Dr. 995

**A Critique of the Council on Economic
Priorities Text "Power Plant
Performance—Nuclear and Coal
Capacity Factors and Economics"**

Ronald L. Simard

MASTER

OAK RIDGE NATIONAL LABORATORY

OPERATED BY UNION CARBIDE CORPORATION FOR THE ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

DISTRIBUTION OF THIS REPORT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
Price: Printed Copy \$4.50; Microfiche \$3.00

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Energy Research and Development Administration/United States Nuclear Regulatory Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Contract No. W-7405-eng-26

Engineering Technology Division

A CRITIQUE OF THE COUNCIL ON ECONOMIC PRIORITIES TEXT
"POWER PLANT PERFORMANCE — NUCLEAR AND COAL
CAPACITY FACTORS AND ECONOMICS"

Ronald L. Simard

Manuscript Completed — April 18, 1977

Date Published — May 1977

NOTICE: This document contains information of a preliminary nature. It is subject to revision or correction and therefore does not represent a final report.

MASTER

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

ep

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	v
ABSTRACT	1
I. EXECUTIVE SUMMARY	1
II. PREDICTION OF NUCLEAR UNIT CAPACITY FACTORS	6
A. The Nuclear Unit Data Base	6
B. The CEP Method of Analysis — BWRs	11
C. The CEP Method of Analysis — PWRs	16
III. PREDICTION OF COAL UNIT CAPACITY PERFORMANCE	21
A. The Coal Unit Data Base	21
B. The CEP Method of Analysis	23
IV. ECONOMIC COMPARISON OF TYPICAL NUCLEAR AND COAL UNITS	24
V. DEFINITION OF STATISTICAL TERMS	32
REFERENCES	39

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

ACKNOWLEDGEMENTS

Forest (Frosty) Miller of the Mathematics and Statistics Research Department of the Computer Science Division carried out regression analyses which allowed a calculation of the associated error bars; he provided valuable assistance in the interpretation and application of the standard statistical techniques used herein.

Charles Bayne of the Mathematics and Statistics Research Department of the Computer Science Division and Howard Bowers of the Engineering Analysis Section of the Engineering Technology Division reviewed the analysis and made helpful comments.

The work was carried out under the sponsorship of the U.S. Energy Research and Development Administration Office of Fossil Energy Program Planning and Analysis and Division of Nuclear Research and Applications, and with the cooperation of the Federal Energy Administration, Fuels and Facilities Division.

A CRITIQUE OF THE COUNCIL ON ECONOMIC PRIORITIES TEXT
"POWER PLANT PERFORMANCE — NUCLEAR AND COAL
CAPACITY FACTORS AND ECONOMICS"

Ronald L. Simard

ABSTRACT

The Council on Economic Priorities (CEP) has assembled capacity factor data on most U.S. nuclear and coal-fired units in operation, performed a regression analysis upon capacity factor as a function of intrinsic unit variables such as size and age, and attempted a comparison of the economics of typical nuclear and coal-fired units using the resultant fitted capacity factors.¹

This report points out inadequacies in the CEP data base, the lack of predictive value in the results of the regression analysis and inconsistencies in the economic evaluation of nuclear and coal-fired units.

I. EXECUTIVE SUMMARY

The Council on Economic Priorities' (CEP) analysis leads to the conclusion that typical nuclear units [1150 MW(e)] operating in 1984 would be economically competitive with typical coal-fired units [600 MW(e)] only in certain restrictive scenarios. The basis of this conclusion is a comparison of the respective generating costs, in mills/kWhr, using capacity factors of most of the operating nuclear and coal units as functions of intrinsic unit variables such as size, age, coal quality, and boiler criticality. This conclusion is not justified because of inadequacies in the data base, misuse of statistical techniques, and questionable assumptions in the economic analysis.

The Nuclear Unit Data Base

The nuclear data base consists of 110 unit-years of operation of both pressurized water reactors (PWRs) and boiling water reactors (BWRs). Most of the units are young; three-fourths of the PWRs and two-thirds of the BWRs have three, or fewer, full calendar years of data. One-third of the units have only first year of operation data which are atypical

of subsequent years because of variations in startup experience and initial cycle refueling schemes.

All of the units are smaller than the 1150 MW(e) unit chosen for the economic analysis. Only 36 of the unit-years are from units larger than 800 MW(e) and these include atypical data from the Palisades and Browns Ferry stations.

Several of the units have operated for various periods at power levels different from the design level used as the basis for calculating capacity factors.

The capacity factors calculated for each full calendar year of operation vary so widely that it is impossible to perform a regression analysis, using intrinsic unit variables, which fits the historical data at levels usually considered acceptable — let alone one which has any significant predictive value in extrapolating to larger or older plants.

Even if the historical data were amenable to a regression analysis which satisfied the usual statistical criteria for predictive value, it would be imprudent to extrapolate from the capacity factor data of 1968 to 1975 when estimating the 1984 performance of a newer-type unit. The newer unit would be operating in a different regulatory climate and would benefit from increased experience with plant construction and operation and the results of productivity improvement programs now gaining momentum.

The Coal Unit Data Base

The coal unit data base describes 250 coal units over 100 MW(e) which accrued 1641 unit-years of operation between 1961 and 1973. Most of these units are small and atypical of the reference plant [600 MW(e)] chosen for the economic analysis, e.g., more than three-fourths of the unit-year data come from units under 400 MW(e), more than three-fifths come from units under 300 MW(e), and more than one-third come from units under 200 MW(e). The data chosen to describe these units in the regression analysis were assembled from information supplied by the Federal Power Commission (FPC) and the Edison Electric Institute (EEI). The data are inconsistent and incomplete.

The "installed generating capacity" as reported to the FPC was used as the basis for calculating capacity factors. However, some utilities report the maximum generator nameplate rating while others use the maximum turbine rating.

Data on coal quality (sulfur, ash, and moisture contents) for the years 1972 to 1973 were taken from one source. Data for the years 1969 to 1971 were taken from a different source. Data for the years 1961 to 1968 were not available and 1969 data were used for these years, thereby ignoring all fuel-switching practices and variation in coal quality prior to 1969.

The calculated capacity factors were converted to capacity performance — a factor which accounts for the loss in capacity due to load-following when system considerations require less generation from a unit than it is capable of supplying. The amount of load-following done by each individual unit could not be determined. However, access to EEI data allowed an estimate of the average capacity lost by units in one of six given size intervals [100-599 MW(e) in 100 MW(e) increments and all units larger than 600 MW(e)] during the years 1968 to 1973. Individual capacity factors were corrected for load-following by applying a correction based upon an average over all units in the given size range for 1968 to 1973, thereby masking variations in load-following practices from year to year, unit to unit, and utility to utility.

Even with the masking of variability due to the assumptions made about capacity performance, coal quality, and unit rating, there remains such a wide variation in the data that it is difficult to perform a regression analysis which fits the historical data at levels usually considered acceptable and which has significant predictive value.

In any case, an extrapolation from the 1961 to 1973 data would fail to account for technological changes, productivity improvements, and the political and environmental issues which are affecting the design, construction, and use patterns of new coal units.

The CEP Methodology

The CEP text attempts a regression analysis of nuclear capacity factors and coal unit capacity performance as functions of intrinsic

unit variables such as size and age. Application of conventional statistical tests reveals that the resultant equations fail to meet the usual criteria for statistical significance and/or predictive value.

Four of the principal equations contain terms which are not statistically significant at the confidence level usually adopted in such tests. Specifically, the functional dependence of

- pressurized water reactor capacity factors upon age,
- all supercritical coal unit capacity performance factors (independent of size) upon age,
- all supercritical coal unit capacity performance factors [units larger than 600 MW(e)] upon coal sulfur content, and
- all coal unit capacity performance factors [units 400-599 MW(e)] upon boiler criticality,

as presented in the CEP analysis, is not justified at the 95% confidence level usually adopted.

None of the principal regression equations explains a significant amount (more than 41%) of the variability in the data. For instance, boiling water reactor capacity factors are treated as functions only of unit size. A conventional statistical technique shows that the regression equation explains only 11% of the variation of individual BWR capacity factors about the mean value. In other words, if capacity factors vary from BWR to BWR, this equation would successfully account for 11% of the variation by attributing it to differences in unit size. The other 89% is due to other factors which influence BWR performance and which do not appear in the CEP equations. The projection of capacity factors as functions of the unit variables appearing in the regression equations is questionable since, according to the CEP analysis, these capacity factors are mainly determined by other considerations.

Further, the regression equations predict only average values of capacity factor. Their poorness of fit and lack of predictive value are underscored by a calculation of the error bars (uncertainties) associated with each predicted factor. This calculation shows that the average over units of given rating can lie within a wide range of values. An individual unit with that same specified rating can lie within an even wider range of values. For instance, the CEP regression fit to boiling water

reactor data predicts, at the 95% confidence level, that any individual 1150 MW(e) BWR can have a capacity factor somewhere between 11% and 72%. Since this covers most of the range of possible values, the prediction is not very useful.

The Economic Evaluation

The economic performance of nuclear and coal units is evaluated by estimating 1984 generating costs as functions of capacity factors. The capacity factor of an 1150 MW(e) nuclear unit and the capacity performance of a 600 MW(e) coal unit are taken from the CEP regression analysis of the historical data. As explained previously, the variations in data and poorness of fit to the data lead to estimates of capacity factors which have such large uncertainties that the resultant comparison of nuclear and coal unit costs as functions of capacity factors is of dubious value. Even if accurate estimates of capacity factors were available, the CEP economic calculations would be questionable because of several assumptions made in choosing the reference coal unit. Specifically, estimates of the relative capital costs of coal vis-a-vis nuclear units are dated and probably too low. Further, coal unit behavior was corrected, in the regression analysis, with considerable effort, to reflect the capacity lost by coal units due to load-following. This correction was ignored, however, in the economic calculations as the reference coal unit was assumed to be fully base-loaded (with no loss of generating capacity due to load-following) and the unit's projected capacity performance, rather than capacity factor, was used in calculating generating costs. Finally, the economic results are extremely sensitive to the choice of reference units. For instance, coal unit capacity performance is "predicted" as a function of unit age, rating, coal quality, etc., and leveled, or averaged, over ten years of operation when appropriate. Changes in the assumptions made about these variables can affect the resultant projection of capacity performance and calculation of generating costs to the point of making the reference coal unit more expensive than the reference nuclear unit; a 599 MW(e) coal unit [different from the reference unit chosen by only 1 MW(e) but similar in other respects] falls

into a different size class from the 600 MW(e) unit. Its capacity performance is projected from a different regression equation and is lower than that of the 600 MW(e) unit. Thus, the CEP estimates that a "typical" 600 MW(e) coal unit operating in 1984 with the required emission controls (scrubbers) would generate cheaper electricity than a typical 1150 MW(e) nuclear unit. However, a "typical" 599 MW(e) coal unit, subject to the same assumptions, would be costlier than the nuclear unit.

II. PREDICTION OF NUCLEAR UNIT CAPACITY FACTORS

A. The Nuclear Unit Data Base

The CEP nuclear data base consists of capacity factors for 24 PWRs and 14 BWRs in operation for at least one full calendar year between 1968 and 1975. The distribution of these 110 unit-years of data is shown in Fig. 1. The "average" unit-year capacity factor calculated by the CEP is the arithmetic mean of 59.3%. A more descriptive value of the "average" might be the median value of 61.4% (see Sect. V).

The "average" plant age calculated by the CEP is the arithmetic mean of 2.9 years. The distribution of plant ages (in terms of full calendar-years of operation) is shown in Fig. 2. This is not a symmetric distribution and a more descriptive value of the "average" plant age in the data base might be the geometric mean of 2.3 years. Note that there are only 2 PWRs older than 5 years; 75% (18/24) of the PWRs have ≤ 3 full calendar-years of data. There are 2 at 4 years, 2 at 5 years, and 2 at 8 years. Note that there are only 2 BWRs older than 5 years; 64% (9/14) of the BWRs have ≤ 3 full calendar-years of operation. Of the 38 units in the nuclear data base, 27, or 71%, have ≤ 3 full calendar-years of operation. These data are inadequate in predicting plant performance out to 10 years of operation, especially since as noted below, first year capacity factors are atypical.

The distribution of plant sizes is shown in Fig. 3. Note that there are only 7 PWRs larger than 850 MW(e). These contribute only 10-unit-years of data. There are only 3 BWRs larger than 850 MW(e). These contribute only 3-unit-years of data and these 3 are all first year of

ORNL DWG 77-7077

Fig. 1. Distribution of CEP nuclear unit-year capacity factors.

Fig. 2. Distribution of nuclear unit ages.

Fig. 3. Distribution of nuclear unit sizes.

operation data. Further, one of the three comes from the Browns Ferry unit (recognized in the text as atypical). None of the units is as large as the 1150 MW(e) reference unit chosen by the CEP for the economic analysis.

Table 3.3 of the CEP text¹ shows that one-third of the units in the CEP nuclear data base have data only from the first full calendar year of operation. The footnote to the table admits that "first-year capacity factors show high fluctuation due to break-in problems and omission of refueling and subsequent down-time in most cases." A plant experiencing start-up and break-in problems can yield an anomalistically low first-year capacity factor. A plant avoiding break-in problems during the first year may yield a very high capacity factor because the first shut-down for refueling is usually scheduled for the second year of operation. Thus, first-year capacity factors are hardly typical of normal operation. And yet, 12 out of 38 units (1/3) in the CEP nuclear data base have only first-year data. If these units were removed, the data base would contain the following anomalies:

1. Only 4 BWRs larger than 700 MW(e). These are all 809 MW(e) and all operated by the same utility — Commonwealth Edison. Their capacity factors are recognized in the text as lower than "average" (possibly due to the Commonwealth Edison practice of occasionally load-following with their nuclear units). The remaining BWRs are all smaller than 700 MW(e) and (with the exception of Monticello) are all located in the East where economic constraints on utility procedures may be unique to the area.
2. Only one PWR larger than 880 MW(e) — viz., Zion 1 whose first full year of operation (1974) was atypical (an availability factor of 57%, a forced outage rate of 15%).
3. Only one Babcock & Wilcox reactor (Oconee 1 with 2 calendar-years of data).
4. Only two Combustion Engineering reactors (one of which is the Palisades unit, recognized in the CEP text as atypical).
5. All other PWRs left in the data base have Nuclear Steam Supply Systems manufactured by Westinghouse.

As shown in the following sections, the scatter of data in the CEP nuclear data base is so wide that it is difficult to perform a significant statistical fit and develop a model which can accurately predict the capacity factors of plants larger or older than those in the data base.

B. The CEP Method of Analysis - BWRs

The CEP analysis of BWR capacity factors provides a poor fit to the data and is inadequate in extrapolating from the data to the performance of the newer, larger BWRs. The CEP admits the "poorness" of fit in the text but claims the analysis is still valuable "as a starting point." However, their extensive use of the results impugns considerably more value to the analysis than is justified.

The "goodness" of a regression analysis fit to data points can be measured by several tests for statistical significance, e.g., t-tests, F-tests, calculation of R^2 (all defined in Sect. V). These tests have been applied to the CEP fits as described below.

An attempt was made by the CEP to fit the BWR unit-year capacity factors as functions of unit size, unit age, and year of commercial installation. The "best" fit was found to be

$$CF = 79.30 - 3.287 \times \left(\frac{MW}{100} \right) \quad (1)$$

with 44 data points and an estimated standard error (e.s.e.) in the coefficient, -3.287 , of 1.46. A t-test (see Sect. V) of the coefficient for 42 degrees of freedom at the 95% confidence level shows the coefficient is significant or, in other words, there is a probability of .95 that the coefficient lies between

$$\begin{aligned} & -3.287 \pm t(42, 95\%) \times \text{e.s.e.} \\ & = -3.287 \pm 2.02 \times 1.46 \\ & = -3.287 \pm 2.95 \end{aligned} \quad (2)$$

If a linear relationship between CF and $\left(\frac{MW}{100} \right)$ were not valid, the expected value of the coefficient of $\left(\frac{MW}{100} \right)$ in the fit would be zero and t would be

$$\begin{aligned}
 t &= \left| \frac{\text{fitted value of coefficient} - \text{expected value}}{\text{e.s.e. in coefficient}} \right| \quad (3) \\
 &= \left| \frac{(-3.287) - (0.0)}{1.46} \right| \\
 &= 2.25
 \end{aligned}$$

and this lies outside the 95% level (2.02) of the t-distribution for 42 degrees of freedom. Thus, the t-test shows the coefficient to be statistically significant. (See also Sect. III.B.)

The F-test would not yield any more information in this case since, for a straight-line fit, the F-test and t-test give numerically equivalent results.

Thus, the t-test (or F-test) shows that it is possible to fit the BWR capacity factor data with the straight line of Eq. (1). However, the calculation of R^2 (see Sect. V) shows the fit to be a poor one. R^2 may be calculated from the standard Analysis of Variance table below.

Source	Degrees of freedom	Sum of squares	Mean square	Calculated F value
Due to regression	1	917.09	917.09	5.05
Deviation from regression	42	7620.44	181.44 (= S^2)	
Total about mean	43	8537.53	198.55	

From the definition of R^2 ,

$$\begin{aligned}
 R^2 &= \frac{\text{sum of squares of regression}}{\text{sum of squares about the mean}} \quad (4) \\
 &= \frac{917.09}{8537.53} \\
 &= .11
 \end{aligned}$$

which means that the fit is capable of explaining only 11% of the variation of individual capacity factors about the mean value. In other words, if capacity factors vary from BWR to BWR, this fit would assign 11% of

the variation to differences in plant size. The other 89% is unexplained. Thus, the text is not justified in using the result of the fit, viz., Eq. (1), in predicting the capacity factor of future BWRs solely as a function of plant size. Similar statements can be made about all the regression equations in the text (see Table 1, Sect. III.B).

This lack of predictive value may also be shown by two other means -- an examination of the calculated F-value and a calculation of the error bars which the text neglected to display.

First, the Analysis of Variance table above shows that the

$$\begin{aligned} \text{calculated F value} &= \frac{\text{mean square of regression}}{\text{mean square about regression}} & (5) \\ &= \frac{917.09}{181.44} \\ &= 5.05 \end{aligned}$$

This is larger than the

$$\begin{aligned} \text{theoretical F value} &= F(1, 42, 95\%) & (6) \\ &= 4.08 \end{aligned}$$

which confirms the results of the t-test as explained above. However, for a regression analysis to have predictive value, one requirement² is that the calculated F-value be at least four times the theoretical F-value.

Second, the text presents only the mean capacity factor, predicted by Eq. (1), for a number of plants of the same size $\left(\frac{MW}{100}\right)$. The author makes no mention of the uncertainty in this value. Figure 4 shows the variations in mean capacity factor and individual capacity factors, among a number of BWRs of the same size, to be expected at the 95% confidence level. The five curves in Fig. 4 are generated by the following equations

$$y(\text{fit}) = 79.30 - 3.287 x \quad (7)$$

Fig. 4. Range of predicted BWR capacity factors.

$$\bar{y} = y(\text{fit}) \pm t(42, 95\%) \cdot [\text{e.s.e. in } y(\text{fit})] \quad (8)$$

$$= (79.30 - 3.287 x) \pm 2.02 \left[\frac{s^2}{n} + \frac{s^2 (x - \bar{x})^2}{\sum_k (x_k - \bar{x})^2} \right]^{1/2}$$

$$= (79.30 - 3.287 x) \pm 2.02 \left[\frac{181.44}{44} + \frac{181.44 (x - 7.11)^2}{84.713} \right]^{1/2}$$

$$= (79.30 - 3.287 x) \pm 2.02 [4.124 + 2.142 (x - 7.11)^2]^{1/2}$$

$$y(\text{individual}) = y(\text{fit}) \pm 2.02 \left[s^2 + \frac{s^2}{n} + \frac{s^2 (x - \bar{x})^2}{\sum_k (x_k - \bar{x})^2} \right]^{1/2} \quad (9)$$

$$= (79.30 - 3.287 x) \pm 2.02 [185.564 + 2.142 (x - 7.11)^2]^{1/2}$$

In these equations,

$$x = \frac{\text{MW}}{100}$$

$\bar{x} = 7.11$ = the arithmetic mean over the 44 x's

$n = 44$ = No. of BWR unit-years

$s^2 = 181.44$ = variance of capacity factors about the fit.

Thus, for 100 BWRs of the same size, say 1150 MW(e), the CEP analysis would predict an average capacity factor of

$$79.30 - (3.287)(11.50) = 41.5.$$

However, 95 out of these 100 capacity factors would be expected to lie anywhere between 11.1 and 71.9. Or, in other words, for a given 1150 MW(e) BWR, the CEP analysis predicts that there is a probability of .95 that its capacity factor for any given year will lie somewhere between 11.1% and 71.9%.

Also shown in Fig. 4 is the straight line $CF = \overline{CF}$ (independent of plant size). This lies almost entirely within the range of average capacity factors predicted by the model.

C. The CEP Method of Analysis - PWRs

As explained in Section B, the CEP fit to BWR unit-year capacity factors is statistically significant at the 95% confidence level usually adopted when using this methodology. However, the CEP fit to PWR unit-year capacity factors is not. The criterion for statistical significance is the t-test (explained in Sect. V) performed on each coefficient in the assumed equation. The CEP text presents the following equation

$$CF = 93.48 - \underset{(2.36)}{4.22} \times \frac{1}{\text{Log } (1 + \text{AGE})} - \underset{(1.14)}{3.368} \times \left(\frac{\text{MW}}{100} \right) \quad (10)$$

as the "best" fit to the data. The estimated standard error (e.s.e.) is shown under each coefficient. The calculated t-values are

$$\frac{3.368}{1.14} = 2.95 \text{ (for the size term)}$$

$$\frac{4.22}{2.36} = 1.79 \text{ . (for the age term)}$$

To satisfy the usual criterion for statistical significance, these should be larger than the theoretical t-value for 63 degrees of freedom and 95% confidence level, viz.,

$$t(63, 95\%) = 1.998 \text{ .}$$

Note that the coefficient of the size term in Eq. (10) satisfies this criterion. The coefficient of the age term does not. In including this term in the equation, the CEP has incorrectly applied a one-sided test rather than a two-sided test (see Sect. III.B). The age term is statistically significant only at the 90% confidence level. All statements in the text about PWR performance changing with unit age are not justified at the significance level usually adopted.

Even if the regression were statistically significant, at the usual level, the same criteria applied to the BWR analysis would reveal its lack of predictive value.

First, the R^2 is only .21 (the size term contributes .19 of this; the age term contributes only .02 — another indication that the age term in the equation should be examined more closely) which means that the regression could explain only 21% of the variation in capacity factors among PWRs of different sizes and ages and implies that 79% is due to other factors than unit size and age.

Second, the calculated F-value for the regression is 9.14. This is less than four times the theoretical F-value (3.15) for 2 and 63 degrees of freedom at the 95% confidence level.

Third, as in the case of BWR capacity factors, the text makes no mention of the uncertainties associated with the regression. Figures 5 and 6 show the variations in mean capacity factor and individual capacity factors, among a number of PWRs of the same size, to be expected at the 95% confidence level. The five curves in Fig. 5 are for AGE = 1 and are generated by the following equations

$$y(\text{fit}) = 93.48 - \frac{4.22}{\text{Log } (1 + \text{AGE})} - 3.368 \left(\frac{\text{MW}}{100} \right) \quad (11)$$

$$= 79.46 - 3.368 \left(\frac{\text{MW}}{100} \right)$$

$$\bar{y}_k = y(\text{fit}) \pm t(63, 95\%) \cdot [\text{e.s.e. in } y(\text{fit})] \quad (12)$$

$$= \left[79.46 - 3.368 \left(\frac{\text{MW}}{100} \right) \right] \pm 1.998 \left[\underline{X}_k' (\underline{S^2 X' X})^{-1} \underline{X}_k \right]^{1/2}$$

$$y_k(\text{individual}) = \left[79.46 - 3.368 \left(\frac{\text{MW}}{100} \right) \right] \quad (13)$$

$$\pm 1.998 \left[\underline{S^2} + \underline{X}_k' (\underline{S^2 X' X})^{-1} \underline{X}_k \right]^{1/2}$$

The five curves in Fig. 6 are for AGE = 10, so 79.46 is replaced by 89.43. In the above equations,

Fig. 5. Range of predicted PWR capacity factors (AGE = 1).

Fig. 6. Range of predicted PWR capacity factors (AGE = 10).

$$S^2 = 226.6$$

$$\underline{X}'_k = \left[1, \left(\frac{MW}{100} \right)_k, \frac{1}{\text{Log } (1 + AGE)_k} \right], \text{ a row vector}$$

$$\underline{X}_k = \text{the transpose of } \underline{X}'_k$$

$$(\underline{S}^2 \underline{X}' \underline{X})^{-1} = \text{the variance-covariance matrix of the regression, viz.,}$$

$$(\underline{S}^2 \underline{X}' \underline{X})^{-1} = \begin{pmatrix} 59.49 & -6.350 & -6.170 \\ -6.350 & 1.300 & -.9890 \\ -6.170 & -.9890 & 5.584 \end{pmatrix}$$

For instance, a number of one-year old 1150 MW(e) PWRs would have an average capacity factor, as predicted by the CEP analysis of,

$$y(\text{fit}) = 79.46 - 3.368 \times 11.5 \quad (11')$$

$$= 40.7 .$$

At the 95% confidence level, this value could actually lie anywhere between 29.7% and 51.8% since

$$\underline{X}'_k = \left[1, 11.5, \frac{1}{\text{Log } 2} \right]$$

$$\underline{X}_k = \begin{bmatrix} 1 \\ 11.5 \\ \frac{1}{\text{Log } 2} \end{bmatrix}$$

and $\underline{X}'_k (\underline{S}^2 \underline{X}' \underline{X})^{-1} \underline{X}_k = 30.46$ (approximately — all calculations were actually done in double precision arithmetic on a computer).

Further, individual values of capacity factor could lie anywhere between 8.7% and 72.8%.

In summary, the CEP regression fit to PWR capacity factors is not statistically significant at the level usually adopted when using this methodology. Even if it were to be accepted and used to predict the performance of units outside the sample, the model would lead to such

empty predictions as "the tenth-year capacity factor for a 600 MW(e) PWR will probably lie between 38% and 100%."

III. PREDICTION OF COAL UNIT CAPACITY PERFORMANCE

A. The Coal Unit Data Base

The CEP coal unit data base describes 250 coal-fired units over 100 MW(e) in rating which accrued at least one full calendar year of operation between 1961 and 1973. Only two of these units operated with scrubbers during that period and most are smaller than the 600 MW(e) reference unit chosen for the economic comparison with nuclear, e.g., more than three-fourths of the unit-years are from units under 400 MW(e), more than three-fifths come from units under 300 MW(e), and more than one-third are from units under 200 MW(e). The 1641 unit-years of data are input to a regression analysis in which the principal independent variates are unit size, age, vintage (period of installation), boiler criticality, coal quality (sulfur and ash contents) and the dependent variate is capacity performance (defined below). Data describing these quantities are taken from inconsistent and/or incomplete sources. Assumptions were made in correcting inadequacies in the data which introduce uncertainties into the regression analysis and contribute to the resultant pooriness of fit.

For instance, capacity factors are calculated on the basis of net electrical generation and installed generating capacity. The installed generating capacity for each unit was taken from forms submitted by the operator to the Federal Power Commission (FPC). However, as the text admits, some utilities report the maximum generator nameplate rating to the FPC, while others report the maximum turbine rating. The text gives an example of a unit for which the two ratings differ by 9%.

There are also inconsistencies in the coal quality assumed in the analysis. Sulfur, ash, and moisture contents for the years 1972 to 1973 were taken from one source. Data for the years 1969 to 1971 were taken from a different source. Data for the years 1961 to 1968 were not

available and the CEP text uses 1969 data for these years as well, thereby neglecting fuel switching practices and variations in coal quality prior to 1969.

Further assumptions were made when correcting the generating performance of the coal units for loss of capacity due to load-following. Nuclear units were assumed to be fully base-loaded; coal units were assumed to load-follow and their calculated capacity factors were converted to capacity performance factors which were used in the regression analysis. The calendar-year capacity performance was defined as

$$\text{capacity performance} = \frac{\text{actual capacity factor}}{\text{optimal capacity factor}}$$

or

$$CP = \frac{CF}{OCF} \quad (14)$$

where the optimal capacity factor is defined as the capacity factor which the unit would have achieved in that year had there been no mechanical outages; it can be calculated as 100% minus the capacity factor lost due to load-following and other system considerations. This capacity can be lost either during shutdown or while operating at a reduced power level. The amount lost can thus be determined by calculating full economy outage hours (FEOH) and equivalized partial economy outage hours (EPEOH). Mathematically,

$$CP = \frac{CF}{1 - (FEOH + EPEOH)/PH} \quad (15)$$

where PH is the number of hours in a year. The utilities which report to the Edison Electric Institute (EEI) report FEOH as Reserve Shutdown Hours, but EPEOH is not available. It is estimated by the CEP from the reported quantities equivalent availability and capacity factor, viz.,

$$EPEOH = (EA - CF) \times PH \quad (16)$$

However, the CEP had access to these quantities only for size groups of units and only for the years 1968 to 1973. Averages over this six-year

period were taken for each size group and were assumed to be representative of operation in all years, viz., average optimal capacity factors (OCF) were used as follows

Size group	OCF
100 - 199 MW(e)	86.55
200 - 299 MW(e)	90.61
300 - 399 MW(e)	92.27
400 - 499 MW(e)	87.35
500 - 599 MW(e)	90.87
600 + MW(e)	92.78

This has the effect of masking variations in load-following practices from year to year, unit to unit, and utility to utility.

For example, the 1962 data from the Sporn No. 5 unit would be entered into the regression analysis as follows.

1. The 1969 values for coal sulfur and ash contents would be used.
2. The 496 MW(e) capacity reported would be assumed to represent the maximum generator nameplate rating and not the turbine rating.
3. The net electrical generation reported for 1962 would be used with this capacity to calculate a capacity factor (68.8%).
4. The capacity factor would be divided by .8735 and "inputted" as a capacity performance of 78.8%.

B. The CEP Method of Analysis

The efficacy of the analysis of coal unit capacity performance can be measured by the same criteria used for the analysis of nuclear capacity factors, specifically the t-test for significance of terms in the regression equations and the R^2 calculation of "goodness" of fit.

Recall that the t-test measures the confidence level at which one is justified in including terms in a regression equation like

$$y = a + bx$$

which has N degrees of freedom. The confidence level usually adopted is 95%. To test if the above linear relationship is justified, one assumes the opposite — that the coefficient b is actually zero. Then the calculated t value for b and its estimated error is compared with the theoretical t value for N degrees of freedom which corresponds to a 95% probability of b different from zero. This is known as a two-sided test. (See Fig. 7.) Throughout the text, CEP has incorrectly applied a one-sided test by using as a reference the theoretical t value which corresponds to a 95% probability of b greater than zero. As a result, the confidence level adopted in the text corresponds to a 90% level as usually defined and many of their regression equations contain terms which are statistically significant at a lower level than is usually considered acceptable (viz., the 95% level).

The table on the next page lists the principal regression equations presented in the text and indicates which contain terms which are questionable in the sense that they cannot be included at the 95% confidence level.

Recall that R^2 measures the amount of variability in the data which is explained by the regression. For instance, Equation 6.12 of the text is able to account for only 11% of the variability in capacity performance of 600 MW(e) (and larger) supercritical coal units. Generally, reported R^2 values in scientific and engineering publications tend to cluster around 90%. As Table 1 shows, the principal CEP equations are characterized by R^2 from 4% to 41%.

As noted in the previous section, some variability has been masked due to assumptions made about rating, coal quality, and load-following. Sufficient variability remains to make it difficult to perform a regression of capacity performance upon the variables chosen which has significant predictive value.

IV. ECONOMIC COMPARISON OF TYPICAL NUCLEAR AND COAL UNITS

The principal comparison is made between 600 MW(e) coal units and 1150 MW(e) nuclear units — sizes assumed to be typical of new plants.

Distribution of t

"Probability = Area in Two Tails of Distribution Outside $\pm t$ -Value in Table "

Degrees of Freedom	Probability									
	0.9	0.7	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	0.510	1.000	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.445	0.816	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.137	0.424	0.765	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.414	0.741	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.408	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.404	0.718	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.402	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.399	0.706	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.398	0.703	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.397	0.700	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.396	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.395	0.695	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.394	0.694	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.393	0.692	1.076	1.345	1.761	2.145	2.624	2.977	3.140
15	0.128	0.393	0.691	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.392	0.690	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.392	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.392	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.391	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.391	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.391	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.390	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.390	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.127	0.390	0.685	1.039	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.390	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.390	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.389	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.389	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.389	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.389	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.388	0.681	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.126	0.387	0.679	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.126	0.386	0.677	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.126	0.385	0.674	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Fig. 7. Percentage points of the t -distribution. Adapted from Ref. 2.

Table 1. Principal regression equations in text

Equation in text	Units described	R ²	Questionable terms
3.1	Pressurized water reactors	.21	$\log (1 + AGE)^{-1}$
3.3	Boiling water reactors	.11	
6.1	All coal unit years	.33	
6.2	All subcritical coal unit years	.33	
6.3	All supercritical coal unit years	.04	AGED
6.7	All coal units 100-399 MW(e)	.26	
6.10	All coal units 400-599 MW(e)	.14	CR
6.12	Supercritical coal units 600 MW(e) +	.11	XSUL
6.16	Subcritical coal units 600 MW(e) +	.41	
6.19	Subcritical coal units 400 MW(e) +	.16	

The basis of the comparison is a calculation of the respective generating costs, in mills/kWhr, for 1984 single-unit operation in the eastern U.S. The total generating costs are functions of capital cost, O&M cost, fuel cost, and capacity factor. The base case assumptions about these parameters lead to the conclusions that

"... PWR's cost will equal supercritical coal units, BWR's will be 12% more expensive than either, based on CEP capacity factor equations. At the 600 MW level, subcritical coal units will be slightly more expensive than supercritical; 2% more if the coal units are scrubber-equipped, 6% more if they are not."

These conclusions are questionable because of the assumptions upon which they are based.

The CEP estimates of capital costs per installed kW are derived from the 1974 report WASH-1345. The 1984 capital costs of \$1020/kW, \$839/kW and \$729/kW for 1150 MW(e) nuclear units, 1150 MW(e) coal units with scrubbers, and 1150 MW(e) coal units without scrubbers, respectively, are derived from 1981 costs in WASH-1345 by escalating at 6% per year for three years and increasing by 50% to compensate for "AEC's historic underestimate of capital costs." Then 600 MW(e) coal unit costs are

scaled up from the 1150 MW(e) costs with a scaling factor of 0.23 (also taken from WASH-1345). There are other, more recent,³⁻⁶ capital cost estimates in use which might be considered more reasonable and which would make the coal units more expensive relative to nuclear.

However, even if the CEP estimates of capital costs, fuel costs, and O&M costs are accepted as representative, certain components of these costs have to be spread over an estimated capacity factor to determine total generating mills/kWhr, and it is at this point that the CEP analysis is particularly questionable.

Figure 8 uses the CEP cost estimates and presents total generating cost as a function of capacity factor. The regression equations discussed in Sects. II and III are used to "predict" the capacity factors as follows,

Unit	Capacity factor (%)
1150 MW(e) BWR	41.5
1150 MW(e) PWR	47.6
600 MW(e) subcritical (no scrubbers)	65.4
600 MW(e) subcritical (with scrubbers)	62.2
600 MW(e) supercritical (no scrubbers)	72.8
600 MW(e) supercritical (with scrubbers)	64.6

where all capacity factors, except for the BWR, have been levelized over 10 years of operation. However, the uncertainties in the regression equations make a comparison of total generating costs meaningless. Returning to the example of Sect. II.C, any individual 1150 MW(e) BWR can have a total generating cost anywhere between 37.2 and 197.5 mills/kWhr. Similar uncertainties for the PWR, supercritical and subcritical coal units, define a large envelope of costs within which any two individual units have an indeterminate relative cost ratio.

Further, the CEP economic analysis is biased toward the reference coal units by the assumptions used to calculate capacity factors. First, the unrealistic assumption is made that the coal units are fully base-loaded, i.e., load-following is ignored and capacity factor is assumed to

Fig. 8. Total generating cost as a function of capacity factor.

be equal to capacity performance. In Chap. 6 of the text, the CEP assumes that the coal units in the data base had forsaken a considerable amount of capacity due to load-following. A considerable effort was made to calculate the average amount lost by units in a given size group. For instance, the "typical" 600 MW(e) unit was assumed to forsake 7% capacity. Capacity factors were corrected and converted to capacity performance and the regression equations "predict" capacity performance. But the predicted capacity performance is not converted back to a capacity factor before use in the economic analysis. This leads to too low an estimate of coal generating costs. The effect of this and other assumptions on coal generating costs vis-a-vis nuclear unit costs has been examined in a sensitivity analysis. The results for an 1150 MW(e) PWR and 600 MW(e) coal units with scrubbers are shown in Tables 2 and 3. The table entries were obtained as follows.

For the 600 MW(e) supercritical unit, the base unit, with scrubbers, is supposed to generate electricity at a capacity factor of 64.6% and a total cost of 52.3 mills/kWhr. The cost is determined from

$$\begin{aligned}
 \text{total mills/kWhr} &= \frac{.16}{8760 \text{ hr/year}} \times 1000 \text{ mills/\$} \times \frac{\text{capital \$/kW}}{\text{CF}} \\
 &+ \frac{1000 \text{ mills/\$}}{8760 \text{ hr/year}} \times \frac{\text{O\&M \$/kW - year}}{\text{CF}} \\
 &+ \text{variable O\&M costs} \\
 &+ \text{fuel costs} \\
 &= \frac{20.40}{\text{CF}} + 20.71
 \end{aligned}$$

A 1% increase in capital cost/kW makes this

$$\text{total mills/kWhr} = \frac{20.57}{\text{CF}} + 20.71$$

A CF = .646 makes this unit 0.5 mills/kWhr costlier than the 1150 MW(e) PWR at 52.1 mills/kWhr.

Table 2. Sensitivity of 600 MW(e) supercritical unit costs^a

Change	Δ (mills/kWhr)	Coal-nuclear ^b
+ 1% in capital cost/kW	+0.3	+0.5
Correction for load following	+2.5	+2.7
- 1% in sulfur content	-0.9	-0.7
+ 1% in sulfur content	+0.9	+1.1
- 1 MW(e) (use of equation 6.10)	+4.8	+5.0
- 1 MW(e) and correction for load following	+8.4	+8.6

^a Assuming use of scrubbers.

^b Differential between resultant coal unit cost and reference 1150 MW(e) PWR, e.g., the reference supercritical unit costs 2.7 mills/kWhr more than the PWR when corrected for load following.

Table 3. Sensitivity of 600 MW(e) subcritical unit costs^a

Change	Δ (mills/kWhr)	Coal-nuclear ^b
+ 1% in capital cost/kW	+0.4	+1.7
Correction for load following	+2.7	+4.0
Operation by Commonwealth Edison	+5.9	+7.2
Operation by TVA	-4.0	-2.7
Use of equation 6.19	+4.7	+6.0
- 1 MW(e) (use of equation 6.10)	+6.2	+7.5
- 1 MW(e) and correction for load following	+10.1	+11.4

^a Assuming use of scrubbers.

^b Differential between resultant coal unit cost and reference 1150 MW(e) PWR, e.g., the reference subcritical unit costs 4.0 mills/kWhr more than the PWR when corrected for load following.

The optimal capacity factor (see Sect. III.A) for units > 600 MW(e) is 92.78. Correction for load following gives a capacity factor of $64.6 \times .9278 = 59.9$ and a resultant cost of 54.8 mills/kWhr.

The capacity performance for the supercritical unit is obtained by a 10-year levelization of equation 6.12 in the text, viz.,

$$CP = 94.95 - 2.394 (6.0) - 3.45 (1.688)$$

$$- 1.94 (XSUL)$$

with XSUL = 3.5 for operation with scrubbers. The effect of $\pm 1\%$ in coal sulfur content is to make XSUL = 2.5 or 4.5. The resultant costs are 51.4 and 53.2 mills/kWhr, respectively.

The reference coal unit chosen [600 MW(e)] lies at the edge of its size group. Its capacity performance is "predicted" by equation 6.12 (supercritical units 600 MW(e) +). A 599 MW(e) supercritical unit would be described by equation 6.10 [all coal units 400-599 MW(e)], namely

$$CP = 65.73 - 1.30 (4.725) + 3.76$$

$$+ 2.82 (XASH) - 3.23 (XSUL)$$

Assuming XSUL = 3.5, as above, and XASH = 2.5, representing 12.5% ash content — the usual CEP assumption, the capacity performance would be 59.1×0.95 (the CEP assumes a 5% loss due to the scrubbers) yielding a cost of 57.1 mills/kWhr.

The optimal capacity factor for units 500-599 MW(e) is 90.87 so a correction for load following would give a CF = 51.0 and a cost of 60.7 mills/kWhr.

For the 600 MW(e) subcritical unit, the base unit, with scrubbers, is supposed to generate electricity at a capacity factor of 62.2% and a cost of 53.4 mills/kWhr. The cost is derived from the equation for supercritical units since the CEP assumes the same capital, O&M, and fuel costs. A 1% increase in capital cost/kW leads to a total generating cost of 53.8 mills/kWhr for CF = 62.2%.

The correction for load following is applied as for the supercritical unit.

The subcritical unit is treated differently, however, in that its capacity performance is "predicted" by a different regression equation — equation 6.16 [subcritical units 600 MW(e) +]. Levelizing over ten years gives

$$CP = 85.15 - 3.284 (6.0) - 9.86 (COMM) + 9.55 (TVA)$$

where COMM (TVA) is 1 if the unit is operated by Commonwealth Edison (TVA) and zero otherwise. If the base case unit were in the mid west or south, the capital, O&M, and fuel cost components would be different, of course, but the table shows that utility ownership could have a significant effect on the "predicted" generating cost.

Subcritical units are also analyzed in the text by examining all units over 400 MW(e). Use of regression equation 6.19 would "predict" a levelized 10-year capacity performance of

$$CP = 75.19 - 2.16 (6.0) + 4.38 (XASH) - 4.52 (XSUL)$$

Using the CEP assumptions of XASH = 2.5 and XSUL = 3.5, this equation would "predict" a 600 MW(e) supercritical CP of $57.4 \times .95$ (loss due to scrubbers) = 54.5 and a resultant cost of 58.1 mills/kWhr.

Equation 6.10 could be used for a 599 MW(e) subcritical unit also; now CR = 0 and the same assumptions used for the supercritical unit would yield a capacity factor of $55.3 \times .95 = 52.5\%$, and a generating cost of 59.6 mills/kWhr.

Correction for load following would make the capacity factor $52.5 \times .9087 = 47.7\%$ and the generating cost 63.5 mills/kWhr.

Thus, the coal unit costs can vary considerably, not only with the assumptions about the reference unit size, coal quality, and use patterns, but also with the regression equation used.

V. DEFINITION OF STATISTICAL TERMS

Analysis of Variance

The Analysis of Variance table is a convenient way of displaying how well a regression analysis fits a number of data points by

separating the variation in data about the mean value into variation explained by the regression and variation of data about the regression.

If the assumed dependent variables, $\{y_i\}$, have an arithmetic mean of $\bar{y}$ and values predicted by the regression analysis of $\{\hat{y}_i\}$, then the following identity

$$y_i - \hat{y}_i = y_i - \bar{y} - (\hat{y}_i - \bar{y})$$

can be manipulated to give

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 ,$$

or in other words, the sum of squares (SS) about the mean = SS about the regression + SS due to the regression. Thus, some of the variation in data points about their mean value can be allocated to the regression and the rest is due to variation of data about the regression. Tests for goodness of fit can then be performed by picking appropriate entries from the following table.

Source	Sum of squares	Degrees of freedom	Mean square
Due to regression	SSR	1	MSR = SSR
Deviation from regression	SSDEV	$n - 2$	MSDEV = $\frac{SSDEV}{n - 2}$
About mean (total, corrected for mean)	$\sum y_i^2 - \frac{1}{n} (\sum y_i)^2$	$n - 1$	—

For the straight line fit to BWR capacity factors, n = number of data = 44, x_i is $\frac{MW}{100}$ for the i -th BWR, y_i is its capacity factor and

$$SSR = \frac{\left[\sum x_i y_i - \frac{1}{n} (\sum x_i) (\sum y_i) \right]^2}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}$$

SSDEV can be calculated by subtracting this from the sum of squares about the mean.

Arithmetic Mean

When data are symmetrically distributed about a central value, a good description of their "average" is the arithmetic mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i .$$

When the distribution is asymmetric, the usual practice⁷ is to use the most frequent value, middle value, or geometric mean.

Capacity Factor

Several different definitions of capacity factor are in use. The CEP text defines the unit-year capacity factor for nuclear units as

$$\frac{\text{net electrical generation in year (MW-hours)}}{\text{net design capacity [MW(e)]} \times 8760 \text{ hours}} \times 100 .$$

This yields capacity factors typically 2-3% lower than those from the Edison Electric Institute and Nuclear Regulatory Commission definition which is based upon the maximum dependable capacity rather than net design rating. For coal units, the maximum generator nameplate rating was used.

Degrees of Freedom

In the Analysis of Variance table, the number of degrees of freedom indicates how many independent pieces of information about the n independent members $y_1, y_2, \dots, y_n$ are needed to calculate the sum of squares. In the regression analyses, the number of degrees of freedom is equal to the number of data points minus the number of coefficients to be determined.

F-test

In the Analysis of Variance table, the mean square explained by the regression and the mean square due to residual variation about the regression are called MSR and MSDEV, respectively. MSDEV is an unbiased estimate of the underlying variability of the data. Their ratio can be used to test whether the coefficients in the regression fit are significantly different from zero given a confidence level. For instance, BWR capacity factor data were fitted by a straight line of the form $y = a + bx$. If this fit were not justified, i.e., if $b = 0$, the ratio

$$F = \frac{MSR}{MSDEV}$$

would follow what's known as an F distribution with 1 and $(n - 2)$ degrees of freedom. Then one can compare this calculated F value with the theoretical value of F for 1 and $(n - 2)$ degrees of freedom at, say, the 95% confidence level by looking up this theoretical F value in a table. In this case, the table would predict that 95% of the F values for 1 and 42 degrees of freedom would lie between 0 and 4.08. Since the calculated F in this case is 5.05, one runs a risk of 5% in assuming the hypothesis is wrong, i.e., $b \neq 0$.

Geometric Mean

When data are not grouped symmetrically about a central value but are distributed as in Fig. 2, an appropriate measure of the "average" is the geometric mean defined as

$$\text{antilog} \left[\frac{1}{n} \sum_{i=1}^n \log (x_i) \right].$$

Median

When data are not grouped symmetrically about a central value but exhibit a skewness, as in Fig. 1, an appropriate measure of the average

is the median, or midpoint, value. In Fig. 1 the arithmetic mean value is 59.3% with 64 unit-years above this value and 46 below. The median value is 61.4% with 55 unit-years above and 55 below. The "average" should be taken to be 61.4%. For example, if the abscissa in Fig. 1 represented Council on Economic Priorities employees and the ordinate their corresponding salaries in thousands of dollars, the CEP employee making \$59,300/year would not likely consider himself "average" when 3 out of every 5 fellow workers makes more than he does. The employee making \$61,400/year is not as likely to resent being called "average" since half his fellow workers make more than he and half make less.

R²

In constructing the Analysis of Variance table, the variation of data about the mean was separated into sum of squares (SS) about the mean = SS due to the regression + SS of data about the regression. R² is defined as the ratio

$$R^2 = \frac{\text{SS due to the regression}}{\text{SS about the mean.}}$$

Obviously, the closer R² is to unity, the better the fit. Since R² measures the proportion of variation about the mean explained by the regression, it is a measure of the amount of variation in data about the mean which is "explained" by the regression.

t-test

Like the F-test, the t-test is made by comparing the ratio of two quantities calculated from the data with a value taken from the tabulated values of a theoretical distribution -- in this case, the t-distribution. The test is made upon each of the coefficients in an assumed regression. First, under the null hypothesis, the coefficient is assumed to be zero. If this were true, the ratio

$$\left| \frac{\text{coefficient}}{\text{estimated standard error in coefficient}} \right|$$

would follow a t-distribution. Second, this statistic is calculated and compared to the theoretical value from a table of t-distribution values for the given number of degrees of freedom and desired confidence level. (Figure 7 shows such a table.) If the table indicates a small probability (usually .05 or less is taken to be small) of exceeding the theoretical value by chance and the calculated value does exceed the theoretical value, then it is assumed that

- a. this discrepancy is not due to chance; that
- b. the reason for the discrepancy is that the hypothesis is wrong, viz.,
- c. the coefficient is indeed different from zero.

The confidence level at which one publishes this conclusion is 100% times (1 - the probability assumed above in using the table).

Note that the appropriate use of the t-test here is a two-sided test: the hypothesis being tested is [coefficient = zero]: the test criterion is the specified probability of the coefficient being different from zero. It is also possible (when physical conditions rule out the possibility of having a negative coefficient, for instance) to perform a one-sided test in which the criterion would be the probability of the coefficient being greater than zero. Examination of Fig. 7 indicates that the 95% confidence level for a one-sided test corresponds to the 90% level for a two-sided test.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

REFERENCES

1. "Powerplant Performance — Nuclear and Coal Capacity Factors and Economics," Council on Economic Priorities, 1976.
2. N. R. Draper and H. Smith, *Applied Regression Analysis*, p. 64, John Wiley and Sons, 1966.
3. "Economic Comparison of Base-Load Generation Alternatives for New England Electric," Arthur D. Little, Inc./S.M. Stoller Corp., March 1975.
4. G. Corey, "A Comparison of the Cost of Nuclear Versus Conventional Electric Generation," presented at MIT, Commonwealth Edison Co., November 1975.
5. T. H. Lee, testimony before the Connecticut PUCA, January 1976.
6. "An Overview of the New York State Energy Picture," The Legislative Commission on Energy Systems, August 1976.
7. R. Langley, *Practical Statistics Simply Explained*, Dover Publications, 1970.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

ORNL/TM-5846
Dist. Category UC-80

Internal Distribution

1. T. D. Anderson	16. M. Levenson
2. S. Baron	17. J. R. McWherter
3. C. K. Bayne	18. F. L. Miller
4. H. I. Bowers	19. M. L. Myers
5. W. E. Cooper	20. J. P. Nichols
6. F. L. Culler	21. H. Postma
7. J. G. Delene	22-121. R. L. Simard
8. G. G. Fee	122. I. Spiewak
9. B. H. Fitzgerald	123. D. B. Trauger
10. E. C. Fox	124. W. J. Wilcox
11. L. C. Fuller	125. ORNL Patent Office
12. M. J. Goglia	126-127. Central Research Library
13. J. F. Harvey	128. Document Reference Section
14. C. R. Hudson	129-131. Laboratory Records Department
15. M. A. Kuliasha	132. Laboratory Records (RC)

External Distribution

- 133. Director, ERDA, ORO
- 134-135. Director, Fossil Energy Program Planning and Analysis, ERDA
- 136-137. Director, Nuclear Research and Applications, ERDA
- 138. Research and Technical Support Division, ERDA, ORO
- 139. B. Baratz, Fossil Energy Program Planning and Analysis, ERDA
- 140. J. Dinwiddie, Fossil Energy Program Planning and Analysis, ERDA
- 141. P. Fine, Division of Safety and Environmental Analysis, NRC
- 142. F. Haines, Office of Fuels and Facilities, FEA
- 143. P. Hedman, Fossil Energy Program Planning and Analysis, ERDA
- 144. C. Komanoff, Council on Economic Priorities, 84 Fifth Ave., New York, N.Y. 10011
- 145. E. Kovacic, Office of Fuels and Facilities, FEA
- 146. R. McFarren, Office of Fuels and Facilities, FEA
- 147. S. E. Miller, Division of Nuclear Research and Applications, ERDA
- 148. R. J. Niebo, Office of Fuels and Facilities, FEA
- 149. S. Strauch, Division of Nuclear Research and Applications, ERDA
- 150. M. Wiener, Division of Nuclear Research and Applications, ERDA
- 151-290. Given distribution as shown under category UC-80.