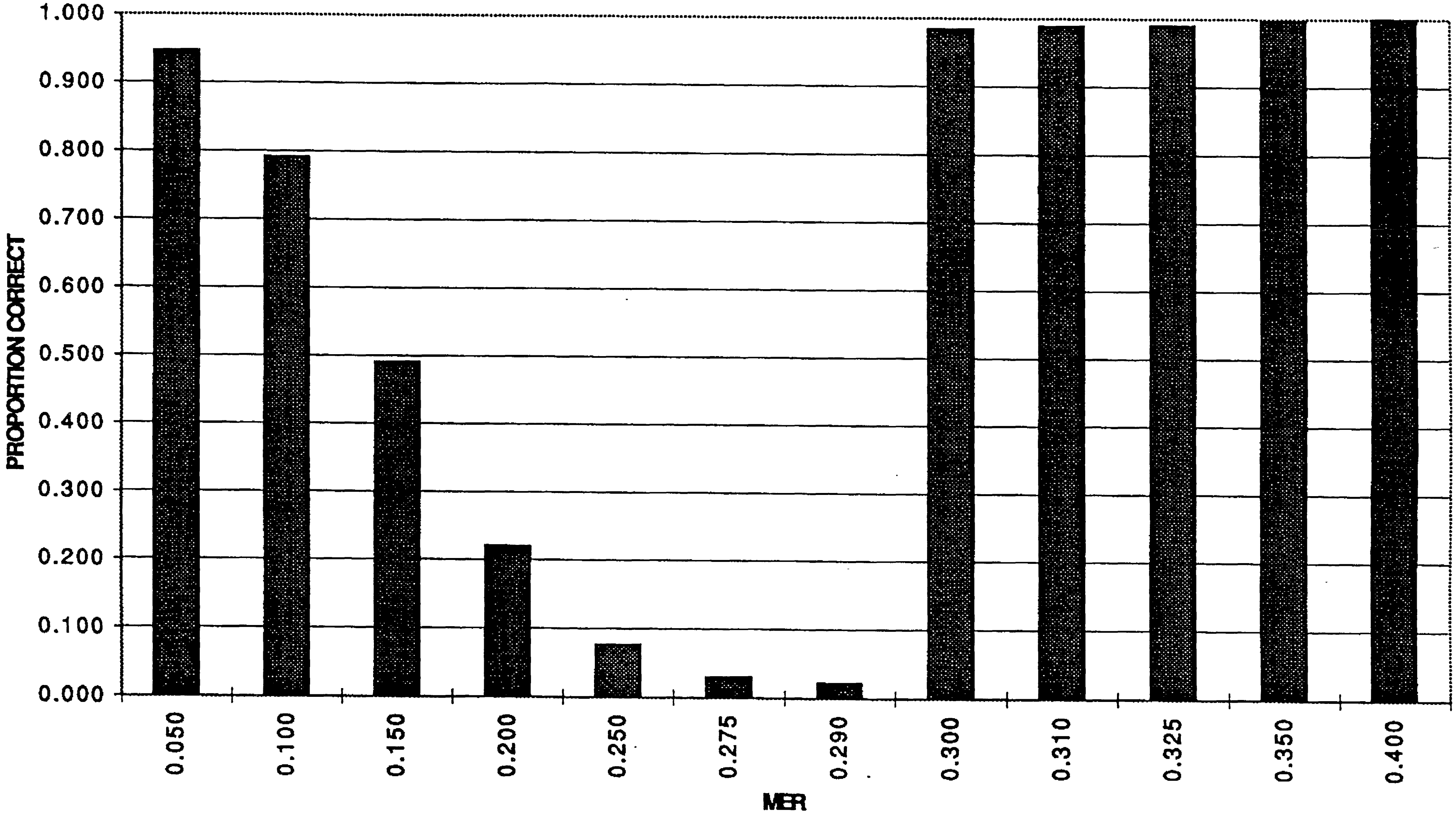

FIGURE 2: TEST LIMITS
ALPHA = .05, BETA = .20

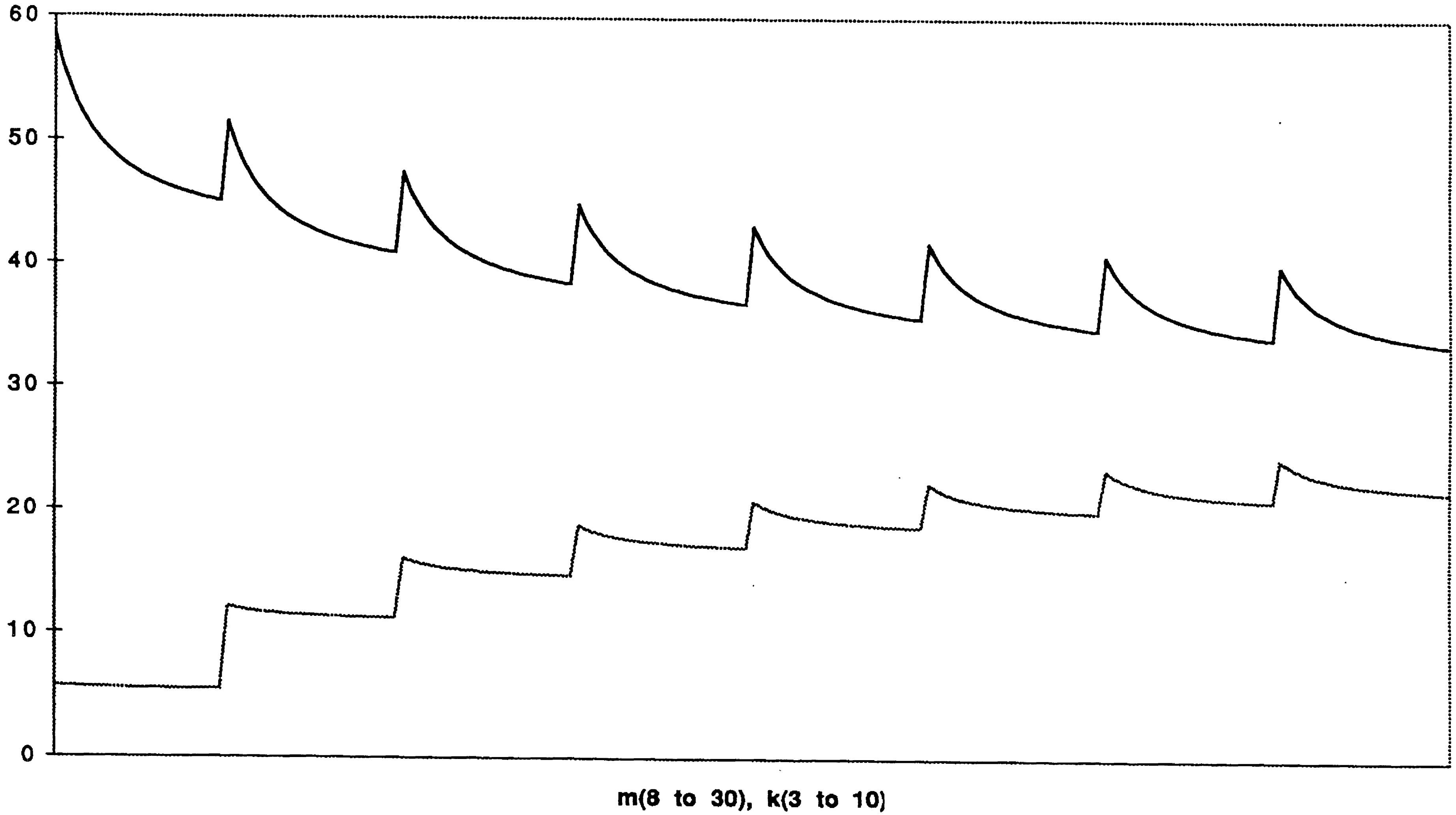

FIGURE 3: TEST LIMITS
REDUCED UPPER LIMIT

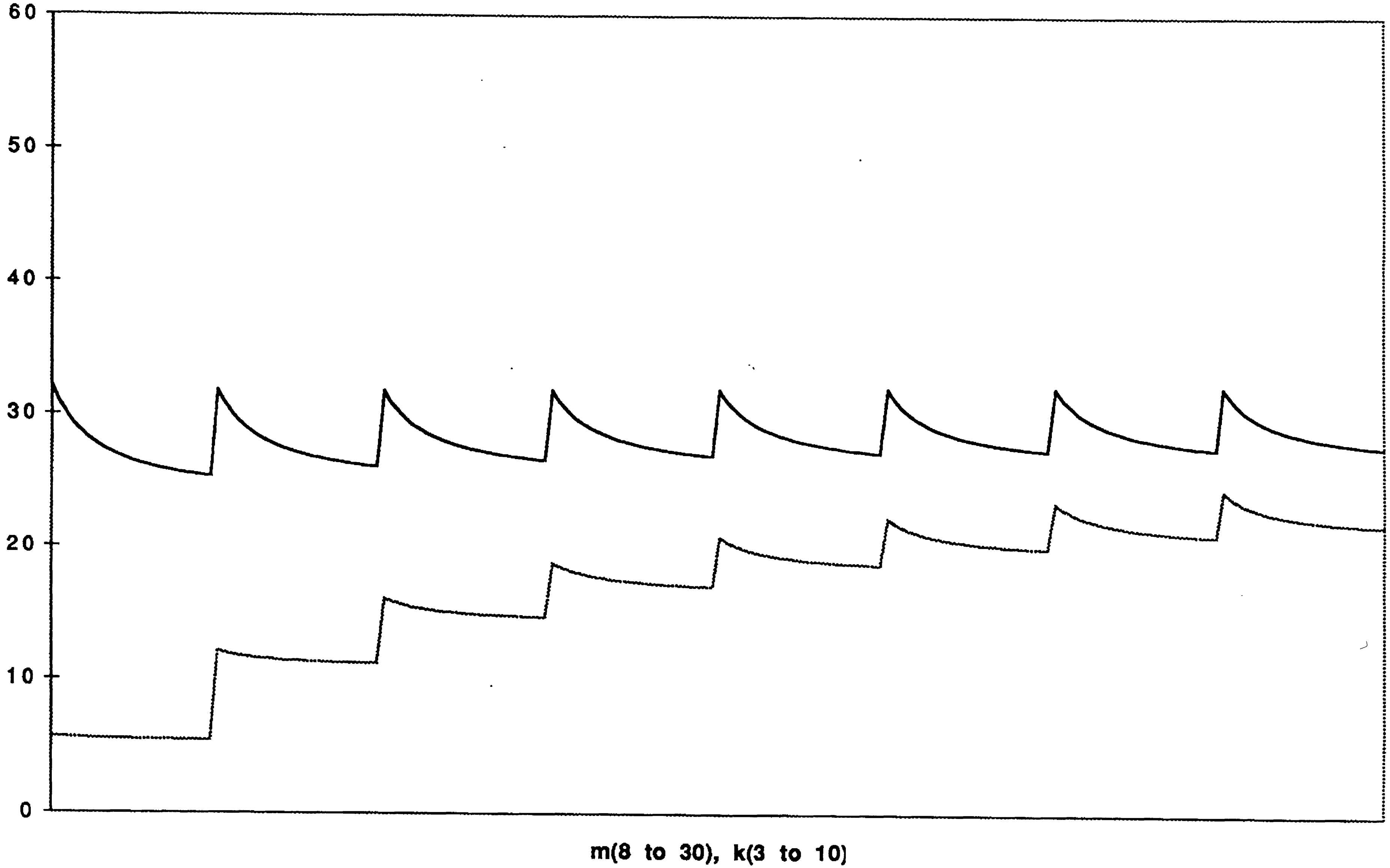

FIGURE 4: REDUCED UPPER LIMIT

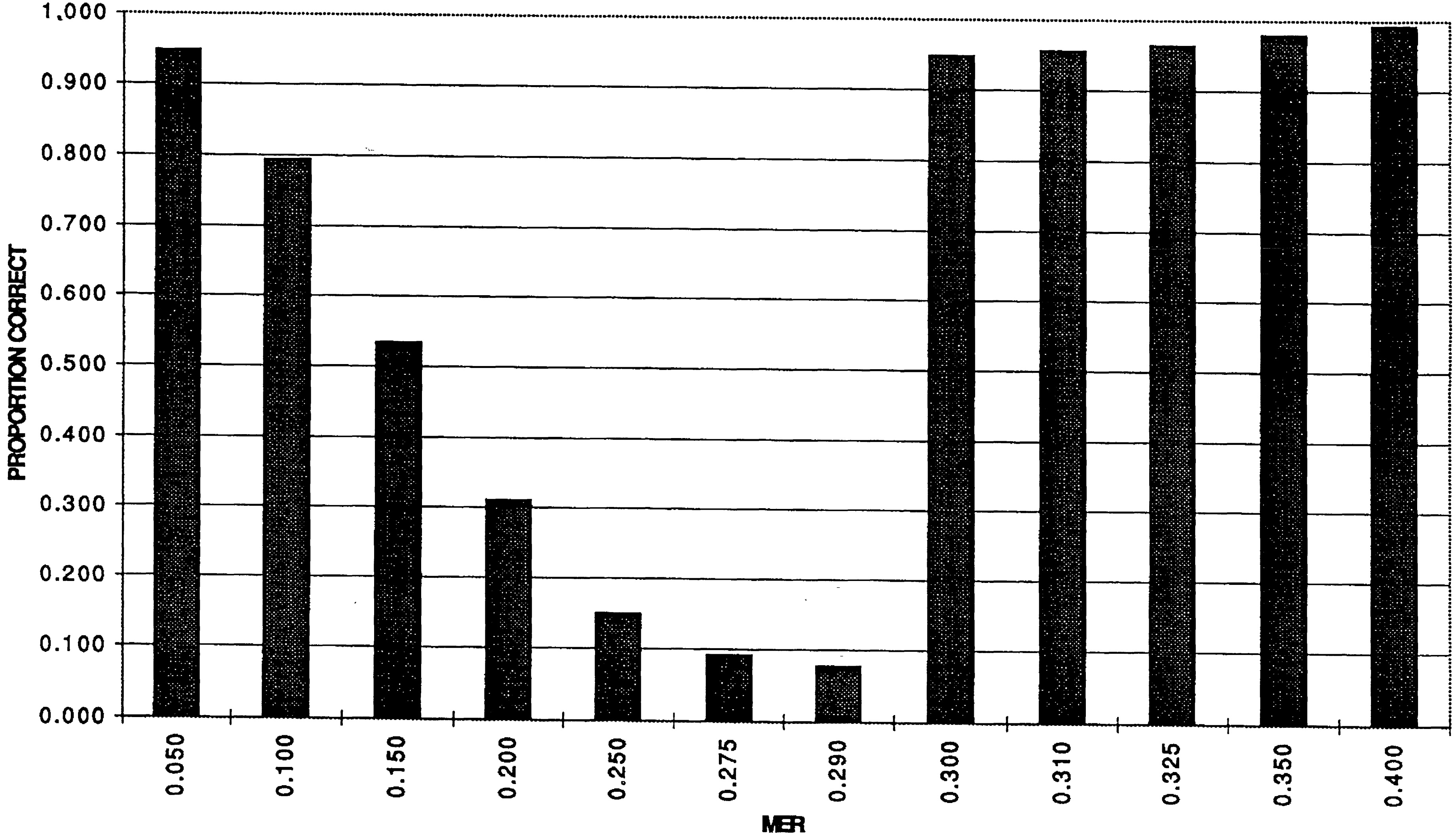

FIGURE 5: COMPARISON OF CONFIDENCE INTERVALS

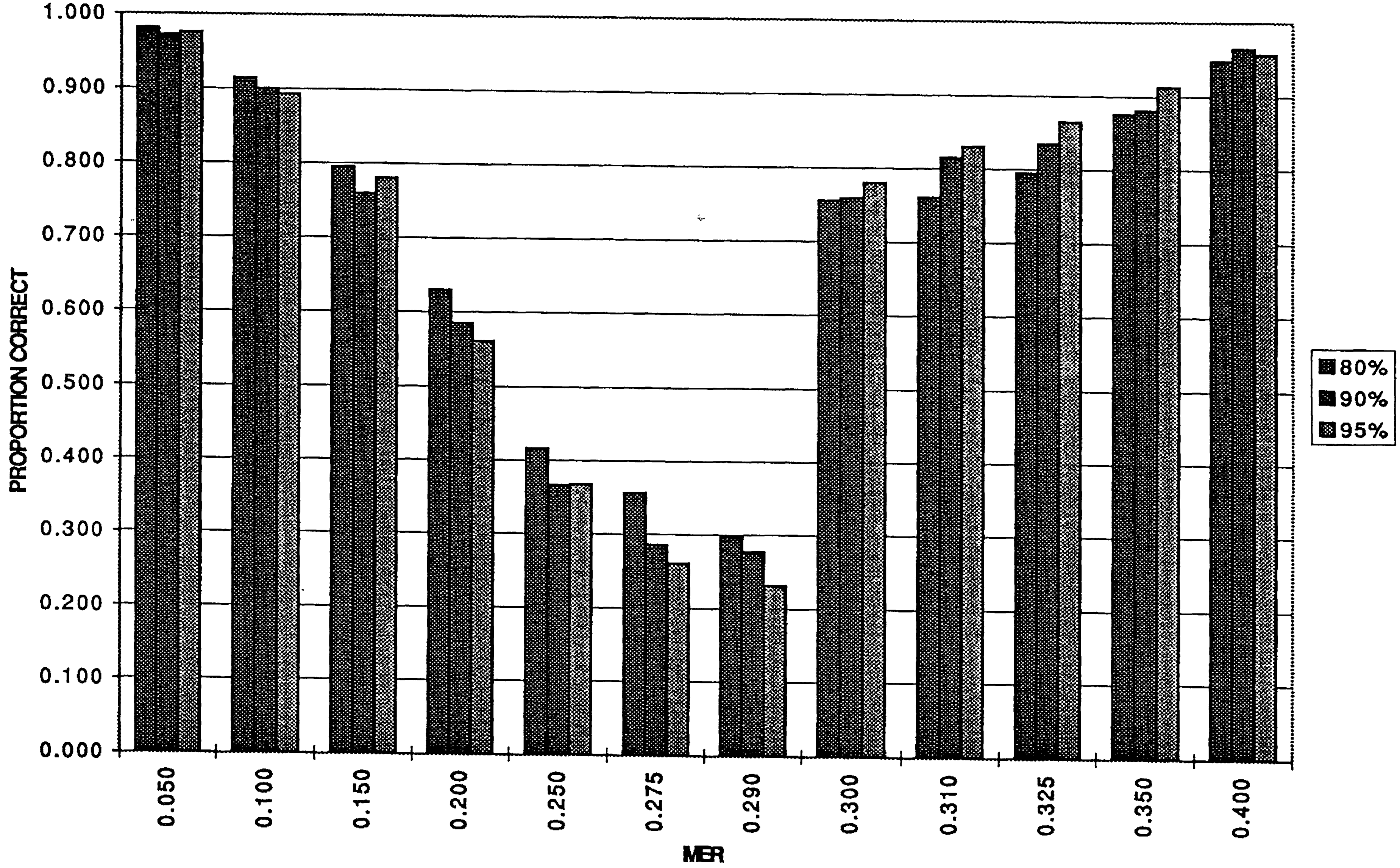

FIGURE 6: BAYES PROCEDURE