Evaluation of the impact of fertilizers and seed quality on winter wheat yield

Marzena Iwańska[1], Zbigniew Laudański[2†], Tadeusz Oleksiak[2]

[1]Department of Experimental Design and Bioinformatics, Warsaw University of Life Science – SGGW; e– mail: m.iwanska123@gmail.com
[2]Plant Breeding and Acclimatization Institute – National Research Institute (IHAR-PIB), Radzików

SUMMARY

The aim of the study was to evaluate the effect of mineral fertilization and seed quality on the yield of winter wheat in production conditions. This assessment is made in terms of the expected probabilities of success in relation to the yield of analyzed cultivars, taking into account the interaction of the factors considered. Analyses were performed on data from 3815 fields. The impact of fertilizers and seed quality was evaluated using logistic regression. Grain yield was transformed into a binomial variable, where values were divided into two classes, i.e. below the mean and above the mean. The results of the analysis proved a significant effect of fertilization rate, which was modified by different seed quality. The highest probability of obtaining yields above the mean was observed for pre-basic and basic seed quality at high fertilizer rates.

Key words: logistic regression, odds ratio, survey research, winter wheat

1. Introduction

In the analysis of agricultural statistical data, modeling issues using regression analysis are significant. We are interested in assessing the degree to which causal factors influence certain final effects. The result is a general assessment of the extent to which a particular factor determines the final result and whether the impact is negative or positive. Solved problems can provide examples of models with methods of linear and non-linear multiple regression. In problems of this type we often deal with differently described causal factors, as well as with final effects (quantitative or categorical variables). In practice, the process of the

applied measurement consists in assigning numbers to certain occurrences or objects so that those numbers reflect a possibly faithfully measured reality, i.e. features of the phenomenon or the object. If we want to determine how a continuous dependent variable is influenced by the causal variables, we usually use a linear multiple regression analysis. However, this method cannot be used in cases where the dependent variable has a binary distribution. In such cases we can apply logistic regression, which in many respects is similar to linear regression. This method is not widespread among researchers in the agricultural sciences, but has found wide application in the economic, social and medical sciences. The inspiration for and main sources used in the preparation of this work, especially in the methodology of logistic regression, were Cox and Snell (1989), Collett (1991), Hosmer and Lemeshow (2000), Agresti, (2002) and Larose (2012), as well as several socio-economic studies: Danieluk (2010), Manso et al. (2010), Mańkowski et al. (2007), Nath et al. (2016), Roszkowska-Mądra et al. (2010) and Finger et al. (2013).

This paper presents the use of logistic regression analysis on production data from studies on the evaluation of agricultural crops, taking into account both quantitative and qualitative variables.

The aim of the study was to evaluate the effect of fertilization rate (NPK – nitrogen, potassium and phosphorus) and seed quality on the grain yield of winter wheat, treated as a binomial variable (with values denoting yields respectively below and above the mean).

2.   Material and methods

The material for statistical analysis consisted of long-term (1992–2003) production data from individual farms carrying out agricultural accounting. These data were obtained as part of research conducted by the Laboratory of Economics of Seed and Plant Breeding IHAR-PIB Radzików. The questionnaire survey was carried out among farms holding agricultural accounting data for the needs of the Institute of Agricultural Economics and Food Economy (IAFE-IERiGŻ). Surveys

were completed by IAFE-IERiGŻ agricultural accountants. Each survey consisted of one farm, and concerned one field, the cultivation of one variety of one species of crops (wheat and other cereal species) and information on the type of seed, characteristics of fields, factors of production and agricultural techniques. Survey data regarding wheat was used to provide and prepare information on 3980 fields.

In this study, results were obtained from the analysis of the yield of winter wheat in the conditions of production for 3815 fields (out of the 3980 fields examined in the survey) that met the criteria, i.e. fields with an area not exceeding 10 ha using a non-zero quantity of mineral fertilization. The obtained yield (dependent variable) was the yield in dt per hectare in the field. For the analyses, the yield was transformed into a binomial variable where yields below the mean were treated as 0 and those above the mean as 1. The quantity of mineral NPK fertilizer, in kg/ha of pure elements, was used as an independent variable for quantitative analysis, and the quality of seed as a nominal variable. In the analyses this variable will be represented by four binary variables, representing respectively Pre-basic, Basic, First Certification and Second Certification, and Non-certified as the reference category, for the individual grades of seed. Each of these binary variable takes the value 1 when the given degree of seed quality applies, and 0 otherwise. The statistical calculations assumed 100 kg/ha as a unit of fertilization for the accurate assessment of the relevant regression coefficients.

3. Multivariate logistic regression

The method of linear regression is used to approximate the relationship between a continuous variable (e.g. in prior analysis) and a set of explanatory variables that can be expressed on different scales. It is often the case that the dependent variable is a qualitative (binary) variable rather than a continuous variable. In such cases, the linear regression model is not adequate, but in its place there can be used the logistic regression model, which in many respects is similar to the linear regression model. Multivariate logistic regression is a method of describing

the relationship between a qualitative dependent variable and a set of explanatory variables. Its graphical presentation in the simplest case is shown in Figure 1. The shape resembles the form of the cumulative distribution function of a normal distribution. Like the distribution function of the normal distribution, the values lie in the interval [0;1] and are determined for all possible values on the axis of real numbers. As a result, the relationship between quantitative and independent variables is presented. Its value is treated as a probability. This function applies, among others, to issues of describing plant growth curves.
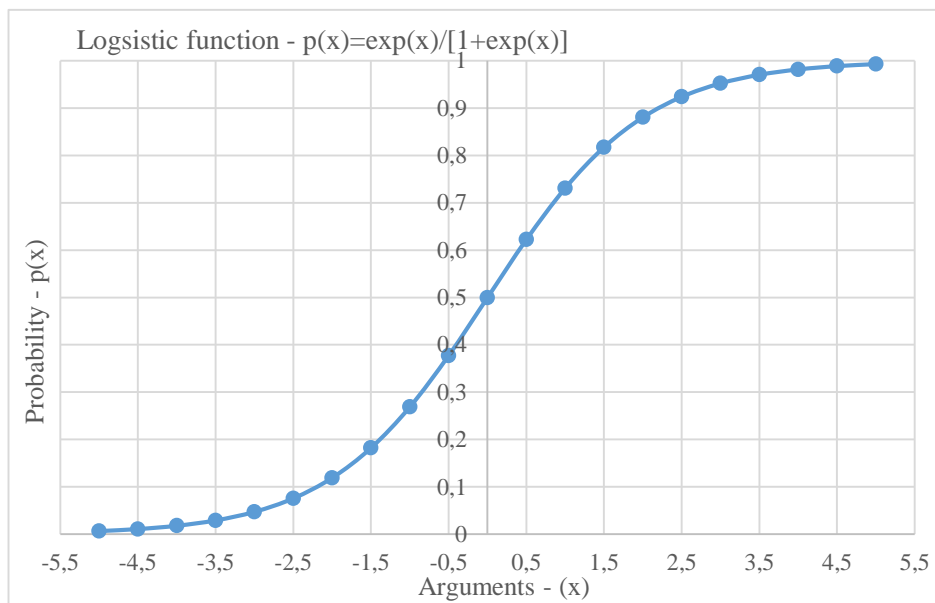

Figure 1. General example of logistic function

The argument of this function usually occurs as a linear combination of independent variables, which we can write:

$$g(x) = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_k \cdot x_k = \beta_0 + \sum_{i=1}^{k} \beta_i \cdot x_i \, .$$

The logistic function of interest to us is written as:

$$p(x) = \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{1}{1+e^{-g(x)}} = \frac{\exp(g(x))}{1+\exp(g(x))} = \frac{1}{1+\exp(-g(x))}.$$

Logistic regression uses a specific way of expressing the likelihood of a random event, called odds. In practice, instead of the classical definition of probability, as the ratio of the number of successes to the number of all possibilities, we calculate the odds expressed as the ratio of the number of successes to the number of failures. This can be expressed as

$$odds = \frac{p(x)}{1-p(x)}, \text{ and as an inverse function: } p(x) = \frac{odds}{1+odds}.$$

If $0 < p(x) < 1$, then $0 < \dfrac{p(x)}{1-p(x)} < +\infty$, whereas $-\infty < \ln\dfrac{p(x)}{1-p(x)} < +\infty$

.

This can be used to estimate the chance. The logarithmic regression method is not limited to the interval $(0;1)$. The transformation function using the logarithm of the odds is called *logit*, which is written:

$$\text{logit}[p(x)] = \ln\frac{p(x)}{1-p(x)} = g(x).$$

Thus, the inverse function has the following equation:

$$p(x) = \frac{\exp[\text{logit}(p(x))]}{1+\exp[\text{logit}(p(x))]} = \frac{1}{1+\exp[-\text{logit}(p(x))]}.$$

The *logit* function, with values from minus infinity to plus infinity, is both a continuous and a monotonic function. Its disadvantage is that it does not exist, for example, in the case of the multiple regression form of analytical solutions for the estimation of the coefficients of logistic regression. In this case we have to use the method of maximum likelihood estimators, by means of which the parameters of maximum likelihood for the observed data are obtained. The values of the estimated coefficients are not subject to interpretation. Instead of them the value of the *odds ratio* (*OR*) is used for inference. OR can be used for comparison

of the effects of different values of independent variables $x = [1, x_1, x_2, \ldots, x_k]$. This can be expressed by the equation:

$$\frac{p(x)}{1-p(x)} = \exp\left(b_0 + \sum_{j=1}^{k} b_j \cdot x_j\right) = \exp(b_0) \cdot \prod_{j=1}^{k} \exp(b_j \cdot x_j) = \exp(b_0) \cdot \prod_{j=1}^{k} \left[\exp(b_j)\right]^{x_j}.$$

The term $\exp(b_j)$ reflects the relative change in the odds ratio of an event under the influence of the factor described by the value of variable X_j, assuming the stability of the other variables in the equation. This value is interpreted by comparison with the value 1; in other words, it is a number that indicates by how many times the chance of success increases with the increase of a given variable by one unit. The signs of the coefficients indicate the directions of the impact of the variable on the probability of success. Three possible cases can be presented with corresponding relationships of the values in question:

- $b_j > 0 \Rightarrow \exp(b_j) > 1$ – it is considered that the factor described by the independent variable X_j has a stimulating effect on the probability (possibility) of the studied phenomenon, at a predetermined impact of other variables included in the equation;

- $b_j < 0 \Rightarrow 0 < \exp(b_j) < 1$ – it is considered that the factor described by the independent variable X_j has a limiting effect on the probability of occurrence of the phenomenon under investigation, with the influence of other variables included in the equation;

- $b_j = 0 \Rightarrow \exp(b_j) = 1$ – it is considered that the factor described by the independent variable X_j has no influence on the probability of occurrence of the phenomenon under investigation, with the influence of other variables included in the equation. Such a factor is eliminated from the model as a minor contributor (not statistically significant).

Practically, testing the significance of the factors in the model leads to study of the significance of the odds ratio in the impact (positive or negative) of a given factor – the independent variable – on the probability of occurrence of the phenomenon studied.

The analyses were conducted using SPSS statistical software.

Assessment of significance of independent variables – model verification

The basic tool for evaluating the significance of all k independent variables in the model is the *likelihood ratio test*. This test verified the global hypothesis:

$$H_0 : \bigwedge_j \beta_j = 0$$

where $j = 1, 2, ..., k$, against the alternative hypothesis:

$$H_1 : \bigvee_j \beta_j \neq 0 .$$

The test statistics take the form

$$\chi^2_{emp} = -2 \cdot \ln \frac{L_0}{L_{FM}} = \left(-2 \cdot \ln L_0\right) - \left(-2 \cdot \ln L_{FM}\right),$$

where L_0 is the value of the likelihood function for a model with no predictors, and L_{FM} is the likelihood for the model being estimated.

For large numbers there is an asymptotic distribution χ^2 with k degrees of freedom.

It should be noted that if the value $-2 \ln L_{(0)}$ decreases (for example, after adding a variable) then the reliability of the model increases.

NB. When building a model, it should be kept in mind that the number of observations is at least ten times greater than the estimated model parameters – $n \geq 10(k + 1)$.

The number L_{FM} represents the *maximum likelihood function*, here for the full model with k independent variables, and the number L_0 represents the *maximum likelihood function* for the model containing only the intercept (the null model), where $\ln L_0 = n_1 \ln n_1 + n_0 \ln n_0 - n \ln n$, where n is the number of observations, n_1 is the number (1) of the dependent variable, and n_0 is the number (0) of the dependent variable. On the basis of the coefficient of determination R^2 in linear regression, describing the degree of fit of the model, there was proposed by Nagelkerke (Hosmer and Lemeshow, 2000) the equivalent:

$$R^2_{Nagelkerke} = \frac{1 - \exp\left[-(2/n)\cdot\left(\ln L_{FM} - \ln L_0\right)\right]}{1 - \exp\left[-(2/n)\cdot \ln L_0\right]}.$$

The values of the model's degree of fit R^2 according to this proposal fall within the closed range [0;1], with values of 1 indicating the perfect fit of the model, while a value of 0 denotes the complete absence of fit.

After verification of the global hypothesis obtained for the regression model, an important task is to assess the statistical significance of each variable appearing in the model. Based on the regression coefficient and its error, we can conclude that the independent variable for which this ratio has been estimated has a significant effect on the dependent variable. For this purpose we use the test proposed by Wald. The null hypothesis has the form

$$H_0\colon \beta_J = 0 \text{ at } H_1\colon \beta_J \neq 0,$$

this can be tested using statistics calculated according to the formula:

$$\chi^2_{Wald} = \left(\frac{b_j}{s_e\left(b_j\right)}\right)^2,$$

which statistic has a χ^2 distribution with 1 degree of freedom, or using the statistics

$$Z_{Wald} = \frac{b_j}{s_e\left(b_j\right)},$$

which has a standardized normal distribution $N(0;1)$. At the same time, the regression coefficient b_j is the estimator of the regression coefficient ($b_j = \beta_j$), and $s_e(b_j)$ denotes its error.

Interpretation of the regression model of independent qualitative variables with two values

Let us introduce the simplest example of the application of the described logistic regression method in the agricultural sciences, using the described data. To this end, our binary dependent variable assumes a value of 1 (success) when the

winter wheat yield (y) per hectare is at least at the average level of the varieties tested in a given year. Otherwise it takes the value 0 (failure). The independent variable (also binary) in our analysis will be represented by the quality of the seed (x), where 1 denotes certified seed and 0 non-certified seed.

The resulting equation for the model of two variables as a function of a *logit* function and *logistic* regression is:

$$\hat{g}(x) = \underset{\substack{s_e(b_0)=0.043 \\ P<0.000001}}{-0.493} + \underset{\substack{s_e(b_1)=0.068 \\ P<0.000001}}{0.974} \cdot x$$

i.e. $\hat{p}(x) = \dfrac{e^{-0.493+0.974 \cdot x}}{1+e^{-0.493+0.974 \cdot x}} = \dfrac{\exp(-0.493+0.974 \cdot x)}{1+\exp(-0.493+0.974 \cdot x)}$.

At a similar level of probability, we find the existence of a constant as well as a regression coefficient, which means that the variable *degree of qualification* is useful for evaluating the behavior of the dependent variable – which is to obtain the appropriate crop yield.

Knowing the equation, we can calculate (Table 1) the probabilities $p(x_i)$ and the odds of obtaining a higher than average yield for a given variety in the year of research using non-qualified (0) and qualified (1) seeds to finally determine the odds ratio, equal to $e^{b_i} = \exp(b_i)$ ($e^{0.974} = \exp(0.974) = 2.648$), with appropriate confidence intervals for a fixed level of significance $\alpha=0.05$.

Table 1. Logistic regression calculations for winter wheat yield (binomial dependent variable 1 – higher than the average; 0 – below the average) for a given variety in the year of research using non-qualified (0) and qualified (1) seeds (x_i)

x_i	$\hat{g}(x_i)$ Logit	$\hat{p}(x_i)$ Probability	$\dfrac{\hat{p}(x_i)}{1-\hat{p}(x_i)}$ Odds	Odds ratio	Confidence intervals
0	–0.493	0.37910	0.6106	0.611	
1	0.481	0.61785	1.6168	2.648	2.318–3.025

From Table 2 it can be seen that the probability of success when using certified seeds (1) according to the model is 0.618 (61.8% of cases). With non-certified seeds it is 0.379 (37.9% cases). The superiority of certified seeds is of course confirmed by other indicators. The chances in the first case are 2.648 times higher (by 164.8%) than the second (odds ratio), or otherwise, the chances of success with non-certified seed are 37.8% of those for the certified seed.

Table 2. Logistic regression equation for the effect of mineral fertilization and seed quality (x) on the yield (as binomial variable, where 0 – below the mean; 1 – above the mean) of winter wheat (y) in production conditions

Model of regression		b_i	Standard error b_i	$\chi^2_{Wald's}$ statistics
Constant	b_0	−1.783	0.092	371.934
Fertilization (NPK) (x_1)	b_1	0.879	0.054	266.602
Pre-basic (x_2)	b_2	1.728	0.432	16.028
Basic (x_3)	b_3	1.042	0.153	46.377
First Certification (x_4)	b_4	0.715	0.096	54.973
Second Certification (x_5)	b_5	1.382	0.240	33.123
Interaction of Second Certification * Mineral fertilization NPK (x_6)	b_6	−0.527	0.126	17.477

Model of regression	p_{value}	Odds ratio $\exp(b_{ii})$	Confidence intervals
Constant	<0.0005	0.168	–
Fertilization (NPK) (x_1)	<0.0005	2.409	2.168–2.678
Pre-basic (x_2)	<0.0005	5.629	2.416–13.116
Basic (x_3)	<0.0005	2.835	2.100–3.827
First Certification (x_4)	<0.0005	2.044	1.692–2.469
Second Certification (x_5)	<0.0005	3.985	2.488–6.381
Interaction of Second Certification * Mineral fertilization NPK (x_6)	<0.0005	0.590	0.461–0.756

Let us introduce the practical meaning of *odds* and *odds ratio*. For this purpose, we will use the cross-table of the values of our variables (Table 3).

Table 3. Cross table of winter wheat yield (y – dependent variable, where 1 – grain yield greater than mean, 0 – grain yield below the mean) and the quality of the seed (x – independent variable, where 1 is certified seed, and 0 uncertified seed)

$y \setminus x$	$x = 0$	$x = 1$	Total
$y = 0$	1402	595	1997
$y = 1$	856	962	1818
Total	2258	1557	3815

According to this table, we have:

$$\hat{p}(y = 0 \,|\, x = 0) = \frac{1402}{2258} = 0.6209, \quad \hat{p}(y = 1 \,|\, x = 0) = \frac{856}{2258} = 0.3791, \quad \text{or}$$

$$odds(y = 1 \,|\, x = 0) = \frac{0.3791}{1 - 0.3791} = \frac{0.3791}{0.6209} = \frac{856}{1402} = 0.6106$$

Similarly we calculate the second value of interest, i.e. $odds(y = 1 \,|\, x = 1) = \frac{962}{595} = 1.6166$. In this way, the odds ratio (*OR*) of obtaining a sufficiently high yield when using certified seed over the use of non-certified seeds is:

$$OR = \frac{1.6168}{0.6106} = 2.648$$

We see, therefore, that logistic regression on such qualitative variables represents an analysis of the numerical table and analyzes the frequency (empirical probability) of occurrence of individual cases.

Dummy qualitative variables – detailed analysis of a model with a quantitative variable

We perform logistic regression analysis of the binary variable Y, determined as 1 (*success*) if the wheat yield (y) of a given variety is at least at the average level of tested varieties in the test year, and 0 (*failure*) otherwise. Based on the analysis we have found significant influences of all variables on the probability of success, as well as for linear regression, including the interaction of second certification with mineral fertilization. The results of the regression analysis of the result

variable Y with respect to causal variables, i.e. seed quality and mineral fertilization, are presented in Table 2 and Figure 2.
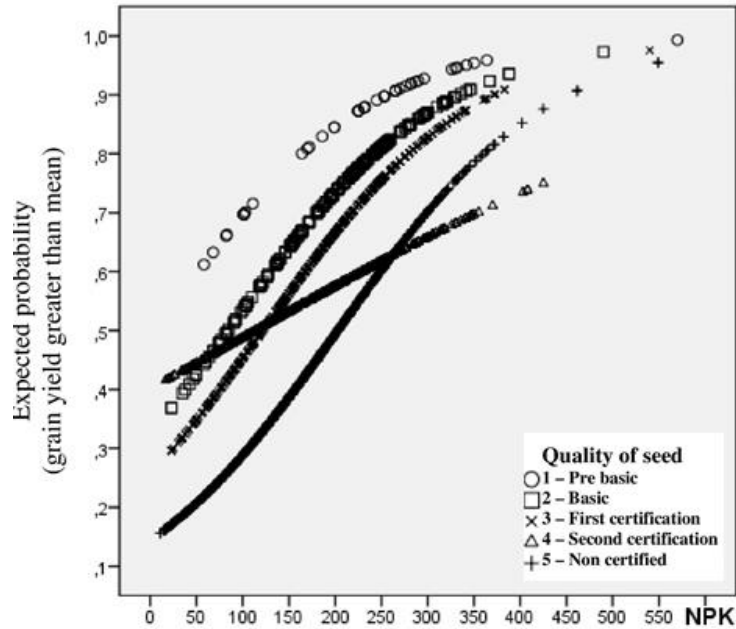

Figure 2. Logistic regression curves for the effect of mineral fertilization and seed quality on the yield of winter wheat (*y* as binomial variable, where 0 – below the mean; 1 – above the mean) in production conditions (1–Pre-basic, 2–Basic, 3–First Certification, 4–Second Certification, 5–Non-certified)

It should be noted that the global hypothesis of non-regression was rejected ($\chi^2_{emp} = 556.406$, but $R^2_{Nagelkerke} = 0.181$).

Another test used to assess the compatibility of the regression model with real data is the compatibility test proposed by Hosmer and Lemeshow (2000) based on the *chi-square* distribution. For our problem the value obtained, $\chi^2_{emp} = 11.544$ with 8 degrees of freedom, does not reject the hypothesis of the model's conformance to the analyzed reality ($P = 0.173$).

The resulting equation *logit* function can be presented as:

$$(\log it(\hat{p}(x)) = \hat{g}(x) = -1.783 + 0.879\, x_1 + 1.728\, x_2$$
$$+ 1.042\, x_3\, 0.715\, x_4\, 1.382\, x_5 - 0.527\, x_6.$$

The logistic regression function can be written according to one of the formulae:

$$\hat{p}(x) = \frac{\exp(\hat{g}(x))}{1 + \exp(\hat{g}(x))} = \frac{1}{1 + \exp(-\hat{g}(x))}.$$

The interpretation of the odds ratio (Table 2) for the case of qualitative characteristics is somewhat obvious, because all indicators for those characteristics relate to preferential characteristics for which this issue is variable – uncertified seed. So we have that Pre-basic seed makes it 5.629 times more likely (by 462.9%) to obtain success in the form of a sufficiently high yield, than in the case of uncertified seed.

Continuing accordingly: Base gives a 2.835 times greater chance (by 183.5%) and First Certification a slightly more than 2 times higher chance (by 104.4%). Second Certification gave a different response, which is presented graphically in Figure 2. In the absence of fertilization the chances to obtain a yield higher than average are 3.985 times greater (by 298.5%), but with an increase of fertilization, these chances gradually decrease. An increase in fertilization by one unit (100 kg) results in a decrease in the chance for non-certified seeds by an average of 41%.

As regards NPK fertilization, the number 2.409 represents the multiplier of the probability of success if the fertilization is increased by one unit (100 kg/ha) relative to no mineral fertilization (an increase by 140.9%). This means that if we use a dose of two units (200 kg/ha), the chance for success increases on average 5.803 times (by 480.3%) compared with no fertilization and 2.409 times compared with a single dose. If the applied dose is 0.5 units (50 kg/ha), then the chance of success relative to no fertilization is an average of 1.552 times higher (by 55.2%).

From the general regression form (Table 2), the logistic regression equation of the correspondingly high crop yields is shown in relation to fertilization for individual seed qualification stages (Table 4).

Table 4. Results of logistic regression presenting the effect of each grade of seed certifications (x) on the yield of winter wheat (y as binomial variable, where 0 – below the mean; 1 – above the mean)

	Logit functions
5 – Non-certified	$\hat{g}(x) = -1.783 + 0.879x_1$
4 – Second Certification	$\hat{g}(x) = -0.400 + 0.352x_1$
3 – First Certification	$\hat{g}(x) = -1.068 + 0.879x_1$
2 – Basic	$\hat{g}(x) = -0.741 + 0.879x_1$
1 – Pre-basic	$\hat{g}(x) = -0.0547 + 0.879x_1$

Tables 2 and 4 above show the significant effect of mineral fertilization on the probability of success in achieving a high yield. The fertilizer effects for all degrees of grain quality in addition to the second sifting are similar (no interaction, see Figure 2). Mineral fertilization stimulates a permanent and similar increase in the chances of achieving a high yield. In contrast, the interaction found for second certification of seeds with mineral fertilization highlights, as in linear regression, another consequence of the screening, namely increasing mineral fertilization. Analysis of the data indicates that the use of seeds with second certification reduced the increase in chances for seeds not subjected to fertilization above 262.4 kg/ha – see Figure 2. At the same time, low fertilization rates resulted in better yields than first certification (for fertilization below 126.8 kg/ha) and basic for very low fertilization rates (up to 64.7 kg/ha).

Summary of consideration

In estimating the logistic regression model, each case of data is modeled as a predicted observation of 0 if the probability value is less than p_{const}. If the probability is greater than p_{const}, then the predicted value is equal to 1. The point p_{const} is called the cut-off point, whose value in the interval (0;1) is arbitrarily determined. If the dependent variable is the number of successes compared to the number of failures, the most commonly accepted value is $p_{const} = 0.5$. On the

other hand, in cases of significant difference between these numbers, it is proposed to accept the mean probability, which is equal to the ratio of the number of successes to the number of all cases. For the analyzed survey data, the value of the cut-off point is given by:

$$p_{const} = 1818 \big/ 3815 = 0.477.$$

We have obtained an estimate of the accuracy of the model forecast, shown in Table 5.

The ratio of the number of observations predicted by the model as a success – 1187/1818, expressed as a percentage (65.3%) – is called the sensitivity or sensivity of the model. However, the proportion of the number of observations that were actually unsuccessful – 1326/1997, expressed as a percentage (66.4%) – is called its specificity. In practice, this means that if the *sensitivity* index is greater than the *specificity* index, then the resulting model better predicts success than failure.

Table 5. Contingency table for observed and expected grain yield (where 1 – grain yield greater than mean, 0 – grain yield below the mean)

Observed \ Expected	0	1	Total	% compatible
0	1326	671	1997	66.4
1	631	1187	1818	65.3
Total	1957	1858	3815	65.9

One may also calculate the percentage of correct prediction of success or failure by the regression model: (1326+1187)/3815, or 5.9%. This is a measure of the accuracy of forecasts, and a measure of the quality of fit of the model. This number (65.9%) is treated as a kind of determination factor. Another way to evaluate the model fit is to use the chi-squared independence test, which verifies the hypothesis of independence of observed results and those obtained from the model. The value is $\chi^2_{emp} = 382.565$. This means that there is a significant

dependence between these results. The Cramer dependency ratio for this important relationship is 0.3167.

4. Conclusions

- The logistic regression of many variables predicts the probability of achieving a certain success, presented by the result variable, for any combination of variants of the yield predictors used as independent variables.

- It is possible to use the regression values obtained using the regression methods discussed in comparative analyses of different variants of selected yielding agents (e.g. varieties) due to certain yield conditions.

REFERENCES

Agresti A. (2002): Categorical Data Analysis, John Wiley & Sons, New Jersey.
Collett D. (1991): Modelling binary data. UK. Chapman & Hall, London.
Cox D.R., Snell E.J. (1989): Analysis of Binary Data. Chapman & Hall, London.
Daniluk B. (2010): Zastosowanie regresji logistycznej w badaniach eksperymentalnych. Psychologia Społeczna, t. 5: 199–216.
Finger R., El Benni N. (2013): Farmers' adoption of extensive wheat production – Determinants and implications. Land Use Policy 30: 206–213.
Hosmer D., Lemeshow S. (2000): Applied Logistic Regression. John Wiley&Sons, New Jersey.
Larose D.T. (2012): Metody i modele eksploracji danych. PWN Warszawa.
Mańkowski D.R., Oleksiak T. (2007): Czynniki determinujące stosowanie kwalifikowanego materiału siewnego w gospodarstwach rolnych. Biul. IHAR 244: 5–9.
Manso M.C., Cerqueira R.M., Fernandes C., Correia M. (2010): Logistic multivariate regression analysis as a tool to predict fibrosis in light-drinking chronic hepatitis C patients. Biometrical Letters 47(1): 33-44.
Nath D.C., Vishwakarma R.K., Bhattacharjee A. (2016): A selection modelling approach to analysing missing data of liver Cirrhosis patients. Biometrical Letters 53(2): 83-103.
Roszkowska-Mądra B., Mańkowski D.R. (2010): Determinanty decyzji rolników o korzystaniu z funduszy Unii Europejskiej i kredytów na działalność rolniczą: Przykład dla rolnictwa z rozwiniętym systemem produkcji mlecznej w województwie podlaskim. Roczniki Nauk Rolniczych, Seria g, 97(1):14–26.