Listy Biometryczne - Biometrical Letters
Vol. 41 (2004), No. 2, 37-49

Analyzing experiments by generalized linear models

Ewa Bakinowska*

Department of Mathematical and Statistical Methods,
Agricultural University of Poznań,
Wojska Polskiego 28, 60-637 Poznań, Poland

SUMMARY

In the paper two generalized linear models, i.e. the log-linear with Poisson distribution and logistic model with multinomial distribution, are applied to the analysis of a real experiment concerning the grain contamination by grain weevils. The comparison of these two approaches as well as their theoretical backgrounds are also presented.

KEY WORDS: estimation, maximum likelihood method, least squares method, weighted least squares method.

1. Introduction

It is known that there are often many alternative approaches to the analysis of a particular set of experimental data. Moreover, even in the same general frames the different methods can be used to establish final conclusions.

In the paper the analyses of a real experiment using two generalized linear models, the log-linear model with Poisson distribution and the logistic model with multinomial distribution, are presented. The unknown parameters in these models were estimated using the least squares method, the weighted least squares method and the Fisher scoring method. The analyzed experiment concerns the grain contamination by grain weevils.

2. Generalized linear model

Let y be a random variable observed in the experiment and let $\mathbf{x}$ be a vector of covariates. The vector $\mathbf{x}$ can be reduced to one covariate only. The generalized

*Partially supported by the State Committee for Scientific Research, grant no. 3 PO6A 021 24

linear model (see McCullagh and Nelder, 1983, 1989) is determined by two following assumptions:

a) a chosen function η, called a link function, of the expectation of y, $E(y) = \mu$, depends linearly on the vector of covariates, $\eta(\mu) = \mathbf{x}^T\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of unknown regression parameters,
b) the variance of the random variable y, $Var(y)$, depends on the expected value μ, $Var(y) = \phi V(\mu)$, where $V(\mu)$ is a variance function and ϕ is a scale parameter.

If a random variable observed is the k-dimensional random vector, $\mathbf{y} = (y_1, ..., y_k)^T$, then the expected value $E(\mathbf{y})$ is the vector

$$\boldsymbol{\mu} = (\mu_1, ..., \mu_k)^T,$$

and $Var(\mathbf{y})$ is the variance-covariance matrix. In this case the link function is a map transforming the vector $\boldsymbol{\mu}$ into a vector $\boldsymbol{\eta}$. The components of the vector $\boldsymbol{\eta} = (\eta_1(\boldsymbol{\mu}), \eta_2(\boldsymbol{\mu}), ..., \eta_t(\boldsymbol{\mu}))^T$ are connected with the covariates by the set of equations

$$\eta_w(\boldsymbol{\mu}) = \mathbf{x}_w^T\boldsymbol{\beta}_w, \qquad w = 1, 2, ..., t,$$

which can be expressed in the matrix form as

$$\boldsymbol{\eta}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}, \tag{2.1}$$

where $\mathbf{X} = diag(\mathbf{x}_w^T)$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_t^T)^T$. Usually it is assumed that the transformation $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu})$ is one-to-one and differentiable.

If in the experiment with the scalar variable y the group of s independent objects, characterized by various values $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_s$ of covariates, is observed, then a dependence of μ on $\mathbf{x}$,

$$\mu = \mu(\mathbf{x}) = \eta^{-1}(\mathbf{x}^T\boldsymbol{\beta}),$$

is of interest. To achieve this, an estimate of the parameter vector $\boldsymbol{\beta}$ is necessary. It can be obtained from the model

$$\begin{pmatrix} \eta(\mu_1) \\ \cdots \\ \eta(\mu_s) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \cdots \\ \mathbf{x}_s^T \end{pmatrix} \boldsymbol{\beta}, \tag{2.2}$$

where $\mu_i = \mu(\mathbf{x}_i)$ for $i = 1, 2, ..., s$.

In case of the vector random variable $\mathbf{y}$, the interest concerns the set of functions $\mu_1(\mathbf{x}), \mu_2(\mathbf{x}), ..., \mu_k(\mathbf{x})$. They can be obtained after estimating the parameter vector

$\boldsymbol{\beta}$ from the model

$$\begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\mu}_1) \\ \cdots \\ \boldsymbol{\eta}(\boldsymbol{\mu}_s) \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \cdots \\ \mathbf{X}_s \end{pmatrix} \boldsymbol{\beta}, \tag{2.3}$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}(\mathbf{x}_i) = (\mu_1(\mathbf{x}_i), ..., \mu_k(\mathbf{x}_i))^T$ and each matrix $\mathbf{X}_i$ has the same form as $\mathbf{X}$ in (2.1), but with elements of $\mathbf{X}_i$ determined by the point $\mathbf{x}_i$.

3. Estimation methods of regression parameters

Estimation of regression parameters can be carried out by various methods (see McCulloch and Searle, 2001). We will present some of them for the model with the vector random variable.

If $\mathbf{y}$ is the random variable on which we can state that it has asymptotically normal distribution,

$$\mathbf{y} \underset{as}{\sim} N(\boldsymbol{\mu}, \phi\mathbf{V}),$$

then the random variable $\boldsymbol{\eta}(\mathbf{y})$ (see Mardia et al. 1979, p. 52; Rao, 1982, p. 398 or Agresti, 1984, p. 247) has also the asymptotically normal distribution,

$$\boldsymbol{\eta}(\mathbf{y}) \underset{as}{\sim} N(\boldsymbol{\eta}(\boldsymbol{\mu}), \phi\mathbf{G}\mathbf{V}\mathbf{G}^T),$$

where $\mathbf{G}$ is a matrix of partial derivatives, $\mathbf{G} = \partial\boldsymbol{\eta}/\partial\mathbf{y}^T$, evaluated at $\mathbf{y} = \boldsymbol{\mu}$. In general the elements of $\mathbf{G}$ are the functions of $\boldsymbol{\mu}$, $\mathbf{G} = \mathbf{G}(\boldsymbol{\mu})$, as well as the element of the variance function $\mathbf{V} = \mathbf{V}(\boldsymbol{\mu})$. The vector $\boldsymbol{\eta}(\mathbf{y})$ is asymptotically an unbiased estimator of $\boldsymbol{\eta}(\boldsymbol{\mu})$. In the frame of the model (2.1), it can be used to estimate $\boldsymbol{\beta}$ by applying the least squares method or the weighted least squares method with weights following from the matrix $\phi\mathbf{G}\mathbf{V}\mathbf{G}^T$.

When s independent objects are observed, then the appropriate normal equations can be written in the form

$$\sum_{i=1}^{s} \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \boldsymbol{\beta} = \sum_{i=1}^{s} \mathbf{X}_i^T \mathbf{W}_i \boldsymbol{\eta}(\mathbf{y}_i), \tag{3.1}$$

where matrices $\mathbf{X}_1, ..., \mathbf{X}_s$ follow from (2.3), $\mathbf{y}_i$ is the response of the observed variable for the i-th object, and $\mathbf{W}_i$ is a matrix of corresponding weights. If $\mathbf{W}_i = \mathbf{I}$ for $i = 1, 2, ..., s$, the equation (3.1) leads to the simple least squares estimator (LS – Least Squares method). Other set of weights is determined by the matrices of the form $\left(\phi\mathbf{G}(\mathbf{y}_i)\mathbf{V}(\mathbf{y}_i)\mathbf{G}^T(\mathbf{y}_i)\right)^{-1}$. Actually, they are inverses of asymptotic covariance matrices estimated by replacing the unknown expected values $\boldsymbol{\mu}_i$ by the vectors $\mathbf{y}_i$, respectively. They lead to the weighted least squares estimator (WLS – Weighted Least

Squares method). Assuming that matrices $\mathbf{X}_i$ are of full column rank, the asymptotic covariance matrices of the LS-estimator and of the WLS-estimator, respectively, have the following forms:

$$\mathbf{V}_{LS} = \left(\sum_{i=1}^{s} \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{s} \mathbf{X}_i^T \left(\phi \mathbf{G}(\boldsymbol{\mu}_i)\mathbf{V}(\boldsymbol{\mu}_i)\mathbf{G}^T(\boldsymbol{\mu}_i)\right) \mathbf{X}_i\right) \left(\sum_{i=1}^{s} \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \tag{3.2}$$

and

$$\mathbf{V}_{WLS} = \left(\sum_{i=1}^{s} \mathbf{X}_i^T \left(\phi \mathbf{G}(\boldsymbol{\mu}_i)\mathbf{V}(\boldsymbol{\mu}_i)\mathbf{G}^T(\boldsymbol{\mu}_i)\right)^{-1} \mathbf{X}_i\right)^{-1}. \tag{3.3}$$

Their estimates can be obtained by substituting $\boldsymbol{\mu}_i$ by $\mathbf{y}_i$.

If the joint distribution of observed variables is known, we can estimate the vector $\boldsymbol{\beta}$ using the likelihood method provided that the transformation $\boldsymbol{\mu} \to \boldsymbol{\eta}$ is invertible. The last assumption allows to determine a gradient $\mathbf{s}_i$ of the logarithm of the likelihood function $l_i = l(\boldsymbol{\mu}_i, \mathbf{y}_i)$ with respect to $\boldsymbol{\beta}$,

$$\mathbf{s}_i^T(\boldsymbol{\beta}) = \frac{\partial l_i}{\partial \boldsymbol{\mu}_i^T} \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}_i^T}\right)^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T},$$

for all observed objects $i = 1, 2, ..., s$. Comparing their sum to the zero vector, provides the maximum likelihood equation which takes the form

$$\sum_{i=1}^{s} \mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{0}.$$

It is a non-linear equation. Its solution can be searched through the Newton-Raphson method. This method simplifies, if the matrix of second derivatives of the likelihood function is replaced by its expectation, which leads to the Fisher information matrix. In result, the subsequent approximations of the solution are related by the equation

$$\boldsymbol{\beta}_{n+1} = \boldsymbol{\beta}_n + \mathbf{F}^{-1}\mathbf{s},$$

where $\mathbf{s} = \sum_{i=1}^{s} \mathbf{s}_i(\boldsymbol{\beta}_n)$ and $\mathbf{F} = \sum_{i=1}^{s} \mathbf{F}_i$, with $\mathbf{F}_i$ being the information matrix corresponding to the i-th observed object. This approach is known as the Fisher scoring method (FS). The estimate of the covariance matrix of the FS estimator has the form as in (3.3), however with $\boldsymbol{\mu}_i$ replaced by $\mathbf{X}_i\boldsymbol{\beta}_{FS}$.

4. A biological experiment

It is known that storage of the damp grain leads to its contamination. One of the causes are grain weevils which leave the corks on the grains. Their activity depends on the natural gas, which is produced by the damp grain. To establish the influence of the gas concentration on the degree of contamination, the following experiment was conducted.

Into nine pots, each containing 100 grains of wheat, ten weevils (5 male and 5 female) were inserted. The pots were filled with gas (1-okten-3-ol). It was supposed that the degree of grain contamination depends on gas concentration. Therefore three concentrations of gas were used: 10 ppm, 50 ppm and 100 ppm, each applied to three pots. The degree of contamination was determined by the number of corks on each grain in each pot.

We can assume that we have $s = 9$ objects (pots), each including $m_i = 100$ units (grains), and nine values of the covariate variable (gas concentration) $x_1 = x_2 = x_3 = 10$, $x_4 = x_5 = x_6 = 50$, $x_7 = x_8 = x_9 = 100$. The results of that experiment are presented in Table 1.

Table 1. The observed grain contamination

Gas concentrations	Numbers of corks					
	0	1	2	3	4	≥ 5
$x_1 = 10$	34	29	17	14	5	1
$x_2 = 10$	35	23	18	15	4	5
$x_3 = 10$	25	34	17	9	6	9
$x_4 = 50$	19	25	17	11	14	14
$x_5 = 50$	20	26	18	9	11	16
$x_6 = 50$	27	21	24	18	4	6
$x_7 = 100$	32	35	12	10	6	5
$x_8 = 100$	34	21	15	14	9	7
$x_9 = 100$	40	29	17	8	3	3

The problem posed in this experiment can be approached by answering two questions, namely, how the average grain contamination depends on the gas concentration and how the probability that the single grain will be contaminated by a given number of corks depends on the gas concentration. Both these questions can be answered using the generalized linear model.

5. Modelling with the use of the Poisson distribution

If the interests of the experimenter are focussed on the average contamination, the data can be analyzed using the generalized linear model with the Poisson distribution.

To this aim let us assume that y_{il}, being the number of corks on the l-th grain in the i-th pot, has the Poisson distribution, $y_{il} \sim P(\lambda_i)$, where $\lambda_i > 0$ is the average number of corks on grains in the i-th pot. Moreover, let y_i denote the average,

$$y_i = \frac{1}{m_i} \sum_{l=1}^{m_i} y_{il} = \frac{1}{100} \sum_{l=1}^{100} y_{il}.$$

In consequence we have

$$E(y_i) = \mu_i = \lambda_i, \quad Var(y_i) = \phi_i V(\lambda_i) = \frac{1}{m_i}\lambda_i = \frac{1}{100}\lambda_i, \tag{5.1}$$

where the scale parameter and the variance function have the forms

$$\phi_i = \frac{1}{m_i}, \quad \text{and} \quad V(\lambda_i) = \lambda_i,$$

respectively.

Since $\lambda_i > 0$, the appropriate link function is the logarithmic transformation,

$$\eta(\lambda_i) = \log(\lambda_i).$$

Moreover, since each three pots were treated by different concentrations of the gas, the numbers λ_i, $i = 1, 2, ..., 9$, can be considered as the values of a function $\lambda(x)$ at points x_i, $i = 1, 2, ..., 9$, i.e. $\lambda(x_i) = \lambda_i$. Finally, since only three different gas concentrations were used, the values of the link function $\eta(\lambda(x))$ can be predicted by the quadratic regression,

$$\log(\lambda(x)) = \beta_0 + \beta_1 x + \beta_2 x^2, \tag{5.2}$$

which leads to the so-called log-linear model (see Agresti, 1984). In consequence, the problem of establishing the average contamination function $\lambda(x)$ consists in estimation of regression parameters $\beta_0, \beta_1, \beta_2$.

The estimation was carried out by three methods described in Section 3. The estimates of parameters $\beta_0, \beta_1, \beta_2$ are presented in Table 2. As a starting point for the Fisher scoring method the estimates following from the LS method were used.

Table 2. Estimates of parameters in log-linear model

Parameters	Method		
	LS	WLS	FS*
β_0	0.208241	0.213454	0.210603
β_1	0.020433	0.020987	0.020732
β_2	−0.000189	−0.000194	−0.000191

*results of the fifth iteration

The asymptotic covariance matrix for the LS estimates was obtained from the formula (3.2). It has the form:

$$\hat{\mathbf{V}}_{LS} = \begin{pmatrix} 478413.066 & -17773.969 & 132.499 \\ -17773.969 & 912.194 & -7.656 \\ 132.499 & -7.656 & 0.069 \end{pmatrix} \cdot 10^{-8}.$$

The estimates of asymptotic covariance matrices of the WLS and the FS estimator are the same up to fifteen decimal places and have following form

$$\hat{\mathbf{V}}_{WLS} = \hat{\mathbf{V}}_{FS} = \begin{pmatrix} 473030.889 & -17538.133 & 130.638 \\ -17538.133 & 896.897 & -7.522 \\ 130.638 & -7.522 & 0.067 \end{pmatrix} \cdot 10^{-8}.$$

The comparison of results following from different methods points a close accordance of the point estimates of regression parameters as well as the estimates of their covariance matrices. Nevertheless, the FS and the WLS methods produce the estimates which are more efficient then those following from the LS method. The estimates of regression parameters can be used to obtain the estimate of the average contamination function $\lambda(x)$. With the FS estimates, it takes the form (see also Figure 1),

$$\hat{\lambda}(x) = \exp(0.210603 + 0.020732x - 0.000191x^2). \tag{5.3}$$

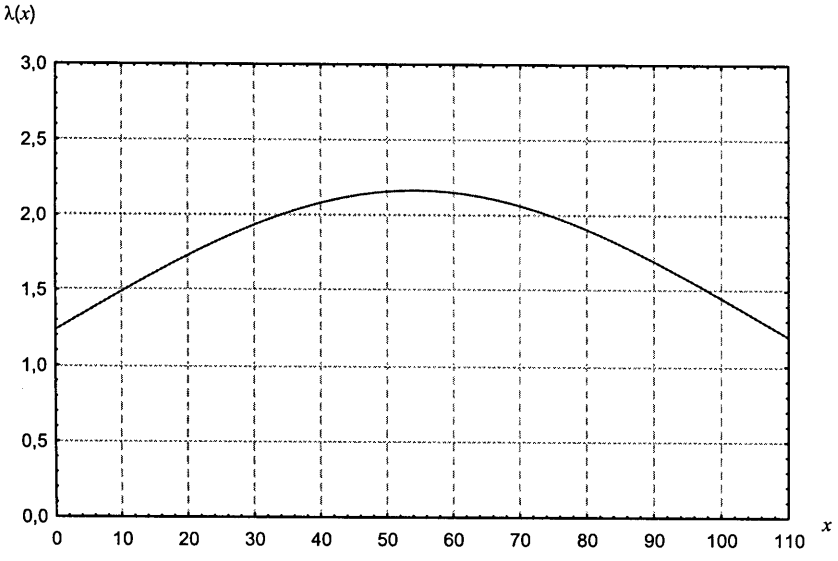

Fig. 1. The average grain contamination

6. Modelling with the use of the multinomial distribution

If the experimenter is interested in estimating the probability that a single grain will be contaminated by a given number of corks, the description of the experiment can be based on the multinomial distribution. To this aim let us assume that each grain was classified to $k = 6$ separate categories corresponding to columns of Table 1. The results of classification for the i-th pot form a random vector $\mathbf{y}_i$ which follows the multinomial distribution, $\mathbf{y}_i \sim M(m_i, \boldsymbol{\pi}_i)$, where $\boldsymbol{\pi}_i = (\pi_{1i}, \pi_{2i}, ..., \pi_{6i})^T$, with π_{ji} being a probability that a grain will be classified to the j-th category, $\sum_{j=1}^{6} \pi_{ji} = 1$. In consequence, using well known results (compare Fisz, 1958, or Mardia et al., 1979), for the vector of frequencies $\mathbf{p}_i$,

$$\mathbf{p}_i = \frac{1}{m_i}\mathbf{y}_i = \frac{1}{100}\mathbf{y}_i,$$

we have

$$E(\mathbf{p}_i) = \boldsymbol{\pi}_i, \qquad Var(\mathbf{p}_i) = \phi_i \mathbf{V}(\boldsymbol{\pi}_i) = \frac{1}{m_i}(\boldsymbol{\pi}_i^{\delta} - \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T),$$

where $\boldsymbol{\pi}_i^{\delta}$ is the diagonal matrix with elements of the vector $\boldsymbol{\pi}_i$ on its diagonal, $\phi_i = 1/m_i = 1/100$ is the scale parameter, while $\mathbf{V}(\boldsymbol{\pi}_i) = (\boldsymbol{\pi}_i^{\delta} - \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T)$ is the variance function.

In experiment under consideration the categories are naturally ordered. In such case, it is more interesting to estimate cumulative probabilities $\gamma_{ji} = \pi_{1i} + \pi_{2i} + \cdots + \pi_{ji}$ that a grain will be contaminated j times at most.

Since $\sum_{j=1}^{6} \pi_{ji} = 1$ and $0 < \pi_j < 1$ the appropriate link functions (see e.g. McCulloch and Searle, 2001) are the logit functions

$$\eta_j(\boldsymbol{\pi}_i) = \log \frac{\gamma_{ji}}{1 - \gamma_{ji}}, \qquad j = 1, 2, ..., 5, \quad i = 1, 2, ..., 9. \tag{6.1}$$

Moreover, since each three pots were treated by different concentration of the gas, the cumulative probabilities γ_{ji}, $i = 1, 2, ..., 9$, can be considered as the values of a functions $\gamma_j(x)$ at points x_i, $i = 1, 2, ..., 9$, i.e. $\gamma_j(x_i) = \gamma_{ji}$. Finally, since only three different gas concentrations were used, the values of the link function $\eta_j(\boldsymbol{\pi}(x))$ can be predicted by the quadratic regression,

$$\log \frac{\gamma_j(x_i)}{1 - \gamma_j(x_i)} = \beta_{0j} + \beta_{1j} x_i + \beta_{2j} x_i^2, \qquad j = 1, 2, ..., 5, \quad i = 1, 2, ...9. \tag{6.2}$$

This transformation leads to the so-called logistic model, which forms the base for estimation of regression parameters β_{0j}, β_{1j}, β_{2j}, $j = 1, 2, ..., 5$. Their estimates obtained by three methods presented in Section 3, are contained in Table 3. As a starting point for the FS method the estimates following from the LS method were

used. Some detailed remarks concerning the estimation methods in generalized linear models with multinomial distribution are presented in the paper by Bakinowska and Kala (2002).

Table 3. Estimates of parameters in logistic model

Parameters	Method		
	LS	WLS	FS*
β_{01}	-0.531303	-0.501710	-0.524021
β_{11}	-0.029065	-0.028644	-0.028862
β_{21}	0.000283	0.000278	0.000281
β_{02}	0.707885	0.686987	0.705642
β_{12}	-0.033354	-0.031905	-0.033192
β_{22}	0.000320	0.000304	0.000317
β_{03}	1.528928	1.514830	1.523048
β_{13}	-0.032931	-0.031970	-0.032604
β_{23}	0.000309	0.000291	0.000302
β_{04}	2.699417	2.509839	2.641406
β_{14}	-0.047500	-0.048641	-0.048742
β_{24}	0.000422	0.000438	0.000432
β_{05}	3.857756	3.141352	3.420443
β_{15}	-0.062756	-0.048222	-0.052360
β_{25}	0.000542	0.000456	0.000476

*results of the sixth iteration

Since the number of estimated parameters is large, below we present estimates of the covariance matrices only for the subvectors $\boldsymbol{\beta}_1 = (\beta_{01}, \beta_{11}, \beta_{21})^T$ and $\boldsymbol{\beta}_2 = (\beta_{02}, \beta_{12}, \beta_{22})^T$:

$$\hat{\mathbf{V}}_{LS}(\boldsymbol{\beta}_1) = \begin{pmatrix} 3553563.133 & -147841.690 & 1139.380 \\ -147841.690 & 8987.398 & -77.048 \\ 1139.380 & -77.048 & 0.689 \end{pmatrix} \cdot 10^{-8},$$

$$\hat{\mathbf{V}}_{WLS}(\boldsymbol{\beta}_1) = \begin{pmatrix} 3492442.879 & -145175.358 & 1118.709 \\ -145175.358 & 8815,689 & -75.583 \\ 1118.709 & -75.583 & 0.676 \end{pmatrix} \cdot 10^{-8},$$

$$\hat{\mathbf{V}}_{FS}(\boldsymbol{\beta}_1) = \begin{pmatrix} 3492190.658 & -145236.714 & 1119.358 \\ -145236.714 & 8825.120 & -75.673 \\ 1119.358 & -75.673 & 0.677 \end{pmatrix} \cdot 10^{-8},$$

$$\hat{\mathbf{V}}_{LS}(\boldsymbol{\beta}_2) = \begin{pmatrix} 3492190.658 & -145236.714 & 1119.358 \\ -145236.714 & 8825.120 & -75.673 \\ 1119.358 & -75.673 & 0.677 \end{pmatrix} \cdot 10^{-8},$$

$$\hat{\mathbf{V}}_{WLS}(\boldsymbol{\beta}_2) = \begin{pmatrix} 3029089.740 & -119287.432 & 906.031 \\ -119287.432 & 6720.310 & -57.202 \\ 906.031 & -57.202 & 0.514 \end{pmatrix} \cdot 10^{-8},$$

$$\hat{\mathbf{V}}_{FS}(\boldsymbol{\beta}_2) = \begin{pmatrix} 3032454.239 & -119413.161 & 906.897 \\ -119413.161 & 6726.755 & -57.248 \\ 906.897 & -57.248 & 0.514 \end{pmatrix} \cdot 10^{-8}.$$

It is easy to note that the matrices $\hat{\mathbf{V}}_{WLS}$ and $\hat{\mathbf{V}}_{FS}$ slightly differ from each other. These differences are justified by different values of the estimates used in the place of $\boldsymbol{\mu}_i$ in formula (2.3).

Some of the searched probability functions, obtained with the use of the FS estimates, have the following forms:

$$\hat{\gamma}_1(x) = \hat{\pi}_1(x) = \frac{\exp(-0.524021 - 0.028862x + 0.000281x^2)}{1 + \exp(-0.524021 - 0.028862x + 0.000281x^2)}, \tag{6.3}$$

$$\hat{\gamma}_2(x) = \hat{\pi}_1(x) + \hat{\pi}_2(x) = \frac{\exp(0.705642 - 0.033192x + 0.000317x^2)}{1 + \exp(0.705642 - 0.033192x + 0.000317x^2)}, \tag{6.4}$$

$$\hat{\gamma}_3(x) = \hat{\pi}_1(x) + \hat{\pi}_2(x) + \hat{\pi}_3(x) = \frac{\exp(1.523048 - 0.032604x + 0.000302x^2)}{1 + \exp(1.523048 - 0.032604x + 0.000302x^2)}, \tag{6.5}$$

where $\gamma_1(x)$ represents the probability that the grain will not be contaminated, $\gamma_2(x)$ – the probability, that the grain will be contaminated once at most, and $\gamma_3(x)$ – the probability, that the grain will be contaminated two times at most. These functions are presented in Figure 2.

7. Comparison of the results

The analyses conducted in the previous section were based on two different models. Hence the obtained estimates of regression parameters can not be directly compared. However, using the estimated average contamination function $\lambda(x)$, obtained with the use of the log-linear model, the probability functions $\gamma_n(x)$, $n = 1, 2, 3$, discussed in Section 6, can be constructed. They are related by the following equations:

$$\begin{aligned} \gamma_1(x) &= \exp(-\lambda(x)), \\ \gamma_2(x) &= \exp(-\lambda(x))(1 + \lambda(x)), \\ \gamma_3(x) &= \exp(-\lambda(x))(1 + \lambda(x) + \frac{1}{2}\lambda^2(x)). \end{aligned}$$

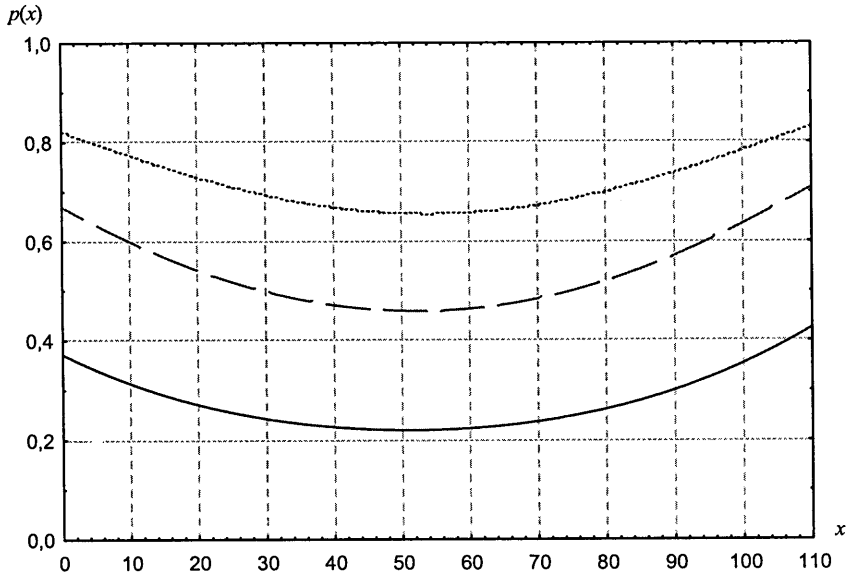

Fig. 2. The grain contamination probabilities. The continous line – the probability that a grain will not be contaminated, the broken line - that the grain will be contaminated once at most, the dotted line – that the grain will be contaminated two times at most.

Their graphs are presented on Figure 3. It is easy to observe, that the graphs are similar. The functions reach their minimum at the same point, for the gas concentration 50 ppm. However, the estimates of probabilities may differ even by 0.1.

On the other hand, in the model with Poisson distribution the emphasis is put on estimation of the average grain contamination, while in the model with multinomial distribution the emphasis is put on estimation of the probability of the grain contamination. Although the results of both analyses are similar, the log-linear model with Poisson distribution contains less parameters which results in more compact form of the model and simplifies the analysis.

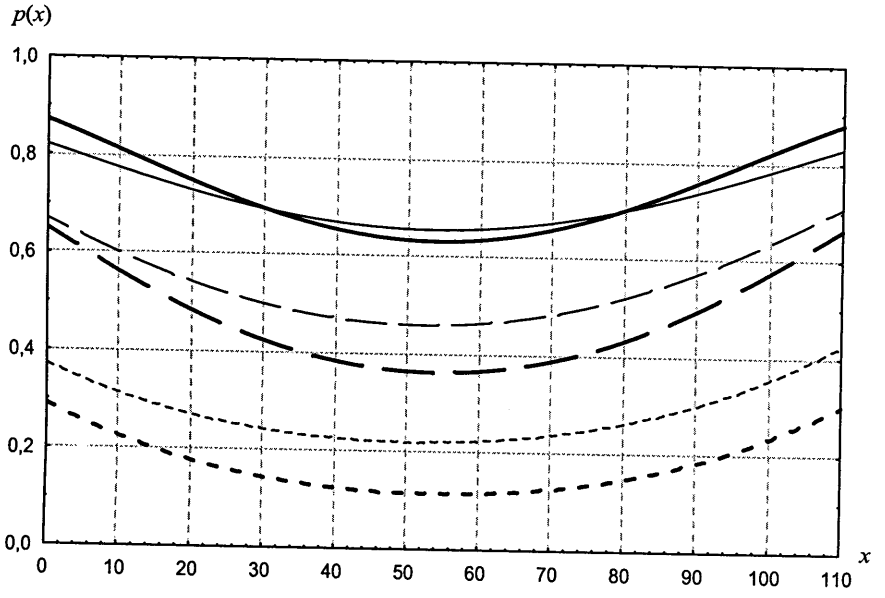

Fig. 3. The grain contamination probabilities. The fat line corresponds to the function following from the model with the Poisson distribution, while the thin line to the model with the multinomial distribution. The dotted line – the probability that a grain will not be contaminated, the broken line – that the grain will be contaminated once at most, the continous line – that the grain will be contaminated two times at most.

References

Agresti A. (1984) *Analysis of ordinal categorical data.* Wiley, New York.

Bakinowska E., Kala R. (2002). Metody estymacji w uogólnionym modelu liniowym, *Colloquium Biometrycze*, **32**, 241-251

Fisz M. (1958) *Rachunek prawdopodobieństwa i statystyka matematyczna.* PWN,Warszawa.

Mardia K. V., Kent I. T., Bibby I. M. (1979) *Multivariate Analysis.* Academic Press, London.

McCullagh P. and Nelder J. A. (1983, 1989) *Generalized linear models.* Chapman and Hall, London.

McCulloch Ch. E. Searle, S. R., (2001) *Generalized, linear, and mixed models.* Wiley, New York.

Rao C. R. (1982) *Modele liniowe statystyki matematycznej.* PWN, Warszawa.

Received 8 December 2004

Zastosowanie uogólnionych modeli liniowych do analizy eksperymentów

STRESZCZENIE

W pracy zastosowano dwa uogólnione modele liniowe, tj. model log-liniowy z rozkładem Poissona i model logistyczny z rozkładem wielomianowym, do analizy rzeczywistego eksperymentu dotyczącego porażenia ziarna przez wołki zbożowe. Przedstawiono teorię określającą omawiane modele, a także dokonano ich porównania.

SŁOWA KLUCZOWE: estymacja, metoda największej wiarogodności, metoda najmniejszych kwadratów, ważona metoda najmniejszych kwadratów.