STANISŁAW MAŃKO
University of Technology and Life Sciences
Bydgoszcz

USE OF STATISTICAL AND MATHEMATICAL METHODS IN TABLE ANALYSES ON THE EXAMPLE OF THE PRODUCTION COSTS AND GROSS MARGINS IN DAIRY COWS REARING IN 2006

Abstract

The paper addresses the problem of the purposefulness of using statistical and mathematical methods in the analysis of production, costs and gross margins. The research shows that the application of variance analysis may increase the level of reliability of conclusions, it does not, however, question most conclusions based on the analysis of table lists. The selection of detailed tests should take account of the nature of the distribution of analyzed variables. Since most variables do not meet the conditions of normal distribution, nonparametric tests should be used in most cases.

Introduction

Most economic studies carried out on the basis of empirical data consist in tabular analysis of results presented in the form of the so called distributive series. If such research is based on bulk data the resulting conclusions should not raise any objections. However, sometimes research samples include up to a hundred or up to a thousand research objects and then it may turn out that the differences between the average results of individual classes are smaller than the differences between observations within a given class. In such circumstances the conclusions on the observed differences may be unjustified. In order to consider the identified differences as significant, it is necessary to support the tables with statistical and mathematical methods of analysis. In the case of biometric research, e.g. within the scope of agricultural experimentation, drawing conclusions with the use of mathematical statistics is an everyday practice. It seems that as regards economic and agricultural research, the use of such methods should also receive more attention.

The Institute of Agricultural and Food Economics – National Research Institute (IAFE-NRI) for many years has carried out research under the Agricultural Products

Data Collection System AGROKOSZTY [4]. This research, first of all, aims at assessing the changes in the level of production and inputs on the basis of reliable source data and the differences in the gross margin of selected production activities for traditional and organic holdings. Because field studies are very labour-consuming and costly, the size of the research samples is usually between one and two hundred observations for individual crop and livestock activity. Until 2005 the results were developed mainly with the use of tables. For the first time the statistical and mathematical methods were applied in the analysis of results for 2006 in the case of costs and gross margins concerning two crop activities, i.e. winter wheat and rye [4, p. 204-225]. The results of these studies confirm that the differentiation of the analyzed variables is so great, that differences between groups are significant only in some cases.

One of the basic production activities examined under the AGROKOSZTY research is dairy cows rearing. Data from 159 holdings raising dairy cows was collected in 2006. Results were analysed according to the level of gross margin calculated per 1 cow annually and at the regional level [4, p. 93]. Under this research, while analysing the value of milk production, inputs, costs and gross margin in milk production, statistical tests were not used to verify differences.

The paper aims primarily at assessing – with the aid of mathematical statistics methods – to what extent the adopted rules of grouping are sufficient to evaluate the actual differences in production, costs and gross margins, as well as examining the possibilities of using parametric and nonparametric tests of significance in these contexts.

In the paper the rules of research results grouping were applied by the team of IAFE-NRI staff under the guidance of A. Skarżyńska [4, p. 22-23]. The statistical analysis includes grouping by quartiles of the gross margin calculated per 1 cow and according to FADN regions. From the wide range of variables characteristic of dairy cows rearing under AGROKOSZTY programme, only some variables were selected, which describe the production conditions and organization, milk yield of cows, production value, prices of products, direct production costs and gross margin (see table 1). Some variables characteristic of specific items of direct costs, which are available in the AGROKOSZTY database, were omitted due to the methodological nature of the study.

Research method

Grouping data according to some defined features is justified when the established groups are characterised by greater homogeneity compared to the entire research sample and the arithmetic means of analytical variables calculated for individual groups differ significantly between each other. Parametric and nonparametric tests based on an analysis of variance are crucial in analysing the differences significance [2, 3, 5].

The initial assessment of differentiation of selected variables with the use of arithmetic mean, the minimum and maximum values, standard deviation and indicator of variation was the starting point for the research carried out with the use of

tests of significance. The analysis of variables distribution includes a nonparametric Shapiro-Wilk test. This test is considered to be the strongest tool among the tests used to verify the normality of variables distribution [1]. The null hypothesis in this test is the following: the variable is characterised by normal distribution. Rejection of the null hypothesis implies adoption of the alternative hypothesis: the distribution of variables deviates from the normal distribution [1, 2, 3, 5]. Moreover, the asymmetry of variables was assessed with the use of the skewness indicator, whereas concentration – with the use of kurtosis. The closer skewness indicator and kurtosis are to zero, the less the distribution of a given variable deviates from normal.

Analysis of variance is the basic method of statistical analysis of significance for different groups of research objects in case of quantitative variables. One of the strongest and most often used tests is the F-Snedecor test [2, 3, 5]. This test belongs to the group of parametric tests, which give correct results for variables characterised by normal distribution. The test was calculated for all analysed variables, however, only the results for the variables, which were characterised by normal distribution, were taken into account in the analysis.

Economic research often concerns variables, which do not have the normal distribution. In such cases it is necessary to use nonparametric tests. When the variables are of quantitative nature and the number of compared groups is more than 2, the H Kruskal-Wallis test may be used [2, 3]. This test made it possible to assess the significance of differences of variables, which were not characterised by normal distribution.

Research results

Table 1 presents a set of variables and their formal and statistical characteristics. The variables describing the size of the holdings, in which the empirical data was recorded (RO02, RO03, RO04), are characterised by an especially high degree of differentiation. A relatively high degree of differentiation is also observed for the number of cows (RO08), estimated costs of non-marketable feed consumption (RO35) and herd replacement costs (RO22). Milk prices are the least differentiated (RO10).

Table 2 presents the basic parameters, which characterise the distribution of the examined features. Using the Shapiro-Wilk test it may be stated that only 5 variables among the analysed ones are characterized by a normal distribution (the significance level of the test SW>0.05):

- milk yield of cows (RO09),
- price of weaned calves (RO11),
- total value of production (RO14).
- value of milk (RO15),
- gross margin (RO45).

In the case of other variables the hypothesis about normal distribution should be rejected with the probability of error below 5% ($p<0.05$). It means that the analysis of variance based on the F-Snedecor test may lead to wrong conclusions despite the fact that the test is highly resistant to variables deviations from the normal distribution.

Table 1

The initial characteristic of differences between the analysed variables

Name of the variable	Sign	Number of observations	Arithmetic mean	Minimum value	Maximum value	Standard deviation	Indicator of variation
Area of arable land (ha)	RO02	159	25.10	2.46	333.95	30.68	122.23
Area of permanent pasture (ha)	RO03	159	8.60	0.00	81.02	11.95	139.07
Utilized Agricultural Area (ha)	RO04	159	33.73	5.58	367.29	36.22	107.37
Soil valuation indicator (points)	RO07	159	0.80	0.06	1.72	0.34	42.68
Average annual number of dairy cows (unit)	RO08	159	19.96	2.46	101.85	16.42	82.27
Milk yield of cows (l/unit)	RO09	159	4853.40	1628.51	8452.27	1463.83	30.16
Price of milk sales (PLN/l)	RO10	157	0.92	0.48	1.19	0.14	14.71
Price of weaned calves (PLN/kg)	RO11	159	9.02	3.77	15.46	2.11	23.37
Selling price of culled cows (PLN/kg)	RO12	116	2.83	0.48	6.69	0.76	27.02
Total value of production (PLN/unit)	RO14	159	5284.43	1719.41	10650.09	1740.41	32.93
including: value of milk (PLN/unit)	RO15	159	4575.12	1238.10	9930.39	1738.84	38.01
Total direct costs (PLN/unit)	RO21	159	2064.52	243.25	4679.04	755.68	36.60
including: herd replacement (PLN/unit)	RO22	159	346.33	0.00	1000.00	214.94	62.06
purchased feed and own marketable feed (PLN/unit)	RO23	159	1170.55	159.67	3176.34	550.08	46.99
non-marketable feed (PLN/unit)	RO35	159	274.98	0.00	1097.81	193.64	70.42
other costs (PLN/unit)	RO40	159	272.66	32.52	951.02	158.94	58.29
Gross margin (PLN/unit)	RO45	159	3219.91	-1595.64	6418.27	1454.86	45.18
Total work inputs (manhour/unit)	RO46	159	183.66	18.87	527.37	93.16	50.72
including: own work inputs (manhour/unit)	RO47	159	176.41	18.87	527.37	95.08	53.90

Source: Own compilation based on the AGROKOSZTY data.

Table 2

Characteristics of analysed variables distribution

Name of the variable	Sign	Number of observations	Shapiro--Wilk test	Significance level of the SW test	Skewness	Kurtosis
Area of arable land (ha)	RO02	159	0.484	0.000	6.928	65.563
Area of permanent pasture (ha)	RO03	159	0.628	0.000	3.072	11.297
Utilized Agricultural Area (ha)	RO04	159	0.546	0.000	5.670	46.351
Soil valuation indicator (points)	RO07	159	0.975	0.006	0.337	-0.655
Average annual number of dairy cows (unit)	RO08	159	0.839	0.000	1.806	4.508
Milk yield of cows (l/unit)	RO09	159	0.994	0.723	0.034	-0.418
Price of milk sales (PLN/l)	RO10	157	0.972	0.003	-0.511	0.597
Price of weaned calves (PLN/kg)	RO11	159	0.992	0.569	0.144	0.038
Selling price of culled cows (PLN/kg)	RO12	116	0.913	0.000	0.515	5.882
Total value of production (PLN/unit)	RO14	159	0.991	0.437	0.204	-0.148
including: value of milk (PLN/unit)	RO15	159	0.986	0.120	0.250	-0.273
Total direct costs (PLN/unit)	RO21	159	0.974	0.004	0.588	0.485
including: herd replacement (PLN/unit)	RO22	159	0.951	0.000	0.212	0.240
purchased feed and own marketable feed (PLN/unit)	RO23	159	0.945	0.000	0.999	1.407
non-marketable feed (PLN/unit)	RO35	159	0.920	0.000	1.250	2.491
other costs (PLN/unit)	RO40	159	0.933	0.000	0.969	1.169
Gross margin (PLN/unit)	RO45	159	0.988	0.195	-0.271	-0.129
Total work inputs (manhour/unit)	RO46	159	0.950	0.000	0.791	0.451
including: own work inputs (manhour/unit)	RO47	159	0.952	0.000	0.775	0.397

Source: Own compilation based on the AGROKOSZTY data.

Table 3

Significance analysis (F-Snedecor test) for data grouped by gross margin

Name of the variable	Sign	Degrees of freedom for a factor	Mean square for a factor	Degrees of freedom	Mean square for an error	F- -Snedecor test	Level of significance
Area of arable land (ha)	RO02	2	5247	156	886	5.923	0.003
Area of permanent pasture (ha)	RO03	2	232	156	142	1.638	0.198
Utilized Agricultural Area (ha)	RO04	2	7567	156	1232	6.144	0.003
Soil valuation indicator (points)	RO07	2	0.246	156	0.116	2.126	0.123
Average annual number of dairy cows (unit)	RO08	2	5264	156	206	25.599	0.000
Milk yield of cows (l/unit)	RO09	2	105951429	156	811912	130.496	0.000
Price of milk sales (PLN/l)	RO10	2	0.570	154	0.011	50.461	0.000
Price of weaned calves (PLN/kg)	RO11	2	3.391	156	4.456	0.761	0.469
Selling price of culled cows (PLN/kg)	RO12	2	0.001	113	0.593	0.003	0.997
Total value of production (PLN/unit)	RO14	2	167219455	156	924023	180.969	0.000
including: value of milk (PLN/unit)	RO15	2	168038081	156	907982	185.068	0.000
Total direct costs (PLN/unit)	RO21	2	2190040	156	550298	3.980	0.021
including: herd replacement (PLN/unit)	RO22	2	182801	156	44448	4.113	0.018
purchased feed and own marketable feed (PLN/unit)	RO23	2	471937	156	224049	2.106	0.125
non-marketable feed (PLN/unit)	RO35	2	139878	156	36185	3.866	0.023
other costs (PLN/unit)	RO40	2	255848	156	22304	11.471	0.000
Gross margin (PLN/unit)	RO45	2	133972620	156	426161	314.371	0.000
Total work inputs (manhour/unit)	RO46	2	73138	156	7852	9.315	0.000
including: own work inputs (manhour/unit)	RO47	2	92343	156	7972	11.583	0.000

Source: Own compilation based on the AGROKOSZTY data.

Table 3 presents aggregated results of variance analysis by the Fisher--Snedecor method for groups separated according to the level of gross margin calculated per 1 cow.

The hypothesis about the absence of differences between mean results for selected data groups with probability lower than 0.001 should be rejected for all variables characterised by normal distribution. Although, for some other variables the level of significance of the F-Snedecor test is lower than 0.05, but the conclusions on significance of differences should be drawn with caution. In line with methodological assumptions, the nonparametric H Kruskal--Wallis test (table 4) was used in order to verify the significance of differences for the remaining variables.

The analysis of H Kruskal-Wallis test (table 4) shows that the classification based on quartiles of gross margin gave positive results for most analysed variables. The differences were insignificant for:

– area of pasture in an agricultural holding (RO03),
– soil valuation indicator (RO07),
– selling prices of culled cows (RO12).

The following variables found themselves at the verge of significance but above the threshold value (p=0.05):

– total direct costs (RO21),
– purchased feed and own marketable feed (RO23),
– non-marketable feed (RO35).

Hence, it may be stated that the data classification method, based on quartiles of the gross margin values, applied for dairy cows rearing under the AGROKOSZTY programme gave correct results. Taking advantage of data on production, costs and gross margin in practice, the holding size, number of cows and milk yield should also be considered. It seems that these features should be additionally taken into account when classifying data characteristic of dairy cows rearing.

The second method of presenting the research results under the AGROKOSZTY programme is grouping by location of a holding in FADN regions [4]. Table 5 presents the results of significance analysis for data classified this way.

Analysis of variables for which normal distribution was confirmed (RO09, RO11, RO14, RO15, RO45) enables us to conclude that only for the price of calves (RO11) the hypothesis on absence of differences between mean results in sub-groups may be rejected (table 5). For the other 4 variables there are no grounds for confirming the regional differences.

The significance analysis carried out by H Kruskal-Wallis method points to significant differences concerning the size of holdings (RO02, RO03, RO04), soil quality (RO07), size of herds in agricultural holdings (RO08), amount of direct costs calculated per 1 cow (RO21), as well as individual cost items: costs of purchased feed and own – potentially marketable – feed (RO23), own non-

marketable feed (RO35), as well as other direct costs (RO40) and work inputs (RO46, RO47). There are no grounds, however, for concluding that there are regional differences within the scope of milk yield of cows (RO09), milk prices (RO10) and culled cows (R012), as well as the value of gross margin (R045). Despite considerable differences in milk production conditions and organisation, as well as the applied production techniques in respective regions, the differences in the level of economic results measured with gross margin per 1 cow are hardly significant. These differences do not have any impact on the level of milk yields of cows.

Table 4

Significance analysis (H Kruskal-Wallis test) for data grouped by gross margins

Name of the variable	Sign	Degrees of freedom	Number of observations	H Kruskal--Wallis test	Level of significance
Area of arable land (ha)	RO02	2	159	20.843	0.000
Area of permanent pasture (ha)	RO03	2	159	3.859	0.145
Utilized Agricultural Area (ha)	RO04	2	159	19.640	0.000
Soil valuation indicator (points)	RO07	2	159	3.051	0.218
Average annual number of dairy cows (unit)	RO08	2	159	56.661	0.000
Milk yield of cows (l/unit)	RO09	2	159	99.860	0.000
Price of milk sales (PLN/l)	RO10	2	157	60.547	0.000
Price of weaned calves (PLN/kg)	RO11	2	159	1.177	0.555
Selling price of culled cows (PLN/kg)	RO12	2	116	0.228	0.892
Total value of production (PLN/unit)	RO14	2	159	113.377	0.000
including: value of milk (PLN/unit)	RO15	2	159	114.048	0.000
Total direct costs (PLN/unit)	RO21	2	159	5.859	0.053
including: herd replacement (PLN/unit)	RO22	2	159	10.492	0.005
purchased feed and own marketable feed (PLN/unit)	RO23	2	159	5.620	0.060
non-marketable feed (PLN/unit)	RO35	2	159	5.835	0.054
other costs (PLN/unit)	RO40	2	159	18.185	0.000
Gross margin (PLN/unit)	RO45	2	159	133.594	0.000
Total work inputs (manhour/unit)	RO46	2	159	17.297	0.000
including: own work inputs (manhour/unit)	RO47	2	159	21.797	0.000

Source: Own compilation based on the AGROKOSZTY data.

Table 5

Significance analysis (F-Snedecor test) for data grouped by regions location

Variable	Sign	Degrees of freedom for a factor	Mean square for a factor	Degrees of freedom for an error	Mean square for an error	F-Snedecor test	Level of significance
Area of arable land (ha)	RO02	3	3812	155	866	4.304	0.006
Area of permanent pastures (ha)	RO03	3	727	155	132	5.524	0.001
Utilized Agricultural Area (ha)	RO04	3	7018	155	1201	5.842	0.001
Soil valuation indicator (points)	RO07	3	0.632	155	0.107	5.890	0.001
Average annual number of dairy cows (unit)	RO08	3	1000	155	256	3.913	0.010
Milk yield of cows (l/unit)	RO09	3	1869238	155	2148086	0.870	0.458
Price of milk sales (PLN/l)	RO10	3	0.008	153	0.019	0.416	0.742
Price of weaned calves (PLN/kg)	RO11	3	28.516	155	3.977	7.171	0.000
Selling price of culled cows (PLN/kg)	RO12	3	0.421	112	0.588	0.716	0.544
Total value of production (PLN/unit)	RO14	3	1784282	155	3053120	0.584	0.626
including: value of milk (PLN/unit)	RO15	3	2997255	155	3024062	0.991	0.399
Total direct costs (PLN/unit)	RO21	3	2593154	155	531917	4.875	0.003
including: herd replacement (PLN/unit)	RO22	3	11955	155	46862	0.255	0.858
purchased feed and own marketable feed (PLN/unit)	RO23	3	1175248	155	285695	4.114	0.008
non-marketable feed (PLN/unit)	RO35	3	203289	155	34289	5.929	0.001
other costs (PLN/unit)	RO40	3	65324	155	24485	2.668	0.050
Gross margin (PLN/unit)	RO45	3	364305	155	2150538	0.169	0.917
Total work inputs (manhour/unit)	RO46	3	44194	155	7991	5.531	0.001
including: own work inputs (manhour/unit)	RO47	3	55632	155	8138	6.836	0.000

Source: Own compilation based on the AGROKOSZTY data.

Table 6

Significance analysis (H Kruskal-Wallis test) for data grouped by regions location

Name of the variable	Sign	Degrees of freedom	Number of observations	H Kruskal--Wallis test	Level of significance
Area of arable land (ha)	RO02	3	159	25.362	0.000
Area of permanent pastures (ha)	RO03	3	159	16.754	0.001
Utilized Agricultural Area (ha)	RO04	3	159	20.187	0.000
Soil valuation indicator (points)	RO07	3	159	15.152	0.002
Average annual number of dairy cows (unit)	RO08	3	159	10.950	0.012
Milk yield of cows (l/unit)	RO09	3	159	2.940	0.401
Price of milk sales (PLN/l)	RO10	3	157	2.451	0.484
Price of weaned calves (PLN/kg)	RO11	3	159	16.942	0.001
Selling price of culled cows (PLN/kg)	RO12	3	116	3.371	0.338
Total value of production (PLN/unit)	RO14	3	159	1.614	0.656
including: value of milk (PLN/unit)	RO15	3	159	2.921	0.404
Total direct costs (PLN/unit)	RO21	3	159	16.356	0.001
including: herd replacement (PLN/unit)	RO22	3	159	0.717	0.869
purchased feed and own marketable feed (PLN/unit)	RO23	3	159	28.881	0.000
non-marketable feed (PLN/unit)	RO35	3	159	16.689	0.001
other costs (PLN/unit)	RO40	3	159	9.823	0.020
Gross margin (PLN/unit)	RO45	3	159	0.305	0.959
Total work inputs (manhour/unit)	RO46	3	159	16.422	0.001
including: own work inputs (manhour/unit)	RO47	3	159	21.443	0.000

Source: Own compilation based on the AGROKOSZTY data.

Conclusions

Tabular analysis is the most popular method of processing empirical data applied in economic and agricultural research. It is a very simple method and for bulk data analysis it allows for drawing correct conclusions. However, on many occasions – because of high level of differentiation between data within separated classes – the conclusions may carry a significant error. The research team under the guidance of A. Skarżyńska decided to apply more precise methods of empirical data analysis in their analyses of the research results for 2006. On the basis of two crop production activities, apart from the traditional analysis, they introduced mathematical statistics methods. The results of the analysis pointed to the fact that it is justified to apply such methods [4]. However, livestock activities have not been analysed under the project so far. The aim of the study conducted on the basis of AGROKOSZTY data on production, costs, gross margins in dairy cows rearing was to verify the usefulness of statistical and mathematical methods in assessing the correctness of conclusions drawn on the ground of tabular analysis.

The statistical analysis shows that most variables relating to results of production, costs and gross margins, as well as milk production conditions and its organisation are characterised by distribution which deviates from normal. In such a case the application of the analysis of variance methods based on parametric tests is useful to a small degree. Nonparametric tests should be used as the basic method of verifying the significance of differences. In case of quantitative variables and if there are more than two classes, the H Kruskal-Wallis test may be useful.

Application of F-Snedecor test in respect to variables characterised by normal distribution and H Kruskal-Wallis test in respect to the other variables proved that most conclusions drawn up by A. Skarżyńska [4. p. 98-103] concerning the differentiation between individual values characteristic of dairy cows rearing against the background of classes selected in line with the level of gross margins, were correct. However, there may be some doubts about the conclusions concerning direct costs, because their differentiation proved to be insignificant (but at the verge of acceptable statistical error) in the light of the applied statistical tests. Taking into account the significance of differences within the scope of variables referring to production conditions, its scale, milk yield of cows against the background of differences in gross margins, it seems that also the impact of the production scale and milk yield of cows on the gross margins should be considered. Here, it would be worthwhile to apply some additional data classification criteria or use descriptive econometric models.

The conclusions concerning regional differences [4, p. 103-113], basically, do not raise any considerable objections, whereas in the light of statistical and mathematical analysis, the regional differences in milk yield of cows cannot be considered as significant.

In order to increase the confidence level of conclusions drawn from the AGROKOSZTY research, it seems justified to support the analysis with meth-

ods of mathematical statistics to a greater extent. It is especially important considering the small research samples of data gathered for individual production activities. The choice of detailed tests should take into account the nature of the distribution of analysed variables. Because most variables do not meet the condition of normal distribution, the nonparametric tests should be used first of all.

Literature:

1. Domański C.: Statystyczne testy nieparametryczne [Statistical nonparametric tests]. PWE, Warsaw 1979.
2. Mynarski S.: Analiza danych rynkowych i marketingowych z wykorzystaniem programu STATISTICA [Analysis of market and marketing data with the use of STATISTICA programme]. Wydawnictwo AE, Kraków 2003.
3. Mynarski S.: Praktyczne metody analizy danych rynkowych i marketingowych [Practical methods of analysing market and marketing data]. Kantor Wydawniczy Zakamycze 2000.
4. Produkcja, koszty i nadwyżka bezpośrednia wybranych produktów rolniczych w 2006 roku [Production, costs and incomes obtained from chosen agricultural products in 2006]. Multi-author work edited by A. Skarżyńska. PW no. 60, IAFE, Warsaw 2007.
5. Steczkowski J., Zeliaś A.: Analiza wariancyjna i kowariancyjna w badaniach ekonomicznych [Analysis of variance and covariance in economic studies]. PWE, Warsaw 1982.