DEPARTMENT OF STATISTICS
University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 85
September 1966

ANALYSIS OF OUTLIERS WITH ADJUSTED
RESIDUALS
by
G. C. Tiao and Irwin Guttman

This research was supported in part by the National Science Foundation
under grant GS-549 and the Wisconsin Alumni Research Foundation.
ANALYSIS OF OUTLIERS WITH ADJUSTED RESIDUALS

by

G. C. Tiao and Irwin Guttman

University of Wisconsin

Summary

Many statistical procedures designed to guard against the occurrence of outliers or spurious observations in normal theory are based upon examining the magnitude of the residuals. A major difficulty involved is caused by the fact that the residuals are correlated. It is shown that one way to avoid such difficulty is to adjust the residuals using information from an auxiliary experiment so that the adjusted residuals become uncorrelated. For the problem of making inferences about the unknown mean of a normal population \( N(\mu, \sigma^2) \) with known \( \sigma^2 \), this leads to a set of estimation procedures by which the observation(s) associated with the largest adjusted residual(s) in magnitude will be excluded. Certain properties of the procedures are discussed and exact numerical results are given for the cases of one and two spurious observations. Generalization to the case of unknown variance and to the general linear model is also given.
1. Introduction

In a paper (1960), Anscombe proposed a set of rules for rejecting outlying observations when sampling from normal populations. Such rules are based essentially on examination of the magnitude of the residuals. The main feature of his analysis is the investigation of the effect of such rules on the sampling properties of the corresponding estimator(s), in contrast to the usual hypothesis testing formulation which is concerned principally with the probabilities of making "right" or "wrong" decisions. This led to the concepts of the "premium" and the "protection" which are measures of the effect of a rejection rule on the mean square error (M.S.E.) of an estimator of the parameter of interest. More specifically, the premium associated with a rejection rule measures the inflation of the M.S.E. when in fact all observations are from a homogeneous source, while protection measures the reduction in M.S.E. if otherwise.

While this approach to the problem seems relevant and appealing, practical application of the resulting procedures has led to exceedingly complicated computational problems. It is because of these difficulties that, even for a simple rule designed to guard against one spurious
observation when sampling from a normal population with known variance, exact results are only known for sample sizes less than five.

In this paper, we propose some procedures for analysing outliers which are also based upon residuals but utilize information obtained from an auxiliary randomized experiment. In so doing, we are able to give exact results for any sample size of the simple situations mentioned above. In addition, we also obtain results for more complex situations.

2. Sampling from a normal population with known variance:

One spurious observation

In this section, we discuss the situation of sampling from a normal population \( N(\mu, \sigma^2) \) with \( \sigma^2 \) known, and propose a rule designed to guard against the possibility that one of the observations might be spurious.

Let \( y_1, \ldots, y_n \) be the observations. When an investigator is concerned with the possibility of spurious observations, it is a common practice to base the analysis on the magnitude of the residuals \( y_i - \bar{y} \), \( i = 1, \ldots, n \), and discard those observations whose residuals are inordinately "large", say more than three or four times the standard deviation \( \sigma \). It is then important to investigate the effect of such practice on the associated estimation procedures of the unknown parameter \( \mu \). In the work of Anscombe previously cited, when guarding against the possibility of one spurious observation, he proposed the following estimation procedure for \( \mu \),
\( \hat{\mu} = \bar{y} - \frac{T}{n-1} \)  \hspace{1cm} (2.1)

with
\[
T = \begin{cases} 
0 & \text{if } |y_i - \bar{y}| < d \quad \text{all } i \\
y_1 - \bar{y} & \text{if } |y_1 - \bar{y}| > d, \quad |y_j - \bar{y}| < |y_1 - \bar{y}| \quad j \neq 1
\end{cases}
\]

and where \( d \) is some appropriately chosen constant, and then initiated an investigation of the premium and protection of this procedure. In general, letting \( \mu_p \) be the estimator of \( \mu \) based upon a certain rejection procedure, say \( p \), the premium is defined as

\[
p_m = \frac{\text{M.S.E.}(\hat{\mu}_p) - \text{M.S.E.} (\bar{y})}{\text{M.S.E.} (\bar{y})} \]  \hspace{1cm} (2.2)

when in fact all observations are from the homogeneous source \( N(\mu, \sigma^2) \) — hereafter referred to as the null case. That is, premium is the price one pays as measured by the inflation in the M.S.E. when the null case holds. Protection, on the other hand, is defined as

\[
\text{Prot.} = \frac{(\text{M.S.E.} (\bar{y}) - \text{M.S.E.} (\hat{\mu}_p))/\text{M.S.E.} (\bar{y})}{\text{M.S.E.} (\bar{y})} \]  \hspace{1cm} (2.3)

when one or more of the observations are in fact spurious. Thus, protection reflects the gain by applying a rule as measured by the decrease in the M.S.E. when one or more observations are spurious.

While conceptually relevant and appealing, numerical computation of the premium turns out to be quite difficult, chiefly due to the fact that the residuals are correlated. The situation is even more complicated for the protection.

This difficulty has motivated us to seek for modifications of the above procedure. It is well known that in the null case, the covariance
matrix of the residuals is $\sigma^2(I_n - \frac{1}{n} \frac{1}{n} \frac{1}{n})$. Suppose we adjust each residual by adding a common quantity $\frac{\sigma}{\sqrt{n}} u$, where $u$ is an independent observation from $N(0, 1)$ — for example $u$ might be obtained from a table of random normal deviates. It is then easy to see that the adjusted residuals

$$z_i = y_i - \bar{y} + \frac{\sigma}{\sqrt{n}} u_i, \quad i = 1, \ldots, n \quad (2.4)$$

are independent $N(0, \sigma^2)$ variables. We may now use as an estimator of $\mu$,

$$\hat{\mu} = \bar{y} - W_1$$

with

$$W_1 = \begin{cases} 0 & \text{if } |z_i| < c \quad \text{all } i \\ \frac{y_i - \bar{y}}{n-1} & \text{if } |z_i| > c \quad |z_i| < |z_1| \\ & i = 1, \ldots, n; \quad j \neq i. \end{cases}$$

where $c$ is some positive constant, the choice of which will be discussed below. It is to be noted that the estimator $\hat{\mu}$ is of precisely the same form as that in (2.1), i.e. $\hat{\mu}$ is either the sample mean $\bar{y}$ if there is no rejection or if the observation $y_1$ is rejected, the sample mean of the remaining $(n-1)$ observations. The difference is of course that we base rejection on the adjusted residuals rather than the residuals themselves. For moderate sample sizes (say $n > 10$) the effect of the adjustment is expected to be negligible since the standard deviation of the residuals is $(\frac{n-1}{n})^{\frac{1}{2}}$ while that of the adjustment is $n^{-\frac{1}{2}}$.

We remark here that the idea of adjusting residuals from normal samples in the above manner has been used by others in different contexts — e.g. Durbin (1961).
Premium of the procedure

We now investigate the properties of the procedure in (2.5) when the null case holds. Since \( \sigma^2 \) is assumed known, with no loss of generality we shall let \( \sigma^2 = 1 \). Thus, the \( z_i \) are independent \( N(0,1) \) variables so that the expectation of the estimator \( \hat{\mu} \) is

\[
E(\hat{\mu}) = \mu - E(W_1) \tag{2.6}
\]

From the definition of \( W_1 \) in (2.5) and noting that

\[
y_i - \bar{y} = z_i - \frac{u}{\sqrt{n}} = z_i - \bar{z} \tag{2.7}
\]

it is easily shown that \( E(W_1) = 0 \), so that \( \hat{\mu} \) is unbiased for \( \mu \). Consequently using the definition in (2.2) the premium of our procedure becomes

\[
p_m = \frac{\text{Var}(\hat{\mu}) - \text{Var}(\bar{y})}{\text{Var}(y)} = nE(W_1^2) \tag{2.8}
\]

where use is made of the fact that \( \bar{y} \) is independent of \( z_1 \).

From (2.4) and (2.5)

\[
E(W_1^2) = \frac{1}{(n-1)^2} \sum_{i=1}^{n} E \left[ (z_i - \bar{z})^2 \mid |z_i| > c, |z_j| < |z_i|, j \neq i \right] \tag{2.9}
\]

and since the \( z \)'s are identically distributed,

\[
E(W_1^2) = \frac{n}{(n-1)^2} E[(z_1 - \bar{z})^2 \mid |z_1| > c, |z_j| < |z_1|, j=2,\ldots,n] \tag{2.10}
\]

It is to be noted that by \( E(g(z) \mid z \in S) \) we mean \( \int_{z \in S} g(z)p(z)dz \), and
This should not be confused with conditional expectation. We can now write

\[(z_1 - \bar{z})^2 = (\frac{n-1}{n})^2 z_1^2 + \frac{1}{n} \sum_{j=2}^{n} z_j^2 + \frac{2}{n^2} \sum_{j=1}^{n} \sum_{j=1}^{n} (z_j - z_m)^2 - \frac{2(n-1)}{n^2} z_1 \sum_{j=2}^{n} z_j \]

(2.11)

On taking expectation, all cross product terms vanish and we have

\[E((z_1 - \bar{z})^2 | \cdot) = (\frac{n-1}{n})^2 A + \frac{n-1}{n^2} B \]

with

\[A = 2 \int_{\infty}^{c} z^2 \phi(z) b(z)^{n-1} dz \]

(2.12)

\[B = \frac{1}{n} [1 - b(c)^n] - 4 \int_{c}^{\infty} z \phi^2(z) b(z)^{n-2} dz \]

where

(1) \ \phi(z) is the density function of an \textit{N}(0, 1) variable

(2) \ \Phi(z) is the cumulative distribution of an \textit{N}(0, 1) variable

and

(3) \ b(z) = [2 \Phi(z) - 1]

Making use of the following two reduction formulae which can be verified by integrating by parts,

\[2 \int_{\theta}^{\infty} z^2 \phi(z) b(z)^{q} dz = 2 \theta \phi(\theta) b(\theta)^{q} + \frac{1}{q+1} [1 - b(\theta)^{q+1}] \]

+ \[4q \int_{\theta}^{\infty} z \phi^2(z) b(z)^{q-1} dz \]

(2.13)

\[4 \int_{\theta}^{\infty} z \phi^2(z) b(z)^{q} dz = \frac{\theta^2}{\pi} - b(\theta)^{q} + 2q \int_{\theta}^{\infty} e^{-z^2} \phi(z) b(z)^{q-1} dz \]

\[\theta > 0, \quad q \geq 1, \]
we obtain

\[ p_m = 2 \phi(c) b(c)^{n-1} + \frac{1}{n-1} \frac{n(n-2)}{(n-1)} e^{-c^2} b(c)^{n-2} \]

\[ + \frac{2n(n-2)}{(n-1)^2} \int_0^\infty e^{-z^2} \phi(z) b(z)^{n-3} \, dz. \]  

(2.14)

Thus, for given \((c, n)\), the corresponding premium may be determined. Alternatively for a given premium and \(n\) we may obtain the necessary \(c\) by iteration. Table 1 gives the values of \(c\) for a combination of values of \(n\) and the premium.

Use of the procedure

We now provide a short summary indicating how the procedure is to be applied. Suppose that the investigator takes a sample of 16 observations and is willing to pay a premium of 2%. From Table 1, \(c = 3.134\). After obtaining the observations \(y_1, \ldots, y_{16}\), he calculates the residuals \(y_i - \bar{y}\). The investigator may then make a random drawing from a table of unit normal deviates to obtain the adjustment \(u/\sqrt{n} = u/4\). After adding \(u/4\) to \(y_i - \bar{y}\) to get the quantities \(z_i\), he then examines \(|z_i|\) to see whether any of them exceeds \(c = 3.134\). If none of them exceeds 3.134, he simply uses \(\bar{y}\) to estimate \(\mu\). If one or more of the \(|z_i|\) exceed 3.134, he discards the observation corresponding to the largest \(|z_i|\) and uses as the estimator the mean of the remaining 15 observations.
Bounds on the premium

In some instances, the investigator may have a preferred value of \( c \) and wish to find the corresponding premium. The first three terms on the right of (2.14) can be readily determined from a table of normal densities and integrals, while the last term must be evaluated by numerical methods (except for \( n = 3 \)). In practice, however, exact evaluation of the last term is rarely needed because useful bounds on the premium may be readily obtained.

Since the last term in (2.14) is positive, a lower bound for the premium is immediately available by ignoring it. Further, since \( e^{-c^2} > e^{-z^2} \) for \( z > c \), we may take \( e^{-z^2} \) as \( e^{-c^2} \) outside the integral sign and evaluate the resulting integral to obtain an upper bound.

Thus, we have

\[
\frac{n(n-2)}{(n-1)} e^{-c^2} b(c)^{n-2} < p_m - \left\{ 2c \phi(c)b(c)^{n-1} + \frac{1}{n-1} [1 - b(c)^n] \right\} \\
< \frac{n(n-2)}{(n-1)} \frac{e^{-c^2}}{\pi}.
\]

(2.15)

While for fixed \( c \) the difference of the extreme ends of (2.15) increases as \( n \) increases, we find that for the range of values of \( p_m \) and \( n \) used in Table 1, these limits are extremely close for any practical purpose.

For example, if \( c = 2.796551 \) and \( n = 7 \), the exact premium is 0.05 and the upper and lower bounds are 0.000726 and 0.000745 respectively, while if \( c = 2.755921 \) with \( n = 40 \), the premium is again 0.05 with the bounds now equal to 0.004992 and 0.006739. Thus, even in extreme
Thus, for given $c$, $\alpha$ is increasing in $n$ and for given $n$, $\alpha$ is decreasing in $c$.

Suppose $c$ is held constant. Since the rule rejects at most one observation, one would then certainly expect to pay a smaller "price", i.e. premium, as $n$ is increased. This is indeed the case. For example, suppose $p_m = 0.02$, $n = 6$ and from Table 1, $c = 3.141$ so that $\alpha = 0.0100604$. If we now let $n$ increase to 7 but keep the rejection probability constant at $\alpha = 0.0100604$, then from (2.16)

$$c = \Phi^{-1} \left[ \frac{1}{2} + \frac{1}{2} (1 - 0.0100604)^{\frac{1}{4}} \right] = 3.186$$

which exceeds the $c$ value 3.138 in the table corresponding to $p_m = 0.02$ and $n = 7$. Since $p_m$ is decreasing in $c$ for given sample size, this implies that the premium at $n = 7$, $\alpha = 0.0100604$ is less than 0.02.

On the other hand, if we kept the premium constant, then the probability of rejection $\alpha$ would certainly be expected to be increasing in $n$. This is exhibited by the entries in every column of Table 1 which were calculated by holding $p_m$ fixed. The decrease in the values of $c$ as $n$ increases simply implies that

$$\alpha(p_m, n+1) - \alpha(p_m, n) = [2 \Phi(c(p_m, n)) - 1]^n - [2 \Phi(c(p_m, n)) - 1]^{n+1} > 0$$

(2.17)

where $\alpha(p_m, n)$ and $c(p_m, n)$ are respectively the values of $\alpha$ and $c$ corresponding to the given values of $p_m$ and $n$. 
Maximum premium

Since for given sample size, the premium is monotonically decreasing in \( c \), it follows that the maximum premium — or the highest price we pay — is the value of \( p_m \) at \( c = 0 \). This corresponds to a rejection with certainty \( (\alpha = 0.0) \), namely we reject with certainty the observation corresponding to the largest \( z_1 \) in magnitude. In this case,

\[
p_m = \frac{1}{n-1} + \frac{2n(n-2)^2}{(n-1)\pi} \int_0^\infty e^{-z^2} \phi(z) b(z)^{n-3} \, dz \quad (2.18)
\]

The entries in the last column of Table 2 give the maximum premium for various \( n \). As expected, it decreases as \( n \) increases. In particular it is less than 0.1 for \( n \) beyond 80 so that it would, for example, be absurd to think of paying a premium of 10% or more for \( n \) larger than 80.

The maximum premium in (2.18) is the sum of two positive quantities the values of which are given in the first two columns of Table 2 for the \( n \) considered. The first term, \( \frac{1}{n-1} \), is precisely the premium corresponding to a procedure in which we randomly discard one of the \( n \) observations. Table 2 therefore shows the premium corresponding to discarding with certainty the observation associated with the largest \( |z_1| \) is substantially larger than that of a random rejection. Even for \( c > 0 \), the premium of the procedure in (2.5) can still exceed that of the random procedure. For instance, with \( n = 30 \) and \( c = 2.77 \), from Table 1, \( p_m = 0.05 \) which is greater than \( \frac{1}{n-1} = \frac{1}{29} \approx 0.0345 \). It can in fact be shown that for every \( n \), there exists a \( \delta \).
such that for $c < \delta$, $p_m > \frac{1}{n-1}$. The above, of course, does not mean that in these situations, the procedure would be inferior to the random rejection procedure. In judging a rejection procedure, not only the premium, but also the protection should both be taken into consideration.

**Protection of the procedure**

Given the premium the investigator is willing to pay, it is only natural for him to inquire what protection is actually afforded by the procedure when there is indeed a spurious observation in the sample. This of course depends upon the nature of the spurious observation. To illustrate the idea, we shall discuss the case, which frequently occurs in practice, that the spurious observation is caused by an exogenous shock to the system inducing a shift in mean. Specifically, then, we assume that the spurious observation is from $N(\mu+a,1)$.

With no loss of generality, we may assume that $y_1$ is from $N(\mu+a,1)$ and the remaining observations from $N(\mu,1)$ (because the rejection rule makes no use of such information). Under this assumption, it is easily seen that $\bar{y}$ and the $z_i$ are all independent with $\bar{y} \sim N(\mu+a, \frac{1}{n})$, $z_1 \sim N(\frac{n-1}{n}a, 1)$ and $z_j \sim N(-\frac{a}{n}, 1)$, $j = 2, \ldots, n$.

Thus, in (2.3), M.S.E. ($\bar{y}$) = $\frac{1}{n} + \left(\frac{a}{n}\right)^2$ so that

$$
\text{Prot} = \frac{\frac{1}{n} + \left(\frac{a}{n}\right)^2 - E(\bar{y} - W_1 - \mu)^2}{\frac{1}{n} + \left(\frac{a}{n}\right)^2} = \frac{\frac{a^2}{n} - \frac{1}{n^2} E(nW_1 - a)^2}{\frac{1}{n} + \left(\frac{a}{n}\right)^2} \quad (2.19)
$$

We can now write
\[ E(nW_1 - a)^2 = E[(nW_1 - a)^2 \mid |z_1| < c, \text{ all } 1] + E[(nW_1 - a)^2 \mid |z_1| > c, \mid z_1 \mid j \neq 1] + (n-1)E[(nW_1 - a)^2 \mid |z_2| > c, \mid z_1 \mid j \neq 2] \]  
(2.20)

Making use of the two identities

\[ \int_{-\infty}^{\infty} t \phi(t+\kappa)dt = \phi(x-\kappa) - \phi(x+\kappa) - \kappa [ \Phi(x+\kappa) - \Phi(-x+\kappa)] \]  
(2.21)

and

\[ \int_{-\infty}^{\infty} t^2 \phi(t+\kappa)dt = (\kappa - x) \phi(x+\kappa) - (x+\kappa) \phi(x-\kappa) + (1+\kappa^2) [ \Phi(x+\kappa) - \Phi(-x+\kappa)] \]

and after some tedious algebraic reduction, we obtain

\[ E(nW_1 - a)^2 = a^2 b(c, f_1) \cdot b(c, f_2) + \int_{C}^{\infty} g(z, a) dz \]  
(2.22)

with

\[ g(z, a) = m_1 + \frac{1}{n-1} m_2 + (1 - \frac{1}{n-1})(m_3 + m_4 + 2m_5 + (n-3)m_6) - 2[m_7 - m_8 - (n-2)m_9] + (n-1)m_{10} + m_{11}, \]

\[ m_1 = [(z^2 + a^2) \gamma_2(z, f_2) - 2az \gamma_1(z, f_2)] b(z, f_1)^{n-1} \]

\[ m_2 = \gamma_2(z, f_2)[f_1 \gamma_1(z, f_1) - z \gamma_2(z, f_1) + (1+f_1^2) b(z, f_1)] b(z, f_1)^{n-2} \]

\[ m_3 = \gamma_2(z, f_2)[\gamma_1(z, f_1) - f_1 b(z, f_1)]^2 b(z, f_1)^{n-3} \]

\[ m_4 = \gamma_2(z, f_1) b(z, f_2)[(1+f_1^2) b(z, f_1) + f_1 \gamma_1(z, f_1) - z \gamma_2(z, f_1)] b(z, f_1)^{n-3} \]
\[ m_5 = \gamma_2(z, f_1)[\gamma_1(z, f_2) + f_2 b(z, f_2)](\gamma_1(z, f_1) + f_1 b(z, f_1))b(z, f_1)^{n-3} \]

\[ m_6 = [\gamma_1(z, f_1) + f_1 b(z, f_1)]^{2}\gamma_2(z, f_1)b(z, f_2)b(z, f_1)^{n-4} \]

\[ m_7 = [z\gamma_1(z, f_2) - a\gamma_2(z, f_2)](\gamma_1(z, f_1) + f_1 b(z, f_1))b(z, f_1)^{n-2} \]

\[ m_8 = [z\gamma_1(z, f_1) - a\gamma_2(z, f_1)](\gamma_1(z, f_1) + f_1 b(z, f_1))b(z, f_2)b(z, f_1)^{n-2} \]

\[ m_9 = [z\gamma_1(z, f_1) - a\gamma_2(z, f_1)](\gamma_1(z, f_1) + f_1 b(z, f_1))b(z, f_2)b(z, f_1)^{n-3} \]

\[ m_{10} = [(z^2 + a^2)\gamma_2(z, f_1) - 2az\gamma_1(z, f_1)]b(z, f_2)b(z, f_1)^{n-2} \]

\[ m_{11} = [(1 + f_2^2)b(z, f_2) + f_2\gamma_1(z, f_2) - z\gamma_2(z, f_2)]\gamma_2(z, f_1)b(z, f_1)^{n-2} \]

where \[ f_1 = \frac{a}{n}, \quad f_2 = \frac{a}{n} - a, \quad b(z, f) = \Phi(z+f) - \Phi(-z+f) \]

\[ \gamma_1(z, f) = \phi(z+f) - \phi(z-f) \text{ and } \gamma_2(z, f) = \phi(z+f) + \phi(z-f). \]

In the special case \( a = 0 \), we have that \( f_1 = f_2 = 0 \), \( b(z, 0) = b(z) \), \( \gamma_1(z, 0) = 0 \) and \( \gamma_2(z, 0) = 2\phi(z) \), from which it can be verified that \( \text{Prot} = -p_m \) as it should be. For any \( n \), all terms in the integrand \( g(z, a) \) are polynomials in \( z \), \( \phi \) and \( \Phi \) and can be easily incorporated into numerical integration routines for calculation on a high speed computer. Table 3 gives the protection corresponding to 0.01 and 0.05 premiums for a combination of values of \( (n, a) \).

Since \( \text{Prot} = -p_m \) when \( a = 0 \), Table 3 shows that, as "a" increases from zero the protection actually decreases at first and then increases monotonically when "a" becomes large. For all cases considered, the increase is the sharpest in the range \( 2 < a < 6 \) and tapers off gradually afterwards. If we think of the break even point...
as the value of "a" which makes the protection just exceed the premium, then both for the 0.01 and 0.05 premiums, such a point occurs in 2<"a"<3 for all the sample sizes considered. This suggests that under the assumption made, the procedure in (2.5) is useful only for large "a", say at least three or four times the standard deviation of the observations as large.

The entries in the table also show that, for fixed "a" (and "a" ≥ 3), the protection decreases as n becomes large. This is intuitively reasonable because when n is large the effect of a spurious observation on the inferences about µ becomes smaller and smaller in any case. Imagine that we have a "crystal ball" procedure which allows us to reject the spurious observation with certainty. The protection is 1-(1-n^{-1})(1+\frac{a^2}{n})^{-1} which is decreasing in n for n>2a^2/(a^2 - 1).

Finally, by comparing the entries in Tables 3(i) and (ii), we note that for a ≥ 8, the protections are almost identical for the two premiums considered.

3. Generalization to several spurious observations

We have shown in the above that using the adjusted residuals one may obtain exact results for any sample size for a procedure which is designed to guard against the possibility of one spurious observation. When we have a moderately sized sample, say n ≥ 20, however, one may wish to protect against the occurrence of more than one spurious observation. In this section, we discuss how the framework of the preceding section may be generalized.
Intuitively, candidates for rejection are the observations corresponding to "large" values of residuals in magnitude. As indicated before, in our procedure one should then reject those observations associated with large absolute values of the adjusted residuals $z_i$. Suppose we wish then to guard against the possibility of $\kappa (\kappa \geq 2$ but small compared with $n$) spurious observations in the sample. We may choose as our estimator

$$\hat{\mu} = \bar{y} - W_{\kappa}$$

with

$$W_{\kappa} = \begin{cases} 
0 & \text{if } |z_{i_1}| < c \quad i = 1, \ldots, n \\
\frac{y_{i_1} + \ldots + y_{i_m} - m\bar{y}}{n-m} & \text{if } |z_{i_j}| > c \text{ and } |z_{i_{m+t}}| < c \\
& j = 1, \ldots, m; \ t = 1, \ldots, n-m \text{ and } m=1, \ldots, (\kappa-1) \\
\frac{y_{i_1} + \ldots + y_{i_\kappa} - \kappa\bar{y}}{n-\kappa} & \text{if } |z_{i_j}| > c \text{ and } |z_{i_{\kappa+t}}| < \min_{j} |z_{i_j}| \\
& j = 1, \ldots, \kappa; \ t = 1, \ldots, n-\kappa
\end{cases}$$

where $(i_1, \ldots, i_n)$ is any $s$-permutation of the integers $(1, \ldots, n)$. In other words, this procedure says that we reject at most $\kappa$ observations and use as the estimator for $\mu$ the sample mean of the remaining ones.

We now discuss certain properties of this procedure following a line of attack similar to that given earlier.

Under the distributional assumptions made in section 2, it is easy to show that in the null case, $\hat{\mu}$ is unbiased for $\mu$ so that the
premium \( p_m(\kappa) = nE(W_k^2) \). Because of the symmetry of the distributional properties of the \( z_i \), we can write

\[
E(W_k^2) = \frac{\kappa - 1}{(n-m)^2} \left[ \sum_{m=1}^{n-m} \frac{m}{(n-m)^2} E[\sum_{i=1}^{m} (z_i - \bar{z})^2 | S_m] + \frac{m}{(n-m)^2} E[\sum_{i=1}^{m} (z_i - \bar{z})^2 | S_k] \right] (3.2)
\]

where \( S_m \) denotes the region \( \{|z_i| > c, i=1, \ldots, m, \text{ and } |z_{m+t}| < c, t=1, \ldots, n-m\} \) and \( S_k \) the region \( \{|z_i| > c, i=1, \ldots, k \text{ and } |z_{k+t}| < \min_{i} |z_i|, t=1, \ldots, n-k\} \). Now,

\[
\left( \sum_{i=1}^{m} (z_i - \bar{z})^2 \right)^2 = \left( \frac{n-m}{n} \right)^2 \sum_{i=1}^{m} z_i^2 + \frac{m^2}{n^2} \sum_{t=1}^{n-m} z_{m+t}^2 + \text{cross products} (3.3)
\]

On taking expectation over the region \( S_m \), all cross product terms vanish and

\[
E[\sum_{i=1}^{m} (z_i - \bar{z})^2 | S_m] = \left( \frac{n-m}{n} \right)^2 m b(c)^{n-m} (1-b(c))^{m-1} [2c\phi(c) + 1-b(c)]
\]
\[
+ \frac{(n-m)m^2}{n^2} b(c)^{n-m-1} (1-b(c))^{m-1} [b(c) - 2c\phi(c)] (3.4)
\]

For the second term on the right of (3.2), we set \( m=\kappa \) in (3.3) and integrate the resulting quantity over the region \( S_k \) to obtain

\[
E[\sum_{i=1}^{\kappa} (z_i - \bar{z})^2 | S_k] = \left( \frac{n-\kappa}{n} \right)^2 \kappa E(z_1^2 | S_k) + \frac{(n-\kappa)\kappa^2}{n^2} E(z_{\kappa+1}^2 | S_k) (3.5)
\]

where

\[
E(z_{\kappa+1}^2 | S_k) = 2\kappa \int_{C}^{\infty} \phi(z)(1-b(z))^{\kappa-1}b(z)^{n-\kappa-1} [b(z) - 2z\phi(z)] \, dz
\]
and
\[
E(z_1^2 | S_\kappa) = 2 \int_0^\infty z^2 \phi(z) [1 - b(z)]^{k-\kappa-1} b(z)^{n-\kappa} \, dz \\
+ \sum_{r=2}^\infty \frac{(k-1)1}{(r-2)1(k-r)!} 4 \int_0^\infty z^2 \phi(z) [1 - b(z)]^{k-r} b(z)^{r-2} b(t)^{n-\kappa-1} \phi(t) \, dt \, dz
\]

Substituting (3.4) and (3.5) into (3.2) and after a little algebraic simplification, we obtain

\[
P_m(\kappa) = [2c \phi(c) + 1 - b(c)] \left\{ \sum_{m=1}^{\kappa-1} \left[ \frac{\sum_{m=1}^{n-1} b(c)^{n-m} [1 - b(c)]^{m-1}}{\sum_{m=1}^{n-1} b(c)^{n-m} [1 - b(c)]^m} \right] \right\} \\
+ [b(c) - 2c \phi(c)] \left\{ \frac{\sum_{m=1}^{n-1} b(c)^{n-m} [1 - b(c)]^{m-1}}{\sum_{m=1}^{n-1} b(c)^{n-m} [1 - b(c)]^m} \right\} \\
+ \left( \sum_{m=1}^{\kappa-1} \frac{2c^2}{n-\kappa} \int_0^\infty \phi(z) [1 - b(z)]^{k-\kappa-1} b(z)^{n-\kappa-1} [b(z) - 2z \phi(z)] \, dz \right) \\
+ \left( \sum_{m=1}^{n-1} \frac{2c^2}{n-\kappa} \int_0^\infty z^2 \phi(z) [1 - b(z)]^{k-\kappa-1} b(z)^{n-\kappa-1} \, dz \right) \\
+ 2 \sum_{r=2}^{\infty} \frac{\kappa}{k-r} \int_0^\infty z^2 \phi(z) [1 - b(z)]^{k-r} b(z)^{n-\kappa-1} \, dz \\
- 2 \left( \sum_{r=2}^{\kappa} \frac{\kappa}{k-r} \right) \sum_{s=0}^{\infty} \left( \frac{\sum_{r=2}(r-2)(r-1)\frac{(-1)^s}{n+s-\kappa+1} b(c)^{n-\kappa+s+1}}{\sum_{r=2}^{\infty} \frac{\sum_{r=2}^\infty \phi(z) [1 - b(z)]^{k-r} b(z)^{r-2-s} \, dz} \right) \\
(3.6)
\]

Using the reduction formulae in (2.13), the premium given in (3.6) may be put in a more convenient form for numerical evaluation. In particular for \( \kappa=2 \), that is, for the rule in which we reject at most two observations, we have
\[ p_m(2) = \frac{2}{n-2} + \left[ 1 + \frac{1}{n-2} + \frac{1}{(n-1)(n-2)} \right] b(c)^n + \left[ \frac{1}{n-1} - \frac{4}{n-2} \right] + 2c \phi(c) \]

\[ + \frac{1}{n-1} - n+2 \right] b(c)^{n-1} + (n-1) \left( \frac{2c}{n-1} - \frac{1}{(n-1)^2} \right) - \frac{e^{-c^2}}{\pi} (n-2) \]

\[ + \left( 1 - \frac{4}{(n-2)^2} \right) b(c)^{n-2} + \frac{e^{-c^2}}{\pi} (n-1)(n-2) \left( 1 - \frac{4}{(n-2)^2} \right) b(c)^{n-3} \]

\[ + \frac{2(n-1)(n-2)}{\pi} \left( 1 - \frac{4}{(n-2)^2} \right) \int_c^\infty e^{-z^2} \phi(z) z^{-3} b(z) dz \]

\[ - (n-2) b(z)^{-n-3} \]

As in the case of a single rejection, it can be verified that

\[ \frac{\partial p_m(2)}{\partial c} < 0 \]

for \( c > 0 \) so that \( p_m(2) \) is monotonically decreasing in \( c \). Table 4 shows the values of \( c \) for a combination of values of \( p_m(2) \) and \( n \).

Similar to the results in Table 1, the change in the value of \( c \) here is again appreciable when the premium is changed. On the other hand, \( c \) is even less sensitive to changes in \( n \) than that for the previous procedure. Each of the entries in Table 4 is larger than the corresponding one in Table 1, implying that the present procedure leads to a higher probability of rejection \( \alpha \) for a fixed premium and a larger premium for a fixed \( \alpha \). This is of course as it should be since the procedure in (3.7) with \( \kappa = 2 \) provides for wider contingencies than those by (2.5).

In so far as the protection of the general \( \kappa \)-rule is concerned, that of course depends upon how one characterizes the nature of the
spurious observations. If these can be assumed to be from $N(\mu+a,1)$, then we could use an argument similar to that which led to (2.22) to obtain the corresponding formulae. The resulting expressions are exceedingly lengthy and shall not be given here.

We remark here in passing that while the results in this and the previous sections provide us with a set of procedures for guarding against the occurrence of one or more spurious observations, they do not tell us anything as to which specific one of the set should be adopted in any given situation. We have said that when the sample size increases, one may wish to protect against more than one spurious observation. This supposition can be interpreted as saying that a priori there is a small chance that each observation could be spurious so that as the number of observations taken is increased, the chances of getting one, two, three,..."bad" observations become more appreciable and wider contingencies should be taken into consideration in selecting the procedure. An analysis of the problem which gives explicit cognizance of such prior consideration is recently given by Box and Tiao (1966) in the Bayesian framework.

4. Some considerations of the general linear model

The method of adjusting residuals can be easily extended to handle the problem of rejecting spurious observations in the general linear model

$$Y = X\theta + \epsilon$$

(4.1)
where \( \mathbf{y} \) is a \( n \times 1 \) vector of observations, \( \mathbf{X} = \{ x_1, \ldots, x_p \} \) an \( n \times p \) matrix of fixed elements, \( \boldsymbol{\theta} \) a \( p \times 1 \) vector of unknown coefficients and \( \mathbf{\epsilon} \) a \( n \times 1 \) vector of disturbances. With no loss of generality, we assume that \( \mathbf{X} \) is of rank \( p \) and that \( \mathbf{X}' \mathbf{X} = \mathbf{D}_p \{ d_{jj} \} \), a \( p \times p \) positive definite diagonal matrix. It is well known that if \( \mathbf{\epsilon} \) has the \( \mathcal{N}(\mathbf{0}, \mathbf{I} \sigma^2) \) distribution, then the residuals \( \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \), where \( \hat{\mathbf{y}} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \), are jointly normal with zero means and covariance matrix

\[
E(\mathbf{e} \mathbf{e}') = \sigma^2 \{ I - \mathbf{X} \mathbf{D}_p^{-1} \mathbf{X}' \} = \sigma^2 \{ I - \sum_{j=1}^{p} \frac{1}{d_{jj}} \mathbf{x}_j \mathbf{x}_j' \} \tag{4.2}
\]

Thus, when \( \sigma^2 \) is known, we may adjust the residual vector \( \mathbf{e} \) such that

\[
\mathbf{z} = \mathbf{e} + \sigma \sum_{j=1}^{p} \frac{1}{\sqrt{d_{jj}}} \mathbf{x}_j u_j
\]

where \( u_j \) are independent drawings from a \( \mathcal{N}(0,1) \) population so that the elements of the vector \( \mathbf{z} \) are independent \( \mathcal{N}(0, \sigma^2) \) variables.

(In the special case that \( p = 1 \), \( x_1 \) is a column vector of ones, \( d_{11} = n \) and the \( z \)'s are precisely those given in section 2.) One may then formulate estimation procedures for \( \boldsymbol{\theta} \) based upon the magnitude of the \( z \)'s analogous to those previously discussed, i.e., we would use the ordinary least squares estimator \( \hat{\boldsymbol{\theta}} \) based on all the observations if no rejections, while if some observations are rejected, estimate \( \hat{\boldsymbol{\theta}} \) from the remaining observations in the usual manner.
5. Sampling from a normal population with unknown variance

One spurious observation

Up to this point we have discussed the situation in which \( \sigma^2 \) is assumed known. When this is not so, the problem becomes more involved in general. The reasons are that the residuals are correlated and also the usual estimators of \( \sigma^2 \) are not independent of the residuals.

However, there are certain situations in which the problem becomes tractable. One which we may investigate is when there is available an independent quantity \( v \) having the \( N(0, \sigma^2) \) distribution. For example, the experimenter might have available two independent observations from \( N(\eta, \sigma^2) \) so that their suitably normed difference will be a \( N(0, \sigma^2) \) variable. If concerned with the problem of estimating a single mean \( \mu \) of a normal population, we may then construct adjusted residuals \( z_i = y_i - \bar{y} + v/\sqrt{n} \), \( i=1, \ldots, n \), so that in the null situations \( z \)'s are independent \( N(0, \sigma^2) \) variables.

In what follows we discuss two cases: (i) there is available an independent estimate \( s_i^2 \) of \( \sigma^2 \) following the \( \sigma^2 \chi^2_v/v \) distribution and (ii) \( \sigma^2 \) must be estimated from the data.

In the first case, we are naturally led to consider the following procedure, analogous to (2.5),

\[
\hat{\mu} = \bar{y} - W_1^* \]

with

\[
W_1^* = \begin{cases} 
0 & \text{if } |z_i| < c_{s_i} & \text{all } i \\
\frac{y_i - \bar{y}}{n-1} & \text{if } |z_i| > c_{s_i} & |z_j| < |z_i| & i=1, \ldots, n; j \neq i
\end{cases}
\]

(5.1)
It is easy to see that when there is no spurious observation, \( E(\hat{\mu}) = \mu \) so that the premium is \( p_m = \frac{n}{n^2} E[W_1^{*2}] \). Now

\[
E(\frac{W_1^{*2}}{c^2}) = \frac{n}{(n-1)^2} E\left[ (\frac{z_1 - \bar{z}}{\sigma})^2 | |z_1| > c \frac{s_1}{\sigma}, |z_j| < |z_1|, j \neq 1 \right]
\]

(5.2)

The variables \( z_1 / \sigma \) are independent \( N(0,1) \) so that conditional on \( s_1 \), the expectation of (5.2) is of the same form as that of (2.10) with \( c \) replaced by \( c \frac{s_1}{\sigma} \). Thus, denoting the expression in (2.14) by \( G(c) \), the unconditional expectation of (5.2) leads immediately to

\[
p_m = \int_0^\infty G(c \frac{s_1}{\sigma}) p(s_1^2 | \sigma^2) \, ds_1^2
\]

(5.3)

To illustrate how uncertainty about \( \sigma^2 \) due to sampling variation of \( s_1^2 \) may affect the premium, we have computed (5.3) for various \( \nu \) using a set of values of \( (c,n) \) from Table 1 all of which correspond to a premium of 0.02 there. The results are given in Table 5.

Since \( s_1^2 \) converges to \( \sigma^2 \) as \( \nu \to \infty \), the value 0.02 is also the asymptotic limit of the entries in each row of the table. We see that sampling variation of \( s_1^2 \) does give rise to a substantial increase in the value of the premium, especially when \( n \) is small. The convergence of the premium to its asymptotic limit of 0.02 appears to be quite slow; even for \( \nu \) as large as 60, it is still more than 25% larger than the limit, a result which seems rather surprising to the authors.
In so far as the protection is concerned, if there is only one spurious observation which can be taken as an $N(\mu + a\sigma, \sigma^2)$ variable, then we may employ the unconditional-conditional argument above to obtain from (2.22) the desired result. We remark in passing that this technique can also be utilized in an analogous manner for the more general rules discussed in Sections 3 and 4.

We now turn to discuss case (ii), namely, no independent estimate of $\sigma^2$ is available and hence it must be estimated from the data. Once we obtain the adjusted residuals, $z_i = y_i - \bar{y} + \nu / \sqrt{n}$, it is then natural to use as estimator $s^2 = \frac{1}{n} \sum z_i^2$. Thus, we use as our estimator for $\mu$,

$$\hat{\mu} = \bar{y} - U$$

with

$$U = \begin{cases} 0 & \text{if } |z_i| < c_s \text{ all } i \\ \frac{y_i - \bar{y}}{n-1} & \text{if } |z_i| > c_s, \ |z_j| < |z_i|, \ i=1,\ldots,n, \ j\neq i \end{cases}$$

It is to be noted that, for this procedure, $c < \sqrt{n}$ because otherwise there will never be a rejection. We now derive an expression for the premium of this rule. Again, because of the symmetry of the distributional properties of the $z$'s, it is easy to verify that $E(\hat{\mu}) = \mu$ and $p_m = \frac{n}{\sigma^2} E(U^2)$. Now

$$E\left(\frac{U^2}{\sigma^2}\right) = \frac{n}{(n-1)^2} E\left[(u_1 - \bar{u})^2 \mid |u_1| > c_s, \ |u_j| < |u_1|, \ j\neq 1\right]$$

(5.5)

where $u_i = \frac{z_i}{\sigma}$, $i=1,\ldots,n$, are independent $N(0,1)$ variables, and $s^* = \frac{S}{\sigma} = \left(\frac{1}{n} \sum u_j^2\right)^{\frac{1}{2}}$. Labelling $S$ as the region defined in (5.5),
we can write
\[ E[(u_1 - \bar{u})^2 \mid S] = \left( \frac{n-1}{n} \right)^2 E(u_1^2 \mid S) + \frac{n-1}{n} E(u_2^2 \mid S) \] \hspace{1cm} (5. 6)

since the cross products have zero expectation. If we make the transformation
\[
\begin{align*}
t_i &= \frac{u_i}{\Sigma u_i} \quad i = 1, \ldots, n-1 \\
x &= \Sigma u_i
\end{align*}
\hspace{1cm} (5. 7)
\]

then, as is well known, (see e.g. Wilks (1962 p.191) or Tiao and Guttman (1965)) (1) the t's follow an (n-1) dimensional Dirichlet distribution and (ii)x has the \( \chi_n^2 \) distribution independent of the t's.

In terms of the transformed variables, the region \( S \) is simply
\[ S: \quad t_1 > \frac{C^2}{n}; \quad t_j < t_1, \quad j = 2, \ldots, n-1; \quad \text{and} \quad 1 - \sum_{j=2}^{n-1} t_j < 2t_1 \] \hspace{1cm} (5. 8)

Noting that \( u_1^2 = t_1 x \) and that \( E(x) = n \), we have
\[ E((u_1 - \bar{u})^2 \mid S) = \left( \frac{n-1}{n} \right)^2 E(t_1 \mid S) + \frac{n-1}{n} E(t_2 \mid S) \] \hspace{1cm} (5. 9)

Thus, we may compute the premium by numerical integration involving a Dirichlet distribution over the rather simple region defined in (5. 8).

The numerical problems involved are, however, somewhat complex and are currently being investigated.

\textbf{Acknowledgement}

The authors are indebted to C. Palit for computing assistance.
References


<table>
<thead>
<tr>
<th>N</th>
<th>PREM</th>
<th>.005</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3.692700</td>
<td>3.379775</td>
<td>3.150256</td>
<td>3.006570</td>
<td>2.899526</td>
<td>2.813132</td>
<td>2.521277</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.584958</td>
<td>3.375805</td>
<td>3.145307</td>
<td>3.000804</td>
<td>2.893005</td>
<td>2.805885</td>
<td>2.516406</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.587865</td>
<td>3.373738</td>
<td>3.142659</td>
<td>2.997640</td>
<td>2.893842</td>
<td>2.801723</td>
<td>2.505334</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.586823</td>
<td>3.372445</td>
<td>3.140948</td>
<td>2.995537</td>
<td>2.886844</td>
<td>2.798820</td>
<td>2.498296</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.586111</td>
<td>3.371541</td>
<td>3.139710</td>
<td>2.993969</td>
<td>2.884935</td>
<td>2.796551</td>
<td>2.493836</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.585587</td>
<td>3.370861</td>
<td>3.138743</td>
<td>2.992709</td>
<td>2.883364</td>
<td>2.794646</td>
<td>2.489877</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.585182</td>
<td>3.370322</td>
<td>3.137947</td>
<td>2.991642</td>
<td>2.882004</td>
<td>2.792968</td>
<td>2.486155</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.584850</td>
<td>3.369877</td>
<td>3.137265</td>
<td>2.990704</td>
<td>2.880785</td>
<td>2.791440</td>
<td>2.484260</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3.584594</td>
<td>3.369498</td>
<td>3.136663</td>
<td>2.989856</td>
<td>2.879663</td>
<td>2.790016</td>
<td>2.479151</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.584354</td>
<td>3.369166</td>
<td>3.136119</td>
<td>2.989733</td>
<td>2.878612</td>
<td>2.788666</td>
<td>2.475765</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3.584153</td>
<td>3.368871</td>
<td>3.135618</td>
<td>2.988338</td>
<td>2.877613</td>
<td>2.787370</td>
<td>2.472411</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.583977</td>
<td>3.368604</td>
<td>3.135150</td>
<td>2.987640</td>
<td>2.876653</td>
<td>2.786115</td>
<td>2.469908</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.583818</td>
<td>3.368358</td>
<td>3.134709</td>
<td>2.986972</td>
<td>2.875724</td>
<td>2.784891</td>
<td>2.467574</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3.583674</td>
<td>3.368129</td>
<td>3.134289</td>
<td>2.986326</td>
<td>2.874819</td>
<td>2.783690</td>
<td>2.465240</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3.583543</td>
<td>3.367915</td>
<td>3.133885</td>
<td>2.986599</td>
<td>2.873933</td>
<td>2.782507</td>
<td>2.463046</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3.583421</td>
<td>3.367713</td>
<td>3.133496</td>
<td>2.985087</td>
<td>2.873061</td>
<td>2.781338</td>
<td>2.460954</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3.583330</td>
<td>3.367520</td>
<td>3.133119</td>
<td>2.984487</td>
<td>2.872201</td>
<td>2.780179</td>
<td>2.458958</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.583200</td>
<td>3.367335</td>
<td>3.132749</td>
<td>2.983896</td>
<td>2.871351</td>
<td>2.779028</td>
<td>2.457053</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>3.583089</td>
<td>3.367167</td>
<td>3.132388</td>
<td>2.983315</td>
<td>2.870509</td>
<td>2.777883</td>
<td>2.445286</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3.583003</td>
<td>3.366986</td>
<td>3.132035</td>
<td>2.982740</td>
<td>2.869673</td>
<td>2.776742</td>
<td>2.443175</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3.582911</td>
<td>3.366819</td>
<td>3.131687</td>
<td>2.982171</td>
<td>2.868841</td>
<td>2.775603</td>
<td>2.441106</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3.582823</td>
<td>3.366657</td>
<td>3.131344</td>
<td>2.981606</td>
<td>2.868013</td>
<td>2.774466</td>
<td>2.439053</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>3.582657</td>
<td>3.366343</td>
<td>3.130671</td>
<td>2.980518</td>
<td>2.866384</td>
<td>2.772192</td>
<td>2.427079</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3.582578</td>
<td>3.366191</td>
<td>3.130339</td>
<td>2.979995</td>
<td>2.865564</td>
<td>2.771053</td>
<td>2.423110</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>3.582501</td>
<td>3.366042</td>
<td>3.130011</td>
<td>2.979538</td>
<td>2.864723</td>
<td>2.769913</td>
<td>2.419139</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>3.582427</td>
<td>3.365895</td>
<td>3.129684</td>
<td>2.979084</td>
<td>2.863903</td>
<td>2.768770</td>
<td>2.415162</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.582354</td>
<td>3.365750</td>
<td>3.129361</td>
<td>2.978586</td>
<td>2.863083</td>
<td>2.767644</td>
<td>2.411261</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>3.582282</td>
<td>3.365607</td>
<td>3.129038</td>
<td>2.977739</td>
<td>2.862262</td>
<td>2.766475</td>
<td>2.407360</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3.582213</td>
<td>3.365466</td>
<td>3.128718</td>
<td>2.977193</td>
<td>2.861414</td>
<td>2.765322</td>
<td>2.404460</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.582144</td>
<td>3.365326</td>
<td>3.128400</td>
<td>2.976648</td>
<td>2.860619</td>
<td>2.764164</td>
<td>2.398895</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>3.582077</td>
<td>3.365188</td>
<td>3.128092</td>
<td>2.976103</td>
<td>2.859795</td>
<td>2.763002</td>
<td>2.394145</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3.582011</td>
<td>3.365051</td>
<td>3.127765</td>
<td>2.975558</td>
<td>2.858970</td>
<td>2.761836</td>
<td>2.388963</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>3.581946</td>
<td>3.364915</td>
<td>3.127450</td>
<td>2.975014</td>
<td>2.858144</td>
<td>2.760644</td>
<td>2.383805</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3.581881</td>
<td>3.364780</td>
<td>3.127135</td>
<td>2.974470</td>
<td>2.857315</td>
<td>2.759487</td>
<td>2.378843</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>3.581818</td>
<td>3.364646</td>
<td>3.126821</td>
<td>2.973926</td>
<td>2.856485</td>
<td>2.758304</td>
<td>2.373785</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>3.581756</td>
<td>3.364513</td>
<td>3.126508</td>
<td>2.973381</td>
<td>2.855652</td>
<td>2.757116</td>
<td>2.368730</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.581694</td>
<td>3.364380</td>
<td>3.126195</td>
<td>2.972836</td>
<td>2.854817</td>
<td>2.755921</td>
<td>2.363684</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1/(N-1)</td>
<td>Δ</td>
<td>Premium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.250000</td>
<td>0.332578</td>
<td>0.582578</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.111111</td>
<td>0.276506</td>
<td>0.387617</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.071429</td>
<td>0.228459</td>
<td>0.299888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.052632</td>
<td>0.195301</td>
<td>0.247933</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.041667</td>
<td>0.171314</td>
<td>0.212981</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.034483</td>
<td>0.153129</td>
<td>0.187612</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.025641</td>
<td>0.127247</td>
<td>0.152888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.020408</td>
<td>0.109573</td>
<td>0.129981</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.016949</td>
<td>0.096644</td>
<td>0.113593</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.014493</td>
<td>0.086725</td>
<td>0.101218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.012658</td>
<td>0.078844</td>
<td>0.091502</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.011236</td>
<td>0.072412</td>
<td>0.083648</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.010101</td>
<td>0.067052</td>
<td>0.077153</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.692949</td>
<td>3.379922</td>
<td>3.150798</td>
<td>3.007510</td>
<td>2.900927</td>
<td>2.815555</td>
<td>2.526643</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.589696</td>
<td>3.376134</td>
<td>3.146116</td>
<td>3.002194</td>
<td>2.905046</td>
<td>2.808695</td>
<td>2.518103</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.588051</td>
<td>3.374183</td>
<td>3.143747</td>
<td>2.999504</td>
<td>2.892098</td>
<td>2.805479</td>
<td>2.513792</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.587059</td>
<td>3.373005</td>
<td>3.142315</td>
<td>2.997877</td>
<td>2.890322</td>
<td>2.803528</td>
<td>2.511147</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.586395</td>
<td>3.372216</td>
<td>3.141355</td>
<td>2.996785</td>
<td>2.889055</td>
<td>2.802215</td>
<td>2.509332</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.585910</td>
<td>3.371650</td>
<td>3.140667</td>
<td>2.995401</td>
<td>2.888226</td>
<td>2.801266</td>
<td>2.507985</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.585553</td>
<td>3.371225</td>
<td>3.140149</td>
<td>2.994509</td>
<td>2.887549</td>
<td>2.800545</td>
<td>2.506926</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.585283</td>
<td>3.370894</td>
<td>3.139744</td>
<td>2.994046</td>
<td>2.887051</td>
<td>2.799976</td>
<td>2.506053</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3.585060</td>
<td>3.370628</td>
<td>3.139419</td>
<td>2.993572</td>
<td>2.886362</td>
<td>2.799511</td>
<td>2.505306</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.584877</td>
<td>3.370411</td>
<td>3.139157</td>
<td>2.993264</td>
<td>2.886285</td>
<td>2.799123</td>
<td>2.504646</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3.584725</td>
<td>3.370229</td>
<td>3.138929</td>
<td>2.992905</td>
<td>2.885990</td>
<td>2.798791</td>
<td>2.504048</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.584596</td>
<td>3.370075</td>
<td>3.138739</td>
<td>2.992604</td>
<td>2.885736</td>
<td>2.798502</td>
<td>2.503493</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.584485</td>
<td>3.369943</td>
<td>3.138575</td>
<td>2.992351</td>
<td>2.885514</td>
<td>2.798246</td>
<td>2.502968</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3.584399</td>
<td>3.369828</td>
<td>3.138432</td>
<td>2.992142</td>
<td>2.885317</td>
<td>2.798016</td>
<td>2.502465</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3.584305</td>
<td>3.369728</td>
<td>3.138306</td>
<td>2.992009</td>
<td>2.885119</td>
<td>2.797780</td>
<td>2.501977</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3.584213</td>
<td>3.369639</td>
<td>3.138194</td>
<td>2.991837</td>
<td>2.884978</td>
<td>2.797614</td>
<td>2.501502</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3.584125</td>
<td>3.369559</td>
<td>3.138094</td>
<td>2.991695</td>
<td>2.884830</td>
<td>2.797435</td>
<td>2.501023</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.584043</td>
<td>3.369488</td>
<td>3.138001</td>
<td>2.991573</td>
<td>2.884693</td>
<td>2.797266</td>
<td>2.500550</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>3.583968</td>
<td>3.369424</td>
<td>3.137921</td>
<td>2.991469</td>
<td>2.884564</td>
<td>2.797106</td>
<td>2.500077</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3.583804</td>
<td>3.369366</td>
<td>3.137845</td>
<td>2.991380</td>
<td>2.884445</td>
<td>2.796953</td>
<td>2.499601</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3.583651</td>
<td>3.369313</td>
<td>3.137776</td>
<td>2.991301</td>
<td>2.884331</td>
<td>2.796806</td>
<td>2.499121</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3.583509</td>
<td>3.369264</td>
<td>3.137712</td>
<td>2.991232</td>
<td>2.884222</td>
<td>2.796663</td>
<td>2.498634</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3.583373</td>
<td>3.369220</td>
<td>3.137652</td>
<td>2.991172</td>
<td>2.884117</td>
<td>2.796524</td>
<td>2.498140</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>3.583248</td>
<td>3.369178</td>
<td>3.137596</td>
<td>2.991120</td>
<td>2.884017</td>
<td>2.796389</td>
<td>2.497639</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3.583131</td>
<td>3.369140</td>
<td>3.137544</td>
<td>2.991070</td>
<td>2.883919</td>
<td>2.796256</td>
<td>2.497128</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>3.583028</td>
<td>3.369104</td>
<td>3.137495</td>
<td>2.991024</td>
<td>2.883824</td>
<td>2.796124</td>
<td>2.496608</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>3.582936</td>
<td>3.369071</td>
<td>3.137448</td>
<td>2.991078</td>
<td>2.883732</td>
<td>2.795994</td>
<td>2.496079</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.582853</td>
<td>3.369040</td>
<td>3.137404</td>
<td>2.991119</td>
<td>2.883641</td>
<td>2.795865</td>
<td>2.495537</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>3.582776</td>
<td>3.369011</td>
<td>3.137362</td>
<td>2.991179</td>
<td>2.883552</td>
<td>2.795737</td>
<td>2.494985</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3.582705</td>
<td>3.368983</td>
<td>3.137321</td>
<td>2.991240</td>
<td>2.883463</td>
<td>2.795608</td>
<td>2.494421</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.582639</td>
<td>3.368957</td>
<td>3.137283</td>
<td>2.991301</td>
<td>2.883376</td>
<td>2.795480</td>
<td>2.493845</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>3.582577</td>
<td>3.368933</td>
<td>3.137246</td>
<td>2.991362</td>
<td>2.883299</td>
<td>2.795352</td>
<td>2.493257</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3.582519</td>
<td>3.368910</td>
<td>3.137210</td>
<td>2.991423</td>
<td>2.883223</td>
<td>2.795223</td>
<td>2.492657</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>3.582463</td>
<td>3.368888</td>
<td>3.137175</td>
<td>2.991488</td>
<td>2.883151</td>
<td>2.795094</td>
<td>2.492044</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3.582411</td>
<td>3.368867</td>
<td>3.137142</td>
<td>2.991552</td>
<td>2.883032</td>
<td>2.794964</td>
<td>2.491417</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>3.582361</td>
<td>3.368847</td>
<td>3.137110</td>
<td>2.991618</td>
<td>2.882946</td>
<td>2.794833</td>
<td>2.490778</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>3.582313</td>
<td>3.368827</td>
<td>3.137078</td>
<td>2.991686</td>
<td>2.882861</td>
<td>2.794701</td>
<td>2.490125</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.582268</td>
<td>3.368809</td>
<td>3.137047</td>
<td>2.991753</td>
<td>2.882775</td>
<td>2.794569</td>
<td>2.489458</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>C</td>
<td>V 10</td>
<td>V 15</td>
<td>V 20</td>
<td>V 30</td>
<td>V 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.142659</td>
<td>0.064234</td>
<td>0.049346</td>
<td>0.041751</td>
<td>0.034228</td>
<td>0.026893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.137265</td>
<td>0.059251</td>
<td>0.046846</td>
<td>0.040220</td>
<td>0.033444</td>
<td>0.026621</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.134709</td>
<td>0.055656</td>
<td>0.044955</td>
<td>0.039039</td>
<td>0.032834</td>
<td>0.026411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.132749</td>
<td>0.052739</td>
<td>0.043342</td>
<td>0.038000</td>
<td>0.032283</td>
<td>0.026218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.129361</td>
<td>0.048159</td>
<td>0.040650</td>
<td>0.035196</td>
<td>0.031289</td>
<td>0.025858</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.126195</td>
<td>0.044645</td>
<td>0.038452</td>
<td>0.034652</td>
<td>0.030400</td>
<td>0.025520</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>