
QR-decomposition from the statistical point of view

Hilmar Drygas¹

Department of Mathematics, University of Kassel, D-34109 Kassel
(Germany) drygas@mathematik.uni.kassel.de

1 Introduction.

In this paper we will deal with the linear model

$$E(y) = \underset{n \times 1}{X} \underset{n \times k}{\beta}, \quad \text{Cov}(y) = \sigma^2 I. \quad (1)$$

We denote by $\text{span}\{z_1, \dots, z_r\}$ the linear subspace spanned by the vectors z_1, \dots, z_r of the vector-space V . For a matrix A we denote the range of the matrix A by $\text{im}(A)$ and its rank by $r(A)$. A g -inverse of A is denoted A^- and the Moore-Penrose of A is denoted by A^+ . P_L denotes the orthogonal projection into the linear subspace L of the vector-space V .

It should be noted that the model (1) is the most general linear model. Further, (1) is understood with respect to a general inner product, i.e.,

$$E(y, a) = (X\beta, a), \quad \text{Var}(y, a) = \sigma^2(a, a) \quad (2)$$

with respect to a given inner product. If with respect to the classical inner product $(x, y)_o = x'y$ and the relation $\text{Cov}(y) = Q$, (Q is nonnegative definite matrix) is correct, then if $\text{im}(X) \subseteq \text{im}(Q)$ with respect to the inner product

$$(x, z) = (x, Q^+z)_o = x'Q^+z, \quad (3)$$

then the relation $\text{Var}(y, a) = \sigma^2(a, a)$ holds. (x, z) is also equal to $(x, \rho)_o$ where ρ is any solution of $Q\rho = z$. This definition does not depend on the selection ρ . If $\text{im}(X) \subseteq \text{im}(Q)$ does not hold, then with respect to the estimation of $E(y)$, we can “regularize” the model by replacing Q

by $W = Q + cXX'$, $c > 0$. Then $\text{im}(X) \subseteq \text{im}(W)$ and for the inner product

$$(x, z) = (x, W^+z)_o = x'W^+z, \quad (4)$$

again $\text{Var}(y, a) = \sigma^2(a, a)$. This is just what $\text{Cov } y = \sigma^2\mathbf{I}$ means.

In the model (1) the BLUE (Best Linear Unbiased Estimator) or GME (Gauss-Markov-Estimator) of Ey is

$$Gy = P_{\text{im}(X)}y, \quad (5)$$

which is the orthogonal projection of y onto $\text{im}(X)$. If $X = (x_1, \dots, x_k)$, $x_i \in \mathbb{R}_{k \times 1}$ and $\beta = (\beta_1, \dots, \beta_k)'$ then

$$X\beta = \beta_1x_1 + \dots + \beta_kx_k. \quad (6)$$

Gy is easy to compute if x_i 's are pairwise orthogonal, i.e., $(x_i, x_j) = 0$ if $i \neq j$. In this case

$$Gy = \sum_{i=1}^k \hat{\beta}_i x_i, \quad \hat{\beta}_i = \begin{cases} \frac{(x_i, y)}{(x_i, x_i)} & \text{if } x_i \neq 0 \\ \text{arbitrary} & \text{if } x_i = 0. \end{cases} \quad (7)$$

Also the measure of determination $R^2 = \|Gy\|^2/\|y\|^2$ can be easily computed. Indeed,

$$\|Gy\|^2 = \sum_{i:x_i \neq 0} \frac{(x_i, y)^2}{(x_i, x_i)} = \sum_{i=1}^k \hat{\beta}_i^2 (x_i, x_i) \quad (8)$$

and

$$R^2 = \sum_{i:x_i \neq 0} \frac{(x_i, y)^2}{(x_i, x_i)(y, y)} = \sum_{i:x_i \neq 0} \text{Corr}^2(x_i, y) \quad (9)$$

where $\text{Corr}(x, y)$ denotes the (empirical) correlation coefficient between x and y .

When x_i 's are not pairwise orthogonal, then by the orthogonalization method due to Erhard Schmidt, late professor at "Humboldt-University at Berlin", it is possible to represent $X\beta$ as follows:

$$X\beta = \sum_{i=1}^s \alpha_i q_i, \quad (10)$$

where $s = r(X)$ and q_i 's are non-zero vectors which are pairwise orthogonal. They are determined by the orthogonalization algorithm

which in the literature is mostly called Gram-Schmidt orthogonalization method. If

$$x_j = \sum_{l=1}^s r_{lj} q_l, \quad j = 1, \dots, k, \quad (11)$$

then it follows that $(x_j, q_l) = r_{lj}(q_l, q_l)$ and

$$r_{lj} = \frac{(x_j, q_l)}{(q_l, q_l)}. \quad (12)$$

Without loss of generality, we can assume that x_1, \dots, x_s are linearly independent. If we then let $q_1 = x_1$, then $r_{11} = 1$ and if also $r_{ll} = 1$, $l = 2, \dots, s$, then

$$q_l = x_l - \sum_{j=1}^{l-1} \frac{(x_l, q_j)}{(q_j, q_j)} q_j. \quad (13)$$

Since $(x_j, q_l) = (q_j, q_l)$ if $l \geq j$, it follows that $r_{ll} = 1$ and $r_{lj} = 0$ if $j < l$. Let

$$Q_1 = (q_1, \dots, q_s), \quad R_1 = (r_{lj}; l = 1, \dots, s, j = 1, \dots, k). \quad (14)$$

Then

$$X = Q_1 R_1 \quad (15)$$

where R_1 is an upper triangular matrix. This representation is called QR-decomposition in numerical analysis (See Björck-Dahlquist (1972) or Lawson and Hanson (1974)). Since $X\beta = Q_1\alpha$, $\alpha = (\alpha_1, \dots, \alpha_s)'$, it follows that $\hat{\alpha} = (Q_1'Q_1)^{-1}Q_1'y$ is the least squares estimator (BLUE, GME) of α . From

$$Q_1\alpha = \sum_{l=1}^s \alpha_l q_l = \sum_{j=1}^k \beta_j x_j = \sum_{j=1}^k \left(\sum_{l=1}^s r_{lj} q_l \right) = \sum_{l=1}^s \left(\sum_{j=1}^k r_{lj} \beta_j \right) q_l = X\beta,$$

it follows that

$$\alpha_l = \sum_{j=1}^k r_{lj} \beta_j = (R_1\beta)_l, \quad l = 1, \dots, s, \quad (16)$$

implying $R_1\beta = \alpha$.

Theorem 1. Let $\hat{\alpha} = (Q_1'Q_1)^{-1}Q_1'y$ and $\hat{\beta}$ be any solution of the equation $R_1\hat{\beta} = \hat{\alpha}$. Then $(l, \hat{\beta})$ is BLUE of (l, β) whenever (l, β) is estimable. There exists at least one solution of $R_1\hat{\beta} = \hat{\alpha}$.

Proof. a) We have the representation $X = Q_1R_1$ and (l, β) is estimable iff $l \in \text{im}(X')$, i.e., $l \in \text{im}(R_1'Q_1')$ or $l = R_1'Q_1'z$ for some z . Then

$$(l, \hat{\beta}) = (R_1'Q_1'z, \hat{\beta}) = (Q_1'z, R_1\hat{\beta}) = (z, Q_1\hat{\alpha}) \quad (17)$$

and this is BLUE of $(z, Q_1\alpha) = (z, X\beta) = (X'z, \beta) = (l, \beta)$.

b) The equation $R_1\beta = \alpha$ is solvable. Let $R = (R_{11}' : R_{12})$ where R_{11} is of order $s \times s$ and R_{12} of order $s \times k - s$. Since R_{11} is upper triangular with diagonal elements equal to one, it follows that the determinant of R_{11} equals one. Hence R_{11} is regular. If we split up $\hat{\beta}$ as $\hat{\beta}' = (\hat{\beta}'_1, \hat{\beta}'_2)$ where β_1 of order $s \times 1$ and β_2 of order $k - s \times 1$, then $\hat{\beta}'_1 = R_{11}^{-1}\alpha$, $\hat{\beta}'_2 = 0$, form a solution of $R_1\beta = \alpha$.

Since R_1 and R_{11} are upper triangular the system $R_1\hat{\beta} = \hat{\alpha}$ and $R_{11}\hat{\beta}'_1 = \hat{\alpha}$ can be easily solved successively as follows:

$$\hat{\beta}'_s = \hat{\alpha}'_s = \frac{(q_s, y)}{(q_s, q_s)} \quad (18)$$

$$\begin{aligned} \hat{\beta}'_l &= \hat{\alpha}'_l - \sum_{j=l+1}^s r_{lj}\hat{\beta}'_j \\ &= \frac{(q_l, y)}{(q_l, q_l)} - \sum_{j=l+1}^s r_{lj}\hat{\beta}'_j, \quad l = s-1, s-2, \dots, 1. \end{aligned} \quad (19)$$

2 Estimable Functions and the Computation of Their Estimates

In the model $E(y) = X\beta$, $\text{Cov}(y) = \sigma^2I$, the least squares estimator of β is

$$\hat{\beta} = (X'X)^{-1}X'y, \quad (20)$$

if X has full column-rank and

$$\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}. \quad (21)$$

Let $X = (X_0 : x_k)$, where X_0 is of order $n \times (k-1)$ and x_k of order $n \times 1$. Then

$$X'X = \begin{pmatrix} X_0'X_0 & x_k'X_0 \\ X_0'x_k & x_k'x_k \end{pmatrix} \quad (22)$$

and

$$(X'X)^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where D is the inverse of the Schur-complement as

$$\begin{aligned} x_k'x_k - x_k'X_0(X_0'X_0)^{-1}X_0'x_k &= x_k'(I - X_0(X_0'X_0)^{-1}X_0')x_k \\ &= x_k'(I - P_{\text{im}(X_0)})x_k \\ &= \|(I - P_{\text{im}(X_0)})x_k\|^2. \end{aligned} \quad (23)$$

Let $q = (I - P_{\text{im}(X_0)})x_k$. Then if $(a, b) = b'a$ is the classical inner product used in this model, then

$$\frac{q'y}{q'q} \quad (24)$$

is an estimator with variance $\sigma^2/(q'q)$ and $q'q$ is the Schur-complement of $x_k'x_k$. Since

$$E(q'y) = q'X\beta = q'(X_0\beta_0 + \beta_k x_k) = q'\beta_k x_k = \beta_k q'x_k = \beta_k q'q, \quad (25)$$

it follows that

$$\frac{q'y}{q'q} \quad (26)$$

is the best linear unbiased estimator (BLUE) of β_k . This result is also correct in more general situations.

Theorem 2. Let the linear model $E(y) = X\beta$, $\text{Cov}(y) = \sigma^2 I$ be given and $l \in \mathbb{R}^k$. Then for $l \neq 0$

- (i) (l, β) is estimable iff $Xl \notin X(l)^\perp$,
- (ii) $(l, \hat{\beta}) = \|l\|^2 \frac{(q, y)}{(q, q)}$ is BLUE of (l, β) , where $q = (I - P_{X(l)^\perp})Xl$,
- (iii) $\text{Var}(l, \hat{\beta}) = \sigma^2 \|l\|^4 \|q\|^{-2}$.

Proof. The proof can already be found in Drygas (1976), but for the sake of completeness, it is repeated here.

- (i) If $Xl \in X(l)^\perp$ then $Xl = Xl_1$ for some $l_1 \perp l$. Then $X(l - l_1) = 0$, $(l, l - l_1) = (l, l) - (l, l_1) = (l, l) \neq 0$. Hence $l \notin \text{im}(X') = (X^{-1}(0))^\perp$.

(ii) Let $q = (I - P_{X(l^\perp)})Xl$. Then

$$E(l, \hat{\beta}) = \|l\|^2 (X'(I - P_{X(l^\perp)})Xl, \beta) / \|q\|^2.$$

Now

$$X'(I - P_{X(l^\perp)})Xl = X'P_{X'^{-1}(\text{span}(l))}Xl = \lambda l \quad (27)$$

for some $\lambda \in \mathbb{R}$ and

$$\begin{aligned} \lambda \|l\|^2 &= \lambda(l, l) = \lambda(l, X'(I - P_{X(l^\perp)})Xl) \\ &= \lambda(Xl, (I - P_{X(l^\perp)})Xl) = \|q\|^2, \end{aligned} \quad (28)$$

i.e., $\lambda = \|q\|^2 \|l\|^{-2}$ and consequently

$$E(l, \hat{\beta}) = (l, \beta).$$

(iii) (c, y) is BLUE of its expectation $(X'c, \beta) = (c, X\beta)$ iff $c \in \text{im}(X)$.

This follows from the linear version of the Lehmann-Scheffé Theorem which says that (c, y) is BLUE iff it is uncorrelated with every linear unbiased estimator of 0. (b, y) is an unbiased estimator of zero iff $X'b = 0$. Thus (c, y) is BLUE iff $(c, b) = 0$ for every $b \in X^{-1'}(0)$, i.e., $c \in (X^{-1'}(0))^\perp = \text{im}(X)$. Since $q \in \text{im}(X)$, the proof is complete.

Corollary 3 *Let x_i be the i -th column of X and X_i be the matrix obtained from X by deleting the i -th column. Then*

$$\hat{\beta}_i = \frac{(y, (I - X_i(X_i'X_i)^{-1}X_i')x_i)}{(x_i, (I - X_i(X_i'X_i)^{-1}X_i')x_i)} \quad (29)$$

is BLUE of β_i .

□

Still the problem of an actual computation of the least squares estimator $(l, \hat{\beta})$ remains. To solve this we need an appropriate form of the Gram-Schmidt orthogonalization method.

Theorem 4. *Let x_1, \dots, x_k be the arbitrary vectors of the inner product vector-space V and let P_i be the orthogonal projection onto $\text{span}\{x_1, \dots, x_i\}$. If $q_1 = x_1$ and*

$$q_i = x_i - P_{i-1}x_i, \quad i = 2, \dots, k$$

then $\{q_1, \dots, q_k\}$ form a system of orthogonal vectors and $\text{span}\{x_1, \dots, x_k\} = \text{span}\{q_1, \dots, q_k\}$.

Proof. Firstly, we show that $\text{span}\{x_1, \dots, x_i\} = \text{span}\{q_1, \dots, q_i\}$, $i = 1, \dots, k$. This assertion is true for $i = 1$ since $x_1 = q_1$. Since by induction assumption $P_{i-1}x_i \in \text{span}\{x_1, \dots, x_{i-1}\} = \text{span}\{q_1, \dots, q_{i-1}\}$ it follows that $x_i = q_i + P_{i-1}x_i \in \text{span}\{q_1, \dots, q_i\}$. But $q_i \in x_i - P_{i-1}x_i \in \text{span}\{x_1, \dots, x_i\}$. Thus $\text{span}\{x_1, \dots, x_i\}$ and $\text{span}\{q_1, \dots, q_i\}$ coincide. Moreover, if $i < j$ then from $q_i \in \text{span}\{q_1, \dots, q_{j-1}\} = \text{span}\{x_1, \dots, x_{j-1}\}$ it follows that

$$\begin{aligned}(q_i, q_j) &= (q_i, x_j) - (q_i, P_{j-1}x_j) = (q_i, x_j) - (P_{j-1}q_i, x_j) \\ &= (q_i, q_j) - (q_i, q_j) = 0.\end{aligned}$$

Clearly, $P_i y = \sum_{j \leq i: q_j \neq 0} \frac{(q_j, y)q_j}{(q_j, q_j)}$ for both sides are equal to q_i if $y = q_i$ and vanish if $y \perp q_1, \dots, q_i$. From the properties of projections it also follows that

$$\|q_i\|^2 = \|x_i\|^2 - \|P_{i-1}x_i\|^2 = \|x_i\|^2 - \sum_{j \leq i: q_j \neq 0} \frac{(x_i, q_j)^2}{\|q_j\|^2}.$$

This is an important formula for making computations.

Now we are in a position to solve the computation-problem for $(l, \hat{\beta})$. Let l_1, \dots, l_{k-1} be an basis of $(l)^\perp$ and perform the Gram-Schmidt orthogonalization procedure for

$$Xl_1, \dots, Xl_{k-1}, Xl.$$

Then

$$\begin{aligned}q_k &= Xl - P_{\text{span}\{Xl_1, \dots, Xl_{k-1}\}}Xl \\ &= Xl - (P_{X(l)^\perp})Xl = (I - P_{X(l)^\perp})Xl.\end{aligned}$$

Thus

$$\frac{\|l\|^2(q_k, y)}{\|q_k\|^2} \tag{30}$$

is the BLUE of (l, β) .

In order to find the BLUE of $\beta_i = (e_i, \beta) = e_i' \beta$, e_i the i -th unit-vector, we have to apply the orthogonalization process to the sequence

$$x_1, \dots, x_{i-1}, x_k, x_{i+1}, \dots, x_{k-1}, x_i.$$

Having found one formula, the other formulae are just obtained by interchanging index k and i .

As an example let us at first consider the case $k = 2$. Then

$$q_1 = x_1, q_2 = x_2 - \frac{(x_2, x_1)}{(x_1, x_1)}x_1, \text{ if } x_1 \neq 0, \quad (31)$$

(if $x_1 = 0$, then $q_2 = x_2$ and we have indeed the case $k = 1$),

$$\|q_2\|^2 = (x_2, x_2) - \frac{(x_1, x_2)^2}{(x_1, x_1)}$$

and

$$\begin{aligned} \hat{\beta}_2 &= \frac{(q_2, y)}{\|q_2\|^2} \\ &= \frac{(x_2, y) - \frac{(x_2, x_1)}{(x_1, x_1)}(x_1, y)}{(x_2, x_2) - \frac{(x_1, x_2)^2}{(x_1, x_1)}} \\ &= \frac{(x_1, x_1)(x_2, y) - (x_2, x_1)(x_1, y)}{(x_2, x_2)(x_1, x_1) - (x_1, x_2)^2}. \end{aligned} \quad (32)$$

(if $q_2 = 0$, then β_2 is not estimable and $\hat{\beta}_1 = \frac{(x_1, y)}{(x_1, x_1)}$).

By interchanging Index 1 and 2 we get $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{(x_2, x_2)(x_1, y) - (x_2, x_1)(x_2, y)}{(x_2, x_2)(x_1, x_1) - (x_1, x_2)^2}. \quad (33)$$

From $R_1\hat{\beta} = \hat{\alpha}$, we get an alternative representation of $\hat{\beta}_1$, namely

$$\hat{\beta}_1 = \frac{(x_1, y)}{(x_1, x_1)} - \frac{(x_2, x_1)}{(x_1, x_1)}\hat{\beta}_2. \quad (34)$$

This gives a possibility to check the computation of $\hat{\beta}_1$ and $\hat{\beta}_2$ computationally and numerically.

Now let us also consider the example $k = 3$

$$q_1 = x_1, q_2 = x_2 - \frac{(x_2, x_1)}{(x_1, x_1)}x_1, q_3 = x_3 - \frac{(x_3, x_1)}{(x_1, x_1)}x_1 - \frac{(x_3, q_2)}{(q_2, q_2)}q_2, \quad (35)$$

$$\|q_3\|^2 = (x_3, x_3) - \frac{(x_3, x_1)^2}{(x_1, x_1)} - \frac{(x_3, q_2)^2(x_1, x_1)}{(x_2, x_2)(x_1, x_1) - (x_2, x_1)^2}. \quad (36)$$

After some rearrangement the formula

$$q_3 = x_3 - \frac{(x_3, x_2)(x_1, x_1) - (x_2, x_1)(x_3, x_1)}{(x_2, x_2)(x_1, x_1) - (x_2, x_1)^2} x_2 + \left(\frac{(x_2, x_1)}{(x_1, x_1)} \frac{(x_3, x_2)(x_1, x_1) - (x_2, x_1)(x_3, x_1)}{(x_2, x_2)(x_1, x_1) - (x_2, x_1)^2} - \frac{(x_3, x_1)}{(x_1, x_1)} \right) x_1 \quad (37)$$

is obtained.

By interchanging Index 3 and 1 we get:

$$q_3^{(1)} = x_1 - \frac{(x_1, x_2)(x_3, x_3) - (x_2, x_3)(x_3, x_3)}{(x_3, x_3)(x_2, x_2) - (x_2, x_3)^2} x_2 + \left(\frac{(x_2, x_3)}{(x_3, x_3)} \frac{(x_1, x_2)(x_3, x_3) - (x_2, x_3)(x_3, x_1)}{(x_2, x_2)(x_3, x_3) - (x_2, x_3)^2} - \frac{(x_3, x_1)}{(x_3, x_3)} \right) x_3 \quad (38)$$

and by interchanging Index 3 and 2:

$$q_3^{(2)} = x_2 - \frac{(x_3, x_2)(x_1, x_1) - (x_3, x_1)(x_2, x_1)}{(x_3, x_3)(x_1, x_1) - (x_3, x_1)^2} x_3 + \left(\frac{(x_2, x_1)}{(x_1, x_1)} \frac{(x_3, x_2)(x_1, x_1) - (x_2, x_3)(x_2, x_1)}{(x_3, x_3)(x_1, x_1) - (x_3, x_1)^2} - \frac{(x_2, x_1)}{(x_3, x_3)} \right) x_1. \quad (39)$$

Of course, $\hat{\beta}_1$ and $\hat{\beta}_2$ can be obtained from $\hat{\beta}_3$ and $\hat{\alpha}_i = \frac{(q_i, y)}{(q_i, q_i)}$, $i = 1, 2$

by solving $R_1 \hat{\beta} = \hat{\alpha}$.

A final remark on the computation. There are, of course, computer-programs such as R, SPSS etc. by which estimates of regression coefficients can easily and efficiently be calculated. But even with a pocket-calculator or similar equipment it is possible to calculate the formula derived in this section.

A pocket-calculator can calculate from a data-array x (an n -dimensional vector x) the (empirical) mean and the (empirical) standard-deviation (SD), if the numbers are plugged in. Moreover, for two arrays x and y the (empirical) linear regression $y = \alpha + \beta x$ can be computed by just plugging in all numbers. The above formulae show that from the inner products (x_i, x_j) and (x_i, y) the regression coefficients estimators can be obtained by some very few elementary calculations. The inner product (x, x) can be obtained from mean and variance. Now the estimator of regression coefficient β in the regression $y = \alpha + \beta x$ is given by

$$\hat{\beta} = \frac{(x, y) - n\bar{x}\bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (40)$$

Thus the inner product (x, y) can be obtained from mean, variance and regression coefficient. Another method is to use the formula

$$(x, y) = \frac{1}{2}((x + y, x + y) - (x, x) - (y, y)) \quad (41)$$

and computing the standard deviation of $x + y$. An alternative to (41) is the Jordan-v.Neumann formula

$$(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2) = \frac{1}{4}((x + y, x + y) - (x - y, x - y)) \quad (42)$$

Here mean and variance of $x + y$ and $x - y$ are needed.

3 Linear Sufficiency

Baksalary and Kala (1981) and Drygas (1983) have introduced the concept of a linearly sufficient statistic $z = Ty$ in the linear model $Ey = X\beta$, $\text{Cov}(y) = \sigma^2Q$. Ty is called linear sufficient if the BLUE of $E(y)$ can be computed from Ty alone. The following theorem was proved:

Theorem 5. *Ty is linearly sufficient iff*

$$\text{im}(X) \subseteq \text{im}(WT'), \quad (43)$$

where $W = Q + cXX'$ is such that $c \geq 0$ and $\text{im}(X) \subseteq \text{im}(W)$.

If we introduce the inner product $(x, y)_{W^+} = (x, W^+y) = x'W^+y$, then WT is just the adjoint T^* of T with respect to this inner product. Indeed, for $x, y \in \text{im}(W)$

$$\begin{aligned} (Tx, y) &= (x, T'y) = (WW^+x, T'y) = (W^+x, WT'y) \\ &= (x, WT'y)_{W^+} = (x, T^*y)_{W^+}. \end{aligned} \quad (44)$$

Thus $\text{im}(X) \subseteq \text{im}(T^*)$ is the more transparent formulation of linear sufficiency.

Let us now return to the model $E(y) = X\beta$, $\text{Cov}(y) = \sigma^2I$ and let $q_1, \dots, q_s, q_{s+1}, \dots, q_k$ be the vectors obtained from the columns of X by applying the Gram-Schmidt orthogonalization procedure to them. Without restricting generality it can be assumed that $q_{s+1} = q_{s+2} = \dots = q_k = 0$. Then let

$$Q = (q_1, \dots, q_s, 0, \dots, 0) = (q_1, \dots, q_k) \quad (45)$$

and

$$Q_1 = (q_1, \dots, q_s) \quad (46)$$

Theorem 6. $Q_1'y$ and $Q'y$ are linearly sufficient statistics.

Proof. Since $\text{im}(X) = \text{im}(Q) = \text{im}(Q_1)$ it follows from theorem 5 that $Q'y$ and $Q_1'y$ are linearly sufficient.

Linear sufficiency means that the BLUE of $X\beta$ can be obtained from $Q'y$ and $Q_1'y$, respectively. Since $Q(Q'Q)^-Q'y$ and $Q_1(Q_1'Q_1)^-Q_1'y$, respectively are the orthogonal projections onto $\text{im}(X)$, this property is evident.

The model $E(y) = X\beta$ can be split up into $X = (X_1 : X_2)$, where $X_2 = X_1A$. Thus $E(y) = X_1(\beta_1 + A\beta_2)$, where $\beta = (\beta_1', \beta_2')$. Then $X_1 = Q_1R_{11}$ and $X_2 = Q_1R_{12}$ and

$$Q_1(Q_1'Q_1)^-Q_1'y = Q_1R_{11}R_{11}^-R_{11}^-(Q_1'Q_1)^-Q_1'y = X_1R_{11}^-(Q_1'Q_1)^-Q_1'y. \quad (47)$$

Thus $R_{11}^-(Q_1'Q_1)^-Q_1'y$ is the BLUE of $\beta_1 + A\beta_2$ in the model $E(y) = X_1(\beta_1 + A\beta_2)$, $\text{Cov}(y) = \sigma^2I$.

A similar representation is more complicated for Q . At first

$$X = QR, \quad R = \begin{pmatrix} R_1 \\ \cdots \\ 0 \end{pmatrix} = \begin{pmatrix} R_{11} & \vdots & R_{12} \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix}. \quad (48)$$

By Searle (1971, p. 4) generalized inverse of R and $Q'Q$ can be found as follows:

$$R^- = \begin{pmatrix} R_{11}^- & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix}, \quad (Q'Q)^- = \begin{pmatrix} (Q_1'Q_1)^- & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix}. \quad (49)$$

Then $RR^-(Q'Q)^- = (Q'Q)^-$ and

$$Q(Q'Q)^-Q' = QRR^-(Q'Q)^-Q' = XR^-(Q'Q)^-Q'. \quad (50)$$

Thus $R^-(Q'Q)^-Q'y = \hat{\beta}$ can be considered as an estimator of β in the sense that $(l, \hat{\beta})$ is BLUE of (l, β) whenever (l, β) is estimable.

It is remarkable to note that in the model $E(Q'y) = Q'X\beta = (Q'Q)R\beta$, $\text{Cov}(Q'y) = \sigma^2(Q'Q)$ there is a linear unbiased estimator of β if and only if $Q'Q$ is regular, i.e., $s = k$. In this case

$$R^{-1}(Q'Q)^{-1}y = \hat{\beta} \quad (51)$$

is the only unbiased estimator. It is at the same time the Aitken-estimator

$$\begin{aligned} & ((Q'Q)R)'(Q'Q)^{-1}((Q'Q)R)^{-1}((Q'Q)R)'(Q'Q)^{-1}y \\ &= (R'(Q'Q)R)^{-1}R'y \\ &= R^{-1}(Q'Q)^{-1}(R')^{-1}R'y \\ &= R^{-1}(Q'Q)^{-1}y. \end{aligned}$$

4 Application: Diabetes Mellitus

Diabetes mellitus is a disease where the autoregulation of metabolism is disturbed. Normally after eating, the content of glucosis in the blood decreases very rapidly after perhaps one hour to a value below 100 mg/dl (5.55 mmol/l). Under diabetes mellitus it takes perhaps 4 hours to reach such a value - even under medicaments. However, by physical training it is possible to get a lower value perhaps already after two hours.

We will here discuss the behavior of glucosis during the night and the early morning. Though there is in general no intensive consumption of food in the late evening and during the night an acceptable value in the morning seems to pose a serious problem.

Some years ago it was said that it can not be recommended to use antidiabetic oral medicaments in the late evening unless you eat regularly during the night. To avoid hypoglycemia during the night (or counter-reactions), it was recommended to eat one bread-unit (12 g carbon-hydrats) just before bedtime. In later years the opinion about this procedure was changed and oral antidiabetic medicaments of an appropriate dose were also recommended before bedtime. There was, however, no recommendation of change of the amount of food that should be consumed at bedtime. The patients continued to eat one bread-unit (BU) just before bedtime. Thus simultaneously measures against too high and too low glucosis-values were taken. This seems to be a rather contradictory proposal. In my opinion with fixed medicaments, the amount of food just before bedtime should depend on the value of glucosis at this time. I finally decided that 100 mg/dl (5.55

mmol/l) and 150 mg/dl (8.32 mmol/l) should be the critical values for this decision. If the glucosis is equal or above 150 mg/dl (8.32 mmol/l) you should not eat anything. If however, the value is equal or below 100 mg/dl, then you should eat one bread-unit (1 BU). If the value is between 100 and 150, then the last meal of day is less accordingly. For example, if the value is 120, then 0.6 BU are eaten.

If you awake during the night between perhaps 1 A.M. and 2 A.M. you repeat the procedure of the evening. If the value is equal or below 100mg/dl, you eat 1 BU and you eat no BU for values equal or above 150 mg/dl. For values between 100 and 150, you eat a fraction of a BU.

I have pursued this method since some time and I have constructed and computed a regression model describing the behavior of the changes of the glucosis-values. We have the following Variables:

- y , change of the glucosis-values, i.e., the difference of the value during the night and the value in the evening or the difference of the values in the morning and value during the night
- x , the amount of food, measured in BU, taken in the evening and during the night, respectively and
- t , time between two measurements.

The model which will be studied in the sequel is as follows:

$$y = \alpha(t - D) + \beta x + \epsilon. \quad (52)$$

Here

$$D = I_{\{x>0\}} = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \end{cases}. \quad (53)$$

The idea behind this modelation is as follows: If you eat an amount x then the glucosis will increase and it will reach its highest point after approximately one hour.

Since the model (52) is difficult to treat by a pocket-calculator, we have changed it to

$$\hat{y} = \frac{y}{(t - D)} = \alpha + \beta \hat{x} + \delta, \quad \hat{x} = \frac{x}{(t - D)}. \quad (54)$$

Table 1. Night/Morning Measurements (Transformed)

Nr.	Date	y	t	x	$\hat{y} = y/(t - D)$	$\hat{x} = x/(t - D)$
1	26.3.07	-23	5.22	0.4	-5.4502	0.0948
2	28.3.07	-42	5.82	0.2	-8.7137	0.0415
3	29.3.07	30	3.12	1.0	14.1509	0.4717
4	30.3.07	-41	5.18	0.4	-9.1086	0.0957
5	31.3.07	-28	6.22	0	-4.5016	0
6	2.4.07	-29	5.28	0.6	-6.7757	0.1402
7	4.4.07	-13	3.53	0.6	-5.1383	0.2372
8	5.4.07	9	5.05	1.0	2.2222	0.2469
9	6.4.07	-48	5.17	0.4	-10.0719	0.0959
10	7.4.07	-18	3.42	0.2	-7.4380	0.0826
m =Mean		-19.7		0.48	-4.1525	0.15065
σ =Standard Deviation		23.4665		0.3293	7.3329	0.1368

Regression $\hat{y} = \hat{\alpha} + \hat{\beta}\hat{x}$, $\hat{y} = -11.2354 + 47.0156\hat{x}$, $r=0.8768090$
 $1 - r^2 = 0.2312059$, $\sum_{i=1}^{10} (\hat{y}_i - \bar{\hat{y}})^2 = 9\sigma_0^2 = 483.9428$,
 $\hat{\sigma}^2 = \frac{1}{8} \times 0.2312059 \times 483.9428 = 13.986304 = (3.7398)^2$.

Table 2. Evening/Night Measurements (Transformed)

Nr.	Date	y	t	x	$\hat{y} = y/(t - D)$	$\hat{x} = x/(t - D)$
1	25./26.3.07	-49	2.98	0	-1.6443	0
2	26./27.3.07	21	3.6	0	5.8333	0
3	27./28.3.07	-39	3.85	0	-10.1299	0
4	28./29.3.07	-85	3.73	0	-22.7882	0
5	29./30.3.07	0	3.57	0.2	0	0.0778
6	30./31.3.07	81	3.55	1	31.7647	0.3922
7	31.3/1.4.07	79	4.77	1	20.9549	0.2653
8	1.4./2.4.07	-38	3.73	0	-10.1877	0
9	2.4./3.4.07	-26	3.8	0	-6.8421	0
10	3.4./4.4.07	-3	5.03	0.6	-0.7444	0.1489
m =Mean		-5.9		0.28	0.62163	0.08842
σ =Standard Deviation		53.9247		0.4237	15.8256	0.1391

Regression $\hat{y} = \hat{\alpha} + \hat{\beta}\hat{x}$, $\hat{y} = -8.1859 + 99.6107\hat{x}$, $r = 0.8758071$
 $1 - r^2 = 0.23296$, $\sum_{i=1}^{10} (\hat{y}_i - \bar{\hat{y}})^2 = 9\sigma_0^2 = 2254.05$
 $\hat{\sigma}^2 = \frac{(1-r^2)}{8} \times \sum_{i=1}^{10} (\hat{y}_i - \bar{\hat{y}})^2 = 65.6378 = (8.1017)^2$

The variation of y is very high, so is the coefficient $\hat{\beta}$.

Since the value for the time t , as we will see later, is almost identical for most values, so it could be hoped that there will be no large difference in the estimated regression coefficients. We have to see whether this is actually the case.

We begin with the Night/Morning values. We now proceed to Table 3 with the original data.

Table 3. Night/Morning Measurements (Original)

Nr.	Date	y	t	x	$t - D$	$y + x$	$y + t - D$	$x + t - D$
1	26.3.07	-23	5.22	0.4	4.22	-22.6	-18.78	4.62
2	28.3.07	-42	5.82	0.2	4.82	-41.8	-37.18	5.02
3	29.3.07	30	3.12	1.0	2.12	31.0	32.12	3.12
4	30.3.07	-41	5.18	0.4	4.18	-40.6	-36.82	4.58
5	31.3.07	-28	6.22	0	6.22	-28	-21.78	6.22
6	2.4.07	-29	5.28	0.6	4.28	-28.4	-24.72	4.88
7	4.4.07	-13	3.53	0.6	2.53	-12.4	-10.47	3.13
8	5.4.07	9	5.05	1.0	4.05	10.0	13.05	5.05
9	6.4.07	-42	5.17	0.4	4.17	-41.6	-37.83	4.57
10	7.4.07	-18	3.42	0.2	2.42	-17.8	-15.58	2.62
m =Mean		-19.7		0.48	3.901	-19.22	-15.791	4.381
s =Standard Deviation		23.4665		0.3293	1.2421	23.7229	22.7683	1.0993
$SQ = 9s^2 + 10m^2$		8837		3.28	166.0631	8759.08	7161.6633	202.8071

$$\begin{aligned} (y, x) &= \frac{1}{2}((y + x, y + x) - (y, y) - (x, x)) = -40.6 \\ (y, t - D) &= \frac{1}{2}((y + t - D, y + t - D) - (y, y) - (t - D, t - D)) = -920.6999 \\ (x, t - D) &= \frac{1}{2}((x + t - D, x + t - D) - (x, x) - (t - D, t - D)) = 16.732 \\ \hat{\alpha} &= \frac{(y, t - D)(x, x) - (y, x)(t - D, x)}{(x, x)(t - D, t - D) - (x, t - D)^2} = \frac{-2340.5765}{264.7272} = -8.8415 \\ \hat{\beta} &= \frac{-(t - D, x)(y, t - D) - (t - D, t - D)(y, x)}{(x, x)(t - D, t - D) - (x, t - D)^2} = \frac{-8662.9888}{264.7272} = 32.7242 \\ &= \frac{(y, x) - (t - D, x)\hat{\alpha}}{(x, x)} = \frac{147.936 - 40.6}{3.28} = \frac{107.336}{3.28} = 32.7244 \end{aligned}$$

Thus the regression function equals $y = -8.8415(t - D) + 32.724x$.

In order to find the measure of determination we must compute (q_1, y) and (q_2, y) . Here

$$\begin{aligned} q_1 &= x, \\ q_2 &= (t - D) - \frac{(t - D, x)}{(x, x)}x \end{aligned}$$

$$\begin{aligned}
(q_1, y) &= (x, y) = -40.6, \\
(q_2, y) &= (t - D, y) - \frac{(t - D, x)}{(x, x)}(x, y) \\
&= -920.6999 + \frac{16.732}{3.28}40.6 = -713.5904, \\
(y, y) &= 8837, \\
(q_1, q_1) &= (x, x) = 3.28, \\
(q_2, q_2) &= (t - D, t - D) - \frac{(t - D, x)^2}{(x, x)} \\
&= 166.0631 - \frac{16.732^2}{3.28} = 80.7095, \\
(q_1, y)^2 &= 1648.36, \\
(q_2, y)^2 &= (713.5904)^2 = 509211.259 \\
R^2 &= 0.05687 + 0.71395 = 0.77082.
\end{aligned}$$

The estimator $\hat{\sigma}^2$ of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{8}(1 - R^2)(y, y) = 253.1580 = (15.9109)^2. \quad (55)$$

Finally we analyze the Evening/Night measurements in their original form, see Table 4 for this.

Table 4. Evening/Night Measurements (Original)

Nr.	Date	y	x	t	$t - D$
1	25./26.3.07	-49	0	2.98	2.98
2	26./27.3.07	21	0	3.6	3.6
3	27./28.3.07	-39	0	3.85	3.85
4	28./29.3.07	-85	0	3.73	3.73
5	29./30.3.07	0	0.2	3.57	2.57
6	30./31.3.07	81	1	3.55	2.55
7	31.3/1.4.07	79	1	4.77	3.77
8	1.4./2.4.07	-38	0	3.73	3.73
9	2.4./3.4.07	-26	0	3.8	3.8
10	3.4./4.4.07	-3	0.6	5.03	4.03
m =Mean		-5.9	0.28		3.461
s =Standard Deviation		53.1247	0.4237		0.5482
$9s^2 = \sum_{i=1}^n (z_i - \bar{z})^2$		26170.90	1.6160		2.4707
$SQ = 9s^2 + 10m^2$		26519	2.4		122.49

$$y = a_1 + b_1x, y = -36.1733 + 108.1188x, r = 0.8496$$

$$(y, x) - 10\bar{x}\bar{y} = 174.7199, (y, x) = 158.2$$

$$y = a_2 + b_2(t - D) = 98.6086 - 30.1961(t - D), r = -0.3070$$

$$(y, t - D) - 10\bar{y}(t - D) = -81.67098, (t - D) = -285.86998$$

$$t - D = a_3 + b_3x = -3.537 - 0.2715x, r = -0.20589$$

$$(t - D, x) - 10(t - D)\bar{x} = -0.4388, (t - D, x) = 9.252, (t - D, x)^2 = 85.5995$$

$$\text{Det} = (x, x)(t - D, t - D) - (x, t - D)^2 = 208.3762$$

$$A_1 = (t - D, t - D)(y, x) - (t - D, x)(y, t - D) = 22022.76414$$

$$A_2 = -(t - D, x)(y, x) + (x, x)(y, t - D) = -2149.7544$$

$$\hat{\beta} = \frac{A_1}{\text{det}} = 105.68752$$

$$\hat{\alpha} = \frac{A_2}{\text{det}} = \frac{[(y, t - D) - \hat{\beta}(x, t - D)]}{(t - D, t - D)} = -10.3167.$$

Thus the (estimated) regression function is

$$y = -10.3167(t - D) + 105.68752x \quad (56)$$

For the computation of the measure of determination let

$$q_1 = (t - D),$$

$$q_2 = x - \frac{(x, t - D)}{(t - D, t - D)}(t - D).$$

Then

$$\begin{aligned}
(q_1, y) &= -285.86998, \\
(q_1, q_1) &= 122.49, \\
(q_2, y) &= 158.2 - \frac{(9.252)(-285.86998)}{122.49} = 179.79523, \\
(q_2, q_2) &= (x, x) - \frac{(x, t - D)^2}{(t - D, t - D)} = 2,4 - \frac{9.252^2}{122.49} = 1.70117.
\end{aligned}$$

Thus

$$\begin{aligned}
R^2 &= \frac{1}{(y, y)} \left(\frac{(q_1, y)^2}{(q_1, q_1)} + \frac{(q_2, y)^2}{(q_2, q_2)} \right) \\
&= 0.025181 + 0.07165366 = 0.7417176.
\end{aligned}$$

In the evening/night-tables, the high variance of y and the high coefficients of the x -variables is very remarkable.

$$\hat{\sigma}^2 = \frac{1}{8}(1 - R^2)(y, y) = 854.2538 = (29.22762)^2.$$

It may perhaps be interesting how large the fastening actual values in the morning have been. These were as follows:

Date	26.3	27.3	28.3	29.3	30.3	31.3	1.4	2.4	3.4	4.4	5.4	6.4	7.4
mg/dl	104	99	97	103	98	124	94	98	98	104	108	86	123

This is very a good result because these values are normal or close to normality. The relative high values on 31th of March and 7th of April can be explained by an additional medicament taken on that day (Saturday).

References

- Baksalary JK and Kala, R (1981) Linear transformation preserving the best linear unbiased estimator in a general Gauss-Markov model. *Annals of Statistics* 9:913-916.
- Björck-Dahlquist (1972) *Numerische Methoden*. R.Oldenbourg Verlag München Wien.
- Drygas, H (1976) Weak and Strong Consistency of the Least Squares Estimator in Regression Models. *Zeitschr. f. Wahrsch.-Th. u. verwandte Gebiete*, 34:119-127.

- Drygas, H (1983) Sufficiency and Completeness in the general Gauss-Markov model, *Sankhya (A)*, Vol. 43, Part 1, 88-98
- Lawson CL and Hanson RJ (1974) *Solving Least Squares Problems*. Prentice-Hall Inc., New Jersey.
- Searle SR (1971) *Linear Models*. Wiley, New York