# Statistical Analysis of Diabetes Mellitus

Hilmar Drygas, University of Kassel, Germany

## 1 Introduction

Diabetes mellitus is a disease where the glucosis-content of the blood does not automatically decrease to a "normal" value between 70 mg/dl and 120 mg/dl (3,89 mmol/l and 6,67 mmol/l) between perhaps one hour (or two hours) after eating. Several instruments can be used to arrive at a relative low increase of the glucosis-content. Besides drugs (oral antidiabetica, insulin) the blood-sugar content can mainly be influenced by

(i) eating, i.e., consumption of the right amount of food at the right time

(ii) physical training (walking, cycling, swimming).

In a recent paper the author has performed a regression analysis on the influence of eating during the night. The result was that one "bread-unit" (12g carbon-hydrats) increases the blood-sugar by about 50 mg/dl, while one hour after eating the blood-sugar decreases by about 10 mg/dl per hour. By applying this result-assuming its correctness - it is easy to eat the right amount during the night and to arrive at a fastening blood-sugar (glucosis-content) in the morning of about 100 mg/dl (5,56 mmol/l).

In this paper we try to incorporate some physical exercise into the model. For every day a number is specified describing the physical activity during the past day. Mostly it consists of the kilometers walked during the past day. It turns out that the estimated regression coefficient associated with the physical exercise is negative. Unfortunately it is not significant. At the beginning of the night it is of moderate magnitude but in the morning it is very small.

The paper starts in the next two sections with some mathematical topics, namely Gram-Schmidt orthogonalization, Gauss-Markov theorem and its application to the estimation of regression coefficients. Then the linear model for the description of the behaviour of glucosis during the night is formulated and some empirical data from 2008 are analyzed. A concluding remark concerns the application of the results and the empirical outcome of the corresponding method.

# 2 Gram-Schmidt orthogonalization

Given vectors $x_1, \ldots, x_k$ of the inner-product vector-space $V$ with inner product $(x, y)$ the following task arises: Determine orthogonal vectors $q_1, \ldots, q_k$ such that $\mathrm{span}\{x_1, \ldots, x_i\} = \mathrm{span}\{q_1, \ldots, q_i\}$ , $i = 1, 2, \ldots, k$ .
From the representation

$$x_i = \sum_{j=1}^{i} \lambda_j \, q_j \tag{1}$$

we get $(x_i, q_j) = \lambda_j (q_j, q_j)$ and hence $\lambda_j$ is arbitrary if $q_j = 0$ and $\lambda_j = (x_i, q_j)/(q_j, q_j)$ if $q_j \neq 0$ . Thus

$$x_i = \sum_{j=1, q_j \neq 0}^{i} \frac{(x_i, q_j)}{(q_j, q_j)} \, q_j \tag{2}$$

and

$$q_i = \left( x_i - \sum_{j=1, q_j \neq 0}^{i-1} \frac{(x, q_j)}{(q_j, q_j)} \, q_j \right) \frac{(q_i, q_i)}{(x_i, q_i)} \tag{3}$$

if $(x_i, q_i) \neq 0$ . Therefore

$$\begin{aligned} q_i \; &= (x_i - P_{\mathrm{span}\{q_1, \ldots, q_{i-1}\}} x_i) \frac{(q_i, q_i)}{(x_i, q_i)} \\ &= \left( P_{\{q_1, \ldots, q_{i-1}\}^{\perp}} x_i \right) \frac{(q_i, q_i)}{(x_i, q_i)} \, , \end{aligned} \tag{4}$$

where $P_M y$ denotes the orthogonal projection of $y \in V$ onto the linear subspace $M \subseteq V$ . Since trivially

$$P_{\{q_1, \ldots, q_{i-1}\}^{\perp}} x_i \perp q_1, \ldots, q_{i-1} \tag{5}$$

it follows that

## 2.1 Theorem:

Let $q_0 = 0$ and $q_i = P_{\{q_0, \ldots, q_{i-1}\}^{\perp}} x_i, i = 1, \ldots, k$ . Then $q_i, i = 1, \ldots, k$ form on orthogonal system of vectors such that $\mathrm{span}\{q_1, \ldots q_i\} = \mathrm{span}\{x_1, \ldots, x_i\}, i = 1, \ldots, k$ .

**Proof:** From $q_0 = 0$ it follows that $q_1 = x_1$ and $\mathrm{span}\{q_1\} = \mathrm{span}\{x_1\}$ . If $\mathrm{span}\{x_1, \ldots, x_{i-1}\}$ and $\mathrm{span}\{q_1, \ldots, q_{i-1}\}$ coincide then it follows that $q_i = x_i - P_{\mathrm{span}\{q_1, \ldots, q_{i-1}\}} x_i = x_i - P_{\mathrm{span}\{x_1, \ldots, x_{i-1}\}} x_i \in \mathrm{span}\{x_1, \ldots, x_i\}$ and $x_i = q_i + P_{\mathrm{span}\{q_1, \ldots, q_{i-1}\}} x_i \in \mathrm{span}\{q_1, \ldots, q_i\}$ . Orthogonality follows from the symmetry of the inner product. □

Moreover $(x_i, q_i) = (x_i - P_{\mathrm{span}\{q_1, \ldots, q_{i-1}\}} x_i, q_i) = (q_i, q_i)$ and $(x_j, q_i) = (q_j, q_i) = 0$ if $i > j$ .

# 3 Gauss-Markov theorem, estimation of regression coefficients

Consider a linear model

$$Ey \in L \quad Cov\ y = Q\,, \tag{1}$$

where $y$ is a $n$-dimensional random vector and $L$ is a linear subspace of the $n$-dimensional vector-space $V$.

The Best Linear Unbiased Estimators (BLUE) or Gauss-Markov Estimators (GME) in this model is a linear mapping $Gy$ from $V$ to $V$ such that $Gy$ is an unbiased estimator of $Ey$ (i.e., $Gl = l$ for all $l \in L$) and $Gy$ possesses smallest Covariance-matrix among all linear unbiased estimators of $Ey$

## 3.1 (Gauss-Markov theorem 1):

$Gy$ is BLUE of $Ey$ in the model $Ey \in L$, $Cov\ y = Q$ iff

(i) $Gy = y \ \forall y \in L$

(ii) $GQy = 0$ if $y \in L^{\perp}$ ($Gy$ is the projection onto $L$ along $QL^{\perp}$).

**Proof:** Drygas (1970), page 55.
Besides this theorem another Gauss-Markov theorem is important. It concerns the estimation of a simple parametric function $(a, Ey)$.

## 3.2 (Gauss-Markov theorem 2):

$(a, y)$ is BLUE of $(Ey, c)$ iff

(i) $a - c \in L^{\perp}$

(ii) $Qa \in L$.

Such an estimator always exists.

**Proof:** From $L \cap QL^{\perp} = \{0\}$ it follows that $L^{\perp} + Q^{-1}(L) = V$.
Therefore for given $c \in V$, $c = b + a$, $Qa \in L$ and $b \in L^{\perp}$. Since $(c - a) = b \in L^{\perp}$ it follows that $(a, y)$ is an unbiased estimator of $(c, Ey)$ and $Qa \in L$. Let $a_1 \in V$ an alternative element of $V$ such that $c - a_1 e L^{\perp}$, i.e., $(a_1, y)$, is an unbiased estimator of $(a, Ey)$. Then

$$\begin{aligned}(Qa_1, a_1) &= \mathrm{Var}(a_1, y) = (Q(a_1 - a) + a, (a_1 - a) + a)) \\ &= (Q(a - a_1), (a - a_1) + (Qa, a) + 2(Qa, (a - a_1))\,.\end{aligned} \tag{2}$$

The letter expression vanishes since $Qa \in L$ and $a - a_1 = (a - c) - (a_1 - c) \in L^{\perp}$. Thus

$$(Qa_1, a_1) = (Q(a - a_1), (a - a_1)) + (Qa, Qa) \geq (Qa, a) \tag{3}$$

with equality iff $(Q(a-a_1),(a-a_1))=0$, i.e., $Q(a_1-a)=0$ or $Qa_1=Qa\in L$.

This Proof ist essential a linear version of the Lehmann-Scheffé theorem which says that an estimator $(a,y)$ is BLUE iff $(a,y)$ is uncorrelated with any unbiased estimator $(d,y)$ of $0$. $(d,y)$ (Schmetterer, 1966, p. 332) is an unbiased estimator of $0$ iff $d\in L^\perp$ and we get the condition $(Qa,d)=0\ \forall d\in L^\perp$, i.e., $Qa\in L^{\perp\perp}=L$. This approach is discussed in some detail in Sengupta/Jammalamadaka (2003). □

We want to apply this theorem to the estimation of $\beta_k$ in the case where

$$L=\{X\beta=x_1\beta_1+x_2\beta_2+\ldots+x_k\beta_k\} \tag{4}$$

where $X=(x_1,\ldots,x_k)$. The case where the estimation of $(l,\beta),l\in\mathbb{R}^k$ is desived can be reduced to this case as follows. Let $l_1,\ldots,l_{k-1}$ be can orthogonal basis of $(l)^\perp$ and $l_k=l$. Then

$$X\beta=\sum_{i=1}^k Xl_i(l_i,l_i)^{-1}(l_i,\beta) \tag{5}$$

as can easily be verified for $\beta=l_i$, $i=1,\ldots,k$. Let

$$z_i=Xl_i(l_i,l_i)^{-1},\ \gamma_i=(l_i,\beta)\,. \tag{6}$$

Then $X\beta=\sum\limits_{i=1}^k z_i\gamma_i$ and the estimation of $\gamma_k$ is desired.

## 3.3 Theorem:

Let $G_1y$ be the BLUE of $Ey$ in the model $Ey\in\mathrm{im(X_1)}$, $\mathrm{X_1=(x_1,\ldots,x_{k-1})}$ such that $\mathrm{im(G_1)\subseteq im(X_1)}$. Then if

$$W=Q+cXX' \tag{7}$$

$c\geq 0$ such that $\mathrm{im(X)\subseteq im(W)}$ it follows that $(a,y)$ is BLUE of $\beta_k$ iff

$$Wa=\lambda(I-G_1)x_k \tag{8}$$

for some $\lambda\in\mathbb{R}$ and $(a,x_k)=1$.

**Proof:** $(a,y)$ is an unbiased estator of $\beta_k=(e_k,\beta)$, $e_k=(0,\ldots,0,1)'$, the $k$-th unit-vector iff $(a,X\beta)=(\beta,e_k)$ for all $\beta\in\mathbb{R}^k$. If we let $\beta=e_i$, the $i$-th unit-vecor, $i=1,\ldots,k-1$ then $(a,x_i)=0$ and if we let $\beta=e_k$ then $(a,x_k)=1$. The optimality condition of theorem 3.2 tells us that

$$Qa\in L\,,\quad \text{i.e}\,,\quad Qa=a_1+\mu x_k\,, \tag{9}$$

where $a_1\in\mathrm{im(X_1)}$ and $\mu\in\mathbb{R}$. This is equivalent to

$$Wa=a_2+\mu_1 x_k \tag{10}$$

where $a_2 \in \mathrm{im}(\mathrm{X}_1)$ and $\mu_1 \in \mathbb{R}$. Since a $\in (\mathrm{im}\,(X_1))^{\perp}$ it follows that

$$G_1 W a = 0 \tag{11}$$

and

$$Wa = (I - G_1)Wa = (I - G_1)(a_2 + \mu_1 x_k) = \mu_1 (I - G_1) x_k \,. \tag{12}$$

□

The question now arises how the equation $Wa = \mu_1(I - G_1)x_k$ can be solved. One attempt may be

$$a = \frac{W^-(I - G_1)x_k}{(x_k, W^-(I - G_1)x_k)} \tag{13}$$

where $W^-$ is a $g$-inverse $W$, i.e, $WW^-W = W$.
This formula is indeed correct if
(i) $W^-$ is $n \cdot n \cdot d$ and $W^-(I - G_1)$ is symmetric

and

(ii) $x_k \notin \mathrm{im}(\mathrm{X}_1)$.

If $x_k \in \mathrm{im}(\mathrm{X}_1)$ then $\beta_k$ is not estimable. Indeed, then $Xl_1 = 0$ for some $l_1 = (l_{11}, \ldots, l_{1k})'$ and $l_{1k} \neq 0$. But $(l_1, l) = l_{1k} \neq 0$ and $l = e_k \notin \mathrm{im}(\mathrm{X}')$, i.e., $(\beta, e_k) = \beta_k$ is not estimable.

If $W^- = W^+$, the Moore-Penrose inverse of $W$ and

$$G_1 = X_1 (X_1' W^+ X_1)^+ X_1' W^+$$

then $W^+ G_1$ is symmetric, $W^+$ is n.n.d and $\mathrm{im}(\mathrm{G}_1) \subseteq \mathrm{im}(\mathrm{X}_1)$, $(\mathrm{I} - \mathrm{G}_1)'\mathrm{W}^-(\mathrm{I} - \mathrm{G}_1) = \mathrm{W}^-(\mathrm{I} - \mathrm{G}_1)(\mathrm{I} - \mathrm{G}_1) = \mathrm{W}^-(\mathrm{I} - \mathrm{G}_1)$. It follows that the dedominator in (12) is equal to

$$((I - G_1)x_k \,,\, W^+(I - G_1)x_k) \,. \tag{14}$$

This expression ranishes iff $W^+(I - G_1)x_k = 0$ or $WW^+(I - G_1)x_k = (I - G_1)x_k = 0$ or $x_k \in \mathrm{im}(\mathrm{X}_1)$. This is just the case when $\beta_k = (\beta, e_k)$ is not estimable.

The question is now to compute a in (12). If $Q = I$, then orthogonalizing $x_1, \ldots, x_k$ by the Gram-Schmidt orthogonaliziation procedure yields $q_k = (I - G_1)x_k$ In the general case the most elegant approach is to change the inner product to $(x, y)_0 = (x, W^+ y)$. Then $Cov\ y = I$ with respect to this inner product and again the Gran-Schmnidt orthogonalization procedure yields to the desired estimator. (See Drygas (2008)).

Perhaps also $\beta_{k-1}, \ldots, \beta_1$ are to be estimated. One way is to change indices. This is not a very economic approach. From

$$x_j = \sum_{i=1}^{j} \frac{(x_j, q_i)}{(q_i, q_i)}\, q_i \tag{15}$$

it follows that

$$X = QR\,, \tag{16}$$

$$Q = (q_1 \ldots q_k)\,,\ R = \left(\frac{(x_j, q_i)}{(q_i, q_i)}\right). \tag{17}$$

The decomposition is called $QR$-decomposition. Since $(x_j, q_i) = 0$ if $i > j$ and $(x_i, q_i) = (q_i, q_i)$ it follows that $R$ is an upper triangular matrix with diagonal elements equal to one. In Drygas (2008) it has been show that the BLUE of $\hat{\beta}$ can be obtained by solving the equation

$$R\hat{\beta} = \hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_k) \tag{18}$$

where

$$\hat{\alpha}_i = \frac{(q_i, y)}{(q_i, q_i)}\,. \tag{19}$$

Since $R$ is uppes triangular, the equation $R\hat{\beta} = \hat{\alpha}$ can succesively be solved beginning with $\hat{\beta}_k$.

There is an alternative more statistical approach to solve this problem. Since

$$E(y - x_k\beta_k) = \sum_{i=1}^{k-1} x_i\beta_i \tag{20}$$

the BLUE of $\beta_{k-1}$ in this model is given by

$$\frac{(q_{k-1}\,,\, y - x_k\beta_k)}{(q_{k-1}\,,\, q_{k-1})} = \tilde{\beta}_{k-1}\,. \tag{21}$$

Unfortunately $\beta_k$ is unknown. But if we replace in (20) $\beta_k$ by its BLUE $\hat{\beta}_k$ the assertion concerning the expectation is still correct. The assention is now that

$$\frac{(q_{k-1}\,,\, y - x_k\hat{\beta}_k)}{(q_{k-1}\,,\, q_{k-1})} = \hat{\beta}_{k-1}\,. \tag{22}$$

is BLUE of $\beta_{k-1}$. Since $\hat{\beta}_k$ is BLUE of $\beta_k$ there exist a vector $a_k \in \mathrm{im(X)}$ such that $\hat{\beta}_k = (a_k, y)$. Thus

$$\hat{\beta}_{k-1} = \frac{(q_{k-1}, y) - (q_{k-1}, x_k)(a_k, y)}{(q_{k-1}, q_{k-1})} \tag{23}$$

$$= \frac{(q_{k-1} - (q_{k-1}, x_k)a_k, y)}{(q_{k-1}, q_{k-1})}\,. \tag{24}$$

Since $q_{k-1} - (q_{k-1}, x_k)a_k \in \mathrm{im\,X}$, this estimator is the BLUE of $\beta_{k-1}$.

## 3.4 Theorem:

Let $\hat{\beta}_i$ , $\hat{\beta}_{i+1}, \ldots, \hat{\beta}_k$ be the BLUE of $\beta_i, \beta_{i+1}, \ldots, \beta_k$ , where $i \geq 2$ .

$$\hat{\beta}_{i-1} = \frac{\Big(q_{i-1} \, , (y - \sum_{j=i}^{k} x_j \hat{\beta}_j)\Big)}{(q_{i-1} \, , \, q_{i-1})}$$

is BLUE of $\beta_{i-1}$ .

**Proof:** Since $E(y - \sum_{j=i}^{k} x_j \hat{\beta}_j) = \sum_{j=1}^{i-1} x_j \beta_j$ and $q_{i-1} = x_{i-1} - P_{\text{span}\{x_1,\ldots,x_{i-2}\}} x_{i-1}$ it follows that $E((q_{i-1} \, , \, y - \sum_{j=i}^{k} x_j \hat{\beta}_j) = \beta_{i-1}(q_{i-1} \, , \, x_{i-1}) = \beta_{i-1}(q_{i-1} \, , \, q_{i-1})$ . Hence $\hat{\beta}_{i-1}$ is an unbiased estimator of $\beta_{i-1}$ .
Since $\hat{\beta}_j$ is BLUE of $\beta_j$ it follows that there is an element $a_j \in \text{im(X)}$ such that $\hat{\beta}_j = (y, a_j)$ . Thus

$$\hat{\beta}_{i-1} = \frac{(q_{i-1} - \sum_{j=i}^{k} (q_{i-1}, x_j) a_j, y)}{(q_{i-1}, q_{i-1})} = (a_{i-1}, y) \, . \tag{25}$$

Since $a_{i-1} \in \text{im(X)}$ , the theorem is proved. □
Theorem 3.4 shows the difference between Gauss-Markow estimation and Least Squares Estimation. While it is possible to find a GME/BLUE of a linear function $(l, \beta)$ by Least Squares we can only obtain an estimator of the whole vector $\beta$ . A minimization of

$$Q = (y - x_1\beta_1 - \ldots x_k\beta_k \, , \, y - x_1\beta_1 - \ldots - x_k\beta_k) \tag{26}$$

with respect to $\beta_1$ would yield

$$\hat{\beta}_1 = \frac{(y - \sum_{i=2}^{k} \beta_i x_i, x_1)}{(x_1, x_1)} \quad \text{if} \quad x_1 \neq 0 \, . \tag{27}$$

If $x_1 = 0$ , then $\hat{\beta}_1$ can be chosen arbirary Plugging $\hat{\beta}_1$ into (24) yields

$$Q = \left\| (y - \frac{(x_1, y)}{(x_1, x_1)} x_1 - \sum_{i=z}^{k} \beta_i (x_i - \frac{(x_i, x_1)}{(x_1, x_1)} x_1) \right\|^2 \, . \tag{28}$$

Now the minimization process can be continued and it is possible to find the least squares estimators by mathematical induction.

## 3.5 Remark:

There is also an easy approach for obtaining the least squares estimator without using the QR-decomposition. The minimizet of $\| \, y - X\beta \, \|^2$ is the orthogonal projection of $y$ onto $\text{im(X)}$ , i.e.

$$X\hat{\beta} = P_{\text{im(X)}} y \, . \tag{29}$$

Let $X\hat{\beta} = x_1\hat{\beta}_1 + \ldots + x_k\hat{\beta}_k$ This is orthogonal projection if

$$(y - X\hat{\beta}, a) = 0 \ \forall a \in \mathrm{im}(X)\,. \tag{30}$$

By taking $a = X\beta$, the normal equations $X'X\hat{\beta} = X'y$ are obtained. (30) is, however, correct if it valid for a basis of im(X). If $x_1, \ldots, x_m$ are linear independent and form a basis of im(X), then by choosing $a = x_1, \ldots, x_m$ the equations

$$X_1'X\beta = X_1'y \tag{31}$$

is obtained, where $x_1 = (x_1, \ldots, x_m)$. Now if we replace $\{x_1, \ldots, x_m\}$ by $\{q_1, \ldots, q_m\}$, an orthogonal basis of $\mathrm{span}\{x_1, \ldots, x_m\}$ then

$$(y - X\hat{\beta}, q_i) = 0\,,\ i = 1, \ldots, m \tag{32}$$

is the necessary and sufficient condition for the Least Squares Estimators. Since $(x_j, q_i) = \delta_{ij}(q_i, q_i)$ for $j \le i$ it follows that we arrive at the triangular equation system

$$y - x_m\hat{\beta}_m - \sum_{j=m+1}^{k} \beta_j(x_j, q_m) = 0 \tag{33}$$

or

$$(y - \sum_{j=m+1}^{k} \beta_j x_j, q_m) = \hat{\beta}_m(q_m, q_m) \tag{34}$$

and

$$(y - \sum_{j=i+1}^{m} \hat{\beta}_j x_j - \sum_{j=m+1}^{k} \beta_j x_j, q_i) = \hat{\beta}_i(q_i, q_i)\,,\ i = m-1, \ldots, 1\,. \tag{35}$$

We see that $\beta_{m+1}, \ldots, \beta_k$ are completely arbitrary. An unique solution is only available if $m = k$.

# 4 Statistical Analysis of Diabetes Mellitus

This section is devoted to the study of the behaviour of blood-sugar during the night. The following strategy is followed to control the blood-sugar and to arrive at a "near normal" value in the morning:

The blood-sugar is measured in the evening just before bedtime. If the blood-sugar is above 150 mg/dl (8,32 mmol/l) nothing is eaten. If the value is 100 mg/dl (5,55 mmol/l) or below then one bread-unit (BE) (12g Carbon-hydrats) is eaten. If the value is between 100 mg/dl (5,55 mmol/l) and 150 mg/dl (8,32 mmol/l) then a smaller amount is eaten. For example if the value is 120 mg/dl (6,66 mmol/l), then 0,6 BE are eaten (linear interpolation). The blood-sugar is again measured during the night at perhaps 2 a.m. or 3 a.m. and it is assumed that the following is approximately correct: one bread-unit increases the blood-sugar within one hour by about 50 mg/dl (2,77 mmol/l). After this hour the blood-sugar decreases by about 10 mg/dl (0,55 mmol/l) per hour. As an example consider the following situation (1.10.2008): At 2.11. a.m. a value of 93 mg/dl (5,16 mmol/l) is measured. The decision was now to eat 0,8 BE. According to the assumption made before the blood-sugar increases to 133 mg/dl (7,38 mmol/l) and will arrive at about 6.11. a.m. at a value of 103 mg/dl (5,72 mmol/l). The actual value at 7,24 a.m. was 108 mg/dl (5,99 mmol/l).

My diabetic career began in 1974 just at the end of the era Nixon. The fastening value was 230 mg/dl (12,76 mmol/l). The proposed therapy consisted of taking one tablet of a very well known sulfonylurea both in the morning and in the evening. A physician at another place declared that a tablet should not be taken in the evening unless some food is eaten during the night. Since the winter-term 1974/1975 I worked at the University of Frankfurt am Main. I decided to consult the endocrinologist Karl Schöffling at the Klinikum of the Johann Wolfgang Goethe-Universität. Concerning the tablets he declared:

"*Man soll nicht mit Kanonen auf Spatzen schießen.*"

As a consequence of this statement the food-strategy was changed according the principle "Eat the right at the right time." Drugs were only occasionally taken until 1982/83. After a new visit in the clinic of Karl Schöffling 1 mg of a not so well known sulfonylurea was taken in the morning. This medication remained valid for a long time until 1994. Karl-Heinz Usadel, the successor of Karl Schöffling at the chair for endocrinology at the JWG-Universität, now increased the sulfonylurea to 3 mg per day. 2 mg should be taken in the morning, 1 mg in the evening. The latter proposal was very surprising to me. The explanation was that the opinion about the drugs has changed. Also during the night insulin is needed. There was, however, no change concerning the food that should be consumed in the evening. Up to now it is recommended that the patient should eat one bread-unit (12 g carbon-hydrats) just before bedtime.

It is a rather contradictory approach to take simulatanmeusly measures against both high and low blood-sugar values. This can't be correct. At least the amount of food consumed just before bedtime should depend on the blood-sugar at this time. In this way I arrived at the

(100, 150)-rule. Together with the (10,50)-rule applied during the night an optimal fastening value can be obtained in the morning besides some exceptional situations (flight, sickness etc.) The drugs in the evening are fixed. They are not subject to any change. Besides 1 mg of a very well known third generation sulfonylurea, 1 mg of a sensitizer and
250 mg of a bigunaid is taken. The following figure shows the successfulness of the method.

Figure 1, Fastening Glucosis 2007 (every 8th day)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 117 | 102 | 114 | 83 | 90 | 105 | 126 | 111 | 91 | 89 | 90 | 98 | 123 |
| 118 | 94 | 110 | 70 | 108 | 94 | 95 | 95 | 100 | 100 | 147 | 95 | 93 |
| 85 | 98 | 123 | 101 | 119 | 104 | 97 | 97 | 99 | 100 | 127 | 151,5 | 102 |
| 101 | 103 | 115,5 | 102 | 126 | 115 | | | | | | | |

Mean $m = 105,369565\ldots$, Standard-Deviation (SD) $s = 15,16884\,, s(46) = 15,003056$
Computed $HbA_{1c} = (m + 86)/33,3 = 5,7468339\,.$

In a recent paper (Drygas (2008)) the author has studied the behavivur of glucosis during the night. The following models were formulated:

$$\text{(I)} \quad y = \alpha + \beta x/(t - D) + \epsilon \tag{1}$$

$$\text{(II)} \quad = \alpha(t - D) + \beta(x) + \gamma\,. \tag{2}$$

Here $y$ is the difference of the glucosis-values either between night-time and the evening of the past day or between the morning-time and the night-time. $x$ is the amount of food consumed in the evening of the past day and during the night, respectively. $t$ is the time passed between the two measurements in the night and in the morning, respectively. $D = I_{\{x>0\}}$, i.e. $D = 1$, if something is eaten and $0$ otherwise. $\alpha$ and $\beta$ are regression parameters and $\epsilon$ and $\gamma$, respectively are the disturbance termes.
The stochastic assumptions are

$$E(\epsilon) = 0, E(\gamma) = 0\,,\ E(\epsilon\epsilon') = \sigma^2 I\,,\ E(\gamma\gamma') = \sigma_0^2 I\,. \tag{3}$$

The estimated parameters were as follows:

Figure 2 Estimated regression coefficients

| | $\alpha$ | $\beta$ |
|---|---|---|
| Evening/Night I | $-8,159$ | $99,6107$ |
| Evening/Night II | $-10,3167$ | $105,68752$ |
| Night Morning I | $-11,2354$ | $47,0156$ |
| Night Morning II | $-8,8415$ | $32,724$ |

It is supposed that these results support the (10, 50)-hypothesis.

The idea behind the models is that the blood-sugar increases by about $\beta$ mg/dl within one hour and decreases thereafter by about $-\alpha$ mg/dl per hour.

In this paper we want to extend this model by the inclusion of physical exercise during the past day. It was observed that the case of intensive physical activity during the past day the glucosis-content was very low during the night.
The new model is

$$y = \alpha(t - D) + \beta x + \gamma B + \epsilon \tag{4}$$

where B ("Bewegung") measures the amount of physical activity during the past day. As already mentioned in the introduction B mostly consists of the kilometers of walking performed during the past day. The data are as follows:

| Evening / Night | $y = \alpha(t-D) + \beta x + \gamma B + \epsilon\,,\ D = I_{\{x>0\}}$ | | | | |
|---|---|---|---|---|---|
| Date | y | t | t-D | x | B |
| 1. 11./12.4.08 | -52 | 3,52 | 2,52 | 1,0 | 10 |
| 2. 12./13.4.08 | + 11 | 4,83 | 3,83 | 0,5 | 7 |
| 3. 13./14.4.08 | + 27 | 4,98 | 3,98 | 1,0 | 7 |
| 4. 14./15.4.08 | - 21 | 3,3 | 3,3 | 0 | 3 |
| 5. 15./16.4.08 | -53 | 5,8 | 4,8 | 0,2 | 9 |
| 6. 16./17.4.08 | +9 | 3,25 | 2,25 | 0,5 | 3 |
| 7. 17./18.4.08 | -33 | 4,57 | 4,57 | 0 | 6 |
| 8. 18./19.4.08 | +31 | 5,2 | 4,2 | 0,8 | 6 |
| 9. 19./20.4.08 | +9 | 2,72 | 2,72 | 0 | 4 |
| 10. 20./21.4.08 | -79 | 4,95 | 4,95 | 0 | 1 |
| Mean $m$ | $-15,1$ | | 3,712 | 0,4 | 5,6 |
| $\sigma = \sigma_9$ | 37,9720006 | | 0,9711711 | 0,4189935 | 2,836273 |
| $9\sigma_g^2 = \sum_{i=1}^{10}(z_i - \tilde{z})^2$ | $12976,914$ | | $8,48856$ | $1,58$ | $72,4$ |
| $(z,z) = 9\sigma_g^2 + 10_m^2$ | 15257 | | $146,278$ | $3,18$ | 386 |
| $y = a_4 + b_4 B,$ | $y = -16,925414 + 0,3259668B\,, r = 0,0243$ | | | | 476 |
| $y = -30,188608$ | $+37,721519x, r = 0.4162289$ | | | | |
| $(y,z)$ | | | $-111,561$ | $-0,8$ | $-822$ |
| $(B,z)$ | | | $208,17$ | $28,6$ | |
| $(x,z)$ | | | $13,86$ | | |
| $q_1 = x,$ | $q_2 = (t-D)-$ | $\frac{(t-D,x)}{(x,x)}x,\ q_3$ | $= B - \frac{(B,x)}{(x,x)}x$ | $-\frac{(B,q_2)}{(q_2,q_2)}q_2$ | |
| $(q_i,q_i)$ | | | $-668,59321$ | $3,18$ | $47,55042$ |
| $(q_i,y)$ | | | $85,869321$ | $-0,8$ | $-164,5260714$ |
| $y = \hat{\alpha}(t-D) +$ | $\hat{\beta}_x + \hat{\epsilon} = -7,786171(t-D) + 33,68431x + \hat{\epsilon}$ | | | | |
| $y = \tilde{\alpha}(t-D) +$ | $\tilde{\beta}_x + \tilde{\gamma}B + \tilde{\epsilon} = -4,4209447(t-D) + 50,15828x - 3,460034B + \tilde{\epsilon}$ | | | | |

The measures of determination are as follows:

$$\hat{R}^2 = \frac{1}{(y,y)}\Big(\frac{(q_1,y)^2}{(q_1,q_1)} + (\frac{(q_2,y)^2}{(q_2,q_2)}\Big) =$$

$$\frac{1}{15257}\Big(\frac{0,64}{3,18} + \frac{(668,59321)^2}{85,869321}\Big) = 0,34129\ldots \tag{5}$$

$$R^2 = \frac{1}{15257}\frac{(q_3,y)^2}{(q_3,q_3)} + \hat{R}^2 = 0,3785\ldots \tag{6}$$

Thus the explanation of the data via both models is still very poor. Moreover, the coefficient $\tilde{\gamma}$ is not significant, i.e., the hypothesis $\gamma = 0$ can not be rejected in the case of a normal distribution of $y$. The test-statistic is

$$\frac{\tilde{\gamma}}{(\mathrm{Var}\,(\tilde{\gamma}))^{\frac{1}{2}}} = \frac{\tilde{\gamma}(q_3,q_3)^{\frac{1}{2}}}{\sigma} \tag{7}$$

which follows a normal distribution $N(0,1)$. Since $\sigma$ is unknown it will be replaced by an estimator $\tilde{\sigma}$. This estimator is obtained from

$$\tilde{\sigma}^2 = s^2 = \frac{1}{7}(1-\tilde{R}^2)(y,y) = 1354,5753 = (36,804528)^2 \tag{8}$$

The test-statistic is therefore

$$t_7 = \frac{\tilde{\gamma}(q_3,q_3)^{\frac{1}{2}}}{s} = \frac{\tilde{\gamma}\cdot 6,8956523}{36,804528} = -0,6482678\ldots \tag{9}$$

This value is not significant for a $t$-distribution with 7 degrees of freedom.

| Night / Morning | $y = \alpha(t-D) + \beta x + \gamma B + \epsilon\,,\ \ D = I_{\{x>0\}}$ | | | | |
|---|---|---|---|---|---|
| Date | y | t | t-D | x | B |
| 1. 12.4.08 | +13 | 5,2 | 4,2 | 0,8 | 10 |
| 2. 13.4.08 | -34 | 3,62 | 3,62 | 0 | 7 |
| 3. 14.4.08 | -25 | 5,23 | 4,23 | 0,2 | 7 |
| 4. 15.4.08 | - 37 | 5,75 | 4,75 | 0,2 | 3 |
| 5. 16.4.08 | + 4 | 2,97 | 1,97 | 0,4 | 9 |
| 6. 17.4.08 | -21,5 | 5,43 | 4,43 | 0,25 | 3 |
| 7. 18.4.08 | -31 | 4,45 | 4,45 | 0 | 6 |
| 8. 19.4.08 | -31 | 3,33 | 3,33 | 0 | 6 |
| 9. 20.4.08 | -38,5 | 4,92 | 4,92 | 0 | 4 |
| 10. 21.4.08 | -14 | 3,75 | 2,75 | 0,14 | 1 |
| Mean $m$ | $-21,5$ | | $3,865$ | $0,199$ | $5,6$ |
| $\sigma = \sigma_9$ | $17,559423$ | | $0,9418451$ | $0,2507965$ | $2,836273$ |
| $\sum_{i=1}^{10}(z_i - \overline{z})^z = 9\sigma_9^2$ | $2775$ | | $7,98365$ | $0,5660899$ | $2,836273$ |
| $(z,z) = 9\sigma_g^2 + 10m^2$ | $7397,5$ | | $157,3659$ | $0,9621$ | $386$ |
| $(y,z)$ | | | $-906,445$ | $-7,735006$ | $-962,5$ |
| $(\beta,z)$ | | | | $14,49$ | $211,33$ |
| $(x,z)$ | | | $7,4365$ | | |
| $q_1 = x,$ | $q_2 = (t-D) - \frac{(t-D,x)}{(x,x)}x,\ q_3 = B - \frac{(B,x)}{(x,x)}x - \frac{(B,q_2)}{(q_2,q_2)}q_2$ | | | | |
| $(q_i,q_i)$ | | | $99,467772$ | $0,9621$ | $68,57568$ |
| $(q_i,y)$ | | | $-846,36015$ | $-7,735$ | $-0,81453$ |
| $y = a_1 + b_1 x$ | $= -38,821274 + 61,91595x, r = 0,8843288$ | | | | |
| $y = a_2 + b_2 B$ | $= -40,179558 + 3,3356354B, r = 0,538786$ | | | | |
| $y = \tilde{\alpha}(t-D)+$ | $\hat{\beta}x + \hat{e} = -8,508882(t-D) + 57,968077x + \hat{e}$ | | | | |
| $y = \tilde{\alpha}(t-D)+$ | $\tilde{\beta}x + \tilde{\gamma}B + \tilde{e} = -8,4970268(t-D) + 58,15905x - 0,0118778B + \tilde{e}$ | | | | |

The coefficient of B is very small and clearly not significant. The measures of determination are

$$\hat{R}^2 = \frac{1}{(y,y)}\Big(\frac{(q_1,y)^2}{(q_1,q_1)} + \frac{(q_2,y)^2}{(q_2,q_2)}\Big) = 0,9735159$$

and

$$R^2 = \hat{R}^2 + \frac{(q_3,y)^2}{(q_3,q_3)}\,\frac{1}{(y,y)} = 0,97351720785\,,$$

respectively.

## 4.1 References

Drygas, H. (1970), The coordinate-free approach to Gauss-Markov estimation, Lecture notes in Operations Research and Mathematical Systems, Springer-Verlag Berlin-Heidelberg-New York.

Drygas, H. (2008), QR-decomposition from the statistical point of view, Recent Advances in Linear Models and Related Areas, Essays in Honour of Helge Toutenburg, Shalabh and Heumann (Eds) p. 293-311, Physica-Verlag, Springer, Heidelberg, 2008.

Schmetterer, Leopold (1966), Einführung in die mathematische Statistik, 2. Auflage, Springer-Verlag Wien-New York.

Seugupta, Debasis and Sreenivasa Rao Jammalamadaka (2003) "Linear Models, An integrated approach" World Scientific, New Jersey-London-Singapore-Hong Kong.