Colloquium Biometricum 41
2011, 87–101

A NEW APPROACH TO PARAMETERS ESTIMATION IN POLYNOMIAL REGRESSION WITH REPLICATIONS

Monika Różańska–Boczula, Mirosława Wesołowska–Janczarek

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Akademicka 13, 20–950 Lublin, Poland
e–mail: monika.boczula@up.lublin.pl; miroslawa.wesolowska@up.lublin.pl

Dedicated to the memory of Professor Wiktor Oktaba

Summary

The paper shows a unified method of estimating parameters of regression with replicated observations, taking into account assumptions of both equal and different variances. In a specific range of values of the independent variable, polynomial regression was considered because of its frequent use in the life sciences. For the considered multivariate approach we present an exemplary procedure (SAS 9.1), which allows one to make appropriate calculations.

Keywords and phrases: polynomial regression, repeated observations, growth curves, repeated measures

Classification AMS 2010: 62J12, 62H12

1. Introduction

Knowledge derived from the analysis and interpretation of results of experiments in which measurements are repeated, has a major contribution to the development of many scientific disciplines. Conclusions drawn of these experiments are characterized by objectivity. Repetitions, which are important in these studies may be of two kinds. In some experiments measurements are taken over the same experimental units. In this case the values of repeated measures, are usually correlated. In other studies, the data are collected as an observations of the dependent (response) variable for the same fixed value of the independent (predictor) variable at the different units. Replications obtained in this way are generally independent results (Tarasińska, 2001). Using them, the researcher can look for the regression curve describing the relationship between the measured features. Regression coefficients are usually estimated by the least squares method which is based on the averages of the replicated observations of response variable. However, this method of parameters estimation requires fulfillment of assumptions of equal error variances for all observations. When the above assumption is not fulfilled we have a regression with replicated observations at different variances. In this case, the least squares method should not be used. Due to the fact that there has not been so far given a uniform method of determining the functional relationship between the measured features regardless of whether the assumption of variance homogeneity occurs or not, this paper tries to analyze various aspects of the growth curves method in order to adapt it to analyze such data. The basis of the considerations is the fact that from the perspective of statistical analysis, replicated observations can be treated as a special case of repeated measurements.

2. Replicated observations in relation to repeated measures

2.1 Replicated observations

Regression with replicated observations occurs when researchers take into account replications, that is independent observations of dependent variable *y* for the same value of the predictor variable *x* (Koronacki, Mielniczuk 2001; Wesołowska–Janczarek, Różańska 2001). Then, we are taking into account *p* fixed values $x_1, x_2, ..., x_p$ for the independent variable *x*. Subsequently we measure the characteristic y for k_i independent units in each point x_i $(i = 1,..., p)$.

The regression equation defines the relationship between *y* and *x*. It takes the form $y_{ij} = f(x_i) + \varepsilon_{ij}$ where $i = 1,..., p$; $j = 1,..., k_i$. It is assumed that the vector $\boldsymbol{\varepsilon}$ of random errors has a normal distribution with expected value $\mathbf{0}$ and variance $\sigma^2 \mathbf{I}_n$, which can be written $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ where $n = \sum_{i=1}^{p} k_i$ is the total number of observations. In general, the vector $\mathbf{x}' = [x_1,...,x_p]$ is being considered and p independent k_i – dimensional vectors $\mathbf{y}_i' = [y_{i1},...,y_{ik_i}]$. Elements of vectors $\mathbf{y}_i$ are independent random variables with the same variance σ^2, i.e. $\mathbf{\Sigma}_{\mathbf{y}_i} = \sigma^2 \mathbf{I}_{k_i}$ (Draper, Smith 1998; Neter et al. 1996). The regression equation of y on x can then be determined taking into account replications averages $\bar{y}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} y_{ij}$ for each of x_i instead of individual y_{ij} $(j = 1,...,k_i)$ and the subsequent values of x_i. Such a procedure requires the equality of variances for all observations.

However, in practice, this assumption may not always be met, and then the covariance matrix takes the form

$$\mathbf{\Sigma}_{\mathbf{y}_i} = \sigma_i^2 \mathbf{I}_{k_i} \quad (i = 1,..., p). \tag{2.1.1}$$

2.2 Repeated measures

Many life sciences experiments are done in accordance with the methodology of repeated measurements. In these studies, the researcher is usually interested in the differences within the same experimental units. Then the results, which he analyzes are strongly correlated, because they are obtained by measurements made repeatedly on the same experimental units. Frequently, measurements that are closer in time, are more strongly correlated than those more distant from one another.

Various methods of data analyses with repeated measures are known (Crowder et al. 1991). In this paper we focus specifically on the method of growth curves (Kshirsagar et al. 1995, Potthoff et al. 1964). As it turns out, this multivariate approach allows to describe and analyze the dynamics of phenomena, which are changing over time and may be helpful in determining the regression function – especially in the case of heteroscedasticity.

3. Considered models – principles and indications

3.1 The growth curves model

The general multivariate linear growth curves model (Baksalary et al. 1978; Kshirsagar, Smith 1995; Potthoff, Roy 1964; von Rosen 1995; Wesołowska–Janczarek 1993) was given as

$$\mathbf{Y} = \mathbf{ABT} + \mathbf{E}$$

Y – the observation matrix ($N \times p$, $N = ak$ $k_1 = ... = k_p = k$), **A** is the $(N \times a)$ known design matrix (without a column of ones), **T** is the known $(q \times p)$ matrix, describing internal structure of observation (Vandermonde matrix)

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_p \\ \vdots & \vdots & \cdots & \vdots \\ t_1^{q-1} & t_2^{q-1} & \cdots & t_p^{q-1} \end{bmatrix},$$

where $t_1, t_2, ..., t_p$ mean time points, **B** is $(a \times q)$ matrix of fixed unknown coefficients of the assessed polynomial curves, and **E** is the $(N \times p)$ matrix of random errors. For the model (3.1.1) is assumed that the observation matrix $\mathbf{Y} \sim N_{N,p}(E(\mathbf{Y}), \mathbf{\Sigma_Y})$, where the expected value has a form

$$E(\mathbf{Y}) = \mathbf{ABT},$$

and

$$\mathbf{\Sigma_Y} = \underset{pp}{\mathbf{\Sigma}} \otimes \mathbf{I}_N = Var(vec\mathbf{Y}). \tag{3.1.1}$$

The covariance matrix (3.1.1) indicates that the rows of matrix **Y** are independent
p–dimensional random vectors with common covariance matrix $\mathbf{\Sigma} > 0$.

Observations for the different experimental units are uncorrelated. Furthermore, $N - r(\mathbf{A}) > p$, where $r(\mathbf{A})$ is rank of the matrix **A**. The estimator of matrix **B** (Potthoff et al. 1964) is of the form

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{T}'\left(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}'\right)^{-1} \tag{3.1.2}$$

where

$$\mathbf{S} = \mathbf{Y}'\left[\mathbf{I}_N - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{Y}$$

if the matrix **A** is full rank.

3.2 A polynomial regression with replications

It is well known, that regression with replicated observations presents relation between independent variable *x* and the mean values of dependent variable $\bar{y}$. This is not a rule, that this relationship defines a straight line. Sometimes, scatter plot for the values of variables $(x, \bar{y})$ shows some "curvature" (Aczel , 2000). In this case, the straight line does not fit to the data. A much better solution is then fit to the data a polynomial of degree higher than one. General form of polynomial regression model with replicated observations with one explanatory variable *x* is given by

$$\bar{y} = \beta_0 + \beta_1 x + ... + \beta_{q-1}x^{q-1} + \varepsilon, \tag{3.2.1}$$

where replications average $\bar{y}$ is the response variable, $\beta_0, \beta_1, ..., \beta_{q-1}$ are parameters, ε is a random error and $q-1$ is a degree of polynomial. The model (3.2.1) can be written in a matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (\bar{y}_1, ..., \bar{y}_p)'$ is *p*–dimensional vector of replications averages of response variable such that $\bar{y}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} y_{ij} \quad (i = 1, ..., p)$,

$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^{q-1} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_p & \ldots & x_p^{q-1} \end{bmatrix}$ is a design matrix $p \times q$ ($r(\mathbf{X}) = q$), whose first column consists of ones, while the others are the values of the independent variable in an appropriate power from 1 to $q-1$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{q-1})'$ is q–dimensional vector of regression coefficients and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$ is p–dimensional vector of random errors such that $\boldsymbol{\varepsilon} \sim N_p\left(\mathbf{0}, \sigma^2 \mathbf{I}_p\right)$.

By $\hat{\boldsymbol{\beta}}$ we define least squares estimator of vector of regression coefficients $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{3.2.2}$$

where $\mathbf{X}'$ is a transpose of a matrix $\mathbf{X}$. The application of the formula (3.2.2) requires that the assumption of equal error variances has been fulfilled for all observations. Thus, for all vectors $\mathbf{y}_i$ a covariance matrix should be

$$\boldsymbol{\Sigma}_{\mathbf{y}_i} = \sigma^2 \mathbf{I}_k . \tag{3.2.3}$$

When the assumption (3.2.3) is not fulfilled, we have a polynomial regression with replicated observations at different variances. In such a situation, to determine estimators of model parameters (3.2.1) we propose to use Potthoff–Roy's method which takes into account all observations of the dependent variable y_{ij} $(i = 1, \ldots, p;\ j = 1, \ldots, k)$ instead of averages.

In further considerations we assume that functions, which are fitting to the experimental data, are higher degree polynomials ($q > 2$) and we assume the same number of independent observations (replications) for each point x_i ($k_i = k$ for all $i = 1, \ldots, p$).

4. Estimation of parameters

4.1 Homogeneity of variances

We suppose that condition (3.2.3) is fulfilled now. Then the estimator (3.2.2) may be obtained by the least squares method.

Let's consider $q-1$ degree (i.e. $q>2$) regression model (3.2.1) with replications and let's assume that the estimators of parameters in this model are respectively $\hat{\beta}_0 = b_0$, $\hat{\beta}_1 = b_1, \ldots, \hat{\beta}_{q-1} = b_{q-1}$. To determine the evaluations of these parameters we need to find $(q-1)$–degree function which minimize $Q = \sum_{i=1}^{p} \left(\bar{y}_i - b_0 - b_1 x_i - b_2 x_i^2 - \ldots - b_{q-1} x_i^{q-1} \right)^2$ for a given sample, where $\bar{y}_i = \frac{1}{k} \sum_{j=1}^{k} y_{ij}$ represents the mean of the replicated observations at the x_i $(i = 1, \ldots, p)$. If we solve the appropriate system of normal equations we will obtain estimators of regression coefficients with replications in the form:

$$b_0 = \frac{1}{\det \mathbf{M}} \left(M_{11} \sum \bar{y}_i + M_{21} \sum \bar{y}_i x_i + \ldots + M_{q\,1} \sum \bar{y}_i x_i^{q-1} \right)$$

$$b_1 = \frac{1}{\det \mathbf{M}} \left(M_{12} \sum \bar{y}_i + M_{22} \sum \bar{y}_i x_i + \ldots + M_{q\,2} \sum \bar{y}_i x_i^{q-1} \right) \quad (4.1.1)$$

$$\vdots$$

$$b_{q-1} = \frac{1}{\det \mathbf{M}} \left(M_{1q} \sum \bar{y}_i + M_{2q} \sum \bar{y}_i x_i + \ldots + M_{qq} \sum \bar{y}_i x_i^{q-1} \right),$$

where $\det \mathbf{M}$ means the determinant of the matrix **M**, $M_{i'j'}$, $i', j' = 1, \ldots, q$ are elements of the adjoint matrix of matrix **M** and

$$\mathbf{M}=\begin{bmatrix} p & \sum x_i & \dots & \sum x_i^{q-1} \\ \sum x_i & \sum x_i^2 & \dots & \sum x_i^q \\ \vdots & \vdots & \dots & \vdots \\ \sum x_i^{q-1} & \sum x_i^q & \dots & \sum x_i^{2(q-1)} \end{bmatrix}.$$

Each of sum relates to the index $i=1,\dots,p$, but for improving the readability of writing these numbers were omitted.

4.2 Heteroscedasticity

In practice, it happens that the method of least squares can not be used to determine polynomial regression equation with replicated observations. The reason may be a departure from the assumptions connected with the homogeneity of variances, the so–called heteroscedasticity. Statistical literature (e.g. Krysicki et al. 1997) proposes several tests to verify the hypothesis of equality of variances of observation vectors $\mathbf{y}_i$ ($i=1,\dots,p$)

$$H_0:\sigma_1^2=\sigma_2^2=\dots\sigma_p^2. \qquad (4.2.1)$$

If we reject this hypothesis, the covariance matrices of vectors $\mathbf{y}_i$ are not the same and have the form (2.1.1), where $k_i=k$ for all $i=1,\dots,p$. In this case, formula (4.1.1) can not be used to estimate the parameters of polynomial regression. Authors of some papers (Horn et al. 1975; Neter et al. 1996; Rao and Subrahmaniam 1971; Wayne et al. 1978) suggest certain solutions for this situation. One of them is the weighted least squares method (WLS). Estimators of regression coefficients are then determined by

$$\hat{\boldsymbol{\beta}}_w=\left(\mathbf{X'WX}\right)^{-1}\mathbf{X'Wy}, \qquad (4.2.2)$$

while **W** define weight matrix

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w_p \end{bmatrix} \tag{4.2.3}$$

and is chosen so that observations with low variance have a large part in determining the regression equation, while those with a large variance respectively less (Draper, Smith 1998; Neter et al. 1996)).

In our work, when the assumption of equal variances of observation vectors is not fulfilled, we propose that a polynomial regression equation is determined by growth curves method. We will take into account a homogeneous group of experimental units and a diagonal covariance matrix form. The availability of replicated observations allows one to estimate variances according to the formula

$$\hat{\sigma}_i^2 = \frac{1}{k-1}\mathbf{y}_i'\left(\mathbf{I}_k - \frac{1}{k}\mathbf{E}_{kk}\right)\mathbf{y}_i \tag{4.2.4}$$

where $\mathbf{E}_{kk}$ is a matrix of $k \times k$ dimension consisting of all ones.

According to the assumptions of the growth curves model, we assume that the observation matrix $\underset{kp}{\mathbf{Y}} = [y_{ji}]$ ($j = 1,...,k$, $i = 1,..., p$) has independent rows with identical covariance matrices

$$\underset{pp}{\mathbf{\Sigma}} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_p^2 \end{bmatrix} = diag\left(\sigma_1^2, \sigma_2^2, ..., \sigma_p^2\right)$$

and expected value $E(\mathbf{Y}) = \mathbf{1}_k \boldsymbol{\beta}'\mathbf{T}$, while $\mathbf{1}_k$ means the k dimensional vector of ones. In order to estimate the parameters of the model (3.2.1), when the condition (3.2.3) is not satisfied, we denote:

$$\boldsymbol{\beta}' = [\beta_0, \beta_1, ..., \beta_{q-1}], \quad \mathbf{T} = \begin{bmatrix} 1 & 1 & ... & 1 \\ x_1 & x_2 & ... & x_p \\ x_1^2 & x_2^2 & ... & x_p^2 \\ \vdots & \vdots & ... & \vdots \\ x_1^{q-1} & x_2^{q-1} & ... & x_p^{q-1} \end{bmatrix},$$

$$\boldsymbol{\Sigma}^{-1} = diag(z_1, z_2, ..., z_p), \quad z_i = \frac{1}{\sigma_i^2}.$$

Finally, having taken into account the above indications and (3.1.2), the searched estimator of vector parameters in case the hypothesis (4.2.1) is rejected has a form

$$\hat{\beta}_0 = \frac{1}{\det \mathbf{K}} \left(K_{11} \sum z_i \bar{y}_i + K_{21} \sum z_i x_i \bar{y}_i + ... + K_{q1} \sum z_i x_i^{q-1} \bar{y}_i \right)$$

$$\hat{\beta}_1 = \frac{1}{\det \mathbf{K}} \left(K_{12} \sum z_i \bar{y}_i + K_{22} \sum z_i x_i \bar{y}_i + ... + K_{q2} \sum z_i x_i^{q-1} \bar{y}_i \right) \qquad (4.2.5)$$

$$\vdots$$

$$\hat{\beta}_{q-1} = \frac{1}{\det \mathbf{K}} \left(K_{1q} \sum z_i \bar{y}_i + K_{2q} \sum z_i x_i \bar{y}_i + ... + K_{qq} \sum z_i x_i^{q-1} \bar{y}_i \right),$$

where

$$\mathbf{K} = \mathbf{T}\boldsymbol{\Sigma}^{-1}\mathbf{T}' = \begin{bmatrix} \sum z_i & \sum z_i x_i & ... & \sum z_i x_i^{q-1} \\ \sum z_i x_i & \sum z_i x_i^2 & ... & \sum z_i x_i^q \\ \vdots & \vdots & ... & \vdots \\ \sum z_i x_i^{q-1} & \sum z_i x_i^q & ... & \sum z_i x_i^{2(q-1)} \end{bmatrix}$$

and $K_{i'j'}$, $i', j' = 1, ..., q$ are elements of the adjoint matrix of matrix **K**.

It is worth noting that when $\sigma_1^2 = \sigma_2^2 = ... = \sigma_p^2 = \sigma^2$ i.e. $z_1 = z_2 = ... = z_p = \frac{1}{\sigma^2}$ and $\sum z_i = \frac{p}{\sigma^2}$ as well as $\mathbf{K} = \frac{1}{\sigma^2}\mathbf{M}$ and $\det \mathbf{K} = \frac{1}{\sigma^2}\det \mathbf{M}$, from the formulas (4.2.5) we obtain (4.1.1). This means that the estimators of parameters obtained by Potthoff–Roy's method coincide with the estimators obtained by least squares method. Moreover, if in the formula (4.2.3) we will take the inverse of variances σ_i^2 equal to z_i as a weight w_i, then the estimator (4.2.2) obtained by the weighted least squares method is identical to (4.2.5) which follows from the growth curves method.

5. Application example

In an experiment conducted at the Department of Agricultural Machinery, University of Life Sciences in Lublin determined change in sugar content in corn grain, depending on harvest date and variety corn (Różańska–Boczula, 2010). The research material consisted of sweet corn ears of Boston, Bonus, and Jubilee varieties (Szymanek et al. 2005). The ears were collected by hand in a random way from different places in plantations four times $(p = 4)$ every two days. The characterization of the test material was determined on the basis of 100 ears. The percentage of sugar content was determined on samples weighing 200 g in 6 replications $(k = 6)$, which gives us 24 $(n = 24)$ observation pairs for both features of each group (i.e. variety). Repetitions in subsequent harvest dates were made on different experimental units thus the data are of replicated observations character.The aim of this study is to determine the functional relationship between sugar content and the date of harvest for each variety.

5.1 Regression with replications at equal variances

Różańska–Boczula (2010) shows that the observations obtained for the different varieties of corn do not lie along a straight line. For this reason, we decided that the model (3.2.1) is appropriate to determine the functional relationship between sugar content and harvest date for each group as well as that the second degree polynomials $(q = 3)$ will be the appropriate functions. As it has already been mentioned, the formula (3.2.2) allows us to estimate the coefficients of regression equations when at subsequent points of measurement x_i covariance matrices of $\mathbf{y}_i$ observation vectors are the same. Hartley's test (e.g. Krysicki et al. 1997), which verifies the hypothesis of variances equality in subsequent dates of corn harvesting, does not reject the null hypothesis (4.2.1) only for Jubilee variety. Hence, the regression equation estimated by the least squares method and the coefficient of determination for this variety are as follows

$$\text{Jubilee} \qquad \bar{y} = 0.048x^2 - 0.469x + 6.004, \quad R^2 = 0.9033.$$

Moreover, the probability value $p < 0.0001$ obtained as a result of variance analysis test and residuals plot show that this model is correctly fitted to the data.

5.2 Regression with replications at different variances

For the other studied corn varieties (Bonus, Boston), it appeared that the assumption of homogeneity of variance of the observation vectors is not fulfilled. In this situation, we propose to determine estimators of regression coefficients by the weighted least squares method or the multivariable growth curves method. In the first case we took the inverse of variances $\hat{\sigma}_i^2$ ($i = 1,...,4$) as weights. For each group $\hat{\sigma}_i^2$ were determined according to (4.2.4). Thus we have received for

$$\text{Bonus variety: } \bar{y} = -0.046x^2 - 0.108x + 6.235, \quad R^2 = 0.9656,$$

and

$$\text{Boston variety: } \bar{y} = -0.03x^2 - 0.257x + 6.591, \quad R^2 = 0.9900.$$

In the second case we got the results using procedure, that have been implemented for this goal in editor of the SAS 9.1 program. Vectors of parameters received by the growth curves method (according to (4.2.5)) are as follows:

$$\hat{\boldsymbol{\beta}}_{Bonus} = \begin{bmatrix} 6.23 \\ -0.102 \\ -0.047 \end{bmatrix}, \qquad \hat{\boldsymbol{\beta}}_{Boston} = \begin{bmatrix} 6.602 \\ -0.266 \\ -0.029 \end{bmatrix}.$$

The method of growth curves has the advantage that it allows to get estimation of regression coefficients simultaneously for all treatment groups.

For this purpose we determine an estimator of the matrix **B** by (3.1.2), assuming $\mathbf{S} = diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2, \hat{\sigma}_4^2)$, where each $\hat{\sigma}_i^2$ ($i = 1,...,4$) is calculated on the basis of observations for all varieties in the *i*–th time. It is worth noting that we use full observation matrix **Y** (not averages) in the calculations. To determine $\hat{\mathbf{B}}$ using SAS 9.1, you can use the following procedure

```
proc iml;
    y1={ observations from the first date of harvest for all groups };
    y2={ observations from the second date of harvest for all groups };
    y3={ observations from the third date of harvest for all groups };
    y4={ observations from the fourth date of harvest for all groups };
        Y=y1||y2||y3||y4;
        war={0.102 0.0127 0.053 0.012};
        SIGMA=diag(war);
        t1={1, 1, 1};
        t2={1, 2, 4};
        t3={1, 3, 9};
        t4={1, 4, 16};
        T=t1||t2||t3||t4;
        a1=J(6,1,1)//J(6,1,0)//J(6,1,0);
        a2=J(6,1,0)//J(6,1,1)//J(6,1,0);
        a3=J(6,1,0)//J(6,1,0)//J(6,1,1);
        A=a1||a2||a3;
        B=inv(A`*A)*A`*Y*inv(SIGMA)*T`*inv(T*inv(SIGMA)*T`);
print B;
run;
```

The result of the above procedure is the following matrix

$$\hat{\mathbf{B}} = \begin{bmatrix} 6.64 & -0.28 & -0.03 \\ 6.18 & -0.103 & -0.07 \\ 6.00 & -0.47 & 0.049 \end{bmatrix}.$$

The first row of the matrix $\hat{\mathbf{B}}$ corresponds to the regression coefficients of Boston variety, the second row is for Bonus variety and the third row is for Jubilee variety. One can note that the last row of the matrix $\hat{\mathbf{B}}$ is identical with the values of the parameters obtained by the least squares method previously (as shown in section 4.2). What's more, the mean determination coefficient

(Wesołowska–Janczarek,2000), that we have got for all these curves (estimated by Pothoff–Roy's method) was equal $\overline{R}^2 = 0.995$.

6. Conclusions

1) Growth curves method, that was used in this paper, is more general than other methods of analysis of replicated observations, because:
   a) it is a multivariate method, including situations of both: equal and different variances.
   b) it allows for simultaneous estimation of parameters for different groups of observations (e.g. for different varieties).
2) As the result of experimental data analysis we have stated what follows:
   a) High determination coefficient for estimated parabolic regression equation for Jubilee variety. The replicated observations of this variety had equal variances in subsequent harvest dates. This allowed to use both least squares method or Pothoff–Roy's method.
   b) For the Boston and the Bonus varieties, where heteroscedasticity was observed, multivariate growth curves method was used. Results are similar to the weighted least square method.
3) The mean determination coefficient that we have got for all curves estimated by Pothoff–Roy's method proves, that those curves have good fit to the experimental data. We can also confirm that date of harvest is a main factor deciding about the sugar content in corn grain.

References

Aczel A. D. (2000). *Statystyka w zarządzaniu*. Warszawa. PWN.

Baksalary J., Dobek A., Kala R. (1978). Estymacja krzywych wzrostu. ABS–65. *Alg. Biom. Statyst*. 7, 81 – 113.

Crowder M.J., Hand D.J. (1991). *Analysis of Repeated Measures*. Chapman & Hall.

Draper N.R., Smith H. (1998). *Applied Regression Analysis*. John Wiley & Sons.

Horn S.D., Horn R.A., Duncan D.B. (1975). Estimating Heteroscedastic Variances in Linear Models. *Journal of the American Statistical Association*, vol. 70, no 350, 380–385.

Koronacki J., Mielniczuk J. (2001). *Statystyka dla studentów kierunków technicznych i przyrodniczych.* WNT, Warszawa.

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (1997). *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. *Cz.2*. PWN, Warszawa.

Kshirsagar A.M., Smith W.B. (1995). *Growth curves*. Marcel Dekker, Inc. New York.

Neter J., Kutner M.H., Nachtsheim C.J., Wassermann W. (1996). *Applied Linear Statistical Models*. Chicago, IL: Irwin (4th ed.).

Potthoff R.F., Roy S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313–326.

Rao J.N.K., Subrahmaniam K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics* 27, 971–990.

Różańska–Boczula M. (2010). *Nowe ujęcie analizy regresji wielomianowej z replikacjami i jej zastosowanie do analizy doświadczeń z inżynierii rolniczej*. Rozprawa doktorska, UP Lublin.

Szymanek M., Dobrzański jr, B., Niedziółka I., Rybczyński R. (2005). *Kukurydza Cukrowa, technologia zbioru, fizyczne właściwości i jakość*. Instytut Agrofizyki im. Bohdana Dobrzańskiego PAN w Lublinie.

Tarasińska J. (2001). Optymalne rozmieszczenie obserwacji w pewnym modelu regresji liniowej. *Colloquium Biometryczne* 31, 47–55.

von Rosen D. (1995). Residuals in the growth curve model. *Ann. Inst. Statist. Math*. 47, No. 1, 129–136.

Wayne A., Fuller W.A., Rao J.N.K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *The Annals of Statistics*. 6, No 5, 1149–1158.

Wesołowska–Janczarek M. (1993). Krzywe wzrostu–przegląd metod i nowe problemy. *Listy Biometryczne* 30, No 2, 73–96.

Wesołowska–Janczarek M. (2000). On The use of determination coefficient to describe goodness of fit of assessed growth curves. *Biometrical Letters* 37, No.1, 13–19.

Wesołowska–Janczarek M., Różańska M. (2001). Regresja z powtarzanymi obserwacjami przy różnych wariancjach. *Colloquium Biometryczne* 31, 137–143.