AD-A047 720 AIR WEATHER SERVICE SCOTT AFB ILL F/G 4/2
TRANSNORMALIZED REGRESSION PROBABILITY.(U)
DEC 76 A R BOEHM
UNCLASSIFIED AWS-TR-75-259 NL

1 OF 2
AD A047 720

END
DATE FILMED
1-78
DDC

CONT

AD A047720

AWS-TR-75-259

COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

TRANSNORMALIZED REGRESSION PROBABILITY

Final rept.,

By
Albert R. Boehm, Capt, USAF
HQ AWS

DDC
DEC 19 1977
DDC FILE COPY

Approved for public release; distribution unlimited

54 p.

PUBLISHED BY
AIR WEATHER SERVICE (MAC)
UNITED STATES AIR FORCE
DECEMBER 1976

014 670

REVIEW AND APPROVAL STATEMENT

This report is approved for public release. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved for publication.

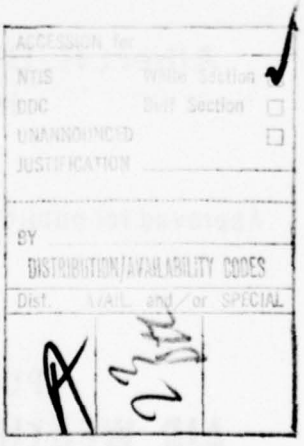
Robert G. Miller

ROBERT G. MILLER
AWS Chief Scientist
Reviewing Officer

FOR THE COMMANDER

Robert M. Gottuso

ROBERT M. GOTTUSO, Col., USAF
DCS/Aerospace Sciences
Hq Air Weather Service



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AWS-TR-75-259	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Transnormalized Regression Probability		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Albert R. Boehm, Capt, USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Hq Air Weather Service Scott AFB, IL 62225		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Hq Air Weather Service Scott AFB, IL 62225		12. REPORT DATE December 1976
		13. NUMBER OF PAGES 52
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Meteorology	Correlation
Statistical Prediction	Transnormalization
Statistical Models	Transnormalized Regression Probability
Regression Probability	
Climatological Distributions	Distribution Functions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Transnormalized Regression Probability (TRP) model uses climatological probabilities and requires only a relatively small data base to calculate the probability of occurrence of a future weather event. The three main procedures that comprise this multivariate transnormalized method are: transnormalization, correlation, and regression conditional probability. The transnormalization process a nonlinear process, transforms the observed or (Cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Cont'd):

raw data into the standard normal variable with the same cumulative climatological probability as the raw predictor. Transnormalization insures that the predictor is normally distributed. The correlation step involves finding the simple correlation between each pair of transnormalized variables. The regression coefficients are calculated using the same steps involved in ordinary multiple linear regression. Conditional probability is calculated using the transnormalized variables in the multivariate normal distribution.

TABLE OF CONTENTS

		Page
Chapter 1	INTRODUCTION	1
2	SOME STATISTICAL FORECASTING TERMS AND VIEWPOINTS	3
2.1	Forecast States and Modes	3
2.2	Objective and Subjective Procedures	3
2.3	Statistical Terms	3
2.4	Probability Terms	4
3	OVERVIEW OF THE MODEL	5
3.1	The Development Mode	5
3.1.1	Transnormalizer Development	5
3.1.2	Correlation Overview	6
3.1.3	Regression Coefficients Overview	6
3.2	The Forecast Mode	7
3.2.1	Transnormalization and the Mean Predictor	7
3.2.2	Conditional Probability	7
3.2.3	Inverse Transnormalization	8
3.2.4	Information Flow in the Forecast Mode	9
4	TRANSNORMALIZATION	11
4.1	Reasons for Transnormalization	11
4.2	Graphical Transnormalization	11
4.3	Two-Stage Transnormalization	11
4.3.1	Histogram Interpolation	12
4.3.2	Rank Order Transnormalization	13
4.3.3	Using a Known Distribution	14
4.3.4	Unknown Distribution — The Gram-Charlier Approximation	14
4.3.5	Unknown Distribution — The Pearson System of Curves	15
4.3.6	Converting Probability to an END	15
4.4	Direct Methods	15
4.4.1	The Cornish-Fisher Expansion	16
4.4.2	Transnormalized Quantile Polynomial Fitting	16
4.4.3	Johnson's Family of Curves	17
4.5	A Transnormalization Example	18
4.6	Selecting a Transnormalization Procedure	19
4.7	Procedures for Inverse Transnormalization	20
4.8	Transnormalization and Marginal Predictors	20
5	CORRELATION	21
5.1	The Pearson Product — Moment Formula	21
5.2	Biserial Correlation	22
5.3	Tetrachoric Correlation	23
5.4	Other Methods of Calculating Correlation	25
5.5	Selecting a Correlation Procedure	25
6	REGRESSION PROBABILITY	27
6.1	The Multivariate Normal	27
6.2	Conditional Probability	28
6.3	Regression Probability for One and Two Predictors	29
6.4	Probability for a Categorical Predictor	30
6.5	Multiple Predictand Categories	31
6.6	Changing Predictand Categories	32
7	EXAMPLES AND APPLICATIONS	33
7.1	Overall Development and Verification Procedures	33
7.2	Probability of Frozen Precipitation	33
7.2.1	Frozen Precipitation with Two Direct Predictors	33
7.2.2	Frozen Precipitation with One Direct and One Marginal Predictor	34
7.2.3	Frozen Precipitation with Unconditional Probability Known for Each Station	35
7.2.4	Probability of Frozen Precipitation with Discriminant Analysis	35
7.2.5	Frozen Precipitation Conclusions	35
7.3	Probability of Ceiling Categories by Model Output Statistics	36
7.3.1	TRP Ceiling Forecast by MOS	36
7.3.2	REEP Ceiling Forecast by MOS	38

TABLE OF CONTENTS (Cont'd)

		Page
Chapter 7.3.3	MDA Ceiling Forecast by MOS.	39
7.3.4	Conclusions for MOS Ceiling Forecasts.	39
8	CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH	41
REFERENCES AND BIBLIOGRAPHY		43
APPENDIX A	PROBABILITY FOR AN UNOBSERVED EVENT	47
APPENDIX B	TAYLOR SERIES ESTIMATES FOR JOHNSON'S LOGNORMAL AND HYPERBOLIC SINE DISTRIBUTIONS.	49
APPENDIX C	CALCULATION OF MOMENTS AND CUMULANTS.	51

LIST OF ILLUSTRATIONS

		Page
Figure 1.	Transnormalization of a Variable.	5
2.	Symbol for a Transnormalizer.	6
3.	Bivariate Normal Conditional Probability.	7
4.	The Inverse Transnormalizer	8
5.	The TRP Forecast Mode Raw Predictors P_1, P_2, ... P_N are Transnormalized into END Variables $\tilde{P}_1$, $\tilde{P}_2$, ... $\tilde{P}_N$.	9
6.	Linearity Caused by Transnormalization.	11
7.	Usable Regions of Transnormalized Polynomials	17
8.	Transnormalized Distribution Approximations to the Pittsburgh Surface Wind Speed.	19

LIST OF TABLES

		Page
Table 1.	Statistical Parameters for Pittsburgh Wind Speed.	18
2.	Frozen Precipitation Verification Results	34
3.	TRP Predictor Correlation Matrix.	36
4.	Ceiling Forecasts Verification Results.	37
5.	REEP Coefficient Matrix	38
6.	MDA Coefficient Matrix.	39

Chapter 1

INTRODUCTION

One of the most important services that the operational meteorologist can perform is to give the probability that a specific weather event will occur. For very long-range forecasts, climatological probabilities generally are used. Climatological probabilities have been prepared for a wide variety of events — high winds, extreme temperatures, low visibility, heavy precipitation, etc. Indeed, climatological probabilities have already been prepared for most operationally important events.

For shorter range forecasts, many statistical methods are available for relating the probability of a future event with observed or forecast weather elements (e.g., Panofsky and Brier, 1965). Using these methods, it generally is desirable to have a large set of data to generate the probability estimates. In some cases, a sufficient data set is difficult to obtain, particularly if some of these data come from a numerical weather prediction model since these models tend to change as improved versions are put into operation.

The purpose of this report is to describe a multivariate, "transnormalized" method that makes explicit use of climatological probabilities and requires a relatively small data base. This method is called the climatologically transnormalized regression probability model, briefly identified as the TRP model.

The TRP model allows predictand categories to be variable and easily changed in operational use. Categorical and continuous variables can be used. Compared with other methods used in statistical forecasting, such as discriminant analysis, TRP is favorably structured for modeling in time and space. Considerable success has been obtained in developing the model for one location and applying it to another. Verification analysis of the model leads to a framework for judging the accuracy and correcting biases in probability forecasts.

The TRP model consists of three main procedures: transnormalization, correlation, and regression conditional probability. These procedures are not new. Transnormalization and correlation were developed by Edgeworth (1898), Pearson (1895), and others before the turn of the century. Although regression probability is a straight-forward result of multivariate normal analysis, its potential when combined with transnormalization and correlation has not been exploited. The model seems particularly well suited for meteorology where the climatology of most elements is known and the correlation between many pairs of variates is known or can be modeled.

The three parts were first combined into meteorological models by McCabe (1968) and Gringorten (1968, 1971, 1972). These models were presented at the First Statistical Conference of the American Meteorological Society in Hartford, Connecticut in 1968. Further development of the TRP model was stimulated by correspondence between Gringorten and the author while the author was stationed in Tokyo, Japan. Optimizing the model and applying it to various operational forecasting problems was judged to be the best application of climatology in Air Weather Service in 1972. Verification of model forecasts was presented at the Third Conference on the Use of Probability and Statistics (Boehm, 1973). Multiple predictor equations for the model were derived in 1973 and final development of efficient transnormalization algorithms was completed in 1974.

This report is organized as follows: First, some statistical forecasting terms are defined, and an overview of the TRP model is presented with an introduction to the three procedures. Next, several transnormalization methods are presented, correlation for continuous and categorized variables is described, and the regression probability equation is derived. Finally, some examples and applications are described.

Preceding Page BLANK - NOT FILMED

Chapter 2

SOME STATISTICAL FORECASTING TERMS AND VIEWPOINTS

2.1. Forecast States and Modes.

Typically, a forecast method goes through four states of implementation: development, testing, evaluation, and operational use. A statistical forecast method is developed by using previous observations to fit parameters of a given statistical model. One important part of the development state is collecting, checking, and building a data base file. This set of previous observations is termed the development or dependent data set. Once the method is developed, it usually is tested on a new data set -- the independent data set. If the procedure performs satisfactorily, it may then be evaluated in an operational forecast setting and finally operationally used to give forecasts to a customer. The four states are desirable but not mandatory, and alternate descriptions are used. See, for example, Panofsky and Brier (1965, p. 174) or Air Weather Service Technical Report 105-38 (1948).

Sometimes it is necessary to differentiate between the development mode and the forecast mode. The development mode uses a statistical procedure to fit parameters. The three stages -- testing, evaluation, and operational use -- utilize the forecast mode of the statistical forecast procedure. Both the development mode and forecast mode use observations as input, but the output of the development mode is parameters, and the output of the forecast mode is a forecast.

2.2. Objective and Subjective Procedures.

Any of the stages or either of the modes may be done objectively or subjectively. An objective procedure is a set of rules that will give the same result if applied by different individuals or computer systems. A subjective procedure depends on the past experience of the individual, and two individuals may get different results. Certain computers can be programmed to use subjective methods; that is, the result at any given time will depend on the previous experience of the computer. For example, refer to Learning Machines (Nilsson, 1965). The term objective forecast is sometimes used synonymously with statistical forecast. However, a nonstatistical method, e.g., dynamic, can also be used to produce objective forecasts.

Any of the stages or either of the modes may be done by automated or manual means. A computer can be used to develop a method that will be used manually, or a person can manually, e.g., graphically, develop a method that will be used on a computer.

2.3. Statistical Terms.

Two other terms, empirical and stochastic, are used in certain cases in place of statistical. An empirical method, to paraphrase Webster's dictionary (Friend and Guralnik, 1959), is based solely on experiment and observations. A statistical method is based on observations, but these are handled according to the science of statistics. A stochastic method involves randomness. Stochastic methods include statistical methods, since collecting observations can be thought of as a random process. However, a stochastic forecast model can be developed without the use of observations, e.g., a model based on Brownian motion theory.

Some statistical procedures require more data than others to obtain the same degree of accuracy. If a procedure needs relatively few data it is said to be an efficient procedure. Efficient procedures which depend on various assumptions in the statistical model are called parametric, since the basic form is assumed and only parameters, which are limited in effect, can be changed. If the basic assumptions are correct or approximately correct, then the procedure will yield reliable results; however, if the assumptions are poor, then a great deal of data will not make the procedure work well. In contrast to parametric methods, nonparametric methods require few assumptions but need more data. (Refer to Kanal and Chandrasekaran, 1971, for further information on the amount of data needed.) More information on statistical methods can be found in Panofsky and Brier (1965), Gringorten

(1955), and Air Weather Service Technical Report 105-38, Short Range and Extended Forecasting by Statistical Methods (1948). Glahn (1975) has given an up-to-date account with an excellent bibliography.

2.4. Probability Terms.

Although probability can be defined several ways (Chacko, 1971), it is most often defined in meteorology as the relative frequency of an event. The relative frequency concept, however, is difficult to extend to rare events which have occurred only once or twice, or perhaps not at all. In this situation, the more natural viewpoint is given by subjective probability which is defined as a personal degree of belief (Savage, 1954). Decision theory requires probability forecasts to optimize use of resources because probability forecasts have more utility than a categorical forecast of the same skill (Thompson and Brier, 1955). The probability notation $Pr(Y)$ denotes the probability of the event Y. In the case of continuous variables the notation $Pr(Y \leq y)$ denotes the probability that the random variable Y, e.g., temperature, will be less than the specified value y, e.g., 32°F. The joint probability of the two events X and Y, $Pr(X,Y)$, denotes the probability that X and Y will both occur.

Conditional probability is the probability of an event based on a given set of information, e.g., the probability of a thunderstorm tomorrow if a thunderstorm occurred today. Conditional probability implies that no other information is given except the specified set. That all information is not given or available is what makes probability a useful concept. If we knew for sure that a thunderstorm would occur tomorrow, we would not need to use probability. Thus, all probability statements are, in fact, conditional probability statements.

Conditional probability uses the notation, $Pr(Y|X)$, which is read: the probability of (the event) Y given that (the event) X has occurred. In forecasting terms, X is the predictor and Y is the predictand. If a second predictor (Z) is added, $Pr(Y|X,Z)$, the conditional probability is different. Thus, the original conditional probability, $(Pr(X|Y)$, could be read: the probability of Y given information about X but not given any information about Z or any other information. In practice, however, some other information is usually implied and not specifically added in the conditional probability notation, e.g., the fact that it is St. Louis, Missouri in July for which the thunderstorm probability is calculated may not appear explicitly in the conditional probability notation. A great deal of confusion in probability forecasting results if the implied information is not specified beforehand.

Chapter 3

OVERVIEW OF THE MODEL

This chapter contains a qualitative description of how the three procedures — transnormalization, correlation, and regression probability — fit together to make an effective forecast method. In brief, equations are known for calculating conditional probability given the multivariate normal distribution. Since many meteorological variables are not normally distributed, they are transformed by a functional relationship into a normal distribution. This process is called transnormalization. Correlation is then calculated between each pair of transnormalized variables. The resulting correlations make up the correlation matrix which is inverted to obtain the regression coefficients. The multivariate normal conditional probability equations are then used to calculate a probability forecast. In this chapter each procedure is described in qualitative manner and mathematical details are presented in the following three chapters.

3.1. The Development Mode.

The development mode consists of three steps: transnormalization, correlation, and calculation of regression coefficients. The dependent set of data and the transnormalizer routine are input to the development mode; the regression coefficients are the output. Correlation is used in the development mode but is not required in the forecast mode.

3.1.1. Transnormalizer Development.

First, a method must be developed to transform the observed or raw predictor into the value of the standard normal variable that has the same cumulative probability as the raw predictor (equivalent normal deviate, END). This transformation

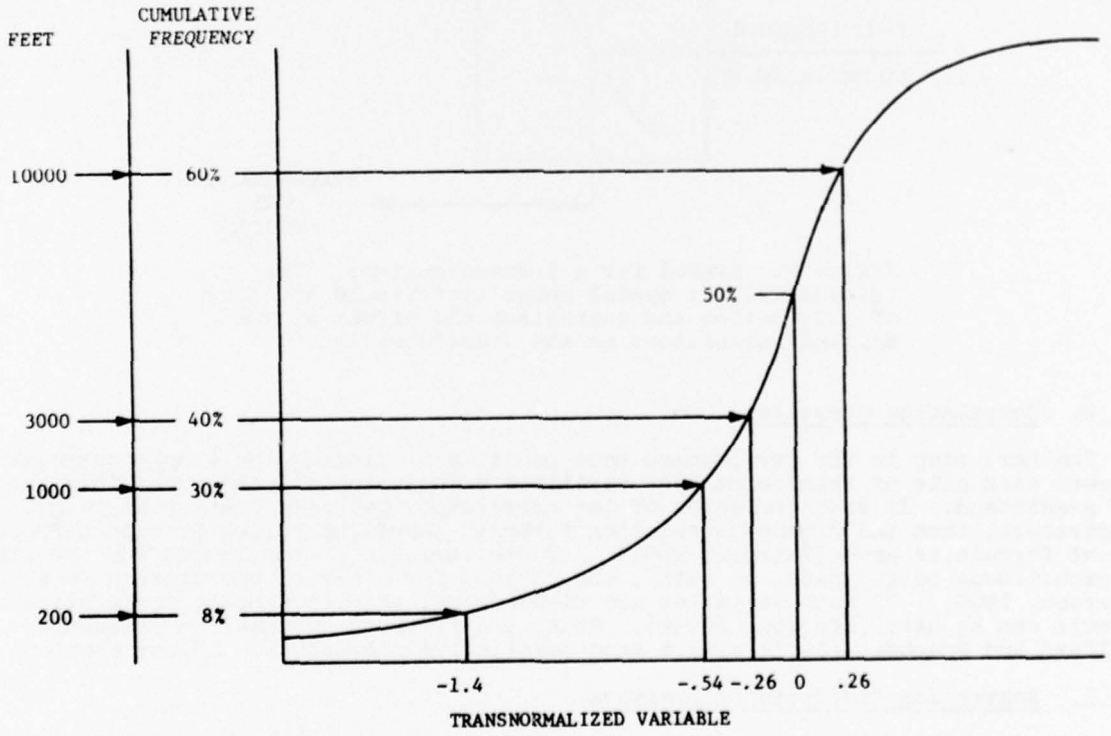


Figure 1. Transnormalization of a Variable. A variable is related by its cumulative probability to an equivalent normal deviate (END).

is based on the predictor's climatological cumulative distribution. For example, suppose a 10,000-ft or less ceiling occurs 60% of the time. A cumulative probability of 60% converts to an END of 0.26 using the normal ogive (see Figure 1). Tables, normal graph paper, and computer algorithms are available to convert from the cumulative frequency to the END. Panofsky and Brier (1965, p. 41) give a description of a graphical procedure.

The term "transnormalized" is used to emphasize that this is a nonlinear transformation. The result is not just a "standard variable" where the mean has been subtracted and the difference divided by the standard deviation. A transnormalizer symbol (see Figure 2) shows that the result depends not only on the raw variable — the direct predictor — but also on marginal predictors. A marginal predictor may be thought of as a condition or constraint on the distribution of the direct predictor; therefore, it changes the shape of the cumulative distribution of the direct predictor. Marginal predictors do not need to be transnormalized and, in fact, usually do not have a distribution per se since the values of marginal predictors are usually selected by man rather than by nature. Typical marginal predictors are time of day, location, time of year, and weather modification effects. Chapter 4 contains further information.

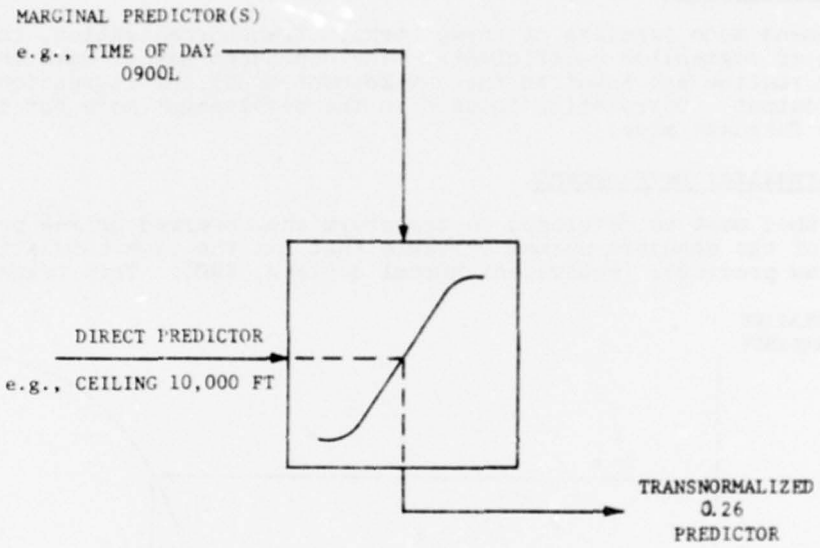


Figure 2. Symbol for a Transnormalizer. The transnormalizer symbol shows pictorially the flow of information and emphasizes the effect of the marginal predictors on the transformation.

3.1.2. Correlation Overview.

The next step in the development mode consists of finding the simple correlation between each pair of transnormalized variables — including all the predictors and the predictand. If both variables of the correlation pair are continuous, e.g., temperature, then the common correlation formula, sometimes called Pearson's Product Moment formula, is used (Pearson, 1895). If one variable is continuous and the other is dichotomous (e.g., rain, no rain), the formula for biserial correlation is used (Pearson, 1909). If both variables are dichotomous, the tetrachoric correlation formula can be used (Pearson, 1913a). Chapter 5 presents correlation formulas. Guilford and Fruchter (1973) give a good descriptive presentation of correlation.

3.1.3. Regression Coefficients Overview.

The regression coefficients are calculated using the same steps as ordinary multiple linear regression. The matrix of simple intercorrelations between predictors is inverted. The inverted matrix is multiplied by the row vector consisting

of correlations between each predictor and the predictand. The resulting vector contains the coefficients needed to calculate conditional probability using multivariate normal distribution. Chapter 6 presents the details and formulas.

3.2. The Forecast Mode.

Given the transnormalizer routines and the regression coefficients, the forecast mode calculates the conditional probability for each set of predictors.

3.2.1. Transnormalization and the Mean Predictor.

In the forecast mode, each direct predictor is first transnormalized with the marginal predictors affecting the transnormalization. The transnormalized predictors are each multiplied by their respective regression coefficients and the results summed to get the "mean" predictor. The term mean predictor is used since the sum just described appears in the same position in the conditional probability equation as the ordinary mean does when a variable is standardized.

3.2.2. Conditional Probability.

Next, the conditional probability is calculated using the multivariate normal distribution. Insight into this calculation can be obtained by considering the simplest case of one predictor. In this case the bivariate normal distribution is used. The bivariate normal is a bell-shaped joint distribution of two variables. Each variable by itself shows a univariate normal distribution, and since trans-

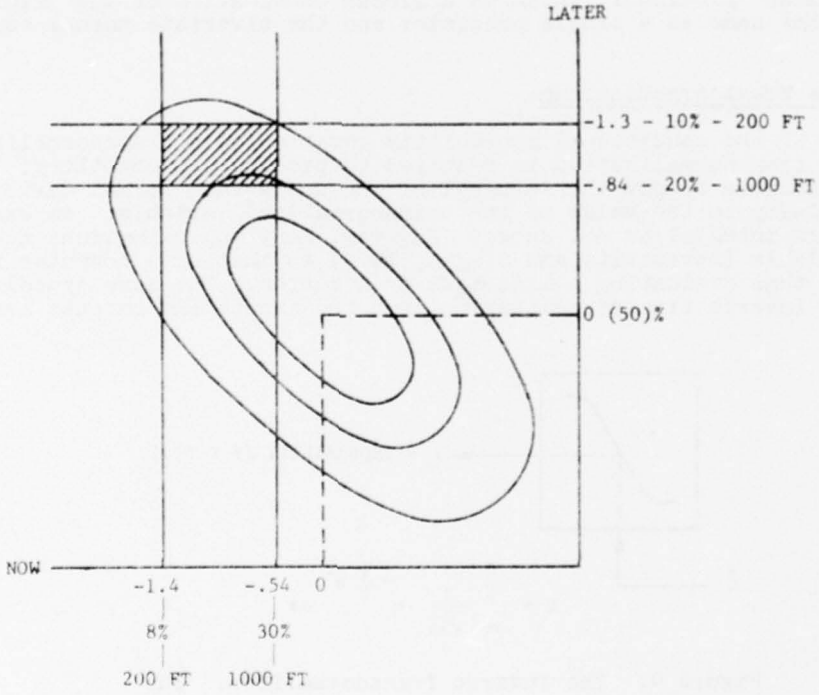


Figure 3. Bivariate Normal Conditional Probability. The bivariate normal is a bell-shaped volume with the maximum density at the means which are equal to zero with transnormalized variables. Isopleths of equal density are ellipses (circles for zero correlation) which become more eccentric as correlation increases. Conditional probability of a predictant category is given by the joint probability (the shaded volume) divided by the probability of the predictor category.

normalized variables are used, the mean of each distribution is zero and the standard deviation is one. If the two variables are correlated, isopleths of equal probability density are ellipses as shown in Figure 3.

If we consider one of the variables as the predictor and the other variable as the predictand, we can calculate the probability that the predictand will be within a specified range, i.e., a category, given that a predictor value has occurred within some specified range. For example, we can calculate the probability that a ceiling 4 hours from now will be in the category 200 feet to 1000 feet given that the ceiling now is in the 200-ft to 1000-ft category. This conditional probability is given by dividing the joint probability of the predictor category and the predictand category by the probability of the predictor category. Unfortunately, the joint probability cannot be calculated except by numerical means such as numerical integration or an infinite series. However, as the predictor category shrinks to a specific predictor value, the conditional probability remains finite, and the limiting equation becomes quite simple (Gringorten, 1972):

$$\overset{\text{"}}{C} = (\overset{\text{"}}{Y} - r\overset{\text{"}}{p})/\sqrt{1 - r^2} \qquad (3.1)$$

where $\overset{\text{"}}{C}$ is the END of the conditional probability

$\overset{\text{"}}{Y}$ is the END of climatological probability of the predictand

r is the correlation between predictor and predictand

$\overset{\text{"}}{p}$ is the END of the predictor

Derivation of this equation is given in Chapter 6. When more than one predictor is used, the "mean" predictor, which is a linear combination of the original predictors, acts the same as a single predictor and the bivariate normal equation can be used.

3.2.3. Inverse Transnormalization.

The result of the conditional probability equation is a transnormalized variable. Inverse transnormalization is required to produce a probability. Inverse transnormalization is simply the integral of the univariate normal distribution from minus infinity to the value of the transnormalized variable. An exact simple formula for this integral is not known. However, many approximations and expansions are available (Abramowitz and Stegun, 1964) so that on a computer it is no more difficult than evaluating a logarithm or a cosine. The same symbol is used for direct and inverse transnormalization, but the inputs and outputs are reversed (Figure 4).

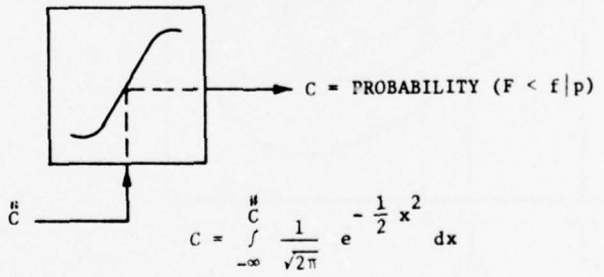


Figure 4. The Inverse Transnormalizer. The inverse transnormalizer symbol shows pictorially the transformation of an END, shown as $\overset{\text{"}}{C}$, into a cumulative probability C. Inverse transnormalization as used here ends with the cumulative probability; no attempt is made to convert the cumulative probability into the equivalent value of the predictand. Inverse transnormalization is simply the normal probability integral function.

3.2.4. Information Flow in the Forecast Mode.

The flow of information and the operations performed on the information are shown in Figure 5. The final probability depends on both direct and marginal predictors. One of the most important effects on the final probability is the effect the marginal predictors have on the climatological probability of the predictand.

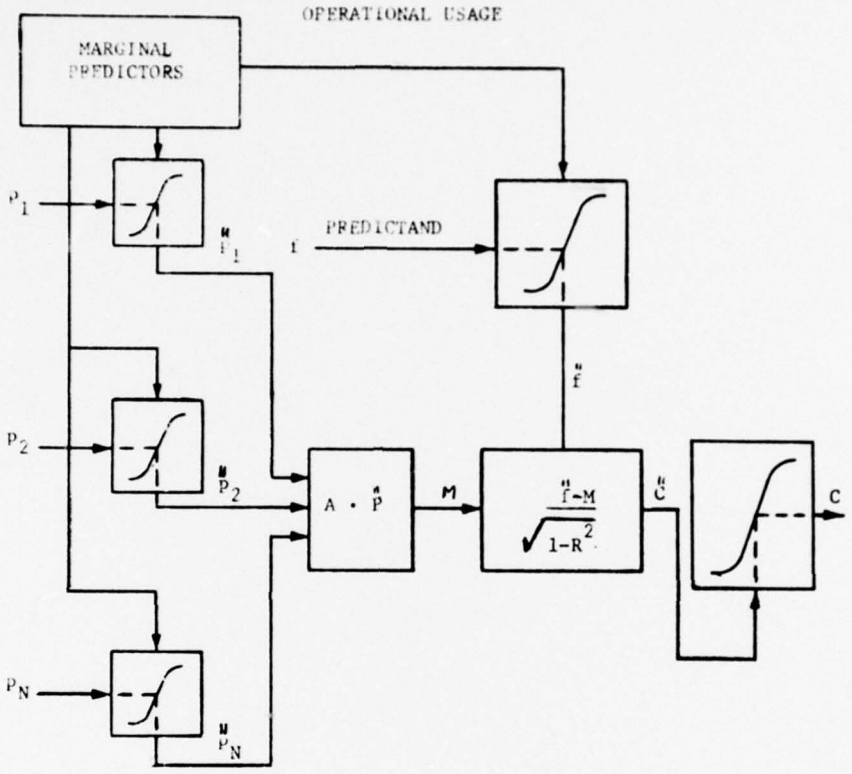


Figure 5. The TRP Forecast Mode Raw Predictors p_1, p_2, ... p_N are Transnormalized into END Variables $\overset{"}{p}_1$, $\overset{"}{p}_2$, ... $\overset{"}{p}_N$. The marginal predictors affect the transnormalizers by changing the shape of the cumulative distributions. Transnormalized variables $\overset{"}{p}$ are multiplied (dot product) by the regression coefficients A to give the scalar M called the mean predictor. The marginal predictors also affect the transnormalization of the predictand. The transnormalized predictand $\overset{"}{f}$ is combined with the mean predictor M and the multiple correlation coefficient R to produce the END of conditional probability $\overset{"}{C}$ which is inverse transnormalized to give the conditional probability C.

Preceding Page BLANK - NOT FILMED

Chapter 4

TRANSNORMALIZATION

4.1. Reasons for Transnormalization.

Transnormalization is the procedure that changes an observed variable into its equavelent normal deviate (END). An END is related to the variable's cumulative frequency of occurrence by use of the cumulative normal distribution. The term "transnormalized" is used to emphasize that this is a nonlinear transformation. The result is not just a standard variable where the mean has been subtracted and the difference divided by the standard deviation.

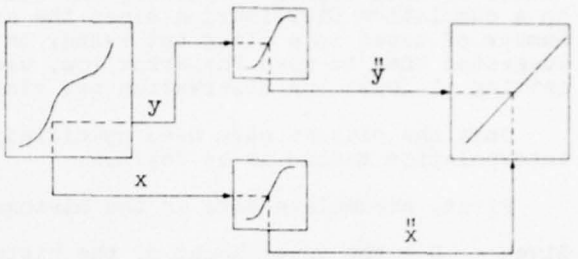


Figure 6. Linearity Caused by Transnormalization. If x and y are related by any arbitrary monotonically increasing function, $y = M(x)$, and both the variables are transnormalized, the resulting relation will be linear. The proof is simple: If events less than x_i occurred n times then n events will also have occurred less than $y_i = M(x_i)$ because x and y are monotonically increasing functions of each other. Thus, x and y will have the same cumulative probability for any x_i and its related y_i, and $\ddot{x}$ and $\ddot{y}$ will be linearly related.

One purpose of the transnormalization process is to insure that the predictor is normally distributed. Another purpose is that many nonlinear effects are taken into account. For example, any continuous one-to-one relationship will become linear if both variables are transnormalized. Of course, not all nonlinear relationships can be taken into account (see Figure 6).

In practice there are several ways to transnormalize a variable. Graphical methods can be accomplished manually. Two-stage methods include histogram interpolation, rank order, and frequency curve fitting. Direct methods use a formula to convert the raw variable directly into a transnormalized variable.

4.2. Graphical Transnormalization.

The graphical method requires first that a series of observations be classified into categories or classes specified by class boundaries. Next, the frequencies in each category or class are converted into a cumulative distribution and graphed as an ogive. Panofsky and Brier (1965) give details and examples of graphing the cumulative distribution.

To use the ogive, enter the graph with the raw value of the predictor and read off the predictor's cumulative probability. Finally, convert the cumulative probability to an equivalent normal deviate (END) by use of normal graph paper or a table of normal cumulative probabilities. The drawbacks of the graphical method are that it is a manual (slow) method and some error is introduced by interpolating on the graph.

4.3. Two-Stage Transnormalization

Two-stage methods consist of an algorithm to convert the raw predictor to its cumulative probability and then a second algorithm to convert the cumulative probability to an equivalent normal deviate. There are two general kinds of algorithms to convert the raw predictor to its cumulative probability: histogram interpolation, which is the automated equivalent of the graphical method, and frequency curve fitting, which fits parameters to a selected distribution or fits the coefficients of a series expansion of a distribution.

4.3.1. Histogram Interpolation.

The histogram interpolation method requires first that class intervals be specified. With frequency histograms the number of classes can be optimized. Fewer, larger classes mean more data in each class, which implies a smaller error due to sampling. Smaller, more numerous classes have greater sampling error, but interpolation error is smaller. Panofsky and Brier (1965, p. 3) give the rule of thumb that the number of classes should equal five times the common logarithm of the number of data values. However, the rule appears not to apply for minimizing error in a cumulative distribution since the error in this case does not depend on the number of cases in a class but rather on the total number below a value. It is suggested that to keep the error low, use as many classes as possible while still getting at least one observation per class.

Once the classes have been specified, the algorithm for using a linear histogram interpolation method is as follows:

First, accumulate data in the histogram.

Given U = the upper bound of the histogram
L = the lower bound
N = the number of classes
T = the total number of observations

Then the class interval (C) is given by:

$$C = (U - L)/N \tag{4.1}$$

An observed value (X) is placed in category I where I is given by:

$$I = (X - L)/C + 1.001 \tag{4.2}$$

where I is truncated to an integer value.

If $I < 1$, set I to 1

If $I > N$, observation is not classified

When an observed value (X) falls on a boundary, the value 1.001 in Equation (4.2) causes X to be placed in the upper class rather than the lower class. This effect and the limits that are imposed on the values of I cause the cumulative probability of a category I (given by Equation (4.3) below) to be the probability that the variables in that category are less than or equal to the upper boundary value of the category. If F(I) is the frequency of category I using the above rules, then the cumulative probability or ogive (O(I)) of category I is given by:

$$O(1) = F(1)/T$$

$$O(I) = [F(I)/T] + O(I - 1), \qquad I = 2, 3, \ldots, N \tag{4.3}$$

Note that O(N) will not equal 1 if some observation were greater than U. To find the cumulative probability (P) equal to or less than some arbitrary value (X), some category (J) is selected first. Next, calculate the fraction (F) between categories:

$$F = G - J \tag{4.4}$$

where $G = [(X - L)/C] + 1$ (G is simply an intermediate result)

and $J = INT(G)$ INT is the integer function, e.g., INT(2.3) = 2

If $J < 1$, set $J = 1$

If $J > N - 1$, set $J = N - 1$

The fraction (F) between categories (Equation (4.4)) is used to linearly interpolate P:

$$P = (1 - F)O(J) + F\ O\ (J + 1) \tag{4.5}$$

For values beyond the observed data, the linear extrapolated probability may be below zero or greater than one. To preclude this from happening, the following algorithm is recommended:

If $P < 0.7/T$, set $P = 0.7/T$

If $P > 1 - 0.7/T$, set $P = 1 - 0.7/T$

The term 0.7/T comes from an extreme value theory using the Poisson distribution. Its derivation is in Appendix A.

4.3.2. Rank Order Transnormalization.

To rank order a set of data means to sort it putting the smallest (or most negative) value first and the largest value last. Once the sort has been completed, the cumulative probability can be directly assigned to each observation. If T is the total number of observations, then the first (lowest) observation has the optimal estimate of cumulative probability of $1/(T + 1)$, the second has $2/(T + 1)$, etc. (Panofsky and Brier, 1965, p. 43).

In practice, ties (two or more observations with the same value) often occur. In this case the straightforward algorithm fails and instead the following algorithm is recommended:

Let T be the total number of observations

S be the number of observations with values less than the Ith observation

H be the number of observations with values greater than the Ith observation

then P(I) is the mean cumulative probability of the Ith observation:

$$P(I) = (S + T - H)/2T \tag{4.6}$$

This algorithm is easy to program, and the mean cumulative probability gives better results for calculating correlation than the less-than-or-equal-to cumulative probability.

The rank order method gives accuracy equal to or better than the histogram method. However, the rank order method has three drawbacks. First, all of the observations must be stored since they are used more than once. The histogram method requires only the frequency in each class to be stored. Second, the rank order method requires a sort where the number of operations (multiplications, additions, etc.) is of the order T^2, whereas the number of operations in the histogram and other methods to be described is of the order T. The difference in the number of operations can make the rank order method expensive in computer time for large amounts of data. Third, the rank order method requires a search when the cumulative probability of a new value is to be calculated. Once the search has found the nearest stored observation greater than the new value, an interpolation scheme similar to Equation (4.5) can be used. The subroutine ATSG in the IBM System/360 Scientific Subroutine Package is efficient in doing the required search.

The above drawbacks limit the usefulness of the rank order method. Storage requirements make it difficult, if not impossible, to program on the small programmable calculators. Nevertheless, it has been used effectively for sets of observa-

tions that number less than 100 and when using only the dependent data since with dependent data, END can be assigned to each observation and no search or interpolation routine is necessary.

4.3.3. Using a Known Distribution.

For certain meteorological parameters, the algebraic form of the climatological distribution is known. For example, the gamma distribution is often used for rainfall. In this case, the parameters of the distribution are fitted or subjectively estimated. Then a procedure must be found to give the cumulative probability (F) for any given value (x). This procedure is equivalent to integrating the frequency distribution f(x):

$$F(x) = \int_{-\infty}^{x} f(x)dx \tag{4.7}$$

If the indefinite integral of f(x) is known, it can be used. If not, rational approximations are sometimes available. For example, Abramowitz and Stegun (1964) give approximations for the gamma distribution, the beta distribution, and others. As a last resort, numerical integration can always be used.

4.3.4. Unknown Distribution — The Gram-Charlier Approximation.

In 1905, Charlier described a series that could be used to approximate a general distribution (Ord, 1972). Usually, only terms using just the first four moments are used. In this case the approximations for the frequency density (g) can be written:

$$g(x) = \Phi(s)\ (1 + K_4/8 - sK_3/6 - s^2K_4/4 + s^3K_3/6 + s^4K_4/24) \tag{4.8}$$

where s is the standardized variable:

$$s = (x - M_1)(M_2)^{-\frac{1}{2}} \tag{4.9}$$

M_1 and M_2 are the first and second central moments (Appendix C), and $\Phi(s)$ is given by the relation:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{EXP}(-z^2/2) \tag{4.10}$$

where z is a dummy variable. K_2, K_3, and K_4 are cumulants calculated from the moments (Appendix C). Equation (4.8) can be integrated by parts to yield the cumulative probability (G):

$$G(x) = \Phi(x)\ (sK_3/6 + sK_4/8 - s^2K_3/6 - s^3K_4/24) + \Phi(s) \tag{4.11}$$

where $\Phi(s)$ is given by the relation:

$$\Phi(\overset{''}{p}) = \int_{-\infty}^{\overset{''}{p}} \frac{1}{\sqrt{2\pi}} \mathrm{EXP}(-\tfrac{1}{2} z^2)\, dz \tag{4.12}$$

where $\overset{''}{p}$ is a dummy variable. Methods of evaluating Equation (4.12) are given in section 4.5. Equation (4.8) can give negative values, particularly for extreme values of s. Thus Equation (4.11) is not always a monotonically increasing function. When this method is used, it should be checked first for significant negative values which should be set equal to zero.

Edgeworth, in 1896, developed a similar expansion (Ord, 1972). However, the Gram-Charlier has a smaller region in which negative values can occur (Barton and Dennis, 1952).

4.3.5. Unknown Distribution — The Pearson System of Curves.

Karl Pearson based a general system of frequency curves on an assumed form of the derivative (f') of an arbitrary frequency curve (Ord, 1972):

$$f' = f\,(x - a)/b - cx - dx^2 \tag{4.13}$$

where x is the dependent variable and a, b, c, and d are parameters.

When Equation (4.13) is integrated, it yields 13 algebraic forms including the normal, the gamma, the inverted gamma, beta, inverted beta, and the exponential distribution. The shapes of the distributions include bell shaped, U shaped, J shaped, and saw tooth. The distributions may be unbounded, bounded at one end, or bounded at both ends. In order to judge which curve should be used, Pearson used the following criterion (K):

$$K = B_1(B_2 - 3)/4(4B_2 - 3B_1)(2B_2 - 3B_1 - 6) \tag{4.14}$$

where

$$B_1 = M_3^2/M_2^3\,, \qquad B_2 = M_4/M_2^2 \tag{4.15}$$

The criterion can range from minus infinity to plus infinity. Various ranges of the criterion indicate which curve should be used (Brooks and Carruthers, 1953, p. 122). For example, K = 1 indicates the inverted gamma should be used.

Once the curve has been selected, parameters of the curve are fitted using the first four moments. Some of the parameters are given by algebraic formulas; others must be fitted by iterative processes.

Elderton (1953) has described in great detail the selection of curves and methods of fitting the parameters. Bouver (1973) describes a set of computer programs to select and fit the Pearson curves. Since each of the 13 algebraic forms require a different fitting algorithm, the procedures are somewhat cumbersome.

4.3.6. Converting Probability to an END.

The first stage of the two-stage method obtains the cumulative probability for a given value of a predictor. The second stage converts this probability into an equivalent normal deviate (END). In other words, one must solve for $\ddot{p}$ given $\Phi(\ddot{p})$ in Equation (4.12). Hastings (1955) has developed rational approximations to solve for $\ddot{p}$. These approximations are also given by Abramowitz and Stegun (1964, Equations 26.2.22 and 26.2.23). One of the rational approximations is:

$$\ddot{p} = t - a + bt/1 + ct + dt\,, \qquad t = \sqrt{\ln\,(1/p^2)} \tag{4.16}$$

$$a = 2.30753,\quad b = 0.27061,\quad c = 0.99229,\quad d = 0.04481$$

The absolute error is less than 0.003. Another algorithm which is handy for the small programmable calculators (Joiner and Rosenblatt, 1971) is:

$$\ddot{p} = 4.91\,[p^{.14} - (1 - p)^{.14}] \tag{4.17}$$

For $0.001 < p < 0.999$ absolute error is less than 0.05.

4.4. Direct Methods.

Direct methods use a function to directly convert the raw predictor value (x) into its END ($\ddot{x}$). Neither the probability density nor the cumulative probability is calculated as an intermediate result. If for some reason, e.g., to check the goodness of fit, probability is needed, the END must be transformed as shown in section 4.5 — inverse transnormalization.

4.4.1. The Cornish-Fisher Expansion.

Cornish and Fisher (1937) developed an infinite series expansion based on cumulants to directly transnormalize a raw predictor. Using just the first four terms of the series gives some desirable smoothing. The END ($\ddot{x}$) is given by:

$$\ddot{x} = A_0 + A_1 s + A_2 s^2 + A_3 s^3 \tag{4.18}$$

where $A_0 = K_3 M_2^{-3/2}/6$

$A_1 = 1 - (7/36)\, K_3^2 M_2^{-3} - K_4 M_2^{-2}$

$A_2 = - K_3 M_2^{-3/2}/6$

$A_3 = K_3^2 M_2^{-3}/6 - K_4 M_2^{-2}/24$

s is the standardized variable $s = (x - M_1)/\sqrt{M_2}$

M_1 and M_2 are the first and second central moments

K_3 and K_4 are cumulants (see Appendix C)

4.4.2. Transnormalized Quantile Polynomial Fitting

If the data has been grouped into classes, the cumulative probability allows these classes to be designated by their quantiles. (Quantiles is a more general name for percentiles.) Let the upper boundary of the ith category be designated by x_i. Also, let $\ddot{x}_i$ be the transnormalized cumulative probability (quantile) of the ith category which can be obtained from Equation 4.16. Then, using the two series (x_i and $\ddot{x}_i$, i = 1, ... N, where N is the number of classes), it is possible to use polynomial regression to get $\ddot{x}$ as a function of x. Standard polynomial regression routines can be used such as documented in the BMD05R Polynomial Regression program (Dixon, 1973). Thus, polynomial equations of the following form can be generated:

$$\ddot{x} = a + bx \tag{4.19}$$

$$\ddot{x} = a + bx + cx^2 \tag{4.20}$$

$$\ddot{x} = a + bx + cx^2 + dx^3 \tag{4.21}$$

Where a, b, c, and d are calculated by polynomial regression. Note that x is the raw variable not the standard variable.

Equation (4.19) is simply a normal distribution with mean - a/b and standard deviation 1/b which can be seen by equating the standard form of a normal equation with mean m and standard deviation s to the linear equation form:

$$\ddot{x} = (x - m)/s = a + bx$$

Nevertheless, it may be necessary to calculate the simple normal equation by regression if the data has been previously categorized in such a fashion that it is difficult to calculate the mean and standard deviation by the usual method of moments.

The Equations (4.20) and (4.21) allow many types of distributions to be fitted. However, it is possible for a negative frequency to be calculated. This will happen when the equations have a negative slope implying that the cumulative frequency is decreasing. This could only occur with a negative frequency (see Figure 7). The slope ($\ddot{x}'$) of Equation (4.20) is:

$$\overset{''}{x}' = b + 2cx \qquad (4.22)$$

The slope is positive (usable range of x) when:

$$-b/2c \leq x \text{ and } c > 0$$

or

$$-b/2c \geq x \text{ and } c < 0$$

Figure 7. Usable Regions (dark lines) of Transnormalized Polynomials. Unusable regions give negative probabilities.

The slope of Equation (4.21) is:

$$\overset{''}{x}' = b + 2cx + 3dx^2 \qquad (4.23)$$

The slope is positive (usable range of X) when:

$$-\infty < x \leq \frac{-c - \sqrt{c^2 - 3bd}}{3d} \text{ and } \frac{-c + \sqrt{c^2 - 3bd}}{3d} \leq x < \infty$$

when

$$d > 0$$

and

$$\frac{-c - \sqrt{c^2 - 3bd}}{3d} \leq x \leq \frac{-c + \sqrt{c^2 - 3bd}}{3d}$$

when

$$d < 0$$

There are several alternatives if the observed range of the variable includes a negative frequency: (a) the method can be discarded and a different method used, (b) the negative frequency set to zero when it occurs, or (c) if the negative value is small in magnitude, it can be used in the case of a predictor with the resultant small error in conditional probability.

4.4.3. Johnson's Family of Curves.

Johnson (1949), to overcome the possibility of negative frequency, advised the use of monotonically increasing functions rather than polynomials for direct transnormalizations. In particular, he found that functions of the form:

$$\overset{''}{x} = a + b \ln \left(\frac{x - L}{U - x}\right) \qquad (4.24)$$

$$\overset{''}{x} = a + b \ln (x + k) \qquad (4.25)$$

$$\overset{''}{x} = a + b \operatorname{Sinh}^{-1} (gx + h) \qquad (4.26)$$

are quite versatile; they can fit any of the Pearson family of curves.

Equation (4.24) is most useful for fitting distributions that are bounded at both end points, e.g., eighths of sky cover or relative humidity. In the bounded distribution, L is the lower bound and U is the upper bound. Numerically, it is wise to make L slightly lower and U slightly higher to keep the argument of the logarithm within limits.

Equation (4.25) gives the well-known lognormal distribution. It is particularly useful for distributions that are bounded at one end, e.g., ceiling, visibility, of rainfall. Often -k is equated to the lower bound or is given a slightly lower value.

Equation (4.26) can be used for unbounded distributions, e.g., vertical motion or pressure.

If U and L, k, or g and h are known, it is easy to calculate a and b by means of linear regression or transnormalized quantiles as described in section 4.4.2. However, if bounds are not known, U and L, k, or g and h require nonlinear fitting. Johnson (1949) and Ord (1972) give several ways of fitting these variables. The following methods were developed by the author using the Taylor series expansion for the logarithm and hyperbolic sine functions (see Appendix B). To fit the log-normal Equation (4.25), a third-order polynomial regression is applied to the transnormalized quantiles (see Equation (4.21)). Thus, k is fitted:

$$k = -2/3\ c/d \tag{4.27}$$

if $x - k < 0$, k is set to the minimum value of x, c and d are the coefficients derived for Equation (4.21). The hyperbolic sine coefficients, g and h, can be fitted using

$$h = \sqrt{\frac{2c^2}{c^2 - 3db}}, \quad g = 3hd/c \tag{4.28}$$

where b, c, and d are coefficients derived for Equation (4.21). If the term in the radical is negative, make it positive before attempting to take the square root and then set h negative.

4.5. A Transnormalization Example.

The above methods were used on a variety of meteorological parameters — temperature, relative humidity, wind speed, eighths of sky cover, 1000-500 mb thickness, etc. — to learn how well the transnormalization procedures fit observed distributions. Figure 8 shows a typical example including various fits to the Pittsburgh 1900 LST surface wind speed for October and November for the years 1961 through 1965. Statistical parameters for the distribution are shown in Table 1.

Table 1. Statistical Parameters for Pittsburgh Wind Speed.

Moments	7.15	18.4	66.0	1626.		
Cumulants		18.3	66.0	628.		
Skewness = .839	Kurtosis = 1.86		Number of cases = 285			
Pearson's B_1 = .696		B_2 = 4.81				

	Constant	1st	2nd	3rd	RMSE	Max Error
Cornish-Fisher Coefficients	0.140	1.10	-0.140	0.040	0.027	0.079
1st-Order Regression	-1.25	0.187	--	--	0.029	0.103
2nd-Order Regression	-1.70	0.305	-0.00490	--	0.026	0.077
3rd-Order Regression	-1.59	0.243	0.00167	-0.000183	0.025	0.084

Usable range of 2nd-order regression fit is from minus infinity to 31.1.
Usable range of Cubic fit is -54.7 to 73.0.

	A	B	C	D	RMSE	Max Error
$\tilde{x} = A+B*\Phi((x-c)/(D-x))$	1.14	0.841	-0.500	25.5	0.044	0.114
$\tilde{x} = A+B*\mathrm{Sinh}^{-1}(c*x+d)$	-0.765	-3.36	-0.0664	0.203	0.027	0.092
$\tilde{x} = A+B*\Phi(x+c)$	-7.36	2.99	6.10	--	0.029	0.067
Gram-Charlier $\Phi(\tilde{x}) = Z(Z+BS+Cs^2+Ds^3)+\Phi(S)$	0.419	0.233	-0.140	-0.078	0.026	0.060

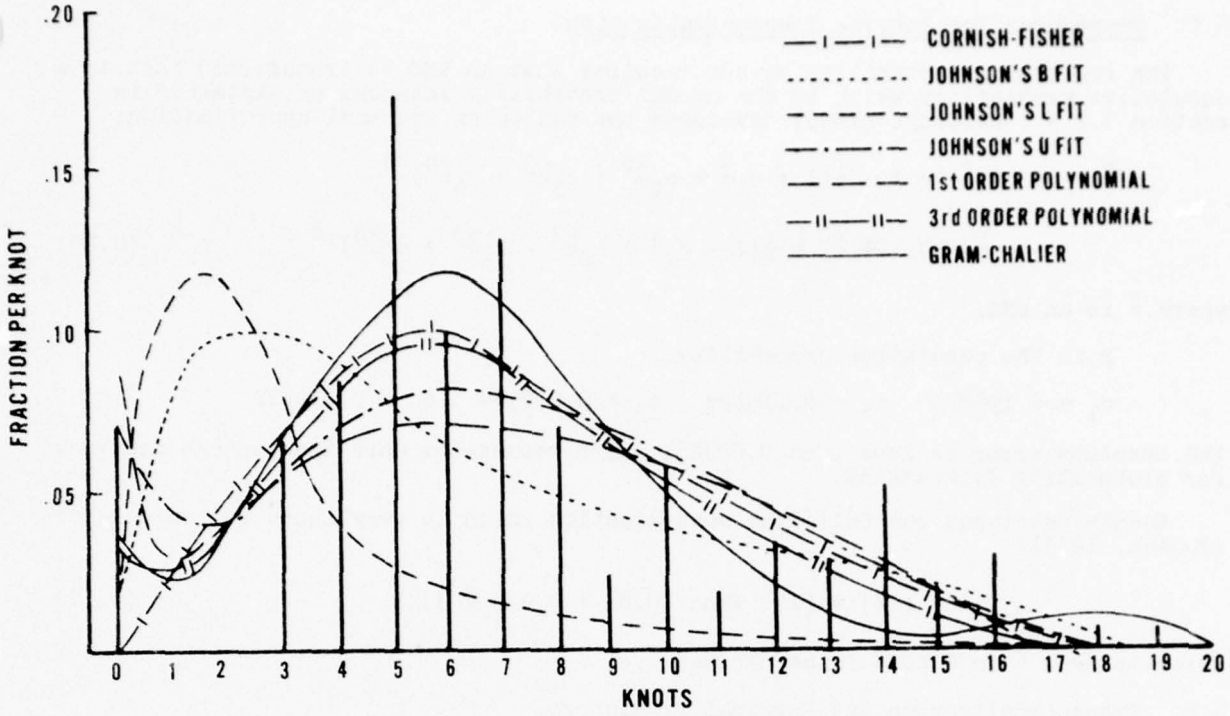


Figure 8. Transnormalized Distribution Approximations to the Pittsburgh Surface Wind Speed. Approximation parameters are given in Table 1. The 2nd-order polynomial graph is nearly identical to the 3rd-order polynomial graph. Some of the approximations are poor due to recording errors of the anemometer and observer.

Some anomalies in the wind speed distribution are typical of observed meteorological distributions in general. Because the anemometer does not respond to a wind less than 3 knots, speed of 1 and 2 knots are not recorded causing a subsequent increase in the number of calm winds. Observer preference for numbers divisible by five shows up in the high observed frequency of 5- and 10-knot winds. Surprisingly, the 15-knot speed does not show a high observed frequency.

The root mean square errors (RMSE) and maximum errors of various transnormalizing procedures are given in Table 1. The term error may be a misnomer since some of the fitted curves may be closer to the actual distribution than the observed frequency which includes instrument and observer bias.

4.6. Selecting a Transnormalization Procedure.

None of the transnormalization procedures work perfectly for all situations. Furthermore, since the data itself contain measurement errors and sampling errors, a good fit on one set of data is not a guarantee that the transnormalizer will do well on more data of the same kind. Selection is, therefore, something of an art. The following strategy is tentatively suggested:

a. If the variable has been known to fit a given distribution, try using the distribution in a two-stage method (section 4.3.3).

b. If the data have been already classified, use histogram interpolation (section 4.3.1) or polynomial regression (section 4.4.2). If the classes are unequally spaced, polynomial regression is the better choice.

c. If the data are continuous with an unknown distribution, the Cornish-Fisher expansion will generally fit the data well (section 4.4.1).

4.7. Procedures for Inverse Transnormalization.

The regression probability method requires that an END be transformed back to a cumulative probability, which is the normal probability integral as explained in section 3.2.3. Hastings (1955) developed the following rational approximation:

$$\ddot{e} > 0: \quad \Phi(\ddot{e}) = 1 - \tfrac{1}{2}(1 + c_1\ddot{e} + c_2\ddot{e}^2 + c_3\ddot{e}^3 + c_4\ddot{e}^4)^{-4}$$

$$\ddot{e} < 0: \quad \Phi(\ddot{e}) = \tfrac{1}{2}(1 - c_1\ddot{e} + c_2\ddot{e}^2 - c_3\ddot{e}^3 + c_4\ddot{e}^4)^{-4} \qquad (4.29)$$

where $\ddot{e}$ is an END.

P is the cumulative probability

$c_1 = 0.196854$ $c_2 = 0.115194$ $c_3 = 0.000344$ $c_4 = 0.019527$

The absolute error is less than 0.00025, which represents more than enough accuracy for probability forecasting.

Muench developed the following approximation which is very handy for calculators (Touart, 1973):

$$P(\ddot{e}) = [1 + \text{Tanh}\ (0.8\ddot{e} + 0.035\ \ddot{e}^3)]/2 \qquad (4.30)$$

Error is less than 0.0004 in magnitude.

4.8. Transnormalization and Marginal Predictors.

Most of the transnormalizing methods discussed the use of three or four parameters to specify the transformation. In general, the value of these parameters will vary depending on the time of day or season of the year; thus, the parameters become functions of the marginal predictors. An example of this can be seen in the visibility transnormalization method used at Cambridge Research Laboratories (now Air Force Geophysics Laboratories)(Chisholm, et al., 1974). In this case the END of visibility ($\ddot{v}$) is given by a lognormal transformation:

$$\ddot{v} = L + K\ \ln(v) \qquad (4.31)$$

where v is visibility and L and K are functions of time of day:

$$K = A \cdot H + B, \qquad L = C \cdot H + D \qquad (4.32)$$

where H is the time before or after sunrise, and A, B, C, D are climatologically determined constants with one set for winter and one for summer. The time of day, as measured by H, will have an important effect on the value of $\ddot{v}$. Climatological probability for any value of visibility for any time of day is stored in only four numbers!

The final point to remember is that the transnormalization parameters, including the marginal predictor transformations, are based on the relatively long climatological data base rather than a specific development data base of paired predictor and predictand observations.

Chapter 5

CORRELATION

Correlation is required to calculate the regression probability coefficients. A specific type of correlation, found in the multivariate normal equation, must be used. In the bivariate normal equation, correlation (ρ) appears in three places (Anderson, 1958):

$$f = \frac{1}{2\pi\sqrt{1 - \rho}} \text{EXP}\left[-\tfrac{1}{2}(x^2 - 2\rho xy - y^2)/(1 - \rho^2)\right] \tag{5.1}$$

where f is the probability density, and ρ is correlation.

If Equation (5.1) is used as the basic model, how can an estimate of ρ be calculated if x or y has been distorted or categorized? Distortion is handled effectively by transnormalization. To handle continuous or categorized variables properly, several methods have been developed to calculate correlation.

5.1. The Pearson Product — Moment Formula.

If x_i and y_i, i = 1, ..., N, are a series of paired observations and x and y are each normally distributed (i.e., have been transnormalized), an estimate of the correlation (r) can be calculated.

$$r = \frac{\frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x\, s_y} \tag{5.2}$$

where the summation is i = 1, ..., N and $\bar{x}$ and $\bar{y}$ are means:

$$\bar{x} = \frac{1}{N}\sum x_i\ , \qquad \bar{y} = \frac{1}{N}\sum y_i \tag{5.3}$$

s_x and s_y are standard deviations:

$$s_x = \left[\frac{1}{N}\sum (x_i - \bar{x})\right]^{\frac{1}{2}}\ , \qquad s_y = \left[\frac{1}{N}\sum (y_i - \bar{y})\right]^{\frac{1}{2}}$$

An equivalent and easier to compute formula is:

$$r = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{\left[N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2\right]^{\frac{1}{2}}\left[N\left(\sum y_i^2\right) - \left(\sum y_i\right)^2\right]^{\frac{1}{2}}} \tag{5.4}$$

If x and y are large numbers with means not equal to zero, Equation (5.4) can have considerable round-off error, but with transnormalized variables, this error is negligible. If x and y have been transnormalized, their means should equal zero and standard deviation equals one. Thus, Equation (5.2) could be written:

$$r = \frac{1}{N}\sum x''_i\, y''_i \tag{5.5}$$

However, experience has shown that because of sampling error (the mean of a given sample will generally not equal zero) and transnormalization error, Equation (5.5) will often give misleading results (e.g., r = 1.2)!

The standard error (E_{ppm}) of r when r is calculated by Equation (5.4) is (Elderton, 1953, p. 193):

$$E_{ppm} = (1 - r^2)/\sqrt{N - 1} \quad \text{(approximation)} \tag{5.6}$$

5.2. Biserial Correlation.

Suppose a normally distributed variable is divided into two categories. If the variable is below some critical value it is placed in category 1, otherwise it is put into category 2: for example, temperature above or below freezing. If the above- or below-freezing category is given for one location and the actual temperature is given at another location, is there a way to calculate correlation between the temperatures at the two locations? The biserial correlation formula (Pearson, 1909) is the answer:

$$r = \frac{(\bar{A} - \bar{B})\, pq}{s\, y} \tag{5.7}$$

where $\bar{A}$ is the mean value of the continuous variable for those cases that are in the category above the critical value

$\bar{B}$ is the mean value of the continuous variable for the cases below the critical value

p is the fraction of cases in the upper category

q is the fraction of cases in the lower caregory (i.e., q = 1 - p)

s is the standard deviation of the continuous variable

y is the ordinate of the normal curve at the equivalent critical value:

$$y = \frac{1}{\sqrt{2\pi}} \text{EXP}\left(-\tfrac{1}{2}\, q''^2\right) \tag{5.8}$$

and q'' is the END for the fraction q.

q'' can be calculated by Equations (4.16) or (4.17).

Guilford (1973) gives an equivalent formula that is slightly easier to compute since only the mean of one category needs to be compiled. The required overall mean of the continuous variable is a by-product of calculating the standard deviation. The alternate formula is:

$$r = \frac{(\bar{A} - \bar{C})}{s}\,\frac{p}{y} \tag{5.9}$$

where $\bar{A}$, p, y, and s are defined in Equations (5.7) and (5.8); $\bar{C}$ is the overall mean of the continuous variable. The standard error (E_b) of the biserial correlation coefficient is (Guilford, 1956, Equation 14.8):

$$E_b = \frac{(\sqrt{pq}/y) - r^2}{\sqrt{N}} \tag{5.10}$$

where p, q, y, and r are defined in Equations (5.7), (5.8), and (5.9); N is the number of cases.

Equations (5.9) and (5.10) are based on a normal (or transnormalized) variable which has been categorized into two parts. If a variable is categorized by nature, e.g., snow versus rain, we assume that there is a continuous and normal underlying variable. In the case of rain versus snow, this variable could be a type of mean temperature in the lowest layers of the atmosphere. The underlying variable for many meteorological categorized variables is obvious but, whether there is an underlying variable is not the key point. The key point is: if an underlying variable is postulated, how good will the probability forecast be? Since a formal answer is not possible the approach must be philosophical. Assume an underlying distribution, then calculate a correlation. This calculation is then used to calculate regression probability. The consistent use of this assumption throughout the calculations avoids some of the potential error. The verifications done by McCabe (1968), Gringorten (1971), Boehm (1973), and Martin, Hull, and Chin (1973) show that

the assumption works well in the cases tested.

5.3. Tetrachoric Correlation.

If both variables are categorized into two categories, the result is a fourfold or tetrachoric table (as seen at the right), where A, B, C, and D are the number of occurrences above or below the critical values of the respective variables.

		VARIABLE 1 (CRITICAL VALUE) BELOW	ABOVE
VARIABLE 2	BELOW	A	B
	ABOVE	C	D

Unfortunately, there is not a simple exact formula for calculating correlations for a fourfold table. The exact expression containing r is (Elderton, 1953, p. 177):

$$A/N = \sum_{i=0}^{\infty} T_i(\overset{''}{x})\ T_i(\overset{''}{y})\ r^i \tag{5.11}$$

where A is the value from the tetrachoric table
N is the total cases, N = A + B + C + D
$\overset{''}{x}$ is the END of (A + B)/N
$\overset{''}{y}$ is the END pf (A + C)/N
T is the ith tetrachoric function

$$T_i(\overset{''}{x}) = \frac{(-1)^{i-1}\phi^{i-1}(\overset{''}{x})}{i!} \tag{5.12}$$

$\phi^K(\overset{''}{x})$ is the Kth derivative of the normal density function:

$$\phi^K(\overset{''}{x}) = \frac{d^K}{dx^K}\frac{1}{\sqrt{2\pi}}\ \text{EXP}\ (-\tfrac{1}{2}\ \overset{''2}{x}) \tag{5.13}$$

$\phi^K(\overset{''}{x})$ is easily calculated by the recurrence equation (Abramowitz and Stegun, 1964):

$$\phi^K(\overset{''}{x}) = -\ \overset{''}{x}\phi^{K-1}(\overset{''}{x}) - (K-1)\phi^{K-2}(\overset{''}{x}) \tag{5.14}$$

for K > 1 and with starting values for K = 1 and K = 0:

$$\phi^0(\overset{''}{x}) = \frac{1}{\sqrt{2\pi}}\ \text{EXP}\ (-\tfrac{1}{2}\ \overset{''2}{x}) \tag{5.15}$$

$$\phi^1(\overset{''}{x}) = -x\phi^0(\overset{''}{x}) \tag{5.16}$$

also,

$$\phi^{-1}(\overset{''}{x}) = \frac{A+B}{N}\ ,\quad \phi^{-1}(\overset{''}{y}) = \frac{A+C}{N} \quad \left[\begin{array}{l}\text{Superscripts here are}\\ \text{index not inverse}\end{array}\right] \tag{5.16}$$

Abramowitz and Stegun give the following series expansion which can be derived from Equations (5.11) and (5.12):

$$A/N = \frac{(A + B)(A + C)}{N^2} + \sum_{i=0}^{\infty} \frac{\phi^i(\overset{\shortparallel}{x})\phi^i(\overset{\shortparallel}{y})r^i}{(i-1)!} \tag{5.18}$$

However, $\phi^i(\overset{\shortparallel}{x})$ alternates in sign and can grow beyond computer limits if $\overset{\shortparallel}{x} > 1$. To solve this problem, the factoral and power of r were incorporated into the difference Equation (5.14):

$$A/N = \frac{(A + B)(A + C)}{N^2} + \sum_{i=0}^{\infty} R_i(\overset{\shortparallel}{x})S_i(\overset{\shortparallel}{y}) \tag{5.19}$$

where $R_i = -xR_{i-1}/(i - 1) - (i - 1)R_{i-2}/i(i - 1)$

$$R_0 = \phi^0(\overset{\shortparallel}{x}), \quad R_1 = -\overset{\shortparallel}{x}\phi^0(\overset{\shortparallel}{x})/2 \tag{5.20}$$

and $S_i = -\overset{\shortparallel}{y}S_{i-1}r - (i - 1)S_{i-2}r^2$

$$S_0 = r\phi^0(y), \quad S_1 = -yr^2\phi^0(y)$$

If the infinite series of Equation (5.19) is truncated after two successive terms of the series are smaller in magnitude than the desired error, possible divergence is prevented and the calculations will remain within computer limits. To solve for r, a method for finding the zeros of a polynomial must be used. Elderton (1953) recommends Newton's method. The Scientific Subroutine Package has a subroutine TETRA which uses r to the seventh power and uses a Newton-Raphson method to find the root. However, seven terms are not sufficient to give adequate accuracy for a large $\overset{\shortparallel}{x}$ or $\overset{\shortparallel}{y}$.

The number of terms must be specified in advance when using any polynomial zero-finding method. To overcome this drawback, an algorithm based on the false position method is recommended. Evaluate the polynomial at two initial guess values, and then use linear extrapolation to find a better estimate.

Let $U(r) = A/N$ as calculated by Equation (5.19).

Convenient first guesses are:

$$r_1 = 0, \quad U(r_1) = (A + B)(A + C)/N^2 \tag{5.21}$$

$$r_2 = \text{Sin}\left[\frac{\pi}{2}\frac{\sqrt{AD} - \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}\right] \tag{5.22}$$

$U(r_2)$ is found using Equation (5.19).

The improved estimate of r is found:

$$r_i = r_{i-1} - (r_{i-1} - r_{i-2})\left[\frac{U(_{i-1})A/N}{U(r_{i-1}) - U(r_{i-2})}\right] \tag{5.23}$$

The process is stopped when:

$$|U(r_i) - A/N| < \epsilon$$

where ϵ is much smaller than the acceptable error, e.g., $\epsilon = 1 \times 10^{-8}$.

Equation (5.22) is sometimes specified as the equation for tetrachoric correlation; however, it is only an approximation. It is accurate when $(A + B)/N = 0.5$ and $(A + C)/N = 0.5$, but for values near one or zero for $(A + B)/N$ and $(A + C)/N$, Equation (5.22) contains sizable error (Castellan, 1966).

The standard error (E_t) of the tetrachoric correlation was first derived by Pearson (1913a):

$$E_t = \frac{1}{X\sqrt{N}} [(a + d)(c + b)/4 + H^2 (a + c)(d + b) + G^2 (a + b)(d + c) + 2GH (ad - bc) - H (ab - cd) - G (ac - bd)]^{-\frac{1}{2}} \quad (5.24)$$

where $a = A/N$, $b = B/N$, $c = C/N$, $d = D/N$; A, B, C, D are the values in the fourfold table which has been arranged so that $A + C > B + C$ and $A + B > C + D$; and

$$G = P [(\overset{''}{x} - r\overset{''}{y})/\sqrt{1 - r^2}]$$

$$H = P [(\overset{''}{y} - r\overset{''}{x})/\sqrt{1 - r^2}]$$

P is the cumulative normal function defined by Equation (4.12) and calculated by Equations (4.29) or Equation (4.30)

$\overset{''}{x}$ is the END of $(A + B)/N$

$\overset{''}{y}$ is the END of $(A + C)/N$

5.4. Other Methods of Calculating Correlation.

There are many ways to calculate correlation beyond the three methods described above. A few of the more prominent ones are mentioned for illustration.

Kendall (1948) gives a generalized equation for correlation. Spearman's rank correlation, Kendall's rank correlation, as well as the Pearson product moment formula can be generated from the general equation. The rank correlation method can give estimates of the bivariate normal correlation (Kendall, 1948, Equation 9.2). However, they suffer from the same drawback as the rank transnormalization method in that they require a sort which is expensive in computer time for large sets of data.

The coefficient of contingency combined with a correction factor (Elderton, 1953) can be used in the case where both variables are split into many categories. Herring (Touart, 1973) used a minimum chi-square fit to calculate correlation in a contingency table. The disadvantage of these two methods is that they weigh each cell in the contingency table equally. If, as is often the case in meteorology, one or several of the cells have only a few cases, the standard error of the method will be large.

5.5. Selecting a Correlation Procedure.

If both variables are continuous and have been transnormalized, the Pearson product moment formula should be used since it has a smaller standard error than the tetrachoric or biserial methods.

Many continuous variables, however, are categorized by the measuring instruments or by the method of encoding the data. For example, sky cover is reported in eighths or tenths rather than a continuous variable from zero to one. Pearson (1913b) calculated the error for various numbers of categories. For five categories or less, there is more than a 5% error in the calculated correlation. Therefore, if there are five or less categories, the biserial or tetrachoric procedure is recommended .

These categories should be divided into two groups so that each has about the same number of cases.

For 6 to 14 equal-size categories, the product moment method can be used, using the mid-points of the classes and the corrections C_x and C_y, which depend on the number of categories for each variable. The information in the table on the right is taken from Pearson (1913b). The corrected value of correlation r' is:

Number of Classes	C_x or C_y
6	.960
8	.977
10	.985
12	.989
14	.992

$$r' = r/C_x C_y \tag{5.25}$$

For 15 or more classes, the Pearson product moment formula could be used directly with little error.

Many meteorological variables are truncated beyond a certain value. For example, visibility is reported only up to 6 miles in the aviation METAR code. All values greater than six are reported only as "six plus." In this case, the biserial or tetrachoric formula will give a better estimate of correlation than the product moment formula because all the values of "six plus" will cause a spike in the distribution. The difference between the tetrachoric and product moment correlation estimates can be large. Tetrachoric correlation is easier to use than the Pearson product moment correlation when handling large data samples which exceed available computer memory.

The product moment formula requires two passes through the data, once for transnormalizing and once for calculating correlation using the transnormalized variables. The tetrachoric formula requires only one pass since the four required values can be accumulated during the transnormalization data pass. The higher standard error of the tetrachoric method is not a significant factor with large data samples. There is, however, one disadvantage. The boundary dividing the two classes must be specified in advance when using the tetrachoric method.

Chapter 6

REGRESSION PROBABILITY

Regression probability is based on the multivariate normal distribution. Details of the multivariate normal distribution can be found in texts by Anderson (1958), Miller (1964), Hogg and Craig (1965), and many others. Analysis of the multivariate normal is such a broad field that only the regression probability equation will be derived. As much as possible, Anderson's notation is used.

The following derivation of the regression probability equation first defines general multivariate normal terminology. The only assumption made in the derivation of the TRP model is: Given that the predictand and the predictor variables have been individually transnormalized, they are distributed joint normally.

6.1. The Multivariate Normal.

Let $\underline{X}$ be a matrix of observations:

$$\underline{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ & & \cdots & \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} \tag{6.1}$$

where the first subscript indicates to which observation the set of measurements belongs, e.g., May 3, May 4, May 5, etc. The second subscript indicates which element is being measured, e.g., temperature, wind speed, pressure, etc.

If $\underline{Z}$ is the matrix of deviation from the mean, $Z_{ij} = X_{ij} - \overline{X}_i$, then the covariance matrix $\underline{\Sigma}$ is defined by (superscript T indicates transpose):

$$\underline{\Sigma} = \underline{Z}^T \underline{Z} \frac{1}{N} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ & & \cdots & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix} \tag{6.2}$$

Any individual element of $\underline{\Sigma}$ is given by:

$$\sigma_{ij} = \frac{1}{N} \sum_{k=1}^{N} Z_{ik} Z_{jk} = \overline{Z_i Z_j} \tag{6.3}$$

which means the matrix $\underline{\Sigma}$ is symmetric: $\sigma_{ij} = \sigma_{ji}$.

Let $\underline{X}$ be divided into two sections:

$$\underline{X}^{(1)} = \begin{bmatrix} x_{11} \\ x_{21} \\ \cdots \\ x_{N1} \end{bmatrix} \qquad \underline{X}^{(2)} = \begin{bmatrix} x_{12} & x_{13} & \cdots & x_{1M} \\ x_{22} & x_{23} & \cdots & x_{2M} \\ & & \cdots & \\ x_{N2} & x_{N3} & \cdots & x_{NM} \end{bmatrix} \tag{6.4}$$

$$\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix} = \left[\begin{array}{c|ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1M} \\ \hline \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2M} \\ & & \cdots & & \\ \sigma_{M1} & \sigma_{M2} & \sigma_{M3} & \cdots & \sigma_{MM} \end{array}\right] \tag{6.5}$$

$\underline{\Sigma}_{11}$ is the variance of the predictand, while $\underline{\Sigma}_{21}$ and $\underline{\Sigma}_{12}$ are the vectors each of which contain the covariances between the predictand and each predictor. $\underline{\Sigma}_{22}$ is the matrix of covariances between predictors. According to Anderson's definition 2.5.1, $\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}$ is the matrix of regression coefficients. That is, in a multiple regression equation:

$$\hat{x}_1 = C + a_2\, x_2 + a_3\, x_3 + \ldots a_M\, x_M \tag{6.6}$$

the a's are given by:

$$\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1} = a_2,\ a_3,\ \ldots,\ a_M = \underline{A} \tag{6.7}$$

Equation (6.6) begins with $a_2\, x_2$ because of the manner in which x has been divided. That is, x_2 is the first predictor. C in the equation is not important in this development since it will equal zero when x_1, x_2, ... x_M are transnormalized variables with means equal to zero. Suppose the distribution of $\underline{X}$ is multivariate normal. Then, the frequency density, $f(\underline{X})$, is given by:

$$f(\underline{X}) = \text{EXP}\ [-\tfrac{1}{2}(\underline{x} - \underline{\mu})^T\ \underline{\Sigma}^{-1}\ (\underline{x} - \underline{\mu})]/(2\pi)^{M/2}\sqrt{|\underline{\Sigma}|} \equiv N\ (\underline{\mu},\ \underline{\Sigma}) \tag{6.8}$$

where $\underline{x}$ is the vector of the ith measurement

$\underline{\mu}$ is the vector of the mean value of the ith measurement

Superscript T indicates transpose

Equation (6.8) describes a multidimensional density with maximum density at the mean. Surfaces of equal density are hyperellipsoids with eccentricity depending on the covariance matrix. The notation $N(\underline{\mu}, \underline{\Sigma})$ emphasizes that the distribution is completely specified by the mean vector $\underline{\mu}$ and the covariance matrix $\underline{\Sigma}$.

6.2. Conditional Probability.

The conditional distribution of $\underline{X}^{(1)}$ given a specified $\underline{X}^{(2)}$ is, by Anderson's theorem 2.5.1:

$$f(\underline{X}^{(1)} | x_2,\ x_3\ \ldots\ x_M) = N[\mu^{(1)} + \underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ (x^{(2)} - \mu^{(2)})\ \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ \underline{\Sigma}_{21}] \tag{6.9}$$

Thus, for x_2, x_3, ... x_M held constant, the distribution of $X^{(1)}$ (the predictand) is normally distributed, N(m, v), with mean m and variance v. If the variables in X have $\mu_i = 0$ and $\sigma_{ii} = 1$ (transnormalization insures that they will), the distribution is called the standard normal and the equations are simplified:

$$m = \mu^{(1)} + \underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ (\underline{X}^{(2)} - \underline{\mu}^{(2)}) \rightarrow (\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1})\ \underline{X}^{(2)} \tag{6.10}$$

$$v = \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ \underline{\Sigma}_{21} \rightarrow 1 - (\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1})\ \underline{\Sigma}_{21} \tag{6.11}$$

and further simplified by substituting Equation (6.7).

$$m = (\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1})\ \underline{X}^{(2)} = A\underline{X}^{(2)} = a_2x_2 \ \ldots\ a_mx_m \tag{6.12}$$

$$v = 1 - (\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1})\ \underline{\Sigma}_{21} = 1 - A\ \underline{\Sigma}_{21} = 1 - (a_2\sigma_{21} + a_3\sigma_{31} + \ldots + a_m\sigma_{m1}) \tag{6.13}$$

or using Anderson's definition 2.5.3 and the fact that $\sigma_{ii} = 1$ for transnormalized variables, the multiple correlation coefficient (R) is given by:

$$R = \frac{\sqrt{\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ \underline{\Sigma}_{21}}}{\sqrt{\sigma_{ii}}} = \sqrt{\underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ \underline{\Sigma}_{21}} \tag{6.14}$$

or

$$R^2 = \underline{\Sigma}_{12}\ \underline{\Sigma}_{22}^{-1}\ \underline{\Sigma}_{21} = \underline{A}\ \underline{\Sigma}_{21} \tag{6.15}$$

Thus, from Equations (6.13) and (6.15), the variance of the conditional distribution can be written:

$$v = 1 - R^2 \tag{6.16}$$

and the standard deviation:

$$s = \sqrt{1 - R^2} \tag{6.17}$$

Using notation Pr(Q) as the probability of Q, and $\Phi(Q)$ as the normal cumulative probability less than Q:

$$Pr(X^{(1)} < x^{(1)} | x_2, x_3, \ldots x_m) = \Phi\left(\frac{x^{(1)} - m}{s}\right) \tag{6.18}$$

or using Equations (6.12), (6.13), and (6.17):

$$Pr(X^{(1)} < x_1^{(1)} | x_2, x_3 \ldots x_m) = \Phi\left(\frac{x_1^{(1)} - a_2x_2 - a_3x_3 - \ldots a_mx_m}{\sqrt{1 - R^2}}\right) \tag{6.19}$$

At this point relabel $X^{(1)}$ as Y to signify that it is the predictand, and relabel x_2 as p_1 to show that it is the first predictor, x_3 as p_2, etc., then:

$$Pr(Y < y | p_1, p_2, \ldots, p_k) = \Phi\left(\frac{\overset{''}{y} - a_1\overset{''}{p}_1 - a_2\overset{''}{p}_2 - \ldots a_K\overset{''}{p}_K}{\sqrt{1 - R^2}}\right) \tag{6.20}$$

where $\overset{''}{y}$ is the transnormalized boundary of the predictand category
$\overset{''}{p}_i$ is the ith transnormalized predictor
a_i are the regression coefficients given by Equation (6.7)
R is the multiple correlation coefficient given by Equation (6.13)
K is the number of predictors, K = m - 1

6.3. Regression Probability for One and Two Predictors.

In the case of a single predictor Equation (6.19) becomes, as shown by Gringorten (1972):

$$Pr(Y < y | x) = \Phi\left(\frac{\overset{''}{y} - r\overset{''}{p}}{\sqrt{1 - r^2}}\right) \tag{6.21}$$

where r is the transnormalized category boundary and is also the correlation between the transnormalized predictor $\overset{''}{p}$ and the transnormalized predictand $\overset{''}{y}$.

Gringorten (1974) also derived the two-predictor case from Equation (6.19):

$$\Pr(Y < y | x_1, x_2) = \Phi\left(\frac{\overset{''}{y} - A_1\overset{''}{p}_1 - A_2\overset{''}{p}_2}{\sqrt{1 - A_1 r_1 - A_2 r_2}}\right) \tag{6.22}$$

where

$$A_1 = (r_1 - r_{12}r_2)/(1 - r_{12}^2)$$

$$A_2 = (r_2 - r_{12}r_1)/(1 - r_{12}^2)$$

and r_1 is the correlation between the first predictor and the predictand

r_2 is the correlation between the second predictor and the predictand

r_{12} is the correlation between the predictors

6.4. Probability for a Categorical Predictor.

Equation (6.20) solves for a predictor equal to a specific value (x). To calculate the probability for a categorical predictor, each value of x in the category must be used — each weighted by the probability that it belongs to the category. Defining:

$$\phi(x) = (1/\sqrt{2\pi})\ \mathrm{EXP}\ (-\tfrac{1}{2}x^2) \quad \text{and} \quad g = (y - \rho x)/\sqrt{1 - \rho^2}$$

then this weight is

$$\phi(x)/\Pr(x_1 \le X < x_2)$$

Thus:

$$\Pr(Y < y | x_1 \le X < x_2) = \frac{\Pr(X, Y)}{\Pr(x)} = \frac{1}{\Pr(X)} \int_{x_1}^{x_2} \phi(x)\, \Phi(g)\, dx \tag{6.22a}$$

or since $d\,\Phi(x) = \phi(x)\,dx$:

$$\Pr(Y < y | x_1 \le X < x_2) = \frac{1}{\Pr(X)} \int_{\Phi(x_1)}^{\Phi(x_2)} \Phi(g)\, d\,\Phi(x) \tag{6.23}$$

Notice that Equation (6.23) is the mean conditional probability for the various values of x in X.

Equation (6.23) can be integrated numerically using the following scheme:

$$\Pr(Y < y | x_1 \le X < x_2) = \frac{1}{N} \sum_{i=1}^{N} \Phi(g_i) \tag{6.24}$$

$$g_i = \frac{y - \rho\Phi^{-1}\,[\Phi(x_i) + (i - \tfrac{1}{2})\,\Delta\Phi(x)]}{\sqrt{1 - r^2}}, \qquad N = \{32\,[\Phi(x_2) - \Phi(x_1)]\}$$

Brackets ({}) are used to show that N is rounded to nearest integer, so that

$$N\Delta\Phi(x) = \Phi(x_2) - \Phi(x_1) \ , \quad \text{or} \quad \Delta\Phi(x) \approx 1/32 \ , \quad \text{i.e., } 3\%$$

Equation (6.24) is an open ended method. Thus, values at end points need not be calculated.

Error is proportional to the second derivative, $d^2\Phi(g)/d\Phi(x)^2$. Error is small, usually much less than 1% for $|\rho| < 0.8$ or for $\Phi(x)$ not close 0 or 1. Error as high as 7% has been observed with $\rho = 0.92$ and $\Phi(x_1) = 0$, $\Phi(x_2) = 0.02$.

If one predictor is categorical (X_c) and there are several other predictors (x_1, x_2 ...) that are exact values, Equation (6.23) is modified thereby:

$$\Pr(Y < y | x_{c1} \le X_c < x_{c2}, x_2, x_3 \ldots) = \frac{1}{P(X_c | x_2, x_3 \ldots)} \int_{\Phi(x_{1c})}^{\Phi(x_{2c})} \Phi(g_c) \, d\Phi(X_c) \quad (6.25)$$

where $$g_c = \frac{y - A_c X_c - A_2 x_2 - A_3 x_3 \ldots}{\sqrt{1 - R^2}}$$

A_c, A_1, etc., are regression coefficients given by Equation (6.7) and R is the multiple correlation coefficient given by Equation (6.15).

Equation (6.25) can be integrated numerically using (6.24) but redefining g_i:

$$g_i = \frac{y - A_c P^{-1} P(x_{1c}) + (i - \tfrac{1}{2}) \Delta P(x_c) - A_2 x_2 - A_3 x_3 \ldots}{\sqrt{1 - R^2}} \quad (6.26)$$

If two predictors are categorical, a double integral (numerically, a double summation) is required to calculate the conditional probability. However, if the simple correlation between the categorical predictor and the predictand is not too large, i.e., less than 0.6, the equivalent normal deviate of the median of the category can be used as an approximation.

6.5. Multiple Predictand Categories.

One of the strengths of the TRP model is the use of calculating conditional probabilities for multiple categories. The mean predictor

$$(m, \ m = a_1 \overset{''}{x}_1 + a_2 \overset{''}{x}_2 + \ldots a_p \overset{''}{x}_p)$$

and the standard deviation

$$(s, \ s = \sqrt{1 - R^2})$$

need only be calculated once for all predictand categories. The conditional probability of the kth predictand category is given by:

$$\Pr(y_{K-1} < Y < y_K | x_1, x_2 \ldots x_p) = \Phi\left(\frac{\overset{''}{y}_K - m}{s}\right) - \Phi\left(\frac{\overset{''}{y}_{K-1} - m}{s}\right) \quad (6.27)$$

where $\overset{''}{y}_K$ is the upper transnormalized bound of the category and $\overset{''}{y}_{K-1}$ is the lower transnormalized bound. For K = 1, Equation (6.27) becomes:

$$\Pr(Y < y_K) = \Phi\left(\frac{\overset{''}{y}_K - m}{s}\right) \quad (6.28)$$

For the uppermost category, Equation (6.27) becomes:

$$Pr(y_{K-1} < Y) = 1 - \Phi\left(\frac{\overset{''}{y}_{K-1} - m}{S}\right) \qquad (6.29$$

6.6. Changing Predictand Categories.

If the predictand categories are changed, it is not necessary to redevelop the regression coefficients; only the new transnormalized predictand boundaries are required. This capability gives the TRP model great flexibility since the method can be developed for a specified category, e.g., rainfall greater than 1/2 inch, and then operationally used for another category, e.g., rainfall greater than 1 inch. Furthermore, this capability allows probabilities to be calculated for any category upon request in a real-time query response mode. Another result of this capability is that probabilities can be forecast in an objective manner for categories that have never occurred — yet.

Chapter 7

EXAMPLES AND APPLICATIONS

Two detailed examples that illustrate the TRP model are contained in this chapter. These examples show the diverse ways the TRP model can be developed and are not intended for operational use. Operational use of the TRP model has been described by Boehm (1973) and Young (1975).

7.1. Overall Development and Verification Procedures.

In both of these examples, data were divided into a dependent data set and an independent data set. Considerable care was taken to keep the independent data from influencing in any way the development of the TRP coefficients. Multiple discriminant analysis (MDA) coefficients were also calculated with the dependent data using the BMD05M computer program (Dixon, 1973).

Both TRP and MDA procedures were verified using a variety of skill scores and diagnostic parameters, but most reliance was placed on the Brier score (Brier, 1950) and per cent correct forecasts using the most probable forecast. The Brier score (PS) is calculated:

$$PS = \frac{1}{K} \sum_{i=1}^{N} \sum_{j=1}^{K} (P_{ij} - d_{ij})^2 \qquad (7.1)$$

where N is the number of categories

K is the number of forecasts

P_{ij} is the probability for category i for the jth forecast

$d_{ij} = 1$ if category i was observed for the jth forecast, otherwise $d_{ij} = 0$

A perfect score is zero; the worst posible score is 2.

7.2. Probability of Frozen Precipitation.

Walker (1974) collected 1000-500 mb thickness and the corresponding type of precipitation for 13 European locations for November and December 1973. For this example, the data were randomly divided into two sets: one set of six locations for the dependent data set (195 cases) and one set of seven locations for the independent data set (272 cases). Precipitation was divided into two categories: category 1 for rain; and category 2 for frozen precipitation consisting of snow, freezing rain, mixed snow and rain, and ice pellets. Days with no precipitation were excluded from the original study so that the forecast is the probability of frozen precipitation given that precipitation occurs. The forecast method was developed three ways (two direct predictors, modeled marginal predictor, and climatology by station), to illustrate the concept of a marginal predictor. Details of the three ways are given in the following sections.

7.2.1. Frozen Precipitation with Two Direct Predictors.

Thickness and station elevation were transnormalized separately and used as direct predictors. The forecast equation was of the form:

$$\ddot{c} = \frac{\ddot{y} - A_1 \ddot{E} - A_2 \ddot{T}}{\sqrt{1 - R^2}} \qquad (7.2)$$

where $\ddot{c}$ is the equivalent normal deviate that specifies the conditional probability of rain, $\Phi(\ddot{c})$, or frozen precipitation, $\Phi(-\ddot{c})$. Note that $\Phi(\ddot{c}) + \Phi(-\ddot{c}) = 1$ because of the symmetry of the normal distribution, $\ddot{y} = 0.4677$; y is the climatological probability ($y = 0.68$) of rain versus frozen precipitation for the locations used in the development sample:

$$A_1 = 0.066, \quad A_2 = -0.897, \quad R = 0.90$$

$\overset{\#}{E}$ is the END of elevation given by the cumulative distribution of station elevations in the development sample. $\overset{\#}{T}$ is the END of thickness. $\overset{\#}{E}$ and $\overset{\#}{T}$ were determined by the histogram interpolation method (section 4.3.1) which proved to give somewhat unsatisfactory results since the elevation of several of the stations in the independent sample were higher than any of the stations of the dependent sample.

Examples: An elevation of 154m and a thickness of 5340m gave a 10% probability of frozen precipitation while a 43m elevation and 5210m thickness gave a 77% probability of frozen precipitation.

The dependent data verified 86.2% correct and with a Brier score of 0.23. The independent data verified 77.2% correct and with a Brier score of 0.40. These verification results can be compared with other results by referring to Table 2.

Table 2. Frozen Precipitation Verification Results.

	Dependent Data		Independent Data	
Method	Per Cent Correct	Brier Score	Per Cent Correct	Brier Score
Discriminant Analysis	86.7	0.19	86.0	0.20
Two Direct Predictors	86.2	0.23	77.2	0.40
One Direct and One Marginal Predictor	88.7	0.20	82.0	0.33
Unconditional Probability Known	89.9	0.18	87.1	0.19

7.2.2. Frozen Precipitation with One Direct and One Marginal Predictor.

Elevation should not be used as a direct predictor because the elevation values have an artificial distribution, i.e., the values were not determined by nature; man selected the station locations (and consequently the elevations) based on various economic and social factors. This point is critical since it governs which predictors should be direct predictors (distribution determined by nature) and which predictors must be marginal (distribution is artificial since values were determined by man).

Using elevation as a marginal predictor, the forecast equation was of the form:

$$\overset{\#}{0} = \frac{\overset{\#}{y} - A_1 \overset{\#}{T}}{\sqrt{1 - R^2}} \tag{7.3}$$

where $\overset{\#}{y}$ is now a function of elevation (E),

$$\overset{\#}{y} = - 0.000487E + 0.5645$$

$A_1 = R = 0.93$

$\overset{\#}{T}$ = END of thickness (transnormalized using direct method, Equation (4.21))

$$\overset{\#}{T} = - 20498. + 116.4T - 0.2205T^2 + 1.393T^3$$

It was determined that the distribution of thickness was not significantly affected by station elevation in the dependent sample although this would not be the case if the stations were farther apart. The third-order polynomial gave the best fit since the thickness distribution was found to be skewed. The higher accuracy

of the direct method over the histogram method used in the previous section accounts for the increase in correlation from 0.90 to 0.93.

Examples: An elevation of 154m and a thickness of 5340m gave an 8% probability of frozen precipitation while a 43m and 5210m thickness gave an 86% probability of frozen precipitation.

The dependent data verified 88.7% correct with a Brier score of 0.20. The independent data verified 82% correct with a Brier score of 0.33 (see Table 2).

7.2.3. Frozen Precipitation with Unconditional Probability Known for Each Station.

Examination of $\bar{y}$ as a function of elevation showed considerable deviation at individual stations. This suggests that various local effects play an important role in the probability of frozen precipitation. Thus, each station's unconditional (climatological) probability of frozen precipitation was used to determine a $\bar{y}$ for that station in Equation (7.3) with the results that the dependent data verified 89.9% correct with a Brier score of 0.18, and the independent data verified 87.1% correct with a Brier score of 0.19. Since the independent data were used to get the unconditional probability of frozen precipitation, the independent results are not entirely independent.

7.2.4. Probability of Frozen Precipitation with Discriminant Analysis.

The BMD05M (Dixon, 1973) computer program was used to obtain discriminant analysis coefficients for the probability of frozen precipitation using thickness and elevation as predictors. The forecast equation is:

$$P = \frac{EXP\ (A_{11} + A_{12}\ E + A_{13}\ T)}{EXP\ (A_{11} + A_{12}\ E + A_{13}\ T) + EXP\ (A_{21} + A_{22}\ E + A_{23}\ T)} \qquad (7.4)$$

where P is the probability of frozen precipitation

E is elevation

T is thickness

and

$A_{11} = -3032.$ $A_{21} = -3221.$

$A_{12} = -0.082$ $A_{22} = -0.087$

$A_{13} = 11.64$ $A_{23} = 12.00$

Examples: An elevation of 154m and a thickness of 5340m gave a 19% probability of frozen precipitation and an elevation of 43m and a 5210m thickness gave a 94% probability of frozen precipitation.

Note that Equation (7.4) cannot be used directly because of numerical overflow. Instead, a value is subtracted from each exponential argument which is equivalent to dividing the numerator and denominator by a constant.

The dependent data verified 86.7% correct with a Brier score of 0.19. The independent data verified 86.0% correct with a Brier score of 0.20.

7.2.5. Frozen Precipitation Conclusions.

The important result of this example is that it shows the importance of local effects in a statistical forecast procedure and the ease with which these effects are taken into account by the TRP model. The small differences between TRP and discriminant analysis may be due to the high climatological probability (39%) of frozen precipitation.

7.3. Probability of Ceiling Categories by Model Output Statistics.

Data for this example were supplied by the Technique Development Laboratory of the National Weather Service. The predictand was the 1800Z ceiling as specified by the five categories with lower limits at 0, 200, 500, 1000, and 2000 feet for Cleveland, Ohio, for December and January. The predictors came from the 0000Z run of the NMC primitive equation (PE) and trajectory models, and the 0600Z ceiling observation. Predictors derived from a numerical weather predictor model are termed Model Output Statistic (MOS) predictors. The MOS predictors were:

Predictor	Model	Valid Time	Predictor
1	PE	1800Z	Precipitable Water
2	PE	1800Z	Boundary Layer Divergence
3	PE	1800Z	Relative Humidity 400-1000 mb
4	PE	1800Z	Precipitable Water (Smoothed Over 9 Points)
5	Trajectory	2400Z	Surface Relative Humidity
6	Trajectory	2400Z	850-mb Relative Humidity
7	PE	2400Z	Relative Humidity 400-1000 mb

The cosine of the day of the year was also included (predictor 8) along with the 0600Z ceiling observation (predictor 9). These predictors were selected at the Technique Development Laboratory by stepwise regression using the winter (October through March) seasons: 1969-1970, 1970-1971, 1971-1972, and 1972-1973. December 1970 and January 1971 (54 cases) were used for independent data and was withheld from the development data base. The Regression Estimate of Event Probability (REEP) development set consisted of all the development data base. Only the December and January development data (162 cases) were used for MDA and TRP.

7.3.1. TRP Ceiling Forecast by MOS.

Correlation between all pairs of the nine predictors was calculated by the product moment formula Equation (5.2) after each variable had been transnormalized by histogram interpolation, Equation (4.5). The predictor correlation matrix is given in Table 3.

Table 3. TRP Predictor Correlation Matrix.

Predictor	1	2	3	4	5	6	7	8	9
1	1.000	-0.581	0.632	0.953	-0.035	0.425	0.623	-0.263	-0.035
2	-0.581	1.00	-0.690	-0.475	-0.044	-0.667	-0.649	-0.043	-0.053
3	0.632	-0.690	1.000	0.527	0.169	0.749	0.902	0.157	-0.033
4	0.953	-0.475	0.527	1.000	-0.059	0.354	0.519	-0.325	-0.071
5	-0.035	-0.044	0.169	-0.059	1.000	0.326	0.015	0.198	-0.088
6	0.425	-0.667	0.749	0.354	0.326	1.000	0.614	0.159	0.020
7	0.623	-0.649	0.902	0.519	0.015	0.614	1.000	0.193	-0.021
8	-0.263	-0.043	0.157	-0.325	0.198	0.159	0.193	1.000	-0.014
9	-0.035	-0.053	-0.033	-0.071	-0.088	0.020	-0.021	-0.014	1.000

The product moment formula, Equation (5.2), also was used to arrive at the correlation between each predictor and the predictand. However, correlations appeared to be unusually low. In particular, the correlation between the 0600Z and 1800Z ceiling observations was -0.13. McCabe (1968) found the autocorrelation of transnormalized ceiling to be 0.45 at 12 hours. The value of -0.13 was found to be

due to large numbers of clear and high ceilings at Cleveland in December and January and the observing bias which often reports high ceilings, e.g., 25,000 feet, as clear at 0600Z (midnight local time). The tetrachoric correlation formula, Equation (5.22), with the dichotomy division at 3000 feet overcame this problem and gave a correlation of 0.479 which is quite close to McCabe's result of 0.45.

Thus, the correlations were recalculated using the tetrachoric formula when both of the variables were ceilings and the biserial formula, Equation (5.7), whenever one of the variables was a ceiling. The predictor intercorrelation matrix was unchanged except for the correlations using predictor 9 — the 0600Z ceiling observation. The predictor 9 intercorrelations became (1) -0.007, (2) -0.003, (3) -0.048, (4) -0.003, (5) 0.157, (6) -0.077, (7) -0.017, (8) -0.038, and (9) 1.000. The correlation between each predictor and the predictand was calculated to be (1) 0.084, (2) -0.284, (3) 0.122, (4) 0.076, (5) -0.028, (6) 0.082, (7) 0.063, (8) 0.101, and (9) 0.479.

Equations (6.27), (6.28), and (6.29) were used for the forecast equation:

$$\overset{''}{C}_i \approx \frac{\overset{''}{Y}_i - M}{\sqrt{1 - R^2}} \tag{7.5}$$

where M is the "mean predictor":

$$M = A_1\overset{''}{P}_1 + A_2\overset{''}{P}_2 + \ldots + A_9\overset{''}{P}_9$$

$A_1 = -0.257$ $A_4 = 0.264$ $A_7 = -0.624$
$A_2 = -0.433$ $A_5 = -0.201$ $A_8 = 0.218$
$A_3 = 0.539$ $A_6 = -0.139$ $A_9 = 0.522$
$R = 0.644$

$\overset{''}{y}_i$ is the END of the ith ceiling category

$\overset{''}{y}_1 = -2.6$ (0.4%), $\overset{''}{y}_2 = -1.97$ (2.5%), $\overset{''}{y}_3 = -1.32$ (9.3%), $\overset{''}{y}_4 = -0.97$ (16.7%)

Example: Predictor values, $P_1 = 22.9$, $P_2 = -26.7$, $P_3 = 99.0$, $P_4 = 20.8$, $P_5 = 82.2$, $P_6 = 86.9$, $P_7 = 88.5$, $P_8 = 0.988$, $P_9 = 300$, gave probabilities of: (1) 0.0007, (2) 0.05, (3) 0.16, (4) 0.15, and (5) 0.63.

Verification results were 83.3% correct and a Brier score of 0.27 on the dependent data and 92.6% correct and a Brier score of 0.14 on the independent data. Results can be compared in Table 4.

Table 4. Ceiling Forecasts Verification Results.

	Dependent Data		Independent Data	
Method	Per Cent Correct	Brier Score	Per Cent Correct	Brier Score
TRP	83.3	0.27	92.6	0.14
REEP	NOT AVAILABLE		75.9	0.35
MDA	67.3	0.50	70.4	0.47

Some of the original predictors were dichotomized to give binary (equals 0 or 1) variables. Some of the variables were used twice with different dichotomies giving a total of 12 new predictors:

	PREDICTOR	TYPE
1.	1000-mb PE relative humidity (1800Z)	Continuous
2.	Ceiling observation (0600Z)	= 1 below 2000
3.	Trajectory relative humidity (2400Z)	Continuous
4.	PE precipitable water (1800Z)	= 1 below 9
5.	Ceiling observation (0600Z)	= 1 below 200
6.	PE relative humidity (2400Z)	= 1 above 95
7.	Relative vorticity (1800Z)	= 1 above -6
8.	PE relative humidity (2400Z)	= 1 above 90
9.	850-mb trajectory relative humidity (2400Z)	= 1 above 85
10.	PE precipitable water (1800Z)	= 1 above 6
11.	Station elevation	= 1 above 1000
12.	Cosine of day of the year	Continuous

7.3.2. REEP Ceiling Forecast by MOS.

Regression Estimate of Event Probabilities (REEP) is one of the main methods used by the National Weather Service to make automated forecasts (Glahn, 1975). The REEP coefficients were calculated and the method verified by the Techniques Development Laboratory. The procedure is included here so that the comparison can be made with TRP results. The REEP forecast equation is of the form:

$$Pr(i) = A_{i0} + A_{i1}p_1 + A_{i2}p_2 + \ldots + A_{i12}p_{12} \qquad (7.6)$$

where Pr(i) is the probability of category i and $A_{ij} \times 100$ is given in Table 5. Verification of the REEP method for the independent data was 75.9% correct with a Brier score of 0.35. Verification for the dependent data was not available.

Table 5. REEP Coefficient Matrix.

Predictor	i = 1	2	3	4	5
j = 0	5.2	18.20	5.07	-33.15	1.03
1	-0.0004	-0.07	0.07	0.18	-0.24
2	0.42	3.98	5.28	12.51	-22.09
3	0.01	0.07	0.18	0.30	-0.55
4	-0.75	-3.95	-5.40	1.22	8.80
5	8.9	9.24	3.69	-12.87	-8.48
6	-1.6	-10.40	4.16	2.3	5.45
7	0.31	1.58	3.64	4.08	-9.63
8	0.27	-1.24	-11.01	1.55	10.43
9	-3.34	-2.34	-3.14	5.33	3.49
10	0.25	-0.30	-2.36	-7.18	9.58
11	-0.82	-5.52	-3.71	3.94	6.12
12	0.14	2.75	3.39	4.45	-10.73

7.3.3. MDA Ceiling Forecast by MOS.

Multiple Discriminant Analysis (MDA) was also used to compare the verification results of the TRP model. However, the lower two categories were grouped together since the lowest category did not occur in the development data base and MDA requires some development data in each category. MDA coefficients were calculated using the BMD05M (Dixon, 1973) computer program. Predictors are the same as in section 7.3.1. The following MDA forecast equation was used:

$$Pr(i) = \frac{EXP\ (A_{i0} + A_{i1}p_1 + A_{i2}p_2 + \ldots + A_{i9}p_9)}{\sum_{j=1}^{9} EXP\ (A_{j0} + A_{j1}p_1 + A_{j2}p_2 + \ldots + A_{j9}p_9)} \qquad (7.7)$$

where $Pr(i)$ is the probability of the ith category

p_j is the value of the jth predictor

and

A_{ij} is as given in Table 6.

Table 6. MDA Coefficient Matrix.

Predictor	i = 1	2	3	4
j = 0	-263.0	-255.0	-249.0	-262.0
1	0.410	0.609	-0.030	0.617
2	0.512	0.452	0.426	0.442
3	0.022	-0.028	0.025	0.022
4	-0.115	-0.358	0.332	-0.465
5	0.034	0.090	0.079	0.074
6	0.0003	0.021	-0.002	0.00001
7	0.072	0.114	0.085	0.098
8	527.0	517.0	509.0	522.0
9	0.014	0.011	0.013	0.016

The coefficient of predictor 8 (cosine of the day of the year) is much larger than the other coefficients because of the small variation in the cosine for December and January.

Because the lowest category did not occur and was given very small probability forecasts by REEP and TRP, it is possible to compare the four category MDA per cent correct with the other five category results. The Brier score, however, needed to be modified for five categories: If the lowest category does not occur and would have near zero probability, then the equivalent five category Brier score is equal to the four category score (set r = d = 0 for the lowest category in Equation (7.1) for proof). A run of the TRP model with four categories verified this result. Thus, the MDA verification results were 67.3% correct with a Brier score of 0.50 for the dependent data and 70.4% correct and a Brier score of 0.47 for the independent data (see Table 4).

7.3.4. Conclusions for MOS Ceiling Forecasts.

The independent data shows improvement of TRP over REEP and MDA; however, the sample size was relatively small and further tests using larger samples are needed before forming definitive conclusions. Part of the improvement may be due to the skewed distribution of ceilings, the number of categories, and relative rarity of the lower four categories. Transnormalization corrects for the skewness and the

rarity of the lower categories. The number of categories is not important in the TRP model since predictand categories are not used in the development of the coefficients. The effect of many categories is exemplified by the number of parameters used in TRP forecast Equation (7.5), in the REEP forecast Equation (7.6), and the MDA forecast Equation (7.7).

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Transnormalized regression probability (TRP) shows considerable potential as an objective probability forecast method. It combines climatology transnormalization with the flexibility of various correlation formulas and the multivariate normal theorems on regression probability. TRP is particularly useful with highly skewed variables such as ceiling or wind speed, either as predictors or as the predictand. It can also be used with a large number of changeable predictand categories. In contrast, REEP and MDA require many coefficients for multiple categories. In the applications that have been tried to date, TRP verified better on independent data than did REEP or MDA.

Transnormalization allows for a systematic approach to diurnal and seasonal effects. Because local effects are to a large extent also taken into account, TRP can be used for several locations using the same coefficients.

TRP makes a distinction between the predictors that are determined by nature (direct predictors) and the ones that are artificially distributed (marginal predictors). This distinction makes TRP particularly well suited for a passive observation science such as meteorology versus a designed experiment science such as chemistry.

A limited number of operational applications of the TRP model have been made. Much more research is needed to determine what conditions produce the best results. The transnormalized correlation matrix lends itself to stepwise regression, although this technique has not yet been tried. Furthermore, since the transnormalized variables all have means of zero and unit variance, the magnitude of a TRP coefficient is related to the importance of a given predictor.

Hermite polynomials of the second kind (Abramowitz and Stegun, 1964) appear to offer a method of increasing the number of predictors to handle types of nonlinearity that straightforward transnormalization cannot handle. In addition, these othogonal polynomials permit interaction between predictors because they incorporate a weighting function that is equal to the normal density function.

Further work is needed to determine which transnormalization methods are best suited for specific meteorological variables.

Preceding Page BLANK - NOT FILMED

REFERENCE AND BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A., 1964: Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series 55, Washington, D.C.

Air Weather Service, 1948: "Short Range and Extended Forecasting by Statistical Methods," Technical Report 105-38, Hq Air Weather Service, Washington, D.C.

Anderson, T. W., 1958: An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York.

Attanasi, E. D., et al., 1973: "Forecasting Work Conditions for Road Construction Activities: An Application of Alternative Probability Models," Monthly Weather Rev., Vol. 101, No. 3, pp. 223-230.

Bartlet, M. A., 1947: "The Use of Transformations," Biometrics, Vol. 3, pp. 39-52.

Barton, D. E. and Dennis, K. E., 1952: The Conditions Under Which the Gram-Charlier and Edgeworth Curves are Positive Definite and Unimodel," Biometrika, 39, pp. 425-427.

Berkson, J., 1955: "Estimate of the Integrated Normal Curve by Minimum Normit Chi-Square With Particular Reference to Bio-Assay," J. Amer. Statist. Assoc., Vol. 50, No. 269, pp. 529-549.

Boehm, A. R., 1973: "Evaluation of Modelled Conditional Probability in Operational Forecasting," Preprints, 3d Conf. on Probability and Statistics in Atmos. Sci., Boulder, CO., American Meteorological Society, pp. 259-262.

Bouver, H., 1973: "Curve Fitting by the Method of Moments," Themis Technical Report Number 29, Technical Report Number 102, University of Georgia (AD-780568).

Box, G. E. P. and Cox, D. R., 1964: "An Analysis of Transformations," J. R. Statist., Soc. B., 26, pp. 211-252.

Brelsford, W. and Jones, R. H., 1967: "Estimating Probabilities," Monthly Weather Rev., Vol. 95, No. 8, pp. 570-576.

Brier, G. W., 1950: "Verification of Forecasts Expressed in Terms of Probability," Monthly Weather Rev., Vol. 78, No. 1, pp. 1-3.

Brooks, C. E. P. amd Carruthers, N., 1953: Handbook of Statistical Methods in Meteorology, Her Majesty's Stationery Office, London.

Bryan, J. G. and Enger, I., 1967: "Use of Probability Forecasts to Maximize Various Skill Scores," J. Appl. Meteorol., Vol. 6, No. 5, pp. 762-769.

Burr, I. W., 1967: "A Useful Approximation to the Normal Distribution Function with an Application to Simulation," Technometrics, 9, pp. 647-751.

Castellan, N. J., 1966: "On the Estimation of the Tetrachoric Correlation Coeffi-cient," Psychometrica, 31, pp. 67-73.

Chacko, G. K., 1971: Applied Statistics in Decision-Making, American Elsevier Publishing Company, Inc., New York, 491 p.

Chisholm, D. A. and Kruse, H., 1974: "The Variability of Visibility in the Hanscom Mesonetwork: A Preliminary Assessment," Environmental Research Papers No. 479, AFCRL-TR-74-0265, AF Cambridge Research Laboratories.

Cornish, E. A. and Fisher, R. A., 1937: "Moments and Cumulants in the Specification of Distribution," Contributions to Mathematical Statistics, reprinted from Extrait de la Revue de l'Institute International de Statistique, Vol. 4, pp.1-14.

Cramer, H., 1946: "Mathematical Methods of Statistics," Princeton Mathematical Series, Princeton University Press, Princeton, NJ.

Dickens, C. W., 1958: "Switching Theory in Objective Forecasting," Masters Thesis, The Pennsylvania State University, University Park, PA.

Dixon, W. J., 1973: BMDP, Biomedical Computer Programs, University of California Press, Berkeley, CA.

Draper, N. R. and Smith, H., 1966: Applied Regression Analysis, John Wiley and Sons, New York.

Edgeworth, F. Y., 1898: "On the Representation of Statistics by Mathematical Formulae," J. Roy. Statist. Soc., 61, p. 670.

Elderton, W. P., 1953: Frequency Curves and Correlations, 4th Edition, Harrem Press, Washington, D.C.

Everitt, P. F., 1909-1910: "Tables of the Tetrachoric Functions for Fourfold Correlation Tables," Biometrika, Vol. 7, pp. 437-451.

Feller, W., 1968: Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition, Wiley, New York.

Feller, E., 1971: Introduction to Probability Theory and Its Applications, Vol 2, 2nd Edition, Wiley, New York.

Fisher, R. A., 1950: "Moments and Cumulants in the Specification of Distributions," Contributions to Mathematical Statistics Paper 30 (with E. A. Cornish), John Wiley and Sons, New York.

Friend, J. H. and Guralnik, D. B., 1959: Webster's New World Dictionary, The World Publishing Co., Cleveland and New York.

Glahn, H. R., 1963: "Decision Theory Concepts Applied to Ceiling Height Prediction," Doctoral Thesis, The Pennsylvania State University, University Park, PA.

Glahn, H. R., 1975: "Making Probability Forecasts by Objective Methods," Aerospace Sciences Review 75-3, prepared for Air Weather Service by USAF Environmental Technical Applications Center, Washington, D.C., pp. 13-23.

Goldberger, A. S., 1964: Econometric Theory, Wiley-Hamilton, New York.

Gringorten, I. I., 1955: "Methods of Objective Weather Forecasting," Advances in Geophysics, Vol. II, H. Landsberg, Ed., Academic Press, Inc., New York.

Gringorten, I. I., 1960: "Extreme Value Statistics in Meteorology — A Method of Application," Air Force Surveys in Geophysics No. 125, Bedford, MA.

Gringorten, I. I., 1965: "A Measure of Skill in Forecasting a Continuous Variable," J. Appl. Meteorol., Vol. 4, No. 1, pp. 47-53.

Gringorten, I. I., 1966: "A Stochastic Model of the Frequency and Duration of Weather Events," J. Appl. Meteorol., Vol. 5, No. 5, pp. 606-624.

Gringorten, I. I., 1968: "Estimating Finite-Time Maxima and Minima of a Stationary Gaussian Ornstein-Uhlenback Process by Monte Carlo Stimulation," J. Amer. Statist. Assoc., Vol. 63, No. 4, pp. 1517-1521.

Gringorten, I. I., 1971: "Modelling Conditional Probability," J. Appl. Meteorol., Vol. 10, No. 8, pp. 646-657.

Gringorten, I. I., 1972: "Conditional Probability for an Exact, Noncategorized Initial Condition," Monthly Weather Rev., Vol. 100, No. 11, pp. 796-798.

Gringorten, I. I., 1974: Private Communication.

Guilford, J. P. and Fruchter, B., 1973: Fundamental Statistics in Psychology and Education, 5th Edition, McGraw-Hill Book Co., Inc., New York.

Hastings, C., 1955: Approximations for Digital Computers, Princeton University Press, Princeton, NJ.

Hogg, R. V. and Craig, A. T., 1965: Introduction to Mathematical Statistics, 2nd Edition, MacMillan, New York.

Johnson, N. L., 1949: "Systems of Frequency Curves Generated by Methods of Translation," Biometrika, Vol. 36, pp. 149-176.

Joiner, B. L. and Rosenblatt, J. R., 1971: "Some Properties of the Range in Samples from Tuckey's Symmetric Lambda Distributions," J. Amer. Statist. Assoc., Vol. 66, pp. 394-399.

Jones, R. H., 1968: "A Nonlinear Model for Estimating Probabilities of k Events," Monthly Weather Rev., Vol. 96, No. 6, pp. 383-384.

Kanal, L. and Chandrasekaran, B., 1971: "On Dimensionability and Sample Size in Statistical Classification," Pattern Recognition, Vol. 3, pp. 225-234.

Kendall, M. G., 1948: Rank Correlation Methods, Griffin, London.

Kendall, M. G., 1949: "Rank and Product Moment Correlation," Biometrika, Vol. 36, pp. 177-193.

Kendall, M. G. and Stuart, A., 1958: The Advanced Theory of Statistics, Hafner Publishing Company, New York.

Martin, D. E., Hull, L. J., and Chin, E., 1973: "Methods of Presenting Climatological Information for Use in Planning and Supporting Weather Sensitive Operation," Technical Report 73-0398, Air Force Cambridge Research Laboratory (AD-766464).

McCabe, J. T., 1968: "Estimating Conditional Probability of Events Having Marked Diurnal Variability," Proc. 1st Statist. Conf., Hartford, CT., American Meteorological Society, pp. 163-172.

Miller, K. S., 1964: Multidimensional Gaussian Distributions, John Wiley and Sons, New York.

Moran, P. A. P., 1948: "Rank Correlation and Product-Moment Correlation," Biometrika, Vol. 35, pp. 203-206.

Murphy, A. H., 1972: "Scalar and Vector Partitions of the Probability Score: Part II, N-State Situation," J. Appl. Meteorol., Vol. 11, No. 8, pp. 1183-1192.

Nilsson, N. J., 1965: Learning Machines: Foundations of Trainable Pattern-Classification Systems, McGraw-Hill Book Company, Inc., New York.

Ord, J. K., 1972: Families of Frequency Distributions, Griffin's Statistical Monographs and Courses Number 30, Hafner Publishing Co., New York.

Panofsky, H. A., and Brier, G. W., 1965: Some Applications of Statistics to Meteorology, The Pennsylvania State University, University Park, PA.

Pearson, K., 1895: "Skew Variation in Homogeneous Material," Philosophical Transactions A, 186, p. 343.

Pearson, K., 1909: "On A New Method of Determining Correlation Between a Measured Character A, and a Character B, of which Only the Percentage of Cases Wherein B Exceeds (or falls short of) A Given Intensity is Recorded for Each of A," Biometrika, Vol. 7, Nos. 1 and 2.

Pearson, K, 1909-1910: "On a New Method of Determining Correlation When One Variable is Given by Alternative and the Other by Multiple Categories," Biometrika, Vol. 7, pp. 248-257.

Pearson, K., 1913a: "On the Probable Error of a Coefficient of Correlation as Found from a Fourfold Table," Biometrika, Vol. 9, p. 22.

Pearson, K., 1913b: "On the Measurement of the Influence of 'Broad Categories' on Correlation," Biometrika, Vol. 9, p. 116.

Pearson, K., 1948: Karl Pearson's Early Statistical Papers, Cambridge University Press.

Peters, C. C. and Van Voorhes, W. R., 1940: Statistical Procedures and Their Mathematical Bases, McGraw-Hill Book Company, Inc., New York.

Plackett, R. L., 1954: "A Reduction Formula for Normal Multivariate Integrals," Biometrika, Vol. 41, p. 351.

Savage, L. J., 1954: The Foundations of Statistics, Wiley, New York.

Schaifer, R., 1959: Probability and Statistics for Business Decisions, McGraw-Hill Book Company, Inc., New York, pp. 441-442.

Siegal, S., 1956: Nonparametric Statistics for the Behavorial Sciences, McGraw-Hill Book Company, Inc., New York.

System/360. Scientific Subroutine Package, Version III, Programmer's Manual. Program Number 306A-03X, 1970.

Thompson, J. C. and Brier, G. W., 1955: "The Economic Utility of Weather Forecasts," Monthly Weather Rev., Vol. 83, No. 11, pp. 249-254.

Touart, C. N., 1973: Synthetic Verification Matrices — 1, Air Force Cambridge Research Laboratories Technical Report 73-0158, Air Force Surveys in Geophysics No. 260 (AD-761485).

Van De Geer, J. P., 1971: Introduction to Multivariate Analysis for the Social Sciences, W. J. Freeman and Co., San Francisco, CA.

Walker, H. C., 1974: "Use of Thickness in Forecasting Rain and Snow in Germany," 2nd Weather Wing Technical Note 74-1, Hq 2nd Weather Wing, APO New York, 09012.

Winkler, R. L. and Murphy, A. H., 1968: "'Good' Probability Assessors," J. Appl. Meteorol., Vol. 7, No. 5, pp. 751-758.

Young, M. J., 1975: "Modeled Peak Gust Conditional Probability Tables," USAFETAC Report 7538, USAF Environmental Technical Applications Center.

Yule, G. U. and Kendall, M. G., 1950: An Introduction to the Theory of Statistics, 14th Edition, Hafner Publishing Co., New York.

Appendix A

PROBABILITY FOR AN UNOBSERVED EVENT

Often it is desirable to have conditional probability estimates for predictor categories that have never been observed. If, in these rare cases, the probability density of a particular frequency being the true frequency is distributed according to the Poisson distribution for zero cases, then:

$$\frac{dPr\ (f|0)}{du} = e^{-u}, \qquad u = mf \tag{A.1}$$

m is the total number of observations, f is the frequency of occurrence for an infinitely long period of record, and $P(f|0)$ is the probability of a particular f being the true frequency given that zero cases have occurred.

Equation (6.22) or Equation (6.23) could now be integrated over all possible values each weighted by the probability of its being the true value:

$$Pr(Y|X) = \int_{x_1}^{\infty} Pr(X_2|0) \left[\frac{1}{Pr(X)} \int_{X_1}^{X_2} \phi(X)\ \Phi(g)\ dx\right] dx \tag{A.2}$$

Instead, a much simpler scheme can be used. Equation (A.1) is integrated to find the median value of u, designated as $\bar{u}$:

$$0.3 = \int_{0}^{\bar{u}} e^{-u}\ du, \qquad \bar{u} = 0.7 \tag{A.3}$$

Since $mf = \bar{u}$, the median frequency $f = 0.7/m$. A frequency higher than the median frequency tends to favor forecasting higher categories, a lower frequency favors lower categories. Thus, using the median frequency is "optimal" in the sense that as more data becomes available, half of the probability forecasts will favor lower categories and half will favor higher categories. In using the median frequency, Equation (6.20) is used directly without any integration.

Preceding Page BLANK - NOT FILMED

Appendix B

TAYLOR SERIES ESTIMATES FOR JOHNSON'S LOGNORMAL AND HYPERBOLIC SINE DISTRIBUTIONS

The Lognormal Distribution

The lognormal distribution is of the Equation (4.25) form:

$$\overset{''}{x} = a + b\,\ln(x + K) \tag{B.1}$$

If K is known, a and b can be easily estimated by the method of transnormalized quantile fitting (Chapter 4) by using a dummy variable, $Q = \ln(x + K)$, and $\overset{''}{x} = a + bQ$. An estimate of K can be obtained by expanding Equation (B.1) in a Taylor series and solving for K in terms of the coefficients of a cubic transnormalized quantile fit:

$$\overset{''}{x} = A + Bx + Cx^2 + Dx^3 \tag{B.2}$$

The Taylor series of Equation (B.1) is:

$$\overset{''}{x} = (a + b\ln K) + (b/K)\,x + (-\tfrac{1}{2}\,b/K^2)\,x^2 + (\tfrac{1}{3}\,b/K^3)\,x^3 \tag{B.3}$$

From Equations (B.2) and (B.3):

$$\frac{C}{D} = \frac{-\frac{1}{2}\,b/K^2}{\frac{1}{3}\,b/K^3} = -\frac{3}{2}\,K \tag{B.4}$$

$$\therefore\ K = -\frac{2}{3}\,C/D \tag{B.5}$$

The Hyperbolic Sine Distribution

Johnson's (1949) hyperbolic sine distribution is of the Equation (4.26) form:

$$\overset{''}{x} = a + b\,\mathrm{Sinh}^{-1}(gx + h) \tag{B.6}$$

Given g and h, find a and b by using a dummy variable, $Q = \mathrm{Sinh}(gx + h)$, and $\overset{''}{x} = a + bQ$. The third-order Taylor series of $\overset{''}{x} = a + b\,\mathrm{Sinh}(Y)$ is:

$$\overset{''}{x} = a + bY - bY^3/6 \tag{B.7}$$

and setting $Y = gx + h$:

$$\overset{''}{x} = a + b(gx + h) - b(g^3x^3 + 3g^2hx^2 + 3gh^2x + h^3)/6 \tag{B.8}$$

or

$$\overset{''}{x} = (a + hb - bh^3/6) + (gb - gh^2b/2)\,x + (-bg^2h/2)\,x^2 + (-bg^3/6)\,x^3 \tag{B.9}$$

From Equations (B.2) and (B.9):

$$\frac{C}{D} = \frac{-bg^2h/2}{-bg^3/6} = 3h/g \tag{B.10}$$

$$\frac{B}{C} = \frac{gb - gh^2b/2}{- bg^2h/2} = \frac{-2}{gh} + \frac{h}{g} \qquad (B.11)$$

Solving in Equation (B.10) and substituting in Equation (B.11):

$$\frac{B}{C} = \frac{-2}{gh} + \frac{C}{3D} \qquad (B.12)$$

Solving gh:

$$gh = \frac{2}{(C/3D) - (B/C)} \qquad (B.13)$$

But from Equation (B.10), $g = 3hD/C$, so that:

$$h(3hD/C) = \frac{2}{(C/3D) - (B/C)}, \quad \text{or} \quad h^2 = \frac{2C^2}{C^2 - 3DB} \qquad (B.14)$$

and

$$h = \sqrt{\frac{2C^2}{C^2 - 3DB}} \qquad (B.15)$$

If the quantity under the radical is negative, h and g will have to change sign to satisfy Equation (B.13). Thus, if:

$$\frac{2C^2}{C^2 - 3DB} < 0, \quad h = -\sqrt{\frac{-2C^2}{C^2 - 3DB}} \qquad (B.16)$$

The above estimates for g, h, and K give approximations that are close enough for most forecast problems in which the error due to other factors is greater than the error introduced by the approximations. However, with highly reliable data, the iterative methods described by Johnson (1949) and Ord (1972) may be worth the extra effort.

Appendix C

CALCULATION OF MOMENTS AND CUMULANTS

If the algebraic form of the frequency distribution is unknown, there are general methods for fitting a distribution. These methods, for the most part, depend first on calculating the moments of the variable. Four moments are usually used as large sampling errors are encountered when using a greater number of moments. Using fewer moments also gives a desirable smoothing to the distribution. Moments are calculated as follows:

Let T = total number of observations

V(I) = value of the Ith observation

Moments about the origin are calculated (summations are from I = 1 to T, A_1 is the first moment, A_2 the second, etc.):

$$A_1 = (1/T) \Sigma V(I)$$

$$A_2 = (1/T) \Sigma V(I)^2$$

$$A_3 = (1/T) \Sigma V(I)^3$$

$$A_4 = (1/T) \Sigma V(I)^4 \qquad (C.1)$$

Raw moments (R) about the mean are:

$$R_1 = 0$$

$$R_2 = A_2 - A_1^2$$

$$R_3 = A_3 - 3A_1A_2 + 2A_1^3$$

$$R_4 = A_4 - 4A_1A_3 - 6A_1^2 A_2 - 3A_1^4 \qquad (C.2)$$

According to Cramer (1946, p. 352), consistent and unbiased estimates of the central moments are:

$$M_1 = 0$$

$$M_2 = R_2T/(T - 1)$$

$$M_3 = R_3T^2/[(T - 1)(T - 2)]$$

$$M_4 = \frac{[R_4(T^2 - 2T - 2) - R_2^2(6T - 9)]\ T}{(T - 1)(T - 2)(T - 3)} \qquad (C.3)$$

Cumulants (K) are also used in describing distributions (Cornish and Fisher, 1937). Cumulants are also called semi-invariants or Fisher's K statistics. Unbiased estimates are calculated:

$$K_1 = A_1, \quad K_2 = M_2, \quad K_3 = M_3$$

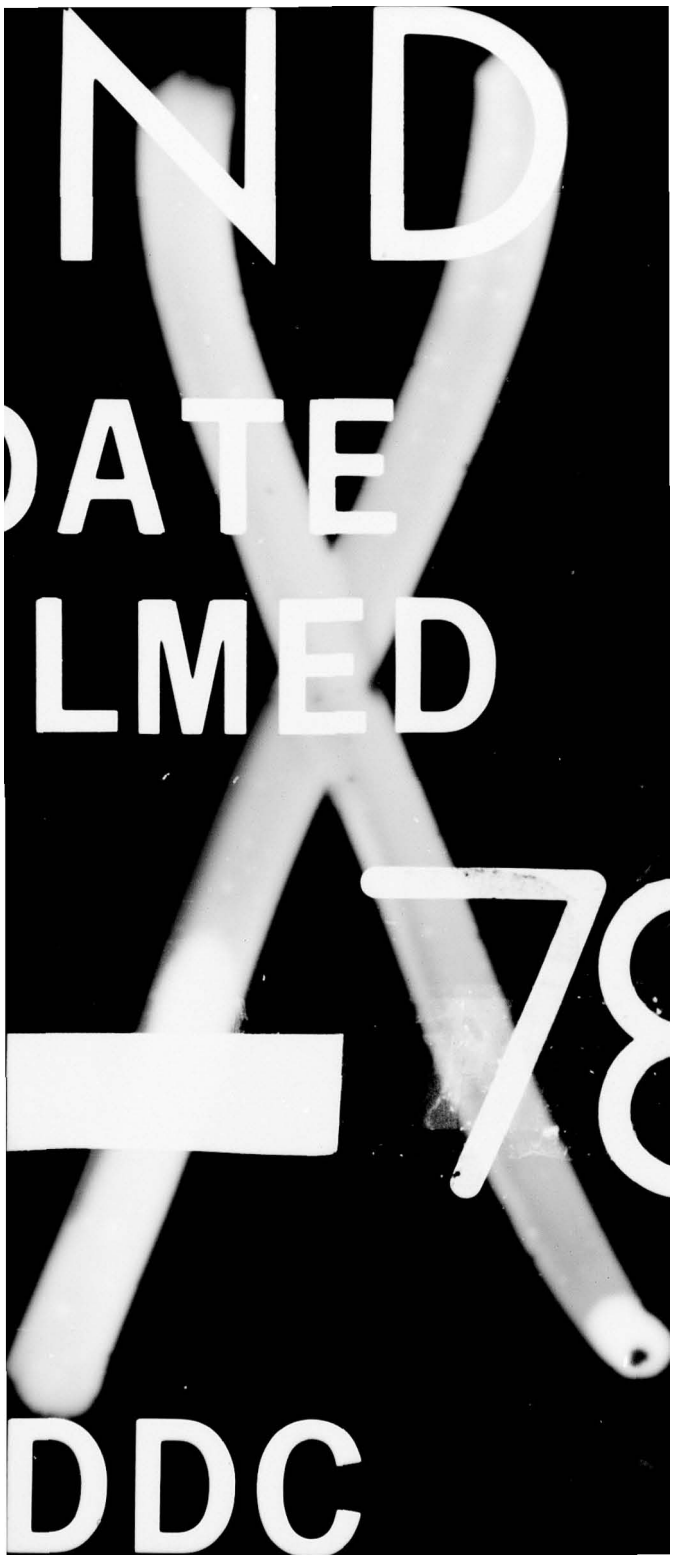
$$K_4 = \frac{R_4(T - 1) - 3(T - 1)\, R_2^2\, T^2}{(T - 1)(T - 2)(T - 3)} \qquad (C.4)$$

Kendall and Stuart (1958) recommend a correction to the cumulants K_2 and K_4, similar to Sheppard's corrections for moments, when the cumulants have been calculated from data that has been lumped into classes of size h. The corrected cumulants (K′) are:

$$K_2' = K_2 - h^2/12$$

$$K_4' = K_4 - h^4/120 \qquad (C.5)$$

The corrected versions should be used when data have been grouped into equal classes.


ND
DATE
LMED
78
DDC

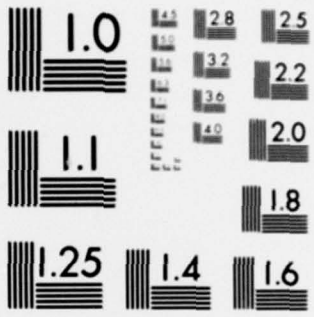
AD-A047 720 TRANSNORMALIZED REGRESSION PROBABILITY(U) AIR WEATHER SERVICE SCOTT AFB ILL A R BOEHM DEC 76 AWS-TR-75-259

2/2

UNCLASSIFIED

F/G 4/2

NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SUPPLEMENTARY

INFORMATION

AD-A047720

ERRATA SHEET

AWS-TR-75-259, "Transonormalized Regression Probability," by Capt Albert R. Boehm, December 1976, 52p (AD-A047720) is changed as follows:

1. Page 11, caption for Figure 6, Line 9 should read:

   occurred less than $y_i = M(X_i)$ because x

2. Page 14, Equation 4.8, third term in parenthesis should be $-sK_3/2$ instead of $-sK_3/6$.

3. Page 14, Equation 4.9, A_1 instead of M_1.

4. Page 14 the line below Equation 4.9

   A_1 and M_2 are the mean and variance (Appendix C),

5. Page 14, Equation 4.11 first term in parenthesis $K_3/6$ instead of $sK_3/6$.

6. Page 15, Equation 4.13, $= f(x-a)/(b-cx-dx^2)$

7. Page 15, Equation 4.14 the numerator should be: $B_1(B_2+3)^2$

8. Page 15, Equation 4.16,

   $\ddot{P} = t-(a+bt)/(1+ct+dt)$...and add: for $0 \leq P < .5$)

9. Page 16, second line after Equation 4.18 should read A1 = ... $+K_4M_2^{-2}/8$ instead of A1 = ...$-K_4M_2^{-2}$.

10. Page 16, Line 5, after Equation 4.18... $s = (x-A_1)/\sqrt{M_2}$.

11. Page 16, Line 6, after Equation 4.18 A_1 and M_2 are the mean and standard deviation.

12. Page 20, after Equation 4.30 add: P. Hicks showed this is mathematically equivalent to:

$$\ddot{P}(e) = 1/(1 + EXP(-1.6\ \ddot{e} - .07\ddot{e}^3)) \quad (4.30b)$$

13. Page 24, line after Equation 5.19 should be:

   $R_i = -xR_{i-1}/(i+1) - (i-1)\ R_{i-2}/[i(i+1)]$.

14. Page 47, Equation A.3 0.5 = instead of 0.3 =.

15. Page 51, Equation C.2 for R_4 third term should be $+6A_1^2A_2$ instead of $-6A_1^2A_2$.

16. Page 52, Equation C.4 numerator should be: $[R_4(T+1) - 3(T-1)R_2^2]T^2$