CHAPTER 6

SOME TRANSFORMATION METHODS IN GOODNESS-OF-FIT*

C. P. Quesenberry
North Carolina State University

6.1 INTRODUCTION

6.1.1 Hypothesis Testing Problems

In this chapter let $X_1, X_2, \dots, X_n$ be identically and independently distributed (i.i.d.) real-valued random variables with a common continuous distribution function (df) F . The classic simple goodness-of-fit problem is to test

$$\begin{aligned} H_0: F &= F_0, \\ H_1: F &\neq F_0, \end{aligned} \tag{6.1}$$

where F_0 is a specific continuous distribution function. The hypothesis testing problem of (6.1) is often not a very useful model in practice. It is more meaningful in many instances to test that the distribution function F has some specified functional form without assuming that the values of all parameters are known.

Let $\theta = (\theta_1, \dots, \theta_p)$ be a vector of real-valued parameters and $\mathcal{F}_0 = \{F_\theta: \theta \in \Omega\}$ be a parametric class of probability distribution functions. Moreover, we assume that Ω is a natural parameter space, i.e., that it contains all points θ for which F_θ is a continuous probability distribution function. Then the composite goodness-of-fit problem is to test

*Work supported in part by National Science Foundation Grant MCS76 - 82652.

$$\begin{aligned} H_0: F \in \mathcal{J}_0, \\ H_1: F \notin \mathcal{J}_0. \end{aligned} \tag{6.2}$$

Let $\tau = (\tau_1, \dots, \tau_{p_1})$ be another vector of real-valued parameters and $\mathcal{J}_1 = \{F_\tau: \tau \in \Omega_1\}$ be another parametric class of continuous distribution functions. The classes $\mathcal{J}_0$ and $\mathcal{J}_1$ are called separate families (Cox (1961)) if the density of an arbitrary member of either class cannot be obtained as the limit of a sequence of densities from the other class. If $\mathcal{J}_0$ and $\mathcal{J}_1$ are separate families, then the separate families testing problem is to test

$$\begin{aligned} H_0: F \in \mathcal{J}_0, \\ H_1: F \in \mathcal{J}_1. \end{aligned} \tag{6.3}$$

In words, this testing problem is to test that the sample is from a member of one class of distributions against the composite alternative hypothesis that it is from a certain alternative class. As a particular example let $\mathcal{J}_0$ be the df's of the scale parameter exponential class of densities given by (6.24) for θ unknown and $\mu = 0$, and let $\mathcal{J}_1$ be the df's of the lognormal class of densities given by (6.35) for $\mu = 0$ and σ^2 unknown. For this choice of $\mathcal{J}_0$ and $\mathcal{J}_1$, (6.3) is to test that the sample is from a scale-parameter exponential distribution against the alternative that it is from a shape-parameter lognormal distribution.

Suppose there are available k independent samples:

$$\begin{array}{l}
X_{11}, X_{12}, \dots, X_{1n_1}; \\
X_{21}, X_{22}, \dots, X_{2n_2}; \\
\vdots \\
X_{k1}, X_{k2}, \dots, X_{kn_k}.
\end{array} \tag{6.4}$$

The several samples goodness-of-fit problem is to test

$$\begin{aligned}
H_0: X_{ij} &\sim F_{\theta_i} \in \mathcal{F}_0; i = 1, \dots, k; j = 1, \dots, n_i; \\
H_1: &\text{Negation of } H_0.
\end{aligned} \tag{6.5}$$

In words, this null hypothesis is that all random variables (rv's) have df's of the same functional form, however, the parameters may change from sample to sample. The testing problem of (6.5) is an important generalization of the classical single sample goodness-of-fit composite null hypothesis testing problem of (6.2). The several samples null hypothesis can also be considered against a corresponding several samples separate families hypothesis testing problem as follows

$$\begin{aligned}
H_0: X_{ij} &\sim F_{\theta_i} \in \mathcal{F}_0; i = 1, \dots, k; j = 1, \dots, n_i; \\
H_1: X_{ij} &\sim F_{\tau_i} \in \mathcal{F}_1; i = 1, \dots, k; j = 1, \dots, n_i.
\end{aligned} \tag{6.6}$$

In the following sections we shall consider some techniques for testing the five types of problems described above. The approach used here may be described as follows for the classical goodness-of-fit problem of display (6.2). The sample $X_1, \dots, X_n$ is transformed to a set of values $U_1, \dots, U_N$, ($N = n - p$, the number of observations minus the number of parameters) in such a way that when H_0 is true, then $U_1, \dots, U_N$ are independently and identically distributed uniform random variables

on the $(0, 1)$ interval -- i.i.d. $U(0, 1)$ rv's. The composite null hypothesis H_0 of (6.2) is then replaced by the surrogate simple null hypothesis that the U 's are i.i.d. $U(0, 1)$ rv's.

The reader who is interested only in the methodology of this approach may wish to go directly to section 6.4, which considers studying the uniformity of the transformed values; and then to section 6.5, where the formulae are given for the transformations for a number of families of distributions. Numerical examples are given in section 6.6 to illustrate the application of this approach to some data sets.

6.1.2 The Transformations Approach

In this section we consider again a single sample $X_1, \dots, X_n$ with parent d.f. F that is assumed under a null hypothesis to be a member of the parametric class $\mathcal{F}_0$ of the last subsection. Consider a set of transformations of the structure

$$\begin{aligned} U_1 &= h_1(X_1, \dots, X_n), \\ U_2 &= h_2(X_1, \dots, X_n), \\ &\vdots \\ U_N &= h_N(X_1, \dots, X_n), \end{aligned} \tag{6.7}$$

where $N \leq n$. Each h_i for $i \in \{1, \dots, N\}$ is a real-valued measurable function. Let P_F denote the probability distribution of $(U_1, \dots, U_N)$ induced from the parent distribution with d.f. F , and recall that a test is similar on H_0 of (6.2) if it has constant probability of rejecting for every $F \in \mathcal{F}_0$. The following are three properties of the U -transformations of (6.7) which we list for consideration.

- (1) There exists a probability distribution Q such that $P_F = Q$ for every $F \in \mathcal{F}_0$, i.e., $(U_1, \dots, U_N)$ has the same distribution for every $F \in \mathcal{F}_0$.

(ii) If $P_F = Q$, then $F \in \mathcal{F}_0$, i.e., $(U_1, \dots, U_N)$ has the characterizing distribution Q only if F is a member of the class $\mathcal{F}_0$.

(iii) There exists a test based on $(U_1, \dots, U_N)$ for testing the null hypothesis H_0 of (6.2) against a particular simple alternative hypothesis that has the same power as the most powerful similar test based on $(X_1, \dots, X_N)$ for this same testing problem.

Condition (i) is very important because it assures that every size α test based on $(U_1, \dots, U_N)$ is also a similar α test for the same null hypothesis.

Condition (ii) is a theoretically interesting property of transformations of the structure of (6.7), however, it should be pointed out that this type of characterization property of transformations has not played an important role in the goodness-of-fit field. The apparent reason why this is so involves the following considerations. The actual test statistic will itself be a real-valued function of the values $(U_1, \dots, U_N)$ of (6.7), i.e., the test statistic is obtained by composition of a real-valued function with the transformations of (6.7). Unless the distribution of the test statistic is also a characterization of the distribution Q in property (ii), then property (ii) for the transformations of (6.7) is of little relevance. In other words, it matters little whether characterization is lost in the first or second step of the transformations. In this context, it should be observed that most of the goodness-of-fit statistics that are important in applied statistics do not characterize a null class of distributions. As a particular example, consider the chi-squared test statistic. Although chi-squared test statistics do not characterize null hypothesis classes of distributions, they have, of course, been and are of great importance in applied statistics. (See Chapter 3)

A number of transformations of the form of (6.7) have been given in the literature for particular parametric families. David and Johnson (1948) considered the probability integral transformation when parameters are replaced by estimates. They showed that the transformed values are dependent, and for many location-scale parameter families that the transformed random variables have distributions that do not depend upon the values of the parameters.

A number of writers have given transformations for particular families of the structure of (6.7) that satisfy condition (i) for transformations. Sarkadi (1960, 1965) gives transformations for the three univariate normal families. Durbin (1961) has proposed a transformation approach that eliminates the nuisance parameters by introducing a further randomization step. Störmer (1964) gives a method for transforming a sample from a $N(\mu, \sigma^2)$ distribution to a sample of $n-2$ values from a $N(0,1)$ distribution. A number of transformations of the structure of (6.7) and satisfying property (i) have been considered in the literature for one and two parameter exponential classes. Two of these transformations are considered by Seshadri, Csörgö and Stephens (1969), and one is shown to have property (ii), also. Csörgö and Seshadri (1970); Csörgö and Seshadri (1971); and Csörgö, Seshadri and Valovsky (1973) have considered this transformations approach; and have given transformations for some particular normal, exponential and gamma families of distributions with properties (i) and (ii). A number of writers have considered the "recursive residuals" for normal regression models, which are of the structure of (6.7) when the variance is known. We shall discuss these in subsection 6.5.7 below.

A general theory for obtaining transformations such that the transformed U's are i.i.d. $U(0,1)$ random variables is given by O'Reilly and Quesenberry (1973), and extended to additional classes by Quesenberry (1975). These authors call the transformations involved conditional probability integral transformations-CPIT's, and we shall in this chapter consider only transformations obtained by this approach. We do not claim that the transformations and resulting tests and other analyses obtained by this approach have advantages over all other transformations for particular classes of distributions. Indeed, in some cases they will be found to give statistics equivalent to those of some other approaches.

6.2 PROBABILITY INTEGRAL TRANSFORMATIONS

6.2.1 Classical Probability Integral Transformations

The well-known probability integral transformation theorem due to R. A. Fisher (1930) can be stated as follows.

Theorem 6.1. If X is a real-valued random variable with continuous df F , then $U = F(X)$ has a uniform distribution on the interval $(0,1)$, i.e., U is a $U(0,1)$ rv.

(Historical aside: Actually, Fisher did not explicitly discuss the transformation in this paper, rather, he used the fact that a continuous distribution function $F(T;\theta)$ of a statistic T is a $U(0,1)$ rv in deriving fiducial limits for a parameter. He also used the result in Theorem 6.2 in "Statistical Methods for Research Workers", Fisher (1932), in his method for combining tests of significance, again, without explicit discussion of the transformation itself.)

Thus if $X_1, \dots, X_n$ are i.i.d. with continuous common df F , then $U_1 = F(X_1), \dots, U_n = F(X_n)$ are i.i.d. $U(0,1)$ random variables. An important generalization of this basic theorem due to Rosenblatt (1952) is given in the next theorem.

If F is a multivariate distribution function we denote by $F(\cdot)$, $F(\cdot|\cdot)$, etc., the usual marginal and conditional df's.

Theorem 6.2. If $(Y_1, \dots, Y_m)$ is a vector of m random variables with absolutely continuous multivariate distribution function F , then the m random variables

$$U_1 = F(Y_1), U_2 = F(Y_2|Y_1), \dots, U_m = F(Y_m|Y_1, \dots, Y_{m-1})$$

are i.i.d. $U(0,1)$ rv's.

As a simple example to illustrate Theorem 6.2, let $m = 2$ and (Y_1, Y_2) have joint density function

$$f(y_1, y_2) = \exp(-y_2), \text{ for } 0 < y_1 < y_2 < +\infty.$$

Then

$$F(y_1) = 1 - \exp(-y_1), \text{ for } y_1 > 0;$$

and

$$F(y_2|y_1) = 1 - \exp(y_1 - y_2), \text{ } 0 < y_1 < y_2 < \infty.$$

Theorem 6.2 says that

$U_1 = F(Y_1) = 1 - \exp(-Y_1)$ and $U_2 = F(Y_2|Y_1) = 1 - \exp(Y_1 - Y_2)$, for $0 < Y_1 < Y_2 < \infty$, are i.i.d. $U(0,1)$ rv's. This result is easily verified directly. Applying the standard transformation of densities gives

$$h(u_1, u_2) = \exp(-y_2) \cdot \exp(y_2) = 1 \text{ for } 0 < u_1 < 1, 0 < u_2 < 1.$$

The generality with which both of the preceding theorems hold should be carefully noted. We shall apply these in particular for cases when F is already explicitly a conditional distribution function.

6.2.2. Conditional Probability Integral Transformations: CPIT's

The model assumptions we make in this chapter are more restrictive than those in O'Reilly and Guesenberry (1973), O-O, but are sufficiently general to cover many important cases. We assume that the parametric class $\mathcal{J}_0$ of distribution functions corresponds to a continuous exponential class (cf. Zacks (1971),

section 2.5), and that T_n is a p -component vector that is a sufficient and complete statistic for $\theta = (\theta_1, \dots, \theta_p)$. Denote by $\bar{F}_n(x_1, \dots, x_n)$ the distribution function of $(X_1, \dots, X_n)$ given the statistic T_n . Then $\bar{F}_n(x_1)$, $\bar{F}_n(x_2|x_1)$, $\dots$, $\bar{F}_n(x_n|x_1, \dots, x_{n-1})$ are the marginal and conditional distribution functions obtained from $\bar{F}_n(x_1, \dots, x_n)$. The following theorem is a direct consequence of Theorem 2.3 of O-Q and can be obtained from Theorem 6.2 above.

Theorem 6.3. The $(n-p)$ random variables

$$\begin{aligned} U_1 &= \bar{F}_n(X_1), \quad U_2 = \bar{F}_n(X_2|X_1), \quad \dots, \\ U_{n-p} &= \bar{F}_n(X_{n-p}|X_1, \dots, X_{n-p-1}) \end{aligned} \quad (6.8)$$

are i.i.d. $U(0,1)$ r.v.'s.

We note that the assumption that $(X_1, \dots, X_n)$ are i.i.d. is not necessary to obtain Theorem 6.3. It is sufficient to require that $(X_1, \dots, X_n)$ have a full-rank absolutely continuous distribution. We give results below in subsection 6.5.3 obtained by applying Theorem 6.3 to the order statistics of a sample from an exponential distribution.

A sequence $(T_n)_{n \geq 1}$ of statistics is said to be doubly transitive if each of the pairs of values (T_n, X_{n+1}) and (T_{n+1}, X_{n+1}) can be computed from the other. For example, if $T_n = \bar{X}_n = (X_1 + \dots + X_n)/n$ then $(\bar{X}_{n+1}, X_{n+1}) = ((n\bar{X}_n + X_{n+1})/(n+1), X_{n+1})$, and $(\bar{X}_n, X_{n+1}) = (((n+1)\bar{X}_{n+1} - X_{n+1})/n, X_{n+1})$; and the sample mean $\bar{X}_n$ is doubly transitive.

The i.i.d. assumptions on $(X_1, \dots, X_n)$ are necessary, in general, for the next theorem, which is obtained from Corollary 2.1 of O-Q. Now, it is often rather difficult to apply Theorem 6.3 directly to obtain explicit formulae for particular parametric families. The next theorem greatly simplifies the task of deriving the actual transformations for many important parametric families.

Theorem 6.4. If $(T_n)_{n \geq 1}$ is doubly transitive, then the $(n-p)$ random variables

$$U_1 = \bar{F}_{p+1}(X_{p+1}), U_2 = \bar{F}_{p+2}(X_{p+2}), \dots, U_{n-p} = \bar{F}_n(X_n) \quad (6.9)$$

are i.i.d. $U(0,1)$ r.v.'s.

6.2.3. Conditional Probability Integral Transformations. Truncation Parameter Families

Some important classes of distributions which are not covered by the transformation theory of the preceding section are the truncation parameter families considered by Quesenberry (1975). For these families we assume that the parent density defined on an interval (a,b) , finite or infinite, is of one of the three forms:

$$f(x; \mu_1, \mu_2, \theta) = c(\mu_1, \mu_2, \theta) h(x, \theta); \quad a < \mu_1 < x < \mu_2 < b; \quad (6.10)$$

$$f_1(x; \mu, \theta) = c_1(\mu, \theta) h_1(x, \theta), \quad a < \mu < x < b; \quad (6.11)$$

$$f_2(x; \mu, \theta) = c_2(\mu, \theta) h_2(x, \theta), \quad a < x < \mu < b. \quad (6.12)$$

Here a and b are known constants; μ , μ_1 and μ_2 are truncation parameters; θ is a p -component parameter vector; and $h(x, \theta)$, $h_1(x, \theta)$, and $h_2(x, \theta)$ are positive, continuous and integrable functions over the intervals (μ_1, μ_2) , (μ, b) and (a, μ) , respectively.

For $X_1, \dots, X_n$ a sample we now set out a particular transformation to another set of values. Let r denote the antirank of $X_{(n)}$, i.e., r must satisfy $X_r = X_{(n)}$; and put

$$W_1 = X_1, \dots, W_{r-1} = X_{r-1}, W_r = X_{r+1}, \dots, W_{n-1} = X_n.$$

Then we shall call $W_1, \dots, W_{n-1}$ the sample with $X_{(n)}$ deleted.

In words: $W_1, \dots, W_{n-1}$ are the sample members that are less than the largest order statistic and subscripted in the same order.

Next, let r denote the antirank of $X_{(1)}$, i.e., r must satisfy $X_r = X_{(1)}$. Define $W_1, \dots, W_{n-1}$ in terms of this X_r in the same manner as above, and these values will be called the sample with $X_{(1)}$ deleted.

Finally, let r_1 and r_2 denote the antiranks of $X_{(1)}$ and $X_{(n)}$, respectively; and put $m_1 = \min\{r_1, r_2\}$, and $m_2 = \max\{r_1, r_2\}$. Put

$$W_1 = X_1, \dots, W_{m_1-1} = X_{m_1-1}, W_{m_1} = X_{m_1+1}, \dots, W_{m_2-2} = X_{m_2-1},$$

$$W_{m_2-1} = X_{m_2+1}, \dots, W_{n-2} = X_n.$$

These values $W_1, \dots, W_{n-2}$ will be called the sample with $X_{(1)}$ and $X_{(n)}$ deleted.

For $X_1, \dots, X_n$ a sample from the density f of (6.10) and $W_1, \dots, W_{n-2}$ the sample with $X_{(1)}$ and $X_{(n)}$ deleted, the next theorem is from Guesenberry (1975).

Theorem 6.5. For fixed $(X_{(1)}, X_{(n)}) = (x_{(1)}, x_{(n)})$, the numbers $W_1, \dots, W_{n-2}$ of the deleted sample are conditionally independent, identically distributed, continuous rv's with common density function

$$g(w, \theta) = \frac{h(w, \theta) I_{(x_{(1)}, x_{(n)})}^{(w)}}{\int_{x_{(1)}}^{x_{(n)}} h(w, \theta) dw} \quad (6.13)$$

where $I_{(x_{(1)}, x_{(n)})}^{(w)} = 1$ if $x_{(1)} < w < x_{(n)}$, and is otherwise zero.

Let T_{n-2} be a complete sufficient statistic [a function of $(W_1, \dots, W_{n-2})$] for θ in the family of $g(w, \theta)$, and let $\bar{G}_{n-2}(w) = P(W_1 \leq w | T_{n-2})$. That is, $\bar{G}_{n-2}$ is the Rao-Blackwell (MVU) estimating df of the df $G(w, \theta)$ corresponding to the density function $g(w, \theta)$.

From Theorem 6.4 above or from Theorem 2 of Guesenberry (1975), the next result follows.

Theorem 6.6. If $(T_j)_{j \geq 1}$ is doubly transitive then the $(n-p-2)$ rv's

$$U_{j-r} = \bar{G}_j(W_j); j = p+1, \dots, n-2; \quad (6.14)$$

are i.i.d. $U(0,1)$ rv's.

Transformations of samples from distributions with densities of the form of f_1 of (6.11) or f_2 of (6.12) can be obtained in a similar fashion. If the sample is from f_1 we fix $X_{(1)} = x_{(1)}$ and let $W_1, \dots, W_{n-1}$ be the sample with $X_{(1)}$ deleted. Then for fixed $X_{(1)} = x_{(1)}$ these values are conditionally independent, identically distributed, continuous rv's with common density function

$$g_1(w) = \frac{h_1(w, \theta) I_{(x_{(1)}, b)}(w)}{\int_{x_{(1)}}^b h_1(w, \theta) dw} \quad (6.15)$$

Finally, if the sample is from f_2 we fix $X_{(n)} = x_{(n)}$ and let $W_1, \dots, W_{n-1}$ be the sample with $X_{(n)}$ deleted. For fixed $X_{(n)} = x_{(n)}$ these values are conditionally independent, identically distributed, continuous rv's with common density function

$$g_2(w) = \frac{h_2(w, \theta) I_{(a, x_{(n)})}(w)}{\int_a^{x_{(n)}} h_2(w, \theta) dw} \quad (6.16)$$

In each of these two cases we apply Theorem 6.4 to the W_j 's to produce $(n-p-1)$ i.i.d. $U(0,1)$ rv's. The transformations for some particular families of truncation parameter distributions will be given in section 6.5.

6.3 SOME PROPERTIES OF CPIT'S

6.3.1 Notation and Terminology

In this section we give a largely descriptive account of some of the more important properties of CPIT's, and for this we shall use some notation and terminology that is not used elsewhere in this chapter.

Denote by $\mathcal{P}_j$ a parametric class of probability measures corresponding to the class $\mathcal{J}_j$ of df's of section 6.1 for $j = 0, 1$. The members of these classes are probability measures on a Borel set I of real numbers and let G denote the Borel subsets of I . Let (I^n, G^n) be the usual product space, and

$$\mathcal{P}_j^n = \{P^n; P^n = P \times \dots \times P, P \in \mathcal{P}_j\}, j = 0, 1.$$

Let $g: I \rightarrow I$ be a strictly increasing 1-to-1 function and $g^n: I^n \rightarrow I^n$ be defined by $g^n(x_1, \dots, x_n) = (gx_1, \dots, gx_n)$. For each function g^n assume there exists a function $\bar{g}: \Omega \rightarrow \Omega$ such that $P_{\bar{g}g^n}(X \in g^n A) = P_g(X \in A)$ for every $A \in G^n$ and $X' = (X_1, \dots, X_n)$.

Let G denote a transformation group on I , G^n the corresponding transformation group on I^n , and $\bar{G}$ the corresponding transformation group on Ω . A transformation group on a space is said to be transitive if its maximal invariant is constant on the space.

The U-values of Theorem 6.4 can be expressed as functions on I^n . The next three theorems are from Quesenberry and Starbuck (1976) (Q-S). In the following $\underline{U}' = (U_1, \dots, U_{n-p})$.

Theorem 6.7. If $\bar{G}$ is a transformation group of strictly increasing functions on I that induces a transitive group $\bar{G}$ on Ω , and conditions for Theorem 6.4 are satisfied, then $\underline{U}$ is equivalent to an invariant statistic, i.e.,

$$\underline{U}(x_1, \dots, x_n) = \underline{U}(gx_1, \dots, gx_n) \text{ a.s. } P^n \forall g \in G.$$

From the distributional result of Theorem 6.3 and Basu (1955, 1960), the next result follows.

Theorem 6.8. For T a complete sufficient statistic for Ω and $\underline{U}$ as given in Theorem 6.3, T and $\underline{U}$ are independent vectors.

The following result from Q-S shows that the U -transformations of section 6.2 are efficient from the power viewpoint.

Theorem 6.9. A most powerful similar test exists for testing $\mathcal{H}_0$ of (6.2) against a simple alternative $F = F_1$, and this test can be expressed as a function of $(U_1, \dots, U_{n-p})$ only.

The import of this result is that all information in the sample about the class $\mathcal{J}_0$ of d 's is also in the U -values $U_1, \dots, U_{n-p}$. This result and Theorem 6.8 shows that the CPIT transformations may be regarded as a technique whereby the information in the sample $(X_1, \dots, X_n)$ can be partitioned into two vectors $(\underline{U}; T)$, and the vector T contains all information about the parameters $(\theta_1, \dots, \theta_p)$, the vector $\underline{U}$ contains all information about the class $\mathcal{J}_0$, and T and $\underline{U}$ are independent. Thus the values $(U_1, \dots, U_{n-p})$ may be used to make inferences about the class of distributions (is it normal? exponential?), the statistic(s) T may be used to make inferences within the class about parameter values (estimate the mean?), and the independence of $\underline{U}$ and T can be exploited to assess overall error rates.

6.3.2 Sequential Nature of CPIT's

From the nature of the transformations in (6.8) and (6.9) it is apparent that, in general, the vector $\underline{U}$ of transformed values is not a symmetric function of the X sample. That is, the vector $\underline{U}$ is not invariant under permutations of the observations. For those cases when the transformations are not permutation invariant, this property requires consideration on a number of points.

One point of concern when the u 's are not permutationally invariant is that a goodness-of-fit test or other analysis may lead to a conclusion that depends upon the presumably irrelevant indexing of the X 's. If two randomly selected orderings are used to compute $\underline{U}$ and then a particular goodness-of-fit test (such as the Neyman smooth test discussed below) is computed on the two $\underline{U}$ vectors, the statistics are identically distributed and dependent, but nonidentical. If we consider these two test statistics, then the situation is similar to that when more than one competing test statistic is computed for the same testing problem. In particular, it is common practice today to compute a number of the goodness-of-fit test statistics, discussed elsewhere in this book, for each sample. A quantity of relevance here is the probability that two tests will agree in their conclusions. Quesenberry and Dietz (1983) considered this probability for Neyman smooth tests made on the U 's from random permutations of a sample. They gave empirical evidence that these agreement probabilities are very high in many cases of interest and are bounded below by the value $2/3$ for all cases considered.

It is possible to obtain CPIT transforms that are invariant under permutations of the original sample. To obtain these transforms recall that in order to apply the approach of Rosenblatt it was necessary to assume only that the rv's $(X_1, \dots, X_n)$ had a full-rank continuous distribution. Thus to obtain permutation invariant transformations we can first transform an i.i.d. sample $(X_1, \dots, X_n)$ to its order statistics, say $(X_{(1)}, \dots, X_{(n)})$, and then make CPIT transformations using Theorem 6.3 on these order statistics. However, for this case Theorem 6.4 is not applicable and it is a difficult task to find the transforms in practice. They have been found only for two rather simple cases of exponential and uniform distributions, that will be considered further below.

The discussion thus far has assumed that the sample is entirely symmetric of given size n . In practice, we often have data that can be naturally ordered by some variable. Perhaps the most common ordering variable is time, and when this is so the observations themselves may arrive sequentially in time. When the data are ordered by time, or some other variable, then the CPIT transformations approach has an especially strong appeal for it will allow the analyst to design tests and other analysis techniques specifically to

detect misspecifications in the i.i.d. model that are related to the ordering of the data. For example, one or more of the parameters might change with time. Some problems of this type are classical ones in statistics - there is a large literature concerned with slippage of normal means and with detecting heteroscedasticity of normal variances. However, we shall not consider these problems in this chapter.

6.4 TESTING SIMPLE UNIFORMITY

6.4.1 Introductory Remarks

The transformations of section 6.2 can be used to construct similar α tests for the testing problems of (6.2) and (6.3) by making size α tests of the surrogate null hypothesis that the U-values are themselves independently and identically distributed as uniform random variables on the interval (0,1), i.e., are i.i.d. $U(0,1)$ rv's. These transformed values of the U's should be studied with care because they contain all test information in the sense of Theorem 6.9. It should also be observed that when the null hypothesis $H_0: P \in P_0$ fails, the distribution of the U's may fail to be i.i.d. $U(0,1)$ in many ways. They may no longer be independent, nor identically distributed, nor uniformly distributed. Moreover, if the model properties of independence or identical distributions of the observations as assumed by the formal goodness-of-fit and separate hypothesis testing problems of (6.2) - (6.6) are violated, then this will also result in transformed values that are not, in general, i.i.d. $U(0,1)$ rv's. Thus we can use these transformed values to study the validity of other model specifications, in addition to violations of an assumed parametric density functional form. Now, the choice of tests to be made will, of course, depend upon the type of violation of model assumptions of concern. In some problems, we may not have reason to be concerned about particular types of model violations, and we would like to perform an analysis with good sensitivity against a wide range of alternatives. Such an analysis can be made by studying the transformed values to determine if these are feasible values for i.i.d. $U(0,1)$ rv's. The analyses which we propose here for this purpose include graphical methods, and two omnibus goodness-of-fit tests on the transformed U values.

Subsequently, we shall use lower case u 's and write $\underline{u} = (u_1, \dots, u_N)'$ for both the random variables and their observed values. Here $N = n - p$, the number of observations minus the number of parameters in the model. Also, when we consider models with normal distributions of the errors in subsections 6.5.5 and 6.5.7 (linear regression models) below, it will be seen that the values u_j are closely related to quantities usually called residuals from the least squares fitted lines from these models. For this reason we shall call $\underline{u}$ the vector of uniform residuals from the parametric model, in general.

Next put

$$z_j = \phi^{-1}(u_j); j = 1, \dots, N; \quad (6.17)$$

for ϕ a $N(0,1)$ df, and ϕ^{-1} its inverse. By the inverse of Theorem 6.1, when the u_j 's are i.i.d $U(0,1)$ rv's, then the z_j 's are i.i.d. $N(0,1)$ rv's. Thus we can also test (6.3) by testing the surrogate null hypothesis that the z_j 's are i.i.d. $N(0,1)$ rv's. We shall consider further tests based on the z_j 's below when we consider particular parametric classes. We shall call the values z_j normal uniform (NU) residuals. A principal reason for considering these NU residuals is that the problem of testing normality is the most extensively studied problem in the goodness-of-fit area. See Chapter 9 of this book, and we will recommend tests of normality below. Hester and Quesenberry (1984) have found that a test based on these NU residuals has attractive power properties for testing for heteroscedasticity, i.e. , for either increasing or decreasing variance for ordered regression data.

6.4.2 Graphical Methods for Symmetric Samples

Graphical methods are very useful for studying the uniformity of the values $u_1, \dots, u_N$ on the unit interval. As a first step the data can be plotted on the $(0,1)$ interval; or, if N is large it will be more convenient to partition the $(0,1)$ interval into a number of subintervals of equal length and to construct a histogram on these subintervals. The data pattern on the unit interval, or the shape of the histogram constructed, conveys important information about the shape of the parent density from which the data was drawn relative to the shapes of the densities of the null hypothesis class. We next consider this in more detail.

In order to study the significance of particular patterns of the u -values on the unit interval, we first suppose that the u 's were obtained from the classical probability integral transformation of Theorem 6.1 by transforming a sample $X_1, \dots, X_N$ using a continuous d.f. $F_0(x)$ with corresponding density function $f_0(x)$. However, if $X_1, \dots, X_N$ is actually a sample from a parent distribution with d.f. $F_1(x)$ and density function $f_1(x)$, then the u 's will not constitute a set of i.i.d. $U(0,1)$ r.v.'s, unless $F_0 = F_1$. Let (a,b) be an interval on the real line where $f_0(x) < f_1(x)$ for every x in (a,b) . Then the expected number of u 's in the interval $(F_0(a), F_0(b))$ will be greater than under uniformity. Conversely, if $f_0(x) > f_1(x)$ for x in the interval (c,d) , then the expected pattern of points in $(F_0(c), F_0(d))$ will be more sparse than under uniformity. Thus the splatter pattern of $u_1, \dots, u_N$ on the unit interval should be interpreted as follows. If the data are too dense on an interval $(F_0(a), F_0(b))$, this implies that the true density function f_1 exceeds the density function f_0 on the interval (a,b) and, conversely, when the data are too sparse on $(F_0(a), F_0(b))$ the true density f_1 is less than f_0 on (a,b) .

Next, suppose that the u 's were obtained from a sample $X_1, \dots, X_n$ by the CPIT transformations of Theorem 6.4 for a parametric class $\mathcal{J}_0$. Then the data patterns have the same interpretations as for the classical probability integral transformations just discussed. This is true because the transforming function used for the i th observation is an estimator of the parent d.f. Indeed, this transforming function is a minimum variance unbiased (MVU) estimator of the parent d.f. based on $(X_1, \dots, X_i)$ whenever the parent d.f. is a member of $\mathcal{J}_0$. (See Seheult and Quesenberry (1971).)

If $u_{(1)}, \dots, u_{(N)}$ are the order statistics of a sample from a $U(0,1)$ parent, then $u_{(i)}$ is a beta random variable with parameters $(i, n-i+1)$. Thus $u_{(i)}$ has mean and variance given by

$$i/(N+1) \text{ and } i(N+1-i)/(N+1)^2(N+2) ,$$

respectively, and its distribution function is the incomplete beta-function. If the points $(u_{(i)}, i/(N+1))$ are plotted on Cartesian axes these points should approximate the line $g(u) = u$ for $0 < u < 1$. Quesenberry and Hales (1980) have given graphs called concentration bands for $N = 2, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, 150, 200, 300, 500$ that are helpful guides for judging the significance of these uniform probability plots.

The discussion made above in this subsection allows us to anticipate the pattern of points that the uniform residuals will make on the unit interval when a particular alternative density is considered. For a given data set this allows a direct interpretation of the data patterns observed in uniform and NU residuals both in histograms and in uniform probability plots. Some examples of data patterns in uniform probability plots will be given in the numerical examples of section 6.6.

6.4.3. Test Statistics for Uniformity and Normality

The statistical literature contains a large number of goodness-of-fit test statistics which can be used to test that a set of values $u_1, \dots, u_N$ constitutes a sample from a $U(0,1)$ distribution. We are here interested in tests which have good power against a wide range of alternative shapes. Monte Carlo power studies of tests of uniformity on which we shall base our choice of test statistics have been given by Stephens (1974),

Quesenberry and Miller (1977), Q-M, and Miller and Quesenberry (1977). We recommend two test statistics for use in testing uniformity of the u -values, and our choices are the U^2 statistic of Watson (1961) and the smooth test of Neyman (1937).

The Watson U^2 statistic was found in Q-M to have good power against a number of classes of alternatives, even for small sample sizes. This statistic is given by

$$U^2 = (1/12N) + \sum_{i=1}^N \{[(2i-1)/2N - u_{(i)}]^2 - N(\bar{u} - 0.5)^2\}, \quad (6.18)$$

where $\bar{u} = (u_1 + \dots + u_N)/N$. Stephens (1970) found empirically that a linear function of U^2 given by

$$U_{MOD}^2 = \{U^2 - 0.1/N + 0.1/N^2\}\{1 + 0.8/N\}, \quad (6.19)$$

has critical values that are approximately constant in N for $N > 10$. Table 1 gives approximate significance points for U_{MOD}^2 that were obtained by Q-M by Monte Carlo methods, and the Stephens approximation for $N > 10$.

TABLE 1
Empirical Critical Values for U_{MOD}^2

N/α	0.1	0.05	0.01
2	0.164	0.181	0.195
3	0.158	0.191	0.242
4	0.152	0.187	0.254
5	0.152	0.185	0.256
6	0.151	0.187	0.260
7	0.151	0.187	0.262
8	0.152	0.188	0.262
9	0.152	0.189	0.270
10	0.151	0.185	0.265
∞	0.152	0.187	0.267

In a classic paper Neyman (1937) proposed a test statistic that is defined as follows. Let π_r denote the r^{th} degree

Legendre polynomial, the first five of which are for $0 < u < 1$;

$$\pi_0(u) = 1 ,$$

$$\pi_1(u) = \sqrt{12} (u - 1/2) ,$$

$$\pi_2(u) = \sqrt{5} (6(u - 1/2)^2 - 1/2) ,$$

$$\pi_3(u) = \sqrt{7} (20(u - 1/2)^3 - 3(u - 1/2)) ,$$

$$\pi_4(u) = 210(u - 1/2)^4 - 45(u - 1/2)^2 + 9/8 .$$

Then put

$$t_r = \sum_{j=1}^N \pi_r(u_j) ; r = 1, \dots, k ;$$

and

$$p_k^2 = (1/N) \sum_{r=1}^k t_r^2 .$$

Miller and Quesenberry (1977), M-Q, showed that each of the tests p_k^2 for $k = 2, 3, 4$ have good power against a number of classes of alternatives. Neyman showed that p_k^2 has an asymptotic $\chi^2(k)$ distribution under the uniformity null hypothesis and a noncentral χ^2 distribution under the alternative.

Following recommendations of M-Q, we shall use p_4^2 as a general omnibus test statistic for the uniformity null hypothesis. Table 2 gives some significance points for p_4^2 obtained by M-Q by Monte Carlo methods.

We shall consider only the statistics U^2 and p_4^2 for testing uniformity because they appear on the basis of the above noted power studies to be two of the best general omnibus tests of uniformity. In those studies it was noted that one weakness shared by many goodness-of-fit tests (including p_4^2) is that for small sample sizes (sometimes for N as large as 15), and some important alternatives, the tests are biased. We have not observed a case where U^2 gives a biased test for any sample size,

nor one for which p_4^2 gives a biased test for N larger than 10. Thus we recommend only U^2 if $N < 10$, and we will compute both U^2 and p_4^2 for larger sample sizes. A practical advantage of p_4^2 is that it has an approximate $\chi^2(4)$ distribution as N increases, and so its observed significance level, p -value, is easily evaluated. These points will be illustrated in the numerical examples of section 6.6.

TABLE 2
Empirical Critical Values for p_4^2

$N \backslash \alpha$	0.1	0.05	0.01
2	7.19	9.52	16.14
3	7.34	9.51	15.80
4	7.46	9.50	15.43
5	7.53	9.49	15.12
6	7.57	9.48	14.86
7	7.60	9.47	14.65
8	7.62	9.47	14.47
9	7.63	9.46	14.32
10	7.64	9.46	14.19
11	7.65	9.45	14.09
12	7.65	9.45	14.00
13	7.66	9.44	13.93
14	7.66	9.44	13.87
15	7.66	9.43	13.82
16	7.66	9.43	13.78
17	7.66	9.43	13.74
18	7.67	9.42	13.71
19	7.67	9.42	13.69
20	7.67	9.42	13.67
30	7.68	9.40	13.58
40	7.68	9.40	13.52
50	7.69	9.40	13.48
∞	7.78	9.49	13.28

Instead of (or in addition to) the graphs and tests using the uniform residuals, we can make graphs and tests using the NU residuals defined in (6.17) above. For most purposes, we feel that the graphs of uniform residuals are more easily interpreted than those of NU residuals. It seems to this writer that it is a bit easier to judge if a histogram agrees with an assumed uniform parent than with a normal-shaped parent. A probability plot of the NU residuals requires normal probability paper, whereas the uniform residuals requires no special paper. Thus the uniform residuals are more convenient to plot with widely available software.

Aside from graphs of NU residuals, we can also make omnibus tests of a wide variety of parametric models by testing the normality of these residuals. There are a number of reasons to consider tests NU residuals in place of, or, in addition to, tests on the uniform residuals. The problem of testing normality for a simple random sample is the most extensively studied problem in the goodness-of-fit field, and excellent tests for this problem are readily available (see Chapter 9 of this handbook). Also, for at least some of the many important problems of testing null hypothesis parametric models with normally distributed errors - such as the multiple samples problems, regression models of section 6.5.7, or the ANOVA model of section 6.5.8 - the NU residuals have a useful tendency to retain data patterns from the original data. For example, Hester and Quesenberry (1984) have exploited this property to construct efficient tests for heteroscedasticity for normal regression models.

Any of the multitude of goodness-of-fit tests for the completely specified null hypothesis testing problem of (6.1) may be used to test normality of the NU residuals, however, it has been shown by both Stephens (1974) and Dyer (1974) that the test for composite normality usually has better power against most alternatives, even when the parameters are known. We have done some simulation work in studying the efficiency of the Anderson-Darling, AD, test on NU residuals, and it appears that this characteristic of goodness-of-fit tests to show better power when parameters are not assumed known can be expected to obtain here, also. Finally, we must choose the particular tests of normality to make on the NU residuals. At the present state of knowledge, we feel that reasonable choices of test statistics would be the Shapiro - Wilk test or the AD test.

Another point that should be recalled in interpreting the analysis of residuals discussed here is the following. We generally consider testing the goodness-of-fit hypothesis testing problem as stated in display (6.2), and this means, of course, that we assume that the observations are i.i.d. However, with real data it will often be the case that we cannot really be sure that these assumptions are valid, and thus in practice in these cases the classical goodness-of-fit null hypothesis of (6.2) should be expanded to include the i.i.d. assumptions. That is, it is desirable to validate the entire model including these sampling assumptions.

6.5 TRANSFORMATIONS FOR PARTICULAR FAMILIES

6.5.1 Introduction

In this section the CPIT transformations are given for a number of parametric families of probability distributions. Lower case letters are used to denote both random variables and their observed values here and in the next section, which considers numerical examples. In all cases we assume that a sample $x_1, \dots, x_n$ is available from an unspecified member of the class of distributions under consideration. We shall denote by $(w_1, \dots, w_{n-1})$ the sample with $x_{(n)}$ deleted, and use similar notation for the other cases when $x_{(1)}$ is deleted or both $x_{(1)}$ and $x_{(n)}$ are deleted. (See section 6.2.3 for details of this transformation.)

6.5.2 Uniform Distributions

Densities:

$$f(x; \mu_1, \mu_2) = (1/(\mu_2 - \mu_1)) I_{(\mu_1, \mu_2)}(x), \quad (6.20)$$

for $-\infty < \mu_1 < \mu_2 < +\infty$.

Case I, $\mu_1 = \mu_{10}$ known, μ_2 unknown.

For $w_1, \dots, w_{n-1}$ the sample with $x_{(n)}$ deleted put

$$u_j = (w_j - \mu_{10}) / (x_{(n)} - \mu_{10}); j = 1, \dots, n-1. \quad (6.21)$$

Case II, μ_1 unknown, $\mu_2 = \mu_{20}$ known.

For $w_1, \dots, w_{n-1}$ the sample with $x_{(1)}$ deleted put

$$u_j = (\mu_{20} - w_j) / (\mu_{20} - x_{(1)}); j = 1, \dots, n-1. \quad (6.22)$$

Case III, Both μ_1 and μ_2 unknown.

For $w_1, \dots, w_{n-2}$ the sample with both $x_{(1)}$ and $x_{(n)}$ deleted put

$$u_j = (w_j - x_{(1)}) / (x_{(n)} - x_{(1)}); j = 1, \dots, n-2. \quad (6.23)$$

For these transformations for uniform distributions, we note that the lack of invariance with respect to permutations of the X's cited in subsection 6.3.2 does not obtain. That is, two permutations of the X's will give the same values for the elements of the u vector, though not in the same order. That is, the ordered values of the components of u will be the same, and any analysis or test that is symmetric in the components of u will be unaffected by a permutation of the X's.

6.5.3 Exponential Distributions

Densities:

$$f(x; \mu, \theta) = (1/\theta) \exp\{-(x - \mu)/\theta\} I_{(\mu, \infty)}(x), \quad (6.24)$$

for $-\infty < \mu < \infty$, and $\theta > 0$.

Case I, μ unknown, $\theta = \theta_0$ known.

Let $w_1, \dots, w_{n-1}$ be the sample with $x_{(1)}$ deleted.

$$u_j = 1 - \exp\{-(w_j - x_{(1)})/\theta_0\}; j = 1, \dots, n-1. \quad (6.25)$$

Note that the remarks about the effect of permuting the X's on the u's above also obtain to these u's, viz., if the X's are permuted we still get the same set of values $\{u_1, \dots, u_{n-1}\}$ from (6.25). For cases II and III that follow, we give the transforms obtained from the sample $X_1, \dots, X_n$ from Theorem 6.4, as (6.26), and, also, transforms obtained by applying Theorem 6.3 directly to the order statistics $(x_{(1)}, \dots, x_{(n)})$, as (6.27); these latter formulae were obtained

by O'Reilly and Stephens (1984).

Case II, θ unknown, $\mu = \mu_0$ known.

Put $x'_i = x_i - \mu_0$, $i = 1, \dots, n$.

$$u_{j-1} = 1 - \left\{ \frac{\sum_{i=1}^{j-1} x'_i}{\sum_{i=1}^j x'_i} \right\}^{j-1}, \quad j=2, \dots, n. \quad (6.26)$$

Next put, $y_i = x_{(i)} - \mu_0$; $i = 1, \dots, n$; $y_0 = 0$.

$$u_j = 1 - \left\{ \frac{1-(n-j+1)y_j/(y_j + \dots + y_n)}{1-(n-j+1)y_{j-1}/(y_{j-1} + \dots + y_n)} \right\}^{n-j}, \quad j=1, \dots, n-1. \quad (6.27)$$

Case III, μ and θ unknown.

First, let $w_1, \dots, w_{n-1}$ be the sample with $x_{(1)}$ deleted and put $w'_i = w_i - x_{(1)}$, $i = 1, \dots, n-1$. Then the transformations for this case are given by (6.26) with x'_i replaced by w'_i and n by $n-1$.

Next, put $y_{i-1} = x_{(i)} - x_{(1)}$, $i = 1, \dots, n$. Then the transformations for this case are given by (6.27) with these y 's, and n replaced by $n-1$.

6.5.4 Pareto Distributions

Densities:

$$f(x; \alpha, \gamma) = (\alpha \gamma^\alpha / x^{1+\alpha}) I_{(\gamma, \infty)}(x), \quad \alpha > 0, \gamma > 0. \quad (6.28)$$

If the transformed values $\ln x_i$; $i = 1, \dots, n$; are considered, then these transformed values are a sample from the exponential family of (6.23) where $\theta = 1/\alpha$ and $\mu = \ln \gamma$.

Case I, γ unknown, $\alpha = \alpha_0$ known.

Let $w_1, \dots, w_{n-1}$ be the sample with $x_{(1)}$ deleted.

$$u_j = 1 - (x_{(1)}/w_j)^{\alpha_0}, \quad (6.29)$$

for $j = 1, \dots, n-1$.

Case II, α unknown, $\gamma = \gamma_0$ known.

Put $x'_i = \ln x_i - \ln \gamma_0$; $i = 1, \dots, n$. Then the CPIT transforms are given by (6.26) using these x'_i 's.

Put $y_i = \ln x_{(i)} = \ln \gamma_0$; $i = 1, \dots, n$. Then the order statistics CPIT is given by (6.27) using these y 's.

Case III, α and γ unknown.

Let $w_1, \dots, w_{n-1}$ be the sample with $x_{(1)}$ deleted, as above in Case I, and put $x'_i = \ln x_{(i)}$ for $i = 1, \dots, n-1$. Then the CPIT transforms are given by (6.26) using these x'_i 's and n replaced by $n-1$.

To obtain the order statistics transforms for this case put $y_{i-1} = \ln x_{(i)} - \ln x_{(1)}$; $i = 1, \dots, n$; and use these y 's in (6.27) with n replaced by $n-1$.

6.5.5 Normal Distributions

Densities:

$$f(x; \mu, \sigma^2) = (1/\sigma\sqrt{2\pi}) \exp\{-(x - \mu)^2/2\sigma^2\}, \quad (6.30)$$

for $-\infty < \mu < \infty$, and $\sigma^2 > 0$.

Case I, μ unknown, $\sigma^2 = \sigma_0^2$ known.

Let $\bar{x}_j = (x_1 + \dots + x_j)/j$ for $j = 1, \dots, n$ and Φ denote the d.f. of a $N(0,1)$ distribution.

$$u_{j-1} = \Phi\{(j-1)^{1/2}(X_j - \bar{X}_{j-1})/j^{1/2}\sigma_0\}; \quad (6.31)$$

for $j = 2, \dots, n$.

Case II, σ^2 unknown, $\mu = \mu_0$ known.

For this case only we put

$$S_j^2 = \sum_{i=1}^j (X_i - \mu_0)^2/j; \quad j = 1, \dots, n;$$

and let G_ν denote a Student-t distribution function with ν degrees of freedom. Then

$$u_{j-1} = G_{j-1}\{(x_j - \mu_0)/S_{j-1}\}; \quad j = 2, \dots, n. \quad (6.32)$$

This can be generalized somewhat, as follows. Suppose that S_v^2 is a mean square estimator of σ^2 that is independent of the x 's, and that vS_v^2/σ^2 has a $\chi^2(v)$ distribution. Then we put

$$S_j^* = \{[(j-1)S_{j-1}^2 + vS_v^2]/(v+j-1)\}^{1/2},$$

and

$$u_{j-1} = G_{v+j-1}\{(x_j - \mu_0)/S_j^*\}; j = 2, \dots, n. \quad (6.32')$$

Case III, μ and σ^2 unknown.

For this case put

$$\bar{x}_j = (1/j) \sum_{i=1}^j x_i, \quad S_j^2 = (1/(j-1)) \sum_{i=1}^j (x_i - \bar{x}_j)^2.$$

Then

$$u_{j-2} = G_{j-2}\{[(j-1)/j]^{1/2}(x_j - \bar{x}_{j-1})/S_{j-1}\}; j = 3, \dots, n. \quad (6.33)$$

Again, let S_v^2 be an independent mean square estimator of σ^2 such that vS_v^2/σ^2 is a $\chi^2(v)$ rv.

Then put

$$S_j^* = \{[(j-2)S_{j-1}^2 + vS_v^2]/(v+j-2)\}^{1/2}$$

and

$$u_{j-2} = G_{v+j-2}\{[(j-1)/j]^{1/2}(x_j - \bar{x}_{j-1})/S_j^*\}. \quad (6.33')$$

In computing the above quantities it is helpful to use the following updating formulae. For $\bar{x}_j$ as defined above, and for

$$(SS_j) = \sum_{i=1}^j (x_i - \bar{x}_j)^2,$$

then

$$\bar{x}_j = [(j-1)\bar{x}_{j-1} + x_j]/j, \quad (SS_j) = (SS_{j-1}) + [j/(j-1)](x_j - \bar{x}_j)^2. \quad (6.34)$$

Youngs and Cramer (1972) and Chan, Golub and Le Veque (1983) discussed these formulae, the latter authors were primarily concerned with numerical accuracy.

6.5.6 Lognormal Distributions

Densities:

$$f(x; \mu, \sigma^2) = (x\sigma\sqrt{2\pi})^{-1} \exp\{-(\ln x - \mu)^2/2\sigma^2\}, \quad (6.35)$$

$-\infty < \mu < \infty$, and $\sigma^2 > 0$.

The values $\ln x_1, \dots, \ln x_n$ constitute a sample for the normal family of (6.30), and transformations for the three cases can be obtained by replacing $(x_1, \dots, x_n)$ by $(\ln x_1, \dots, \ln x_n)$ in the transformations for the corresponding normal cases above.

6.5.7 Normal Linear Regression Models

For $y_n = (y_1, \dots, y_n)'$ a vector of independent rv's, X_n an $n \times p$ matrix of full rank ($n > p$), and $\beta = (\beta_1, \dots, \beta_p)'$ a vector of p parameters, we consider the family of normal distributions

$$y_i \sim N(x_j' \beta, \sigma^2); j = 1, \dots, n; \quad (6.36)$$

for x_j' the j th row of X_n . The matrix X_j denotes the matrix consisting of the first j rows of X_n . We shall give transformations here for two cases, viz., for σ^2 known and unknown.

Case I, $\sigma^2 = \sigma_0^2$ known.

For this case it is readily verified that the statistic $t_n' = X_n' y_n$ is a complete sufficient statistic for β . Moreover, the UMVU estimating distribution function is itself the df of a normal distribution with

$$\text{mean} = x_j'(X_j'X_j)^{-1}X_j'y_j, \text{ variance} = \sigma_0^2[1 - x_j'(X_j'X_j)^{-1}x_j].$$

Using this result with Theorem 6.4 gives the following results. Put

$$A_j = \frac{y_j - \bar{x}_j'(X_j'X_j)^{-1}X_j'y_j}{\sigma_o[1 - \bar{x}_j'(X_j'X_j)^{-1}\bar{x}_j]^{\frac{1}{2}}} \quad (6.37)$$

and

$$u_{j-p} = \phi(A_j), \quad j = p + 1, \dots, n. \quad (6.38)$$

Certain well-known updating formulae [Placket (1950), Bartlett (1951)] are very convenient for the computation of A_j as well as for the development below. For

$b_j = (X_j'X_j)^{-1}X_j'y_j$, the least squares estimate of β from the first j observations,

$S_j^2 = y_j'[I - X_j(X_j'X_j)^{-1}X_j']y_j$, the usual least squares sum of squares for residuals from the first j observations, then the updating formulae are

$$(X_j'X_j)^{-1} = (X_{j-1}'X_{j-1})^{-1} - \frac{(X_{j-1}'X_{j-1})^{-1}\bar{x}_{j-1}\bar{x}_{j-1}'(X_{j-1}'X_{j-1})^{-1}}{1 + \bar{x}_{j-1}'(X_{j-1}'X_{j-1})^{-1}\bar{x}_{j-1}} \quad (6.39)$$

$$b_j = b_{j-1} + (X_j'X_j)^{-1}\bar{x}_j(y_j - \bar{x}_j'b_{j-1}) \quad (6.40)$$

$$S_j^2 = S_{j-1}^2 + \sigma_o^2 A_j^2 \quad (6.41)$$

Using these relations, A_j can be written in the alternative form

$$A_j = \frac{y_j - \bar{x}_j'b_{j-1}}{\sigma_o[1 + \bar{x}_j'(X_{j-1}'X_{j-1})^{-1}\bar{x}_j]^{\frac{1}{2}}} \quad (6.37')$$

Put $w_j = \sigma_o A_j$, and these w_j 's are the quantities sometimes called recursive residuals, and have been considered by a number writers including Hedayat and Robson (1970); Brown, Durbin and Evans (1975), and, recently, by Galpin and Hawkins (1984). These writers have generally not assumed that σ^2 is known. Then in the form (6.37') the w_j 's can be shown to be i.i.d. $N(0, \sigma^2)$, which

also follows easily from the CPIT results given above.

Case II, σ^2 unknown.

From O-Q, Example 4.3, we put

$$B_j = \frac{(j - p - 1)^{\frac{1}{2}}(y_j - x_j' b_j)}{\{[1 - x_j'(X_j' X_j)^{-1} x_j] s_j^2 - (y_j - x_j' b_j)^2\}^{\frac{1}{2}}} \quad (6.42)$$

and

$$u_{j-p-1} = G_{j-p-1}(B_j); j = p + 2, \dots, n. \quad (6.43)$$

Using the updating formulae again, we express B_j in the alternative form

$$B_j = \frac{(j - p - 1)^{\frac{1}{2}}(y_j - x_j' b_{j-1})}{s_{r-1} [1 + x_j'(X_{j-1}' X_{j-1})^{-1} x_j]^{\frac{1}{2}}} \quad (6.42')$$

Note that the quantity $(y_j - x_j' b_{j-1})$ in the numerator of (6.37') and (6.42') is the residual of y_j from the least squares line fitted using the first $j - 1$ points. This is a normal rv with mean zero, and by examining (6.37') and (6.42'), we see that A_j and B_j are the standardized and Studentized forms of this rv, respectively. Moreover, by again using the updating formulae, we can show that

$$B_j = (j - p - 1)^{\frac{1}{2}} A_j / \left(\sum_{i=1}^{j-1} A_i^2 \right)^{\frac{1}{2}} = (j - p - 1)^{\frac{1}{2}} w_j / \left(\sum_{i=1}^{j-1} w_i^2 \right)^{\frac{1}{2}}, \quad (6.44)$$

for w_j the recursive residual defined above.

It should be noted that the formulae given above in subsection 6.5.5 for the univariate normal distribution for the cases when σ^2 is known are, of course, special cases of the formulae given in this section for the univariate regression model.

6.5.8 Normal Analysis of Variance Model

The material in this section is largely from Quesenberry, Giesbrecht and Burns (1983), QGB. We consider k mutually independent samples, x_{ij} ; $i = 1, \dots, k$; $j = 1, \dots, n_i$; as in (6.4); and the family of distributions

$$x_{ij} \sim N(\mu_i, \sigma^2); i = 1, \dots, k. \quad (6.45)$$

In words, this is a fixed-effects, one-way, normal errors, analysis-of-variance model. Define

$$n = n_1 + \dots + n_k, v_{ij} = n_1 + \dots + n_{i-1} + j - i - 1$$

$$\bar{x}_{ij} = (x_{i1} + \dots + x_{ij})/j, SS_{ij} = \sum_{\alpha=1}^j (x_{i\alpha} - \bar{x}_{ij})^2$$

and for $v_{ij} > 0$,

$$A_{ij} = \frac{[(j-1)v_{ij}/j]^{\frac{1}{2}}(x_{ij} - \bar{x}_{i(j-1)})}{\left[\sum_{\alpha=1}^{i-1} SS_{\alpha n_{\alpha}} + SS_{i(j-1)}\right]^{\frac{1}{2}}} \quad (6.46)$$

(Remark: Readers familiar with the QGB paper will note that the updating formulae in (6.34) above have been used to rewrite the formula given for A_{ij} in that paper in a more convenient form.)

Case I, $\sigma^2 = \sigma_0^2$ known.

$$u_{ij} = \Phi\{[x_{ij} - \bar{x}_{ij}][j/(j-1)]^{\frac{1}{2}}/\sigma_0\} = \Phi\{[x_{ij} - \bar{x}_{i(j-1)}][(j-1)/j]^{\frac{1}{2}}/\sigma_0\} \quad (6.47)$$

Case II, σ^2 unknown.

$$u_{ij} = G_{v_{ij}}(A_{ij}) \quad (6.48)$$

for $i = 1$ and $j = 3, \dots, n_1$; and $i = 2, \dots, k$; $j = 2, \dots, n_i$.

A case can arise which is, in a sense, intermediate between cases I and II above. Suppose there is an external mean square estimator S_v^2 of σ^2 available such that vs_v^2/σ^2 is a $\chi^2(v)$ rv. Then

put

$$v_{ij}^* = v_{ij} + v, S^* = \left\{ \left(\sum_{\alpha=1}^{i-1} SS_{\alpha n_{\alpha}} + SS_{i(j-1)} + vS_v^2 \right) / v_{ij}^* \right\}^{\frac{1}{2}},$$

and

$$u_{ij} = G_{v_{ij}^*} \{ [(j-1)/j]^{\frac{1}{2}}(x_{ij} - \bar{x}_{i(j-1)})/S^* \}. \quad (6.49)$$

The question can be raised as to what is to be gained by using (6.49) rather than (6.48) since, under the null hypothesis model, we obtain the same number of i.i.d. $v(0,1)$ rv's. The

answer is that when some of the model assumptions are incorrect the anomalous data patterns resulting should be more distinct and analyses based on them more sensitive. Another point is that even under the null hypothesis the dependence of the u 's upon the ordering of the data is weakened. This point can be seen by observing in (6.49) that as $v \rightarrow \infty$ the distribution function in (6.49) converges to Φ and thus (6.47) is the limiting case of (6.49).

6.6 NUMERICAL EXAMPLES

6.6.1 Introduction

In this section we use some of the transformations of the last section in numerical examples. In all of these examples the computations were performed using programs written by the author for this purpose.

6.6.2 Salinity Data

A large scale program by North Carolina State University to study environmental impact in the Cape Fear Estuary includes a sampling over time by weeks of the larval density in the estuary. Accompanying data includes salinity measured in parts per thousand (ppt) at the time of collection of the larval specimen. The salinity data consists of several samples, and each sample consists of measurements made over a relatively short time period (approx. 24 hrs.), and the samples are taken at intervals of at least one month apart. It appears reasonable to consider each sample to be an independent random sample from a common parent distribution, and to assume that the functional form of the parent density is the same for all samples. However, the means and variances can be expected to vary widely from sample to sample.

There are available six samples, given in Table 3, for studying the functional form of the parent distributions.

TABLE 3

Obs.	Sample Number					
	1	2	3	4	5	6
1	18.3	19.8	11.1	27.0	13.1	9.1
2	16.9	22.0	12.7	26.4	12.6	13.6
3	15.9	22.5	10.2	28.5	18.7	16.4
4	16.0	21.3	8.6	25.5	19.4	14.3
5	15.8	20.0	7.9	28.1	17.4	14.0
6	14.5	19.1	10.9	26.3	16.9	10.2
7	18.6	18.8	13.7	26.6	15.5	11.7
8	21.8	22.4	12.4	24.5	16.7	12.6
9	18.3	22.9	12.9	27.8	17.7	12.0
10	18.5	21.5	10.1	28.7		11.1
11	17.1	21.4	9.9	28.7		
12	16.6	20.9		26.8		

Normal Analysis

We consider the null hypothesis H_0 of (6.5) with F_{θ_j} the df of a $N(\mu_j, \sigma_j^2)$ distribution, i.e., we consider the null hypothesis that each sample is from a normal parent, but allow each of these parents to have different means and variances (cf. Quesenberry, Whitaker and Dickens (1976)).

We have transformed each of the samples of Table 3 using the transformations of equation (6.33). Since a sample of size N_i gives $n_i - 2$ transformed values, there are a total of $N = n_1 + \dots + n_6 - 12 = 54$ transformed u-values. These 54 u-values are given in Table 4, and are graphed against the expected values

$(1/(N + 1)) = (1/55)$ in Figure 1.

TABLE 4
Pooled and Ranked u-values, Normal Analysis

.0380	.1431	.2625	.3487	.4572	.5908	.6861	.7918	.9080
.0734	.1630	.2676	.3884	.4716	.6030	.6989	.8066	.9106
.0988	.1823	.2892	.3894	.4779	.6196	.7223	.8197	.9274
.1012	.1972	.2918	.3971	.5638	.6288	.7607	.8620	.9366
.1159	.2177	.3139	.4201	.5690	.6380	.7797	.8770	.9765
.1233	.2450	.3473	.4364	.5705	.6641	.7908	.8926	.9922

$U_{MOD}^2 = 0.013$					$P_4^2 = 1.631$			
---------------------	--	--	--	--	-----------------	--	--	--

FIG. 1: SALINITY DATA. NORMAL ANALYSIS

The values of the modified Watson statistic U_{MOD}^2 and the Neyman smooth statistic p_4^2 are both much too small for significance at the 10 per cent level. The p_4^2 statistic is approximately a $\chi^2(4)$ rv and the observed significance level is

$$P(p_4^2 > 1.631) \doteq P(\chi^2(4) > 1.631) = 0.80 .$$

Consider Figure 1 and recall that the points should approximate the line if the underlying parent family is normal. The fit is excellent and from this and the small values of U_{MOD}^2 and p_4^2 we conclude that normal distributions fit this data very well, indeed.

Even though we have concluded that normal distributions fit this data well, we shall in the remainder of this subsection consider fitting two other two-parameter families to the data in order to illustrate the use of the transformations.

Uniform Analysis

We have transformed each of the samples of Table 3 using the transformations of (6.23) for a two-parameter uniform family. The 54 u-values obtained are given in Table 5 and are graphed against the null hypothesis expected values in Figure 2. This uniform probability plot is typical of the pattern we obtain when normal data is transformed using the transformations for a uniform family. This S-shaped pattern can be anticipated from the discussion of data patterns in subsection 6.4.2, since both distributions are symmetric and on the range of the sample the normal density has thinner tails near the extremes, and is higher than the uniform density in the center.

The value of U_{MOD}^2 in Table 5 falls between the upper 10 and 5 per cent points, and the observed significance level of p_4^2 is

$$P(p_4^2 > 6.610) \doteq P(\chi^2(4) > 6.610) = .158 .$$

From these statistics and from Figure 2 it is clear that the uniform distributions do not fit the data well, and certainly not as well as normal distributions. It should be borne in mind here that the uniform distribution as an alternative to normality is one that is rather difficult to detect.

TABLE 5
Pooled and Ranked u-values, Uniform Analysis

.0732	.2055	.3288	.3973	.5122	.5517	.6324	.7500	.8621
.0735	.2381	.3448	.4265	.5172	.5616	.6341	.7759	.8780
.1207	.2439	.3562	.4286	.5205	.5952	.6585	.7805	.8971
.1507	.2740	.3562	.4524	.5205	.6029	.6712	.7857	.9024
.1781	.2877	.3793	.4795	.5476	.6098	.7059	.8276	.9524
.1918	.2927	.3966	.5000	.5479	.6164	.7123	.8571	1.0000

$U_{MOD}^2 = 0.169$					$p_4^2 = 6.610$			
---------------------	--	--	--	--	-----------------	--	--	--

FIG. 2: SALINITY DATA. UNIFORM ANALYSIS

Exponential Analysis

We have transformed each of the samples of Table 3 using the transformations of (6.27) for two-parameter exponential distributions, see Case III in subsection 6.5.3. The 54 pooled and ranked values obtained from these transformations are given in Table 6. These values are plotted in Figure 3 against the expected values. This probability plot is typical of the pattern obtained when normal data is transformed using exponential family transformations - recall the discussion in subsection 6.4.2.

TABLE 6

Pooled and Ranked u-values, Exponential Analysis

.1199	.3843	.4254	.4857	.5489	.6009	.6753	.7266	.7952
.1780	.3871	.4318	.5182	.5550	.6186	.6870	.7301	.8066
.1790	.3967	.4397	.5313	.5617	.6189	.7073	.7574	.8102
.2864	.4138	.4546	.5322	.5719	.6294	.7131	.7606	.8233
.3345	.4146	.4684	.5365	.5926	.6439	.7174	.7619	.9134
.3539	.4175	.4771	.5386	.6000	.6725	.7256	.7794	.9242

$U_{MOD}^2 = .747$	$p_4^2 = 28.260$
--------------------	------------------

This value of U_{MOD}^2 is highly significant since the upper 1 percent point for U_{MOD}^2 is .267. Moreover, $P(p_4^2 > 28.260) \doteq P(\chi^2(4) > 28.260) \doteq 0.00001$. From these values and Figure 3 it is clear that the two-parameter exponential distributions fit this data very poorly.

6.6.3 Simulated Data

The appendix gives three samples of 100 observations each that have been drawn from normal, exponential and uniform distributions. These samples are named NOR, NEX and UNI; which we shall abbreviate further to N, E and U in this section. For each of these three samples we have computed 98 u-values using the appropriate transformations for two-parameter normal, exponential and uniform families. This gives 9 sets of u-values. We have partitioned the (0,1) interval into 10 subintervals of equal lengths, and denote by M_i the number of u's that fall in the subinterval $((i-1)/10, i/10]$ for $i=1, \dots, 10$, for each set of u-values. These values and those for U_{MOD}^2 , p_4^2 and the observed significance

FIG. 3: SALINITY DATA. EXPONENTIAL ANALYSIS

level of p_H^2 , called Sig., are given in Table 7. The label (N,E) on row four means the normal sample was subjected to exponential transformations, for example.

The first three rows of Table 7 give results when the sample from a distribution is subjected to the transformations for a class to which that distribution belongs. The expected cell frequencies are 9.8, and none of the test statistics are significant at any of the usual levels. The other six rows all give level .05 significant tests for both U_{MOD}^2 and p_H^2 . All except (U,N) have observed significance levels for both statistics that are very small, indeed. A more detailed analysis could be carried out by plotting the 98 u-values for each case as was done in the last example in Figures 1, 2 and 3.

TABLE 7
Analysis of Simulated Samples

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	U_{MOD}^2	P_4^2	Sig.
(N,N)	12	12	5	10	7	8	17	3	13	11	.068	0.818	.94
(U,U)	10	10	14	10	5	9	8	8	11	13	.040	1.547	.82
(E,E)	13	11	10	10	10	8	10	5	13	8	.093	3.521	.47
(N,E)	0	3	4	9	12	15	27	20	8	0	1.587	63.041	2.4E(-12)
(U,N)	14	15	9	11	9	7	1	8	18	6	.221	13.174	.01
(E,U)	54	21	12	7	3	0	1	0	0	0	3.340	281.416	4.5E(-12)
(N,U)	3	8	13	12	18	19	9	10	5	1	.770	30.046	4.8E(-6)
(U,E)	8	5	8	10	9	13	13	21	11	0	.416	22.083	1.9E(-4)
(E,N)	0	12	27	17	7	8	5	7	6	9	.767	42.259	1.5E(-8)

6.6.4 The Bliss Data

The appendix gives data from Bliss (1967) on the body weight in grams of 21-day-old white leghorn chicks at two dosage levels of vitamin D. We have randomized each of the samples labeled series A and series B. Table 8 gives the data in the order in which it was analyzed here. Table 8 gives also the transformed u-values when the samples are subjected to two-parameter normal, uniform and exponential analyses. The test statistics U_{MOD}^2 and P_4^2 as well as the observed significance level, Sig., of P_4^2 have been computed for each sample as well as for the pooled samples and are given in Table 9. Graphs of the pooled u-values are given in Figures 4, 5 and 6 for the normal, uniform and exponential classes, respectively.

The probability plot in Figure 6 as well as the test statistics for the exponential case in Table 9 easily eliminate the exponential class from consideration for fitting this data. Comparison of the graphs in Figures 4 and 5 suggests that the normal class fits better than the uniform class, and this conclusion is supported by the values of U_{MOD}^2 and P_4^2 computed for the pooled u-values in Table 9. The statistic U_{MOD}^2 is significant at the .05 level for the uniform class but is much too small for significance at even

TABLE 8

<u>Sample A</u>				<u>Sample B</u>			
<u>Ranked u-values</u>				<u>Ranked u-values</u>			
Data	Normal	Uniform	Exp.	Data	Normal	Uniform	Exp.
210				196			
190				236			
182	.060	.053	.111	246	.022	.147	.184
230	.067	.105	.191	187	.024	.216	.281
236	.077	.228	.356	193	.083	.224	.318
214	.145	.263	.386	231	.122	.319	.411
246	.244	.298	.404	199	.161	.405	.418
186	.365	.298	.500	147	.165	.422	.456
168	.419	.351	.563	177	.225	.457	.489
162	.524	.404	.604	232	.263	.491	.496
196	.563	.474	.625	155	.273	.491	.531
226	.731	.509	.715	195	.296	.543	.534
156	.789	.561	.758	179	.372	.560	.562
202	.825	.614	.758	208	.422	.569	.590
190	.868	.649	.761	167	.445	.595	.616
236	.891	.649	.768	130	.616	.629	.643
220	.901	.702	.777	225	.621	.672	.654
320	.909	.702	.798	183	.724	.819	.667
242	.918	.754	.801	187	.727	.871	.669
270	.980	.789	.868	203	.809	.879	.735
				156	.822	.914	.744

the .10 level for the normal class. The probability plot of Fig. 4 suggests that the true distribution (assuming that the i.i.d. assumptions hold, of course) has a density with slightly thicker tails than a normal density, but it does look pretty symmetric. Figure 5 suggests that the true density has thinner tails on its interval of support (where the density is positive) than the tails of the uniform density on its interval of support, but again the density appears to be symmetric.

6.6.5 Regression Data

In this section we give three examples to illustrate the use of the transformations of (6.37) to test the regression model assumptions of (6.36). It should be borne in mind that we are testing all of the assumptions of the normal linear regression model.

TABLE 9
Test Statistics for Bliss Data

Transformation	Sample	U_{MOD}^2	P_4^2	Sig.
Normal	A	.128	4.707	.32
	B	.060	4.329	.36
	Pooled	.068	3.257	.52
Uniform	A	.110	4.347	.36
	B	.135	4.259	.37
	Pooled	.200	7.634	.11
Exponential	A	.224	9.852	.04
	B	.423	15.550	.004
	Pooled	.470	18.333	.001

FIG. 4: BLISS DATA. NORMAL ANALYSIS

FIG. 5: BLISS DATA. UNIFORM ANALYSIS

Fisher's Data

In this example we consider data given by Fisher (1958), p. 137, on the relative effects of two nitrogenous fertilizers in maintaining yields in bushels of wheat over a period of thirty years. Fisher uses a simple linear normal regression model for this data, and suggests that a more complex (curved) regression line might give a better fit.

We have applied the regression transformations of (6.43) to these data in the order that they are given by Fisher. Here $n = 30$ and $p = 2$, and the number of u-values is the number of observations less the number of parameters in the model (two regression coefficients and the variance), i.e., $N = n - p - 1 = 30 - 3 = 27$. The ranked u-values and test statistics u_{MOD}^2 and p_1^2 are given in

FIG. 6: BLISS DATA. EXPONENTIAL ANALYSIS

Table 10, and the ranked u-values are plotted against expected values in Figure 7.

The value $U_{MOD}^2 = 0.117$ is much less than the upper 10 per cent point of 0.152 given in Table 1, and is not significant at this level. Using the chi-squared approximation to the distribution of p_4^2 we obtain $P(p_4^2 > 12.16) \doteq P(\chi^2(4) > 12.16) \doteq 0.016$, and this value is significant at the 2% level but not at 1%. The graph of Figure 7 also raises doubts about the appropriateness of the normal linear model. This pattern suggests that the error distribution is not symmetric and is skewed to the right.

Wallace and Snedecor Data

Wallace and Snedecor (1931) give 25 observations on a dependent variable Y, and five independent variables (X_1, X_2, X_3, X_4, X_5).

TABLE 10
Ranked u-values for Fisher Data

.0001	.0870	.3685	.6630
.0107	.1419	.4275	.7599
.0163	.1532	.4384	.7848
.0414	.2273	.5188	.9046
.0635	.2275	.5899	.9640
.0636	.2702	.5946	.9842
.0768	.2940	.6403	
$U_{MOD}^2 = 0.117$		$p_4^2 = 12.16$

FIG. 7: FISHER DATA. REGRESSION ANALYSIS

Ostle (1954), p. 220, also gives these data and uses them in an example to illustrate a standard least squares analysis of a normal multiple linear regression model. We have applied the transformations of (6.43) to these data in the order given by Ostle. Here $n = 25$ and the number of independent variables is 5, so the number of u-values is $25 - 7 = 18$.

TABLE 11
u-values for Wallace-Snedecor Data

.029	.266	.687	.933
.034	.350	.717	.961
.106	.534	.819	.992
.200	.603	.826	
.262	.641	.886	
$U_{MOD}^2 = .063$		$p_4^2 = 2.389$	

The values of $U_{MOD}^2 = .063$ and $p_4^2 = 2.389$ are much less than the upper 10 per cent points for these statistics, and Figure 8 also implies that the normal linear regression model fits this data quite well.

Snedecor-Cochran Data

Snedecor and Cochran (1967), p. 140, give the initial weight x and the gain in weight y of 15 female rats on a high protein diet, from the 24th to 84th day of age and suggest considering a normal linear regression model. The ranked u-values for this data are given in Table 12, and are plotted against expected values in Figure 9.

TABLE 12
u-values for Snedecor-Cochran Data

.057	.147	.211	.453
.101	.179	.258	.498
.138	.201	.429	.631
$U_{MOD}^2 = 0.225$		$p_4^2 = 9.464$

FIG. 8: WALLACE-SNEDECOR DATA. REGRESSION ANALYSIS

The number of observations here, 15, is rather small and gives only 12 transformed u-values for testing the model. With so little data only rather large violations of the model will be detected. Recall that for this small number of u's that many tests are biased for some alternatives. The U_{MOD}^2 statistic is unbiased for all known cases and is our preferred test statistic here. The observed value of $U_{MOD}^2 = .225$ here falls between the upper 5 and 1 percent values, and $P(p_{14}^2 > 9.464) \approx .05$. The pattern of points in Figure 9 also raises doubts about the adequacy of fitting this data with a normal simple linear regression model. We conclude, even with so few observations, that the normal simple linear regression model assumptions are unwarranted.

FIG. 9: SNEDECOR-COCHRAN DATA. REGRESSION ANALYSIS

6.6.6 Further Examples in the Literature

Some of the transformations of section 6.5 have been used in numerical examples in the literature. Quesenberry, Whitaker and Dickens (1976) used the transformations (6.33) to study the normality of eleven samples of peanut aflatoxin data. Quesenberry, Giesbrecht and Burns (1983) used the transformations (6.33) and (6.48) to study the analysis of variance model assumptions for small samples (size 4) of tall fescue obtained as part of a forage management study. The variable considered was neutral detergent fiber.

The CPIT transformations for multivariate normal models were given by Rincon-Gallardo, Quesenberry, and O'Reilly (1979). Applications of these formulae for multiple multivariate samples were considered by Rincon-Gallardo and Quesenberry (1982).

(Author's note: This chapter was written in 1977 and revised in 1984.)

BIBLIOGRAPHY

- Anderson, R. L. and Bancroft, T. A. (1952). Statistical Theory in Research. McGraw-Hill, New York.
- Bartlett, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis, Ann. Math. Stat. 22, 107-11.
- Basu, D. (1955, 1960). On statistics independent of a complete sufficient statistic. Sankhyā 15, 337-80 (1955); 20, 223-6.
- Bliss, C. J. (1946). Collaborative comparison of three ratios for the chick assay of vitamin D. J. Assoc. Off. Agr. Chemists 29, 396-408.
- Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of the regression relationship over time. J. R. Statist. Soc. B37, 149-92.
- Chan, T. F., Golub, G. H. and LeVeque, R. J. (1983). Algorithms for computing the sample variance: analysis and recommendations. American Statist. 37(3), 242-7.
- Cox, D. R. (1961). Tests of separate families of hypotheses. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1, Berkeley, University of California Press, 105-23.
- Csörgö, M. and Seshadri, V. (1970). On the problem of replacing composite hypotheses by equivalent simple ones. Rev. Int. Statist. Inst. 38, 351-68.
- Csörgö, M. and Seshadri, V. (1971). Characterizing the Gaussian and exponential laws via mappings onto the unit interval. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 18, 333-39.
- Csörgö, M., Seshadri, V. and Yalovsky, M. (1973). Some exact tests for normality in the presence of unknown parameters. J. Roy. Statist. Soc. B 35, 507-22.
- David, F. N. and Johnson, N. L. (1948). The probability integral transformation when parameters are estimated from the sample. Biometrika 35, 182-92.
- Durbin, J. (1961). Some methods of constructing exact tests Biometrika 40, 41-55.
- Dyer, A. R. (1974). Hypothesis testing procedures for separate families of hypotheses. Jour. Am. Stat. Assn. 69, 140-5.

- Fisher, R. A. (1930). Inverse Probability. Proceedings of the Cambridge Philosophical Society 36(4), 528-35.
- Fisher, R. A. (1932). Statistical Methods for Research Workers. 4th Ed., Oliver and Boyd, Ltd.
- Galpin, J. S. and Hawkins, D. M. (1984). The use of recursive residuals in checking model fit in linear regression. American Statist. 38(2), 94-105.
- Hedayat, A. and Robson, D. S. (1970). Independent Stepwise Residuals for Testing Homoscedasticity. Jour. Am. Stat. Assn. 65, 1573-81.
- Hester, R. A., Jr. and Quesenberry, C. P. (1984). Analyzing uniform residuals for heteroscedasticity. Institute of Statistics Mimeo Series No. 1639, North Carolina State University.
- Miller, F. L., Jr. and Quesenberry, C. P. (1979). Power studies of some tests for uniformity, II. Commun. Statist. B8(3), 271-290.
- Neyman, Jerzy (1937). 'Smooth' test for goodness of fit. Skandinavisk Aktuarietidskrift 20, 149-99.
- O'Reilly, F. and Quesenberry, C. P. (1973). The conditional probability integral transformation and applications to obtain composite chi-square goodness of fit tests, Ann. Statist. 1, 74-83.
- O'Reilly, F. J. and Stephens, M. A. (1984 p. 34MS). Characterizations of goodness of fit tests. J. R. Statist. Soc. 44(3), 353-60.
- Ostle, B. (1954). Statistics in Research. Iowa State University Press, Ames, Iowa.
- Plackett, R. L. (1950). Some theorems in least squares, Biometrika 37, 149-57.
- Quesenberry, C. P. (1975). Transforming samples from truncation parameter distributions to uniformity. Comm. in Statist. 4, 1149-55.
- Quesenberry, C. P. and Dietz, E. J. (1983). Agreement probabilities for some CPIT-Neyman smooth tests. J. Statist. Comput. Simul. 17, 125-31.
- Quesenberry, C. P., Giesbrecht, F. G. and Burns, J. C. (1983). Some methods for studying the validity of model assumptions for multiple samples. Biometrics 39, 735-9.

- Quesenberry, C. P. and Hales, C. (1980). Concentration Bands for Uniformity Plots. J. Statist. Comput. Simul. 11, 41-53.
- Quesenberry, C. P. and Miller, F. L., Jr. (1977). Power studies of some tests for uniformity. J. Statist. Comput. Simul. 5, 169-91.
- Quesenberry, C. P. and Starbuck, R. R. (1976). On optimal tests for separate hypotheses and conditional probability integral transformations. Comm. in Statist. A5(6), 507-24.
- Quesenberry, C. P., Whitaker, T. B. and Dickens, J. W. (1976). On testing normality using several samples: an analysis of peanut aflatoxin data. Biometrics 32(4), 753-9.
- Rincon-Gallardo, S. and Quesenberry, C. P. (1982). Testing multivariate normality using several samples: applications techniques. Commun. Statist. - Theor. Meth. 11(4), 343-58.
- Rincon-Gallardo, S., Quesenberry, C. P. and O'Reilly, F. J. (1979). Conditional probability integral transformations and goodness-of-fit tests for multivariate normal distributions. Ann. Statist. 7(5), 1052-57.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. Ann. Math. Stat. 23, 470-72.
- Sarkadi, K. (1960). On testing for normality. Publications of the Mathematical Institute of the Hungarian Academy of Sciences. Vol. 3, 269-75.
- Sarkadi, K. (1965). On testing for normality. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, University of California Press, 373-87.
- Seheult, A. H. and Quesenberry, C. P. (1971). On unbiased estimation of density functions. Ann. Math. Stat. 42, 1434-38.
- Seshadri, V., Csörgö, M. and Stephens, M. A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. J. Roy. Statist. Soc. B 31, 499-509.
- Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods. 6th Ed., Iowa State University Press, Ames.
- Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. J. R. Statist. Soc. B 32, 115-22.
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. Jour. Am. Stat. Assn. 69, 730-7.

- Störmer, Horand (1964). Ein test zum erkennen von Normalverteilungen. Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete 2, 420-28.
- Wallace, H. A. and Snedecor, G. W. (1930). Correlation and machine calculation. (Revised edition) Iowa State College Official Publication, Vol. 30, No. 4.
- Watson, G. S. (1961). Goodness-of-fit tests on a circle. Biometrika 48, 109-14.
- Youngs, E. A. and Cramer, E. M. (1971). Some results relevant to choice of sum and sum-of-product algorithms. Technometrics, 13, 657-65.
- Zacks, S. (1971). The Theory of Statistical Inference. John Wiley and Sons, New York.