TECHNICAL REPORT NO. 3/87

Components of variance estimation in a calibration problem where both variables are subject to error.

D R Fox

Department of Mathematics and Statistics

January 1987

Components of variance estimation in a calibration problem where both variables are subject to error.

David R Fox

Department of Mathematics and Statistics

Curtin University of Technology

ABSTRACT:

In this paper we examine a particular calibration problem of the form $Y = f(X)$ where measurements on Y contain two sources of variability, namely instrument error in the determination of Y and uncertainty in X. Given certain distributional assumptions, together with simplifying approximations, maximum likelihood estimates for the two error variance components may be obtained.

Furthermore, it is shown how the iterative parameter estimation procedure can be readily adapted for use with the MINITAB statistical package or alternatively, using the GLIM software. A practical application of the procedure is given.

KEYWORDS:

Calibration, components of variance, generalized linear models.

INTRODUCTION

This analysis arose in the context of an experiment designed to test the accuracy with which a vehicle's speed could be determined by an observer in an overhead aircraft.

The standard procedure is for the airborne observer to time the vehicle below as it crosses two marked lines on the road. These lines are usually separated by either 500m or 1000m.

It is then a simple matter to convert this recorded time into an average speed for the measured distance.

The problem of assessing the accuracy of the procedure and/or of varying the method of calculation falls under the umbrella of statistical calibration. The broader (and as yet largely unresolved) issues surrounding the so-called calibration problem will be discussed in a separate paper. For the moment, we are concerned with obtaining estimates for the error variances associated with :

(i) the taking of time measurements,

and

(ii) the ability of a driver to maintain constant speed.

Given two marked lines separated by some distance d metres we can compute a vehicle's speed using equation (1).

$$\text{recorded speed (km/hr)} = \frac{3.6\ d}{t} \tag{1}$$

where t is measured time in seconds.

In the present context it is preferable to rearrange the terms in equation (1) to express time as a function of speed since it is the former which is actually measured. Therefore, in general terms we have as our model $y = f(X)$ where y is the time required to cover a prescribed distance when travelling at some speed X.

A simple check on the accuracy of the airborne procedure is provided by taking time measurements on a vehicle whose speed can otherwise be accurately determined and comparing the actual speed with that calculated using equation (1). In setting up such an experiment the problem becomes one of determining to what extent discrepencies between the calculated speed using equation (1) and the assumed speed are attributable to the two sources :

(i) error in time readings due to

  (a) observer's reaction time

  and

  (b) observer's ability to accurately judge the crossing of the two lines.

(ii) driver error (a driver told to travel at 80 km/hr say, will have trouble in maintaining a speedometer reading of *exactly* 80 km/hr).

Thus, whilst the *true* time (y) is precisely determined by the *actual* speed (X), in practice the the *measured* time Y is recorded and it is this measurement which is subject to error. Furthermore, the actual or true vehicle speed X is never known - the driver is simply instructed to travel at some nominal speed x.

We therefore have that

$$X_i = x_i + U_i \qquad \text{and} \qquad Y_i = y_i + V_i$$

where

x_i is some nominal speed

X_i is the actual speed

U_i is a random error reflecting the drivers inability to maintain constant speed x_i.

y_i is the actual time required when travelling at X_i .

Y_i is the measured time as made by the airborne observer.

V_i is a random error associated with taking time measurements.

Given certain assumptions concerning the distributions of U_i and V_i together with the other simplifying approximations, we now derive a procedure for estimating the components of variability σ_u^2 and σ_v^2 .

2. Estimating the components of variability

Given the previous definitions we have

$$Y_i = f(X_i) + V_i$$

$$= f(x_i + U_i) + V_i$$

and using a first-order Taylor approximation

$$Yi \simeq f(X_i) + U_i f'(x_i) + V_i \tag{2}$$

In what follows we shall assume

$$U_i \sim N(0, \sigma_u^2) \quad \text{and} \quad V_i \sim N(0, \sigma_v^2)$$

and thus : $E[Y_i] = f(x_i)$

$$Var[Y_i] = \sigma_v^2 + \sigma_u^2 [f'(x_i)]^2 \tag{3}$$

Also let $Y_i - f(x_i) \simeq u_i f'(x_i) + V_i = \varepsilon_i$

and so $E[\varepsilon_i] = 0$

$$Var\ [\varepsilon_i] = \sigma_v^2 + \sigma_u^2 [f'(x_i)]^2 = \sigma_i^2$$

We shall call σ_i^2 the "effective variance" at point x_i . An unbiassed estimator of σ_i^2 is provided by s_i^2 where

$$s_i^2 = \sum \frac{(E_i - \bar{E}_i)^2}{n - 1}$$

and E_i is the difference in a measured time and a calculated time for the *assumed* speed x_i (there being n determinations at each x_i).

Now $\frac{(n-1)s_i^2}{\sigma_i^2}$ follows a Chi-square distribution with $k = n - 1$ degrees of freedom, and thus s_i^2 has the p.d.f.

$$g_{s_i^2}(s_i^2) = \frac{m^m}{2^m \Gamma(m)} (s_i^2)^{m-1} \left(\frac{1}{\sigma_i^2}\right)^m e^{-m \frac{s_i^2}{\sigma_i^2}} \tag{4}$$

where $m = \frac{k}{2}$ and $s_i^2 > 0$

For convenience, let $K = \frac{m^m}{2^m \Gamma(m)}$ in equation (4).

We assume that at each x_i $i = 1, \ldots, N$, n replications are available. This effectively determines the value of m and so the p.d.f. $g(\cdot)$ is parameterized by σ_i^2 .

The likelihood function is thus

$$L(\sigma_i^2; s_i^2) = \prod_{i=1}^{N} \left\{ K(s_i^2)^{m-1} \left(\frac{1}{\sigma_i^2}\right)^m e^{-m \; s_i^2/\sigma_i^2} \right\}$$

$$= K^N e^{-m \sum_{i=1}^{N} \frac{s_i^2}{\sigma_i^2}} \prod_{i=1}^{N} \left\{ (s_i^2)^{m-1} \left(\frac{1}{\sigma_i^2}\right)^m \right\} \tag{5}$$

and the log-likelihood

$$\ln L = N \ln K - m \sum_{i=1}^{N} \left(\frac{s_i^2}{\sigma_i^2}\right) - m \sum_{i=1}^{N} \ln \sigma_i^2 + (m - 1) \sum_{i=1}^{N} \ln (s_i^2) \tag{6}$$

Now, in the context of vehicle speed estimation we have $f(x_i) = \frac{c}{x_i}$ where c is a constant depending on the separation of the two marked lines.

Therefore, by equation (3) : $\sigma_i^2 = \sigma_v^2 + \frac{c^2}{x_i^4} c_u^2$

and hence $\frac{\partial \sigma_i^2}{\partial \sigma_v^2} = 1$ and $\frac{\partial \sigma_i^2}{\partial \sigma_u^2} = \frac{c^2}{x_i{}^4}$

We see that equation (6) is a function of σ_u^2 and σ_v^2 . By taking partial derivatives of equation (6) with respect to both σ_u^2 and σ_v^2 and setting these to zero, we obtain the m.l.e.'s $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$.

Now,
$$\frac{\partial \ln L}{\partial \sigma_v^2} = \frac{\partial \ln L}{\partial \sigma_i^2} \cdot \frac{\partial \sigma_i^2}{\partial \sigma_v^2}$$

$$= \left\{ m \sum_{i=1}^{N} \frac{s_i^2}{\sigma_i^4} - m \sum_{i=1}^{N} \frac{1}{\sigma^2{}_i} \right\} \times 1 \qquad (7)$$

and

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{\partial \ln L}{\partial \sigma_i^2} \cdot \frac{\partial \sigma_i^2}{\partial \sigma_u^2} = m \sum_{i=1}^{N} \frac{c s_i^2}{x_i^4 \sigma_i^4} - m \sum_{i=1}^{N} \frac{c}{x_i^4 \sigma_i^2} \qquad (8)$$

Setting equations (7) and (8) to zero and letting $w_i = \frac{1}{\sigma_i^2}$ we obtain :

$$\sum_{i=1}^{N} w_i^2 s_i^2 = \sum_{i=1}^{N} w_i \qquad (9a)$$

and

$$\sum_{i=1}^{N} \frac{w_i^2 s_i^2}{x_i{}^4} = \sum_{i=1}^{N} \frac{w_i}{x_i{}^4} \qquad (9b)$$

Equations (9a) and (9b) are the maximum likelihood equations which will need to be solved iteratively. We first solve directly using Newton's Method for simultaneous non-linear equations and then demonstrate how the problem can be cast in the context of a Generalized Linear Model and thus amenable to solution using the GLIM software (the two approaches are mathematically equivalent).

3. Iteratively re-weighted least squares method for m.l.e's

Let $$f^* = \sum_{i=1}^{N} (w_i^2 s_i^2 - w_i)$$

and $$g^* = \sum_{i=1}^{N} \left(\frac{w_i^2 s_i^2}{x_i^4} - \frac{w_i}{x_i^4}\right)$$

where $$w_i = \frac{1}{\sigma_v^2 + \sigma_u^2 \frac{c^2}{x_i^4}}$$

and suppose $\delta\sigma_v^2$ and $\delta\sigma_u^2$ are the increments in σ_v^2 and σ_u^2 respectively required to reach the true solution, then :

$$\begin{bmatrix} \frac{\partial f^*}{\partial \sigma_v^2} & \frac{\partial f^*}{\partial \sigma_u^2} \\ \frac{\partial g^*}{\partial \sigma_v^2} & \frac{\partial g^*}{\partial \sigma_u^2} \end{bmatrix} \begin{bmatrix} \delta\sigma^2 \\ \delta\sigma_u^2 \end{bmatrix} = \begin{bmatrix} -f \\ -g \end{bmatrix}$$

Now ,

$$\frac{\partial f^*}{\partial \sigma_v^2} = \sum_{i=1}^{N} \{ w_i^2 - 2s_i^2 w_i^3 \}$$

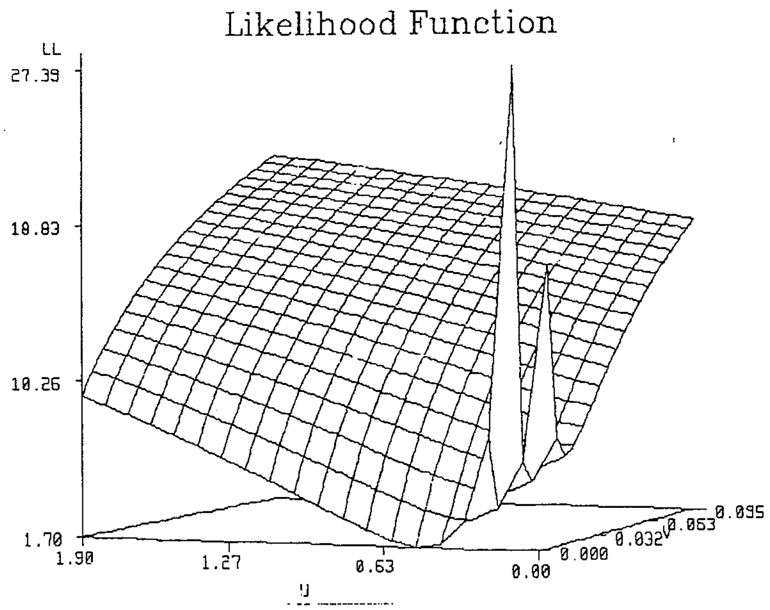
$$\frac{\partial f^*}{\partial \sigma_u^2} = \sum_{i=1}^{N} \{ \frac{c^2}{x_i^4} [w_i^2 - 2s_i^2 w_i^3] \}$$

$$\frac{\partial g^*}{\partial \sigma_v^2} = \sum_{i=1}^{N} \{ \frac{1}{x_i^4} [w_i^2 - 2s_i^2 w_i^3] \}$$

$$\frac{\partial g^*}{\partial \sigma_u^2} = \sum_{i=1}^{N} \frac{c^2}{x_i^8} [w_i^2 - 2s_i^2 w_i^3] \}$$

3.1 Choosing initial estimates for $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$

Convergence of the method described in the previous section may very much depend on the choice of initial estimates $\hat{\sigma}_{vo}^2$ and $\hat{\sigma}_{uo}^2$. An example of the log-likelihood function is depicted below :



The above plot reveals the difficulties likely to be encountered with convergence if the initial estimates are far removed from the global maximum.

We have already that $f^*(\sigma_v^2;\sigma_u^2) = \sum_{i=1}^{N} w_i^2 s_i^2 - w_i = 0$

$$\Rightarrow \sum_{i=1}^{N} w_i^2 s_i^2 = \sum_{i=1}^{N} w_i$$

This equation is obviously true for the choice $w_i = \frac{1}{s_i^2}$, although in practice this will not hold for all i since

$$\frac{1}{w_i} = \sigma_i^2 \quad \text{(by definition)}$$

and we do not expect $\sigma_i^2 \equiv s_i^2$ due to the sampling variation in s_i^2 .

Nevertheless, since s_i^2 is our sample estimate of σ_i^2, the choice $w_i = \frac{1}{s_i^2}$ is a sensible one and at least affords a starting point in the search for those $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ which satisfy (9a) and (9b).

It is suggested that initial estimates $\hat{\sigma}_{vo}^2$ and $\hat{\sigma}_{uo}^2$ be obtained by taking two arbitary x_i's (of reasonable separation) and solve the resulting pair of linear equations.

eg.

x_i	85	135
s_i^2	0.0441	0.001631

Thus :

$$\hat{\sigma}_{vo}^2 + 0.0621\hat{\sigma}_{uo}^2 = 0.0441$$

and

$$\hat{\sigma}_{vo}^2 + 0.0098\hat{\sigma}_{uo}^2 = 0.001631$$

=> $\hat{\sigma}_{vo}^2 = -0.0066$ and $\hat{\sigma}_{uo}^2 = 0.8167$

(assume $\hat{\sigma}_{vo}^2 = 0$)

The iterative procedure has been implemented on computer using the macro facility of the MINITAB package. A listing of the MINITAB routine may be found in Appendix A. An example of the procedure is now given.

4. MINITAB implementation

We now demonstrate how the iterative procedure for obtaining the mle's may be implemented in MINITAB. In the following example we have generated a set of 'measured' times according to the model outlined in §1. Thirteen assumed speeds were used ranging from 80 km/hr to 140 km/hr. At each of these speeds 30 replicates were obtained using errors generated from the following distributions :

$$U \sim N(0, 3.8^2) \qquad \text{and} \qquad V \sim N(0, 0.75^2)$$

Analysis of the restuling data provided the following sample standard deviations at each speed:

assumed speed x_i	80	85	90	95	100	105	110	115
standard deviation s_i	1.133	1.237	0.995	1.008	0.960	0.969	0.938	0.981

120	125	130	135	140
0.774	0.707	0.700	0.839	0.713

Sample output corresponding to the MINITAB analysis of the above data is shown on the next page.

```
MTB > # Simulation of components of variance estimation
MTB > # 30 observations at each speed using following parameters ;
MTB > #
MTB > #                    sigma-squared U = 14.44
MTB > #
MTB > #                    sigma-squared V = .5625
MTB > let k1=3.8
MTB > let k2=.75
MTB > exec 'simul'



 ROWS: C1

         C6
    STD DEV

 80    1.133
 85    1.237
 90    0.995
 95    1.008
 100   0.960
 105   0.969
 110   0.938
 115   0.981
 120   0.774
 125   0.707
 130   0.700
 135   0.839
 140   0.713
 ALL   3.106

MTB > set c1                     # put assumed speeds into C1
DATA> 80:140/5
DATA> set c3                     # put std. deviations into C3
DATA> 1.133,1.237,.995,1.008,.96,.969,.938,.981,.774,.707,.7,.839,.713
DATA> print c1 c3
 ROW     C1      C3

   1     80   1.133
   2     85   1.237
   3     90   0.995
   4     95   1.008
   5    100   0.960
   6    105   0.969
   7    110   0.938
   8    115   0.981
   9    120   0.774
  10    125   0.707
  11    130   0.700
  12    135   0.839
  13    140   0.713

MTB > raise c3 2 c3              # convert entries in C3 to variances
MTB > let k1=.1                  # set initial estimate for sigma-squared V
MTB > let k2=.1                  # set initial estimate for sigma-squared U
MTB > exec 'macl' 5              # perform 5 iterations of mle routine
   SUM     =      929.73
   SUM     = 0.000010355
   SUM     =      -19183
   SUM     =     -674.87
   SUM     = -0.00020829
   SUM     =-0.000010277

 MATRIX M4

 0.0453770
 0.0878439


   SUM    =       403.88
   SUM    =0.0000045172
   SUM    =      -5883.2
   SUM    =      -205.70
   SUM    =-0.000063486
   SUM    =-3.11524E-06
 MATRIX M4

 0.062446
 0.177431


   SUM    =       173.15
   SUM    =0.0000019502
   SUM    =      -1840.7
   SUM    =      -63.711
   SUM    =-0.000019664
   SUM    =-9.56465E-07
 MATRIX M4

 0.081465
 0.364177


   SUM    =       72.607
   SUM    =8.276335E-07
   SUM    =      -595.65
   SUM    =      -20.270
   SUM    =-6.25628E-06
   SUM    =-2.99900E-07
 MATRIX M4

 0.096459
 0.747441


   SUM    =       29.268
   SUM    =3.410009E-07
   SUM    =      -204.65
   SUM    =      -6.7597
   SUM    =-2.08632E-06
   SUM    =-9.75033E-08
 MATRIX M4

  0.09377
  1.49094


MTB > note . . . Convergence reached after another 10 iterations
MTB > print k1 k2            # print final estimates
K1      0.453761
K2      14.1403
```

After 15 iterations we obtain the following parameter estimates :

$$\hat{\sigma}_u^2 = 14.403 \qquad (\hat{\sigma}_u = 3.76)$$

and

$$\hat{\sigma}_v^2 = 0.453761 \qquad (\hat{\sigma}_v = 0.67)$$

These are in good agreement with the actual values of 14.44 and 0.5625.

4.1 Application to air surveillance method

The air surveillance method of checking vehicle speeds has already been outlined in §1. We now apply the components of variance estimation procedure to actual test data collected in a controlled experiment. Details of the experimental set-up may be found in Fox (1985). The assumed speeds and sample variances are given below :

Assumed speed x_i	Sample variance s_i^2
85	0.0441126
90	0.0483296
95	0.0216649
100	0.0162971
105	0.0057608
110	0.0063680
115	0.0183088
120	0.0007868
125	0.0080210
130	0.0028164
135	0.0016314
140	0.0027468

Part of the MINITAB output using the above data follows.

```
      let k1=0          # initial estimate for sigma-squared V
MTB > let k2=.2         # initial estimate for sigma-squared U
MTB > NOTE . . . Column C1 is column of speeds
MTB > NOTE . . . Column C3 is column of sample variances
MTB > print c1 c3
 ROW  speed    variance

   1     85   0.0441126
   2     90   0.0483296
   3     95   0.0216649
   4    100   0.0162971
   5    105   0.0057608
   6    110   0.0063680
   7    115   0.0183088
   8    120   0.0007868
   9    125   0.0080210
  10    130   0.0028164
  11    135   0.0016314
  12    140   0.0027468

MTB > noecho            # surpresses display of intermediate calculations
MTB > exec 'macl' 10    # execute MLE routine 10 times
   SUM      =       3145.8
   SUM      = 0.000024886
   SUM      =     -3002834
   SUM      =       -48399
   SUM      =    -0.014938
   SUM      = -0.00034145
 MATRIX M4

-0.0004310
 0.0917384

      .
      .
      .

   SUM      = -0.00018311
   SUM      =-6.57252E-13
   SUM      =     -594519
   SUM      =     -6038.3
   SUM      =  -0.0018637
   SUM      =-0.000023096
 MATRIX M4

-0.000000000
-0.000000020

MTB > NOTE . . . Convergence reached after 10 iterations
MTB > print k1 k2           # display MLE's
K1       -0.00367924
K2        0.672324
MTB > NOTE . . . Regard negative variance as essentially zero
MTB >
```

After 10 iterations we obtain the following estimates :

$$\hat{\sigma}_u^2 = 0.672324$$

and $$\hat{\sigma}_v^2 = -0.00367924$$

We thus regard $\hat{\sigma}_v^2$ as being essentially zero. The implication of this result is that virtually all error associated with the air surveillance procedure has arisen from variation in vehicle speed rather than an inability to make accurate timing measurements.

This observation supports the previously reported conclusion [FOX (1985)] that the airborne method of vehicle speed determination is both accurate and precise.

5. Re-formulation as a Generalized Linear Model

We now take an alternative, although mathematically equivalent approach to the problem by casting it in the framework of a generalized linear model and obtaining parameter estimates via the GLIM software.

5.1 Error distribution and link function

We start with the model : $\theta_i = \sigma_v^2 + \sigma_u^2 \dfrac{c^2}{x_i^4}$ (10)

and data Y_i, $i = 1, \ldots, N$. (θ_i was previously identified as σ_i^2 and the Y_i as s_i^2).

Following from §2, the error distribution for the Y_i is of the form :

$$f_{Y_i}(y_i;\theta_i) = K y_i^{m-1} \left(\frac{1}{\theta_i}\right)^m e^{-my_i/\theta_i} \qquad (11)$$

where K and m are as previously defined.

The log-likelihood function is also as before :

$$\ln L(\theta_i;y_i) = N \ln K - \sum_{i=1}^{N} my_i/\theta_i - m\sum_{i=1}^{N} \ln \theta_i + (m-1) \sum_{i=1}^{N} \ln y_i \qquad (12)$$

Now, equation (10) is of the form $\theta_i = \beta_0 + \beta_1 x_i'$ where $\beta_0 = \sigma_v^2$, $\beta_1 = \sigma_u^2$ and $x_i' = \dfrac{c^2}{x_i^4}$.

Writing $f_Y(\cdot)$ in the exponential form

$$f_Y(y;\theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\}$$

we obtain $a(y) = my$; $b(\theta) = \dfrac{1}{\theta}$

$c(\theta) = n \ln \theta$; $d(y) = (m-1) \ln y$

and hence using standard results

$$E[a(y)] = E[mY] = mE[Y] = \frac{m/\theta}{1/\theta^2} = m\theta$$

$$\Rightarrow \quad E[Y] = \theta$$

and $\quad Var\,[mY] = m^2\, Var\,[Y]$

$$= \frac{[(\frac{2}{\theta^3})(\frac{m}{\theta}) \;-\; \frac{m}{\theta^2}\frac{1}{\theta^2}]}{(\frac{1}{\theta^6})}$$

$$= m\theta^2$$

$$\Rightarrow \quad Var\,[Y] = \frac{\theta^2}{m}$$

Now, since $\mu_i = E[Y_i] = \theta_i$ and $\theta_i = \eta_i$ where η_i is the linear predictor, we see that we have a GLM with error distribution specified by equation (11) and identity link function.

5.2 Deviance

The (scaled) deviance D, is defined as $2[\ln l_f - \ln l_c]$ where l_f and l_c are respectively the likelihoods under the full and current models. Let $\tilde{\theta}_i$ be the parameter estimate under a full model and $\hat{\theta}_i$ be the parameter estimate obtained from the current model.

It is easily shown that

$$D = 2[\sum_{i=1}^{N} \frac{my_i}{\hat{\theta}_i} + m\sum_{i=1}^{N} \ln \hat{\theta}_i - m\sum_{i=1}^{N} \ln \tilde{\theta}_i - \sum_{i=1}^{N} \frac{my_i}{\tilde{\theta}_i}] \qquad (13)$$

We are now in a position to set-up the required macros in preparation for implementation using the GLIM software.

5.3 GLIM implementation

In GLIM, four macros must be specified which assign values to the system vectors %FV, %DR, %VA and %DI.

Thus, using the results of §5.1 and §5.2 we have :

```
%FV  =  %LP
%DR  =  1        (Since link is identity function)
%VA  =  (%LP**2)/%M      (%M= m)
%DI  =  2* %M*((%YV/%FV - 1) + (%LOG(%FV/%YV)))
```

5.4 Example

Output from the GLIM analysis of the 'synthetic' data of §4 is given below.

```
? $UNITS 13
? $C
? $C  SIMULATION OF COMPONENTS OF VARIANCE ESTIMATION
? $C 30 OBSERVATIONS AT EACH SPEED USING FOLLOWING PARAMETERRS :
? $C
? $C                   SIGMA-SQUARED U = 14.44
? $C
? $C                   SIGMA-SQUARED V = 0.5625
? $C
? $DATA X Y
? $READ
? 80 1.133
? 85 1.237
? 90 .995
? 95 1.008
? 100 .96
? 105 .969
? 110 .938
? 115 .981
? 120 .774
? 125 .707
? 130 .7
? 135 .839
? 140 .713
? $CALC Y=Y**2$
? $ERI
? $LOOK X Y$
   1    80.00      1.284
   2    85.00      1.530
   3    90.00     0.9900
   4    95.00      1.016
   5    100.0     0.9216
   6    105.0     0.9390
   7    110.0     0.8798
   8    115.0     0.9624
   9    120.0     0.5991
  10    125.0     0.4998
  11    130.0     0.4900
  12    135.0     0.7039
  13    140.0     0.5084
? $EVAY
? $YVAR Y
? $CALC X1=1800**2/X**4
? $CALC %M=(30-1)/2
? $MAC M1
? $CALC %FV=%LP
? $ENDMAC
? $MAC M2
? $CALC %DR=1
? $ENDMAC
? $MAC M3
? $CALC %VA=(%LP**2)/%M
? $ENDMAC
? $MAC M4
? $CALC %DI=2*%M*((%YV/%FV-1)+(%LOG(%FV/%YV)))
? $ENDMAC
? $CALC %LP=%YV
? $OWN M1 M2 M3 M4$
? $FIT$
 CYCLE  DEVIANCE     DF
    3    22.69       12

? $FIT X1$
 CYCLE  DEVIANCE     DF
    3    5.471       11

? $D ME$
Y-VARIATE  Y
ERROR OWN      LINK OWN
M1
M2
M3
M4

LINEAR PREDICTOR
%GM X1

       ESTIMATE      S.E.      PARAMETER
   1  0.4543      0.6675E-01  %GM
   2  14.12       2.676       X1
  SCALE PARAMETER TAKEN AS   0.4974
```

The parameter estimates are as previously obtained :

$$\hat{\sigma}_u^2 = 14.12$$

and $$\hat{\sigma}_v^2 = 0.4543$$

Applying the procedure now to the experimental data of §4.1 we obtain the following results.

```
fox > GLIM
GLIM 3.12 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

? $C  COMPONENTS OF VARIANCE ESTIMATION FOR ACTUAL TEST DATA
? $UNITS 12
? $DATA X Y
? $READ
? 85 .0441126
? 90 .0483296
? 95 .0216649
? 100 .0162971
? 105 .0057608
? 110 .006368
? 115 .0183088
? 120 .0007868
? 125 .008021
? 130 .0028164
? 135 .0016314
? 140 .0027468
? $YVAR Y
? $MAC M1
? $CALC %FV=%LP
? $ENDMAC
? $MAC M2
? $CALC %DR=1
? $ENDMAC
? $MAC M3
? $CALC %VA=(%LP**2)/%M
? $ENDMAC
? $MAC M4
? $CALC %DI=2*%M*((%YV/%FV-1)+(%LOG(%FV/%YV)))
? $ENDMAC
? $CALC X1=1800**2/X**4
? $OWN M1 M2 M3 M4
? $CALC %LP=%YV
? $CALC %M=(3-1)/2
? $C  ***************  FIT NULL MODEL  *************************************
? $SCALE %M
? $CALC %LP=%YV
? $FIT$
----- CURRENT DISPLAY INHIBITED
        SCALED
 CYCLE  DEVIANCE     DF
   3     15.73       11

? $C  ***************  ADD COVARIATE TO MODEL  *****************************
? $CALC %LP=%YV
? $FIT X1$
----- CURRENT DISPLAY INHIBITED
        SCALED
 CYCLE  DEVIANCE     DF
   3     5.164       10

? $D ME
Y-VARIATE  Y
ERROR OWN      LINK OWN
M1
M2
M3
M4

LINEAR PREDICTOR
%GM X1

       ESTIMATE      S.E.      PARAMETER
   1 -0.3518E-02  0.3306E-02  %GM
   2  0.6614      0.2930      X1
   SCALE PARAMETER TAKEN AS   1.000

? $D V
(CO)VARIANCE MATRIX
  1  1.0930E-05
  2 -8.9494E-04  8.5861E-02
        1            2
   SCALE PARAMETER TAKEN AS    1.000
```

After 3 iterations of the GLIM macro we obtain

$$\hat{\sigma}_u^2 = 0.6614$$

and $$\hat{\sigma}_v^2 = -0.003518$$

We note that $\hat{\sigma}_v^2$ is one standard deviation from zero and as such $\hat{\sigma}_v^2 = 0$ is plausable.

6. Conclusions

We have demonstrated that given certain assumptions about the nature of the error distributions involved together with the use of some simplifying approximations, the two components of variation associated with the estimation of vehicle speeds from airborne observation are indeed estimable.

Analysis of previously obtained data suggests that by far the greater source of variability is the inability of the driver to maintain constant speed. The contribution to the total "effective" variance due to timing inaccuracies has been shown to be negligable.

These results, together with previous findings, give support to the contention that the airborne procedure of speed determination is both accurate and precise.

Appendix A : MINITAB listing for parameter estimation macro A.1

```
ty mac1.mtb
let c2=1/(k1+((1800/c1**2)**2)*k2)
let c4=c3*c2**2-c2
sum c4 k4
let c5=(c2/c1**4)*(c2*c3-1)
sum c5 k5
let c10=c2**2*(1-2*c3*c2)
sum c10 k6
let c7=((1800/c1**2)**2)*c10
sum c7 k7
let c8=c10/c1**4
sum c8 k8
let c9=((1800/c1**4)**2)*c10
sum c9 k9
set c98
k6 k8
set c99
k7 k9
copy c98 c99 m1
let k4=-k4
let k5=-k5
set c100
k4 k5
copy c100 m2
invert m1 m3
mult m3 m2 m4
print m4
copy m4 c12
let k10=c12(1)
let k11=c12(2)
let k1=k1+k10
let k2=k2+k11
end
```

REFERENCE

Fox, D.R. A Statistical Appraisal of vehicle speed determination from airborne observation.
Report prepared for WA Police Department, November 1985.