

ANALYSIS OF SELECTED GRASSPEA MUTANTS BY GENERALIZED LINEAR MODEL

Jan Bocianowski¹⁾, Ewa Bakinowska¹⁾, Wojciech Rybiński²⁾

¹⁾Department of Mathematical and Statistical Methods
University of Life Sciences in Poznań
Wojska Polskiego 28, 60-637 Poznań, Poland
e-mails: ewabak@up.poznan.pl, jboc@up.poznan.pl

²⁾Institute of Plant Genetics, Polish Academy of Science
Strzeszyńska 34, 60-479 Poznań, Poland
e-mail: wryb@igr.poznan.pl

Summary

Grasspea (*Lathyrus sativus* L.) is characterized by low habitat requirements, high content of protein and favorable composition of fatty acid in seed. Simultaneously, grasspea possesses a lot of unfavorable traits from an agronomy point of view. Essential is to modify this plant on the genetic level for example through mutation induction. The object of investigation constituted three grasspea mutants and their initial cultivar Derek, analyzed for eight characters in Experiment Station in Cerekwica. The field trial was conducted according to method of randomized block design with three replication. All mutants were obtained after grain irradiation of cultivar Derek with laser light and two chemomutagen treatments: N-nitroso-N-methylurea (MNU) and sodium azide (NaN_3). The aim of the work was to compare the variety Derek with its mutants with respect to two selected traits: the number of pods per plant and the number of branches per plant. Both considered traits are treated as discrete, so in the analysis the logistic model which belongs to the class of generalized linear models was applied.

Key words and phrases: generalized linear model, grasspea mutants, logistic model

Subject classification AMS 1993: 62F10, 62J02, 62F12

1. Introduction

The genus *Lathyrus* is large with 187 species and subspecies being recognized (Alkin et al. 1983). However, only one species – *Lathyrus sativus* – is widely cultivated as a food crop, while other species are cultivated to a lesser extent for booth food and forage (Campbell, 1997). *Larhyrus sativus* (grasspea) has a long history of cultivation in part of Europe, North Africa and Asia, a specially of certain region in India, Pakistan, Bangladesh and Ethiopia. According to Milczak et al. (1997), grasspea first appeared in Poland (in the region Podlasie) accompanying lentil as weed as far back as the 17th century, together with Tatar settlements. In 1991, Milczak et al. (2001), through selection from local grasspea population, there were developed two promising lines with symbols 'Der' and 'Kra' which, after a series multi-year field experiments were registered as original cultivars – Derek and Krab.

Lathyrus sativus is used as a forage for livestock and as a pulse for human consumption; it is nutritionally on a par with other grains legume species, containing up to 30% crude protein (which is high in lysine), 0.6% fat and about 60% carbohydrate (Hartman et al., 1974). The seeds contain also an interesting fatty acid composition (linoleic acid – 38-56%, linolenic acid – 6-8%), dietary fibre (NDF – 130-170g, ADF – 76-85g) and minerals: Ca, P and microminerals (Grela and Winiarska, 1997). The grasspea is favoured for its ability to mature and produce a yield in times of drought when the other crops have failed (Kearney and Smartt, 1995). Apart from unquestionable advantages this plant is also characterized by number of less favourable features, such as strong lodging, indeterminate character of growth, or excessively long period till ripeness, as well as the presence of anti-nutritional substances in seeds (Rybiński et al., 2004). Apart from effects of recombination the genetic improvement of a number of unfavourable characters was made by use of mutation induction and obtained mutants were the main objects of presented paper.

This paper presents the comparison of selected variety of grasspea with its three mutants with respect to number of branches per plant and number of pods per plant. Both traits, number of branches and number of pods, are represented by the discrete random variable. In that case the methods of standard linear model are insufficient. Therefore to the statistical analysis the logistic model which belongs to the class of generalized linear model was used.

2. Research material

The initial material for mutagenic treatment was constituted by seeds of Polish grasspea (*Lathyrus sativus* L.) cultivar – Derek with use of helium-neon laser light and N-nitroso-N-methylurea (MNU). Selection of mutated genotypes was performed in generation M_2 and in M_3 , the progeny of obtained forms was sown, together with the initial cultivar Derek, in order to verify of the changes observed in generation M_2 . After multiplication of stabilized mutants, three of them (D4, D11 and D13) as well as initial cultivar Derek (control) were the objects of field trial set up according to the random block method (with three replications) performed on the Experimental Field of Institute of Plant Genetics in Cerekwica in 2003. The seeds were placed in experimental plots with spacing 15×30 cm. Determination performed after the harvest covered eight qualitative traits. The paper describes the statistical analysis performed in independent manner according two of estimated characters: pods number and number of lateral branches per plant. Each of studied units was classified with respect to three separate categories. For the first trait (number of branches per plant) categories “small” means 4-9 branches, “medium”: 10-15, and “large”: 16-25. And the second trait (number of pods per plant) categories “small” means less than 70 pods, “medium”: 70-100, and “large” more than 100 pods.

3. Description of method

The “visible” traits like for example number of branches per plant, number of pods per plant and number of seed per plant are usually the results of some genetic tendency continuous in character. These tendencies are represented by continuous random variable which is hidden and only the results of classification of observed units (with respect to a symbolic ordinal scale) are analyzed in detail. We classify the observed units to ordinal, separate and usually symbolic categories. And for example, the number of seed per plant can be described as small, medium and large (three symbolic categories). The results of such classification are connected with some discrete random variable.

The relation of hidden continuous random variable with a discrete is determined by borders of successive categories. In literature the borders are called thresholds (see Misztal et al., 1989) or cutpoints (see Miller et al., 1993).

For an experimenter who carries out an experiment, it may be interesting to estimate unknown probabilities of successes of individual categories or cumula-

tive probabilities. But sometimes the possibility of the comparison of such probabilities for various treatments is more interesting.

In order to obtain searched probabilities one can use the generalized linear models (see McCullagh and Nelder, 1983, 1989). One of them is so called probit model (see McCullagh and Nelder 1989, p.31), in which we assume that hidden continuous random variable is standard normal distributed. However, searching for estimators of parameters in such a model means encountering numerical problems connected with the inversion of a cumulative density function. It is easier to find the solution assuming a standard logistic distribution (see Rao and Toutenburg, 1999, p. 316), which has a simpler form of an inverse function. Such assumption leads to the model which is called the logistic model (see Agresti, 1984, p.104).

Let us assume that in experiment there are s independent treatments, each represented by fixed number of units m_i . The studied units are classified to k separate categories. Let π_{ji} be the probability of the belonging of the studied unit to the j -th category corresponding to i -th treatment. The logistic model can be now written in the following form (see Miller et al., 1993, compare Bakinowska and Kala, 2007)

$$\log \frac{\gamma_{ji}}{1 - \gamma_{ji}} = \theta_j + \tau_i, \quad j = 1, 2, \dots, k-1, \quad i = 1, 2, \dots, s, \quad (3.1)$$

where θ_j is border (cutpoint) of j -th category, τ_i is the effect of i -th treatment (in result $\theta_j + \tau_i$ means the cutpoint of j -th category for i -th treatment), and γ_{ji} is the j -th cumulative probability corresponding to units of i -th treatment,

$$\gamma_{ji} = \pi_{1i} + \pi_{2i} + \dots + \pi_{ji}, \quad j = 1, 2, \dots, k-1.$$

The results of classification of studied units are usually modeled with the use of multinomial distribution, which is determined by probabilities π_{ji} , $j = 1, 2, \dots, k$, summing up to one, $\sum_{j=1}^k \pi_{ji} = 1$, and the fixed number of units m_i . Our aim is to estimate the unknown cumulative probabilities in model (3.1) based on the experimental data. The observed frequencies, which are natural estimators of unknown probabilities π_{ji} , $j = 1, 2, \dots, k$, will be used in estimation by weighted least squares method.

4. Analysis of experiment

The model described in previous section will be used to analysis of data presented in “**Research material**”. For each trait we have $s=4$ treatments (Derek, D4, D11, D13) and $k=3$ separate categories to which homogeneous units are classified. The data are presented in Tables 1 and 2.

Table 1. Data concerning number of pods per plant for cultivar Derek and its mutants

Treatments	Categories		
	small	medium	large
Derek	1	78	71
D4	8	100	42
D11	34	112	4
D13	2	121	27

Table 2. Data concerning number of branches per plant for cultivar Derek and its mutants

Treatments	Categories		
	small	medium	large
Derek	24	90	36
D4	42	82	26
D11	43	85	22
D13	63	77	10

The parameter θ_j in model (3.1) can be interpreted as the average, with respect to varieties, value of the j -th cutpoint, because effects τ_i sum to zero (compare McCullagh and Nelder 1989, p. 176):

$$\sum_{i=1}^4 \tau_i = 0. \quad (4.1)$$

Remembering about above equality (4.1) we can consider the model (3.1) with smaller number of parameters. The cutpoints for researched trait for next treatments can be written as:

$$\text{Cutpoint of Derek: } \theta_j + \tau_1 = \theta_j + \rho_1 + \rho_2 + \rho_3, \quad j = 1, 2,$$

$$\text{Cutpoint of D4: } \theta_j + \tau_2 = \theta_j - \rho_1, \quad j = 1, 2,$$

$$\text{Cutpoint of D11: } \theta_j + \tau_3 = \theta_j - \rho_2, \quad j = 1, 2,$$

$$\text{Cutpoint of D13: } \theta_j + \tau_4 = \theta_j - \rho_3, \quad j = 1, 2.$$

Using the weighted least squares method, based on the data presented in Table 1, we have obtained the estimates of unknown parameters θ_j $j=1,2$ and ρ_l , $l=1,2,3$ for number of pods per plant:

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \end{pmatrix} = \begin{pmatrix} -3.006 \\ 1.462 \\ 0.414 \\ -1.817 \\ 0.039 \end{pmatrix}$$

and similarly for number of branches per plant

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \end{pmatrix} = \begin{pmatrix} -0.956 \\ 1.749 \\ 0.064 \\ -0.034 \\ -0.669 \end{pmatrix}$$

Therefore, searched treatment effects for number of pods per plant are

$$\begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \hat{\tau}_4 \end{pmatrix} = \begin{pmatrix} -1.364 \\ -0.414 \\ 1.817 \\ -0.039 \end{pmatrix}$$

and for number of branches per plant:

$$\begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \hat{\tau}_4 \end{pmatrix} = \begin{pmatrix} -0.639 \\ -0.064 \\ 0.034 \\ 0.669 \end{pmatrix}.$$

Based on these estimates we can obtain the estimates of cumulative probabilities. According to the model (3.1) the searched probabilities are of such form

$$\hat{\gamma}_{ji} = \frac{\exp(\hat{\theta}_j + \hat{\tau}_i)}{1 + \exp(\hat{\theta}_j + \hat{\tau}_i)}, \quad j = 1, 2, \quad i = 1, 2, 3, 4.$$

The results are presented in Table 3.

Table 3. Estimates of cumulative probabilities for number of pods

	Derek	D4	D11	D13
$j=1$	$\hat{\gamma}_{11}=0.0125$	$\hat{\gamma}_{12}=0.0317$	$\hat{\gamma}_{13}=0.2333$	$\hat{\gamma}_{14}=0.0454$
$j=2$	$\hat{\gamma}_{21}=0.5244$	$\hat{\gamma}_{22}=0.7404$	$\hat{\gamma}_{23}=0.9637$	$\hat{\gamma}_{24}=0.8058$

And similarly for number of branches per plant the results are presented in Table 4.

Table 4. Estimates of cumulative probabilities for number of branches

	Derek	D4	D11	D13
$j=1$	$\hat{\gamma}_{11}=0.1687$	$\hat{\gamma}_{12}=0.2650$	$\hat{\gamma}_{13}=0.2846$	$\hat{\gamma}_{14}=0.4286$
$j=2$	$\hat{\gamma}_{21}=0.7522$	$\hat{\gamma}_{22}=0.8435$	$\hat{\gamma}_{23}=0.8560$	$\hat{\gamma}_{24}=0.9181$

However, in Figures 1 and 2 relation between estimated cumulative probabilities $\hat{\gamma}_{ji}$ obtained using the model (3.1), and cumulative probabilities

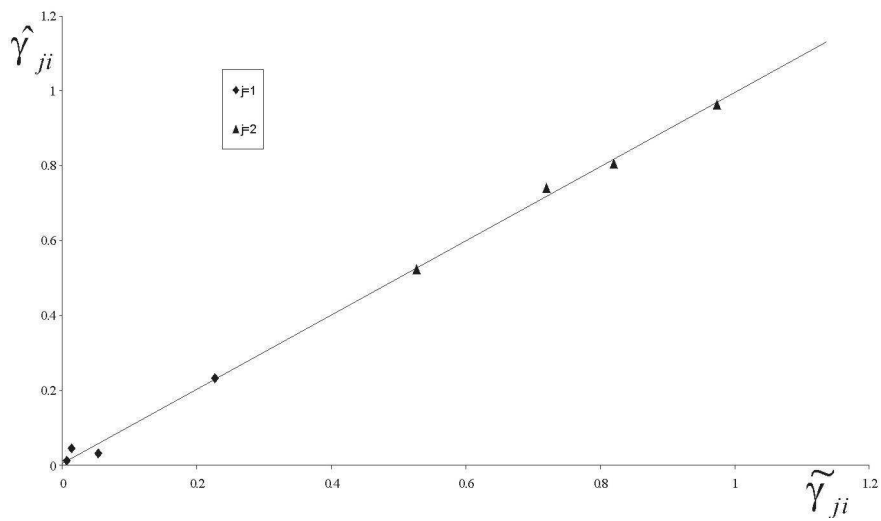


Fig. 1. Relation between estimated cumulative probabilities $\hat{\gamma}_{ji}$ obtained using the model (3.1), and cumulative probabilities $\tilde{\gamma}_{ji}$ for number of pods per plant

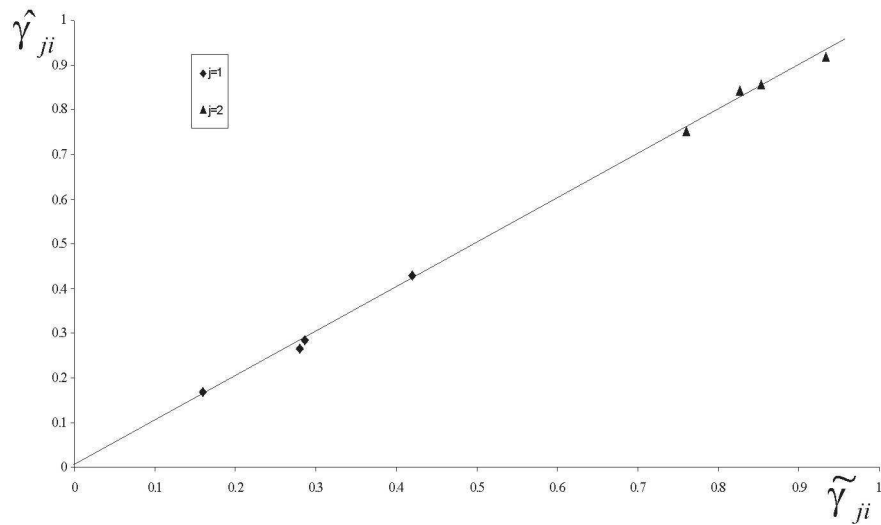


Fig. 2. Relation between estimated cumulative probabilities $\hat{\gamma}_{ji}$ obtained using the model (3.1), and cumulative probabilities $\tilde{\gamma}_{ji}$ for number of branches per plant

$\tilde{\gamma}_{ji} = \tilde{\pi}_{1i} + \dots + \tilde{\pi}_{ji}$ as the natural estimators of γ_{ji} (where $\tilde{\pi}_{ji}$ are the observed frequencies) for number of pods and for number of branches, respectively, is presented.

The analysis of Polish cultivar Derek and their three chosen mutants by using logistic model showed that the most distanced from initial form mutant with respect to number of branches per plant was D13, but with respect to number of pods per plant was D11 mutant.

5. Discussion

In the last time the increasing interest to use of vegetarian diet caused in Poland a visible propagation of consumption such vegetable crop as grasspea, characterized by high nutritional value as well as taste seeds (Milczak et al., 1997). The observed tendency for consumption a slight transformed foods may cause, that not fully ripened grasspea seeds could be very attractive for packing trade as chilled fruits and sterilized tinned food (Kmieciak et al., 2001).

Grasspea apart from very interesting characters, important from agronomical point of view, is characterized by a few undesirable traits (Campbell et al., 1994, Hanbury et al., 2000). The improvement of this above mentioned traits may be obtained through use of mutation induction led to broader genetic variability. The papers presented by Nerkal (1976) and Singh and Chaturvedi (1987) indicate on such possibility.

The obtained results should be treated with great caution as the effect of the initial analysis of research material. Drawing the firm conclusions two aspects has to be taken into account: (i) proposed categories (small number, medium number and large number) are conventionally only and the experimenter (farmer) as the cutpoints can accept any values appropriate his opinion and in practice, (ii) experimental treatments should be examined in the context a lot of traits, because for example large number of pods and/or large number of branches not always connected with its large grain weight, and in consequence it is not good material to further researches.

References

- Agresti A. (1984). *Analysis of ordinal categorical data*. Wiley, New York.
- Alkin R., Macfarlane T. D., White F. A., Bisby F. A., Adey M. E. (1983). Names and synonyms of species and subspecies in the Viciae. Issue 2. *Viciae Database Project Publication No. 2*, Southampton.
- Bakinowska E., Kala R. (2007). An application of logistic models for comparison of varieties of seed pea with respect to lodging. *Biometrical Letters* 44(2), 143-154.
- Campbell C. G. (1997). *Grass pea. Lathyrus sativus L. Promoting the conservation and use of underutilized neglected crops*. Ed. by IPIGRI and IPK 1997, 5-92.
- Campbell C. G., Mehra R. J., Agrawal S. K., Chen Y. Z., El Moneim B., Khawaja H. I. T., Yadov C. R., Tay J. U., Araya W. A. (1994). Current status and future strategy in breeding grasspea (*Lathyrus sativus L.*). *Euphytica* 37, 167-175.
- Grela E. R., Winiarska A. (1997). Chemical composition and nutritive value of flat pea (*Lathyrus sativus L.*). Int. Conf. "Lathyrus sativus – cultivation and nutritive value in animals and humans", Lublin-Radom, 9-10 June 1997, *Proc. of Conf.*, 49-55.
- Hanbury C. D., White C. I., Mullan B. P., Siddique K. H. M. (2000). A review of the potential of *Lathyrus L.* and *L. cicere L.* grain for use as an animal feed. *Animal Feed Science and Technology* 87, 1-27.
- Hartman C. P., Divakar N. G., Nagaraja Rao U. N. (1974). A study on *Lathyrus sativus*. *Ind. J. Nutr. Dietet.* 11, 178-191.
- Kearney J., Smartt J. (1995). *The grasspea – Lathyrus sativus (Leguminosae-Papilionoideae)*. In: Langman P (ed) *Evolution of Crop Plants*, London, 266-270.
- Kmiecik W., Lisiewska Z., Korus A. (2001). Effect of maturity degree of grass pea seeds (*Lathyrus sativus L.*) on the quality of canned products. *Veg. Crops Res. Bull.* 55, 97-107.
- McCullagh P., Nelder J. A. (1983). *Generalized linear models*. Chapman and Hall, London.
- McCullagh P., Nelder J. A. (1989). *Generalized linear models*. 2nd ed. Chapman and Hall, London.
- Milczak M., Pędziński M., Mnichowska H., Szwed-Urbaś K. (1997). Creative breeding of chickling vetch (*Lathyrus sativus L.*) summation of first-stage of investigation. Int. Conf. "Lathyrus sativus – cultivation and nutritive value in animals and humans", Lublin-Radom, 9-10 June 1997, *Proc. of Conf.*, 13-22.
- Milczak M., Pędziński M., Mnichowska H., Szwed-Urbaś K., Rybiński, W. (2001). Creative breeding of grasspea (*Lathyrus sativus L.*) in Poland. *Lathyrus Lathyrism Newsletter* 2(2), 85-88.
- Miller M. E., Davis Ch. S., Landis, J. R. (1993). The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics* 49, 1033-1044.
- Misztal I., Gianola D., Foulley J. L. (1989). Computing aspects of a nonlinear method of sire evaluation for categorical data. *Journal of Dairy Science* 72, 1557-1568.
- Nerkal Y. S. (1976). Mutation studies in *Lathyrus sativus*. *Indian Journal of Genetics and Plant Breeding* 36(2), 223-229.
- Rao C. R., Toutenburg H. (1999). *Linear Models*. 2nd ed. Springer-Verlag, New York.

- Rybiński W., Szot B., Pokora L. (2004). Estimation of genetic variation of traits and physical properties of seeds for grass pea mutants. *International Agrophysics* 18, 339-346.
- Singh M., Chaturvedi S. N. (1987). Effectiveness and efficiency of mutagen alone or in combination with dimethyl sulphoxide in *Lathyrus sativus*. *J. of Agric. Sci.* 57, 503-507.

ANALIZA WYBRANYCH MUTANTÓW ŁĘDŹWIANU SIEWNEGO PRZY ZASTOSOWANIU UOGÓLNIENEGO MODELU LINIOWEGO

Streszczenie

Łędźwian siewny (*Lathyrus sativus* L.) charakteryzuje się niskimi wymaganiami środowiskowymi, wysoką zawartością białka oraz korzystnym składem kwasów tłuszczowych w ziarnie. Równocześnie, łędźwian posiada dużo niekorzystnych cech z agronomicznego punktu widzenia. Dodatkową zmienność genetyczną wprowadzaną do nowych odmian łędźwianu siewnego można wywołać poprzez mutagenezę. Materiał do badań stanowiły trzy mutanty łędźwianu oraz ich forma wyjściowa Derek, analizowane pod względem ośmiu cech na polu doświadczalnym Instytutu Genetyki Roślin w Cerekwicy. Doświadczenie przeprowadzono w układzie bloków losowanych w trzech powtórzeniach. Mutanty otrzymano w wyniku traktowania nasion odmiany Derek światłem lasera i chemomutagenami N-nitroso-N-metylomocznikiem (MNU) i azydkiem sodu (NaN_3). Celem pracy było porównanie odmiany Derek z jego mutantami pod względem dwóch wybranych cech: liczby strąków z rośliny oraz liczby rozgałęzień z rośliny. Obie rozważane cechy traktowane były jako dyskretne, co umożliwiło zastosowanie modelu logistycznego, który należy do klasy uogólnionych modeli liniowych.

Słowa kluczowe: model logistyczny, mutanty łędźwianu, uogólniony model liniowy

Klasyfikacja AMS 1993: 62F10, 62J02, 62F12