Dick Smith

FINAL REPORT

A SURVEY AND CRITICAL REVIEW OF U.S. OIL SPILL DATA RESOURCES WITH APPLICATION TO THE TANKER/PIPELINE TRANSPORT CONTROVERSY

Robert J. Stewart

port to

Dr. Robert K. Davis
U.S. Department of the Interior
Washington, DC

under

Purchase Order A1821
Requisition No. 12345-4150-PDB

September 29, 1976

Revised October 7, 1976

November 29, 1976

ERRATA

"A Survey and Critical Review of U.S. Oil Spill Data Resources with Application to the Tanker/Pipeline Controversy", By Robert J. Stewart, 7 October 1976, Martingale Inc., Cambridge, MA

Page(line)	Was	Should be
ii(15)	4.3	4.4
(line below)	4.4	4.3
1(12)	.. Groussard OTC..	.. Broussard OTC...
28(8)	.. availble..	.. available..
46(8)	inports	imports
49(1)	TABLE 4.3	TABLE 4.4
50(1)	TABLE 4.4	TABLE 4.3
69(5)	963,834	160,639
(line below)	36,000	6,000
71(2 from bot)	216×10^{-6}	42×10^{-6}
72(bottom)	216×10^{-6}	42×10^{-6}

1. INTRODUCTION

There is a growing body of opinion that tankers are, in general, a less desirable transport mode for crude oil than are subsea pipelines. Opinions to this effect have played a role in legal controversies relating to the Department of the Interior's lease schedule. Policy makers are considering measures that would force developers to use pipelines in bringing oil to shore. And even the National Academy of Sciences has opined that pipelines are the preferred transport mode.

It is difficult to identify any one paper or report as the foundation for this belief, but one of the earlier papers that states this conclusion explicitly is the Milz and Broussard OTC paper of 1972. This paper has been referenced several times in recent years. A review of the paper reveals that the basis for their statement was a rather superficial examination of some U.S. Geological Survey and Department of Transportation statistics coupled with an unsubstantiated and probably incorrect assertion that there was a total of "13,000 miles of trunk lines in the marine environment" in 1972. Nowhere did they address the issue of the quality and degree of completeness of the data used to generate the statistics. Nor did they address the subtle analytical problems that accompany such comparisons. Their whole argument required one small paragraph.*

---

* In fairness to these authors, it should be noted that the relative merits of tankers and pipelines was an issue very much tangential to their central theme.

In the interim, a number of related papers have drawn the same conclusion. It is not our contention that all of these papers were equally glib, because we have by no means examined all of the literature. However, of the rather substantial number we have reviewed, none does a significantly better job of addressing the problems associated with using the existing statistics to compare tankers with pipelines. As a general rule, the papers fail to state whether the sample population used to generate the statistics was in fact the entire population or whether it was a subset of the population. Further, if the latter is the case and if the paper mentions this, then it fails to tell how the sampling process might bias the conclusions.

Within this body of literature, as if these omissions weren't enough, there is also a rather distressing reliance upon the ratio of volume spilled per volume handled (historically) as a parameter characterizing alternative transportation modes. Such approaches might have considerable usefulness in cases where the number of incidents is large and where the volume spilled per incident is a random variable with a standard deviation that is small compared to the mean. However, for the highly skewed and squashed out distribution we find in the oil spill business, and for applications where the sample size is small (in Milz and Broussard's paper only 4 pipeline spills were identified), the use of average volume statistics is likely to be very misleading, as we show in the following section.

Thus the purpose of this study is to examine both the data that has become available in recent years and the analytical techniques applied to the data to see if there is indeed a reasonable basis within the data for the preference given pipelines over tankers. Despite the author's predilection to discount the previous literature on the subject, no preference was established initially in support of either the pipeline case or the tanker case. We simply wanted to look at the data to see what could, or could not, be said.

2. AVERAGE VOLUME COMPARISON TECHNIQUES

As we mentioned in the introduction, a substantial portion of the existing literature (e.g., impact statements and the like) make use of comparisons between modes based in some fashion upon an average spill volume statistic. This may be hidden in the analysis, but whenever one sees statements like ".0006 percent of the oil handled by system 'R' will be spilled," or "X will spill Q percent less than Y," then one has entered the world of the average statistic. This author and Professor J. W. Devanney III of M.I.T. have frequently argued against the use of such a statistic in the evaluation of oil spill risk, but apparently this hasn't fallen upon the right ears. It is still possible to find statements like "Pipelines will spill (on the average) 40% less oil in the life of the field than will the alternative, tankers."

We would have no argument if this statement were indeed supported by the data. In fact, we would be very interested in such conclusions. However, such statements are usually based upon small samples and superficial estimation techniques, and as a result, the validity of the 40% value is questionable. In the event that such analyses lead to the adoption of a policy that discriminates against tankers, then we may be imposing economic and perhaps environmental penalties that are not at 1 consistent with the data or in the interest of the public or the developer.

The primary cause of the difficulty in conceptually handling an average spillage statistic springs from the possible variability in the value of the sum of several random numbers. Most people, professionals as well as laymen, expect such sums to exhibit nice statistical properties. This may be due to a popular misconception regarding the universality of the Law of Large Numbers which may lead to a belief in the general normality of sums of random variables (of random variables that have a second moment, that is). However, the asymptotic character of the proof of the Law of Large Numbers requires very large numbers of summands and there is no basis for a belief in the general applicability of the law to small samples.

Furthermore, there are classes of random variables that do not have first and second moments and that are yet of value in looking at oil spill statistics.* Such distributions would completely fail to comply with the "law" of large numbers. Moreover, one needn't look at obscure distribution (like the "Stable Laws") before one uncovers random phenomena that exhibit highly irregular sums.

An example can best illustrate this point. Assume that the distribution on the volume, v, of oil spilled in any one

---

* See Paulson, for example, for an application of Stable Law distributions to some oil spill data.

incident is given by the Gamma distribution:

$$f(v) = \frac{\lambda(\lambda v)^{R-1} e^{-\lambda v}}{\Gamma(R)} \tag{2.1}$$

where R and λ are the parameters determining the character of the distribution on v.

Assume that the volume spilled in any one incident is independent of the volume spilled in any other incident. If our sample comprises 'N' incidents, then the distribution on the sum, y, of these 'N' values of v, ($y = \sum_{i=1}^{N} v_i$), is given by the related Gamma function:

$$f(y) = \frac{\lambda(\lambda y)^{NR-1} e^{-\lambda y}}{\Gamma(NR)} \text{ .*} \tag{2.2}$$

If we now take the average of these spills, $z = \frac{y}{N}$, we find that the average is distributed like:

$$f(z) = \frac{\beta(\beta z)^{NR-1} e^{-\beta z}}{\Gamma(NR)} \tag{2.3}$$

where $\beta = N\lambda$.

This distribution has the following moments:

1. Mean $= \bar{z} = \frac{NR}{\beta} = \frac{R}{\lambda}$ (2.4a)

2. $\overline{z^2} = \frac{NR(NR+1)}{\beta^2}$ (2.4b)

---

* This relationship between the distribution of the sum and the underlying distribution is a property of the Gamma distribution and is not generalizable to other probability density functions (PDFs).

3. $\overline{z^3} = \frac{NR(NR + 1)(NR + 2)}{\beta^3}$ (2.4c)

4. $\text{Variance} = \sigma_z^2 = \frac{NR}{\beta^2} = \frac{R}{N\lambda^2}$ (2.4d)

5. $\text{Skewness} = \gamma_I^z = \frac{2}{(NR)^{1/2}}$ (2.4e)

If we write these in terms of the mean, $\overline{v}$, variance, σ_v^2, and skewness, $\gamma_1^v = \frac{2}{R^{1/2}}$, of the underlying distribution on v, we find:

$$\overline{z} = \overline{v} \tag{2.5a}$$

$$\sigma_z^2 = \frac{1}{N} \sigma_v^2 \tag{2.5b}$$

$$\gamma_1^z = \frac{1}{N^{1/2}} \gamma_1^v \tag{2.5c}$$

Thus, increasing the number of samples causes the variance to decrease like $\frac{1}{N}$, which is the same behavior found in normal distributions, and what we expect based on the law of large numbers. However, unlike the normal distribution, the distribution of the mean, z, can be highly skewed.* This suggests that a substantial portion of the distribution of z may lie in regions very much removed from $\overline{z}$, and this means that the probability of determining $\overline{z}$ to good accuracy may be correspondingly small with a small sample.

---

* If γ_1^v has a skewness of 16, then we required an enormous number of samples (eg., 250) to bring the skewness of the mean, γ_1^z down to a value of order 1.

More exactly, the probability of finding a z more than three times larger than $\bar{z}$ is given by the integral

$$P[z \geq 3\bar{z}] = \int_{3\bar{z}}^{\infty} \frac{e^{-\beta z}\beta(\beta z)^{NR-1}}{\Gamma(NR)}\, dz = \frac{\Gamma(NR, 3NR)}{\Gamma(NR)}. \quad (2.6)$$

Similarly, the probability of finding a z that is less than one-third the value of $\bar{z}$ is given by

$$P[z \leq \tfrac{1}{3}\bar{z}] = \int_{0}^{\frac{1}{3}\bar{z}} \frac{e^{-\beta z}\beta(\beta z)^{NR-1}}{\Gamma(NR)}\, dz = 1 - \frac{\Gamma(NR, \frac{1}{3}NR)}{\Gamma(NR)} \quad (2.7)$$

Making a few substitutions and combining the results above, we can readily show the probability that the value of z (the sample mean) will fall within a factor of 3 of the desired value, $\bar{z}$, is:

$$P[\tfrac{1}{3}\bar{z} \leq z \leq 3\bar{z}] = \frac{\Gamma(NR, \frac{1}{3}NR) - \Gamma(NR, 3NR)}{\Gamma(NR)} \quad (2.8)$$

where $\Gamma(a,x)$ is the incomplete Gamma function (See Abramowitz, pp 260-263)

We can see that this probability is solely a function of NR. That is, the probability that the estimated mean value will fall in the range $(\frac{1}{3}\bar{z}, 3\bar{z})$ will be a function of the skewness of the underlying distribution (ie., $R = 4/(\gamma_1{}^{v})^2$) and the number of samples. To provide a more concrete understanding of the problem we have evaluated (approximately) the probability for various values of (NR) using Figure 6.3 of Abramowitz for small values of NR and the Table of the CDF of X^2 in Benjamin and Cornell for larger values. The results are tabulated below.

TABLE 2.1

PROBABILITY THAT THE SAMPLE MEAN WILL FALL WITHIN A FACTOR OF 3 OF THE REAL MEAN

NR	$P\ (\frac{1}{3}\bar{z} \le z \le 3\bar{z})$
.5	.44
1.0	.64
2.0	.85
3.0	.91

Thus, for example, (NR) must fall in the range of .5 to 1.0 if we are to have a 50% chance (P = .5) of estimating $\bar{z}$ within a factor of three.

We can use this result to estimate the desired sample size. If we assume that the population from which the sample is drawn has a skewness of twenty, then we need about 100 samples (i.e., $R = 4/20^2 = .01$, $N \cong 1/.01 = 100$) to have a 50% chance at estimating $\bar{z}$ within a factor of three. Obviously as the sample gets larger, (NR) increases and we have a correspondingly greater chance of estimating the mean value to any specified degree of accuracy. Unfortunately, we cannot arbitrarily draw more samples in the oil spill problem and so our estimate of the mean is restricted in its accuracy by phenomena beyond our control. It seems clear that if we cannot be ery confident of estimating the average spill size within a factor of three, with ten or even one hundred samples, then we must be suspicious of the reliability of the comparison of two such means. In the tanker/pipeline controversy the data employed in such comparisons are usually drawn from records of major spill events involving either loss of life or large spill volumes. These records are then usually culled further to include only spills over some arbitrary threshold like 1000 BBLS. The assumption here is that large spills will generally be reported in the literature and so we can assume that these large spill records are exhaustive compilations.

In a typical application, average spill sizes, Z_i, are calculated based on the samples available for the alternative modes (i = 1,2), and an average number of spills, η_i, is estimated for the life of the development based on some exposure parameter suitable for the alternatives being considered (e.g. total production in barrels from an offshore petroleum development). The average spillage per mode, Σ_i, is then calculated based on the product of the average number of spills in the life of the development times the average volume spilled per event ($\Sigma_i = \eta_i Z_i$; $i = 1,2$).

The average number of spills, η_i, is a randomly varying parameter, and so the variability in Σ_i will be a complex function of the η_i and Z_i PDFs.* If we attempt to incorporate

* Generally,

$$f_\Sigma(\sigma) = \int_{-\infty}^{\infty} \left|\frac{1}{y}\right| f_\eta\left(\frac{\sigma}{y}\right) f_Z(y)\ dy\ .$$

where both f_η and f_Z are continuous functions (if either f_η or f_Z are discontinuous, the formula is qualitatively different).

the additional variability due to the η_i's, we find that the mathematics become more and more cumbersome and the number of assumptions required to pose the problem increases. Neither of these consequences is desirable in view of the qualitative nature of this discussion. We shall assume, therefore, that the η_i's are deterministic and that the distribution on Σ_i is given by:

$$f_{\Sigma_i}(\sigma) = \eta_i^{-1} f_{Z_i}\left(\frac{\sigma}{\eta_i}\right) \tag{2.9}$$

A comparison of alternatives based on the projected average spillage parameters devolves in this case to a comparison of the sort

$$\frac{\eta_1}{\eta_2} Z_1 \overset{?}{>} Z_2 \qquad \text{where } \frac{\eta_1}{\eta_2} \text{ is known} \tag{2.10}$$

The number of spill events in the large spill pipeline data base is in the tens or twenties and the number of tanker spills is in the hundreds. Both populations exhibit large skewness, and so one problem that must be addressed is what is the effect of the different sample sizes on the comparison shown in (2.10). That is, for the skewed distributions we find in the oil spill business, does the sample size play a role in the comparison, and if so, does it bear upon the validity of any conclusions we might draw from such comparisons?

In terms of the classical statistical literature, we can phrase this concern in terms of the probability that we will rongly accept the hypothesis that $\eta_1\overline{Z}_1$ is less than $\eta_2\overline{Z}_2$ when in fact $\eta_2\overline{Z}_2$ is equal to (or less than) $\eta_1\overline{Z}_1$. Stated in this form we have a Type II error. Since η_1 and η_2 are free variables, it is useful to consider some particular set of values of these parameters to make the analysis less abstract. A reasonable selection is $\eta_1 = \eta_2$, although the reader is cautioned to remember this specification. In any particular application, η_1 will not equal η_2, and a specialized analysis would necessarily follow.

We can make the analysis specific by now asking what is the probability that the mean of a sample drawn from one population will be greater than the mean drawn from another population given the parameters characterizing each population. To simplify the calculations, we will assume each population is distributed like a Gamma variable. We will also assume that both populations have the same λ factor, although we will allow R to vary. This problem may sound somewhat peculiar, even trivial, because both estimates of the mean are unbiased and thus we might expect that populations having equal means will always have a probability of .5 of one sample mean exceeding the other. On reflection, however, this is obviously fallacious, as counterexamples can be readily constructed.

Letting Z_1 and Z_2 be the observed sample means, we find the following distributions for Z_1 and Z_2:

$$f(Z_1) = \frac{(N_1\lambda)\;(N_1\lambda Z_1)^{N_1R_1-1}e^{-N_1\lambda Z_1}}{\Gamma(N_1R_1)} \qquad (2.11)$$

$$f(Z_2) = \frac{(N_2\lambda)\;(N_2\lambda Z_2)^{N_2R_2-1}e^{-N_2\lambda Z_2}}{\Gamma(N_2R_2)} \qquad (2.12)$$

Here N_i is the number of samples drawn from population "i".

The probability that Z_2 will be greater than or equal to Z_1 is given by the integral:

$$P[Z_2 \geq Z_1 \mid N_1, N_2, R_1, R_2, \lambda] = \int_0^\infty dz_1 \left[\int_{z_1}^\infty dz_2 \, \frac{e^{-N_2 \lambda Z_2} (N_2\lambda)(N_2 \lambda Z_2)^{R_2 N_2 - 1}}{\Gamma(R_2 N_2)} \right]$$

$$\cdot \frac{(N_1\lambda)(N_1\lambda Z_1)^{R_1 N_1 - 1} e^{-N_1 \lambda Z_1}}{\Gamma(R_1 N_1)} \qquad (2.13)$$

This can be evaluated in terms of Gauss' Hypergeometric function, ${}_2F_1(a,b;c;d)$, with the result*

$$P[Z_2 \geq Z_1 \mid \ldots] = \frac{1}{R_1 N_1} \left(\frac{N_1}{N_1+N_2}\right)^{R_1 N_1} \left(\frac{N_2}{N_1+N_2}\right)^{R_2 N_2} \frac{\Gamma(R_1 N_1 + R_2 N_2)}{\Gamma(RN_1)\Gamma(R_2 N_2)} \cdot$$

$${}_2F_1\left(1, R_1N_1+R_2N_2; 1+R_1N_1; \frac{N_1}{N_1+N_2}\right). \qquad (2.14)$$

This can be evaluated for general values of the arguments R_1N_1, R_2N_2 and $\frac{N_1}{N_1+N_2}$, but the formula involves an infinite series which is rather difficult to work with. It serves our purpose just as well to consider a special case, the case of $R_1N_1 \equiv 1$. This corresponds to drawing 10 samples from a reasonably skewed population ($R = .1$, $\gamma_1^v = \sqrt{40} = 6.32$). We might,

---

* See Gradshteyn, equation 6.455(1).

for example, consider this to be a model of the pipeline data base in view of both the small number of samples and the value of the skewness.

In this case the probability that the value Z_2 will exceed or equal Z_1 is given by:

$$P[Z_2 \geq Z_1 \mid N_1 R_1 = 1] = 1 - \left(\frac{N_2}{N_1+N_2}\right)^{R_2 N_2} = 1 - \left(\frac{N_2/N_1}{1+N_2/N_1}\right)^{\frac{R_2}{R_1}\frac{N_2}{N_1}} \quad (2.15)$$

But R_2/R_1 is the ratio of the population means, so let $R_2/R_1 = \beta$, and note that when β is less than 1, then population 2 has a smaller mean than population 1, and so on.

Notice that in the case $\beta = 1$,

$$P[Z_2 > Z_1] = 1 - \left(\frac{N_2/N_1}{1+N_2/N_1}\right)^{\frac{N_2}{N_1}} . \quad (2.16)$$

Only in the event that N_2 equals N_1 does this equation show that we have a 50/50 chance of finding $Z_2 > Z_1$. In fact, we find that as we draw more and more samples from population 2, we improve our chances that Z_2 will equal or exceed Z_1. Worse yet, if we draw a very large number of samples from population 2, our chances are asymptotically .632, $(1-\frac{1}{e})$, that Z_2 will be greater than or equal to Z_1.* This despite the fact that $\overline{Z_1} = \overline{Z_2}$.

---

* The probability that Z_2 will equal Z_1 is vanishingly small and so we can interpret $Z_2 \geq Z_1$ as $Z_2 > Z_1$ for any practical purpose.

If we specialize the problem by assuming

$$\frac{N_2}{N_1} >> 1$$

then (2.16) shows that this Type II error will occur with probability .632 when $\bar{Z}_2 = \bar{Z}_1$. We can specialize (2.15) for the general case of β less than or equal to 1 and $N_2/N_1 >> 1$, with the result

$$P\{Z_2>Z_1 | R_1N_1=1;\ \frac{N_2}{N_1}>>1;\ 0\leq\beta\leq1\} = 1-e^{-\beta}=P\{\text{Type II error}\} \quad (2.18)$$

This is shown on Figure 2.1. Notice that for small βs the probability of observing a Type II error is nearly proportional to the ratio of the population mean values. That is, if population 2 had an average that was one-tenth that of population 1, we would still have a probability of about one-tenth of finding $Z_2 \geq Z_1$ due solely to the differences in sample sizes.

Many comparisons require Z_2 to be some substantial fraction greater than Z_1 in order to avoid this type of error. This can still lead to erroneous conclusions. For example, if we set the value of the lower limit of integration within the bracket in (2.13) to $2Z_1$ then we have the probability that Z_2 will be twice Z_1 given the distribution and sample size parameters R_1, R_2, N_1, N_2, and λ. (Note that this is the technique to be used in determining the probability of $\Sigma_2 \geq \Sigma_1$ when $\eta_1=2\eta_2$.) If we again specialize the problem for $R_1N_1=1$, then

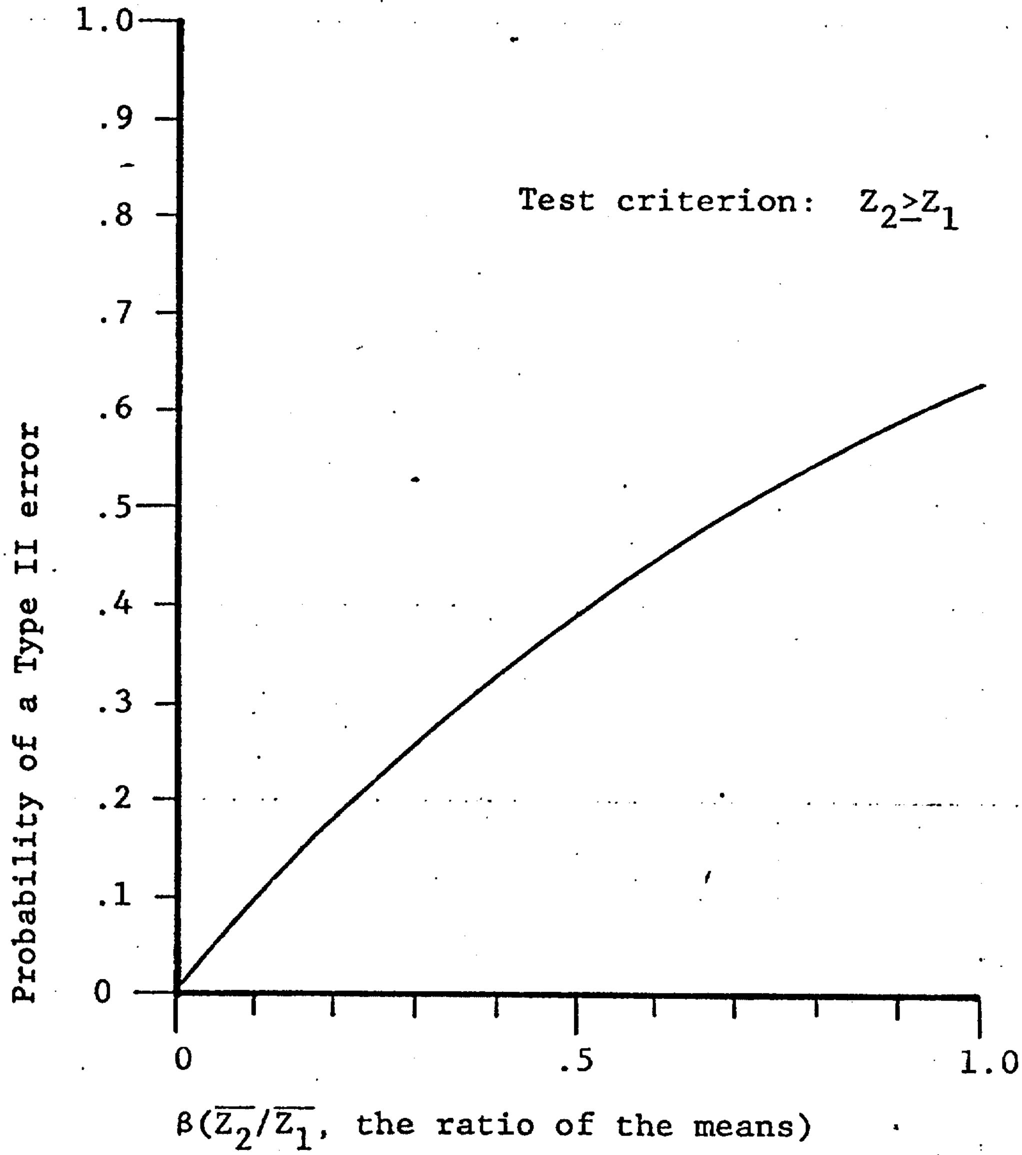

Figure 2.1--Probability of erroneously concluding that the mean of Population 2 is larger than that of Population 1 when $R_1N_1=1$, $N_2/N_1>>1$, and the test requires $Z_2 \geq Z_1$.

$$P\{Z_2 \geq 2Z_1 | R_1N_1=1\} = 1 - \left\{\frac{2\frac{N_2}{N_1}}{1+2\frac{N_2}{N_1}}\right\}^{\beta\frac{N_2}{N_1}} \qquad (2.17)$$

Specializing this further for the case $N_2/N_1 >> 1$, we may readily determine that the probability of erroneously concluding that $\overline{Z}_2$ was greater than $\overline{Z}_1$ based on a $Z_2 \geq 2Z_1$ test is:

$$P\{Z_2 \geq 2Z_1 | R_1N_1=1;\ \frac{N_2}{N_1} >> 1;\ 0 \leq \beta \leq 1\} = 1-e^{-\frac{\beta}{2}} = P\{\text{Type II error}\} \qquad (2.18)$$

This function is shown in Figure 2.2. As we can see, little has been gained by this rather severe criterion. Imposing such conditions on Z_2 also increases the probability of committing a Type I error (rejecting the correct hypothesis). Thus, average volume comparisons lead to greater and greater difficulties as we try to accommodate processes with substantial skewness.

Since these results are not intuitive, it behooves us to state when such comparisons ought to be acceptable, as they are used frequently and with much success. One ready example is the case where both populations are normally distributed. In this event, the sample mean is also normally distributed with an ever-decreasing variance and zero skewness. If we visualize the joint distribution of Z_1^n and Z_2^n under these circumstances, we see that a contour map of the joint PDF would consist of a number of ellipses constructed about the locus of the means $(\overline{Z_1^n}, \overline{Z_2^n})$. As the

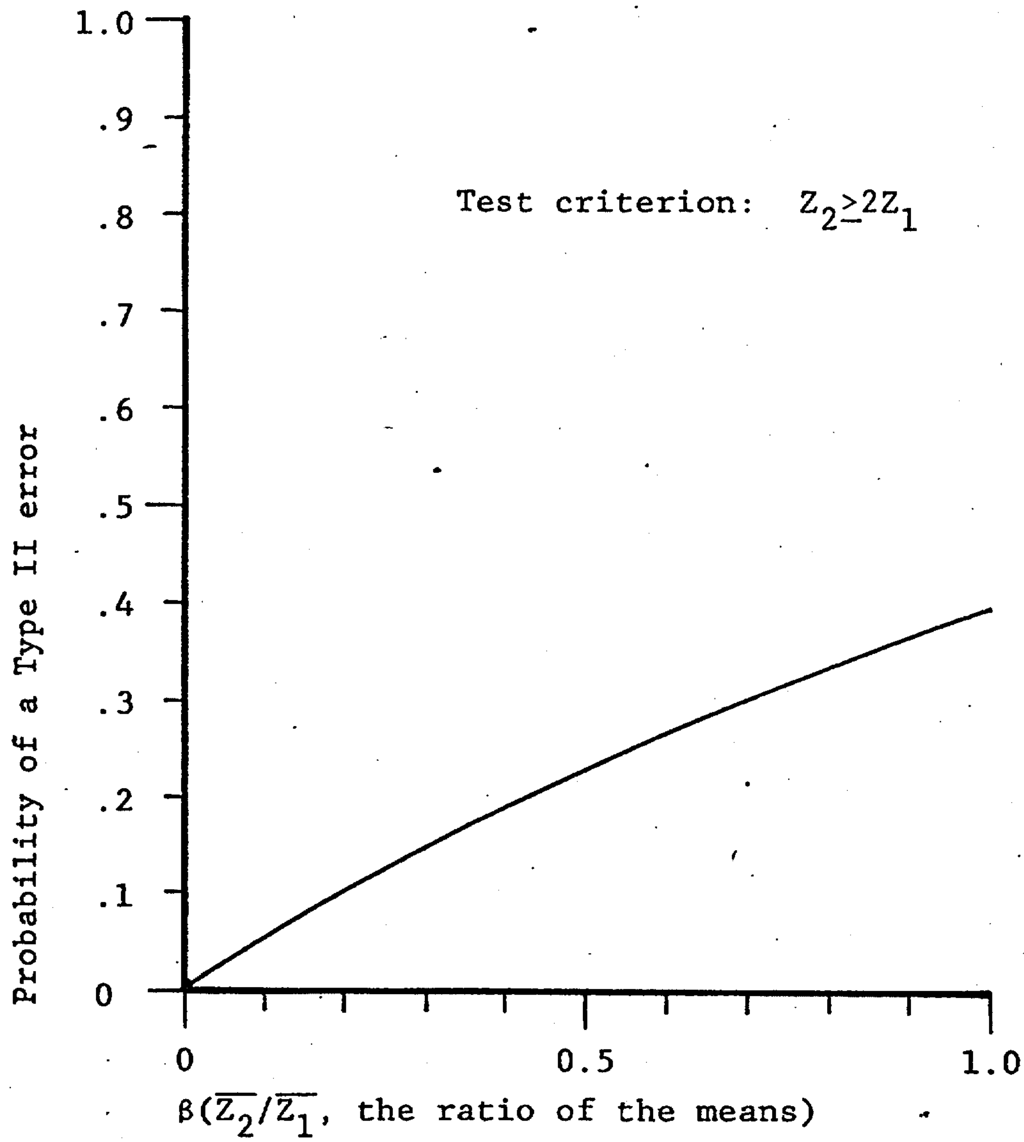

Figure 2.2--Probability of erroneously concluding that the mean of Population 2 is larger than that of Population 1 when $R_1N_1=1$, $N_2/N_1>>1$, and the test requires $Z_2 \geq 2Z_1$.

number of samples used to calculate Z_2 is increased, the axis of the ellipse in the Z_2 direction simply becomes smaller.

If we consider the case $\overline{Z_1^n}=\overline{Z_2^n}$, then the probability that Z_2^n is greater than Z_1^n would be .5 irrespective of the sample size since the line $Z_1=Z_2$ would be a line of symmetry in the (Z_1,Z_2) plane. Conversely, the skewed Gamma distributions do not have such lines of symmetry (except as N_1R_1 and N_2R_2 become much greater than unity or when $N_1R_1=N_2R_2$), and so the portion of the distribution lying on one side or the other of the line $Z_1=Z_2$ varies with N_2.

If the reader is still disconcerted at this point, and perhaps skeptical of the generality of the results, which we readily admit are outlined only for the case R_1N_1 equal to unity and $\eta_1=\eta_2$, we apologize. The extension of these results to the more general case can be done based on the equations presented here, but it is well beyond the scope of the present study to attempt such an exhaustive survey. Our main hope is that in the absence of such a conclusive investigation, the reader will at least apply some small fraction of his present skepticism to average volume comparisons. Particularly when these comparisons are made without sufficient consideration of the treacherous properties of these joint statistics.