

STATISTICAL STUDY OF SOME INDICATORS IN THE LIVESTOCK SECTOR

Ramona-Maria DIMITROV

ORCID: <https://orcid.org/0000-0002-3478-9420>

University of Craiova, Craiova, Romania

Email: ramona.dimitrov@yahoo.com

DOI: 10.52846/MNMK.22.1.01

Abstract:

In this paper, some indicators covering milk production, weight gain and feeding efficiency are studied, such as the duration of lactation, total amount of milk, daily amount of milk, percentage of fat, protein, lactose in milk, for a representative sample taken from S.C.D.A Simnic. In the first part, a descriptive statistical analysis is made of the indicators in the zootechnical field presented, followed by econometric methods regarding confidence intervals for the average of the entire population, at a threshold of 5%. Also, the correlation matrix between the indicators is analyzed and three validated linear regression models are presented with the interpretation of the results. These results help to optimize management in the zootechnical field by making decisions based on quantitative analyses.

Keywords: *livestock management, statistical methods, correlation and regression.*

1. Introduction

The use of statistical methods in agriculture and animal husbandry is of significant importance for effective resource management, informed decision making and improved yield and productivity. Agricultural statistics allow field personnel to monitor and evaluate production, identify trends, and make comparisons between different periods or batches. It helps to identify factors that can negatively or positively influence production. Statistics provide data on the demand for resources such as water, energy, animal feed, and agricultural inputs. This helps in proper resource planning to ensure efficient and sustainable management. Weather and climate statistics help farmers plan their activities according to weather conditions and take measures to reduce weather-related risks such as drought, floods or extreme temperatures. Statistics can also be used to evaluate the performance of different plant varieties or animal breeds according to characteristics such as yield, disease resistance or the quality of meat and dairy products. This helps in selecting the best genetic lines to get better yields. Moreover, it helps to efficiently manage stocks of agricultural products and plan their distribution to markets or consumers. In animal husbandry, statistics can be used to track animal health and quickly detect disease outbreaks. This allows rapid interventions to prevent the spread of disease and maintain animal health.

Animal husbandry is one of the important branches of agriculture, having as its object the growth, reproduction, breeding and exploitation of animals. Thus, raising and improving animal breeds is necessary to provide the population with animal products, the food and light industry with raw materials, and agriculture with organic fertilizers.

Statistics also play a significant role in livestock management (Cucu and Maciuc 2004, Maciuc 2006), contributing to effective decision-making to improve performance and productivity in animal husbandry. Thus, it enables the collection and analysis of animal performance data such as milk production, weight gain and feed efficiency. This data helps identify individuals or groups of animals performing above or below average and can provide clues to the genetic quality of the herd. Statistically analyzed data can contribute to the optimization of nutrition and can be used to monitor feed consumption, animal nutritional needs and feeding efficiency. This helps optimize the diet to ensure maximum growth and production at minimum cost. Statistics can also be used to track animal reproductive cycles, reproductive rate, and reproductive success. This helps in planning breeding programs and optimizing breeding yield. Last but not least, by evaluating costs and profitability, statistics allow the calculation of production costs, including costs for food, care and health. These data are essential to assess the profitability and efficiency of livestock operations.

In conclusion, statistical methods play an important role in livestock management, helping managers make effective decisions and optimize operations to achieve better yields, improved animal health, and increased profitability in animal husbandry.

In the present paper we will deal with the statistical study of some quantitative and qualitative indicators regarding a sample of 31 cattle (females), out of a total of 98, and the production of dairy products at S.C.D.A Simnic.

In the first years (1958-1962), the research-development activity had as its general objective the establishment of the productive effect and the way of using hybrid maize, green and "pickled" in the nutrition of different species of animals. Along the way (1962-1977) the area of concern was expanded to include food works, feed technology, animal maintenance and exploitation technology, breeding and animal reproduction. Studies were undertaken on sheep, pigs and bulls. It was necessary to solve some new problems (the development of intensive animal breeding systems), but also to reconsider the old conceptions of animal breeding. The results of research and collaborations with other units served the preparatory stage for the concentration and specialization of animal production in industrial breeding systems. Next, emphasis was placed on the study of the possibilities of forming new breeds of animals and the genetic consolidation of some types, families and zootechnical lines with high productive potential, but also with valuable technological attributes in sheep, pigs and bulls.

Starting with 1977, the research-development activity had as its general objective the continuous improvement of the productivity and efficiency of the growth of the Holstein Frize breed. The specific objectives were aimed at solving the complex physiological, ecological, ethological, ameliorative, reproductive, technological and sanitary-veterinary problems of the breeding of this breed imported from Denmark in 1977 and 1978. The broadening of the genetic base and the change of the genetic structure of the imported herd was done continuously through methods and specific means (selection, migration, mating management),

the beneficial effects accumulating over time and manifesting in the conditions of the milk production system at S.C.D.A. Simnic.

At the beginning, the "Şimnic" dairy cow type was formed and consolidated with good morpho-productive characteristics (body weight 650-660 Kg, productive potential 8000-9000 l udders with a fat content of 3.95-4.06%, age at first calving 25-27 months and docile temperament).

Due to the numerical increase of very valuable phenotypes, the farm is integrated into the National Bull Breeding Program as an "elite farm" with the objective of activity being the directed production of generations of bulls intended for selection, with a view to their use in the insemination of Romanian cow herds. Among the bulls "built" at Şimnic, we mention here LINCOLN ET 1, PENSTAR 3, ARLINDA JET 8.

Starting with the year 1990, the intensive use of intensively improving milk production bulls (imported) was imposed for the insemination of females. The genetic material used determined the variation of the productive and functional characteristics of the herd from S.C.D.A. Simnic. The Holstein Friesian progeny in which genes of North American origin predominated were phenotypically different from the progeny in which genes of European origin predominated. Taking into account the future diversification of cow's milk production systems, it was considered that matching the type of cow with the production system in which it is raised is of great importance for optimizing profitability. For this purpose, starting from 2004, the cattle herd from S.C.D.A. Simnic served as the basis for the formation of two genetic lines with different body development at maturity.

The first genetic line was called the Holstein Friza Şimnic line with high live weight at maturity (Holstein Friza Şimnic G.V.M.) and the second, the Holstein Friza Şimnic line with low live weight (Holstein Friza Şimnic g.v.m.). Holstein Friza line Şimnic G.V.M. it is intended for 'high input' systems where the focus is on technical efficiency and managerial decisions are based on their direct impact on profitability. Holstein Friza line Şimnic g.v.m. it is suitable for 'low input' systems where the focus is on maximizing the use of fodder from pastures and hayfields. The formation of the two genetic lines was based on the existence of interactions between genotype and environment.

Holstein Friza line Şimnic G.V.M. has a productive potential of over 9,000 kg of milk, the percentage of milk protein of 3.3 - 3.4%, and of fat of 3.6 - 3.8%. Holstein Friza line Şimnic g.v.m. it has a productive potential of 7,500-8,000 kg of milk, the milk protein percentage is 3.2-3.3% and the fat percentage is over 4.5%. The HF Şimnic G.V.M line has a higher stature, a higher live weight, a calving interval greater than 360 days, mobilizes a lot of energy from body reserves in the first part of lactation, a longer postpartum period, with a negative energy balance and uses less energy to restore body reserves during lactation, compared to the Holstein Friza Line Şimnic g.v.m.

Starting from 2015, research began on improving the robustness of the animals and the quality of the milk produced, research carried out within a project called "Improving a line of cows from the Holstein Friza breed towards the simultaneous improvement of robustness and milk quality"

In this paper, some indicators covering milk production, weight gain and feeding efficiency are studied, such as the duration of lactation, total amount of milk, daily amount of milk, percentage of fat, protein, lactose in milk, but also the weight of cattle. We also have four nutrition indicators given by the vital functions

nutritional units, vital functions digestible protein (proportional to weight), but also lactation nutritional units and lactation digestible protein (proportional to the daily amount of milk).

The work is structured as follows. In the second section, a presentation of the specialized literature in the field is made, highlighting the research undertaken previously. The third section contains the research methodology, mentioning the main statistical indicators used in the analysis, as well as the statistical tests used to validate the econometric models. The fourth section contains the novelty of the work. In the first part, a descriptive statistical analysis of the previously presented zootechnical indicators is performed, followed by econometric methods regarding confidence intervals for the average of the entire population, at a threshold of 5%, as well as correlation and regression. The work ends with conclusions. Analyzed data were obtained from S.C.D.A. Siminic (Annex 1).

2. Literature review

Anghel, Anghelache and Panait (2017) studied the results that were recorded by the member countries of the European Union in the field of agriculture from the period before 2017, highlighting those of Romania and analyzed, among others, certain aspects regarding the evolution agricultural production and animal production. Anghelache and Dumitrescu (2015) carried out research, which largely focused on the analysis of agricultural production by types of crops, animals, reproduction, but also the size of agricultural holdings. Anghelache et al. (2014) presented the evolution of agricultural production in Romania, analyzing the two main components of agriculture, namely, plant production and animal husbandry.

In their work, Herrero et al. (2013) studied the main systems of raising and exploiting animals: intensive closed, mixed (plant culture and animal breeding) and open (pasture), which totals a herd of over 17 billion animals. of animals contribute by providing meat, milk, eggs to ensure the nutritional value of food providing approximately 13% of the energy and 28% of the protein consumed worldwide. Moreover, in developed countries, these percentages increase to 20% for energy and 48 % for protein (FAO, 2009). Farkas et al. (2023) conducted research using statistical data from the period 1990-2022, a literature review and 66 interviews conducted in the southern region of Hungary. The results showed a decrease in the number of livestock and small farms in rural settlements, correlated with an intense concentration of farms, which negatively affects income generation opportunities.

Herath, Weersink and Carpentier (2005), study, among others, the factors that affect the state's annual share of the national inventory for each of the cattle and milk sectors, using data from the 48 states from 1976 to 2000. The results indicate that differences in the severity of environmental regulations faced by livestock producers had a significant influence on production decisions in the dairy sector.

Robinson and others (2014) show that animal husbandry contributes directly to the livelihoods and food security of almost a billion people. This paper describes the current approach in detail and presents new global distribution maps at 1 km resolution for cattle, pigs and chickens. These digital layers are made available to the public via the Livestock Geo-Wiki (<http://www.livestock-geo-wiki.org>), as are maps of other livestock types as they are produced.

Sandu (2015) evaluates the Romanian livestock sector, studying the evolution of the main indicators of animal production, through the lens of the availability of agri-food products of animal origin. The analysis of the main indicators of livestock production is based on statistical data from the National Institute of Statistics. It was found that animal husbandry presents serious structural problems due to excessive property fragmentation, low productivity and high self-consumption in farms. The level of livestock production is low, caused by the sharp and constant reduction of livestock numbers, poor performance and lack of competitiveness. Grosu and others (2003), Korkmaz and others (2012) use regression in the statistical study of indicators in the livestock sector.

3. Research methodology

The statistical analysis of some quantitative and qualitative indicators in the zootechnical field was carried out by calculating the following numerical indicators: selection mean, selection dispersion, standard deviation and coefficient of variation. To check whether or not the data series follow a normal distribution, the Kolmogorov-Smirnov, Shapiro-Wilk and Jarque-Bera tests are used (for more details see Georgescu, 2005, Maddala, 2001, Wooldridge, 1999). The Kolmogorov-Smirnov and Jarque-Bera tests are mainly used to test large data sets while the Shapiro-Wilk test is more meaningful for a smaller sample, such as 50 observations or less. If the probability attached to the test, for any of the tests, is above 0.05, the data are normally distributed, at a significance threshold of 5%.

For a series Y_t , $t = 1, \dots, n$, the Shapiro-Wilk test (Shapiro & Wilk, 1965) regarding the normality of the distribution is defined as follows:

$$W = \frac{\sum_{t=1}^n (p_t y_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2},$$

where y_t represent the values of the initial series Y_t , which are ordered in ascending order ($y_1 \leq y_2 \leq \dots \leq y_n$), $\bar{y}$ is the mean of the respective series and p_t are weights. The null hypothesis H_0 of normality of the distribution is rejected if $W \leq W_\alpha$, where the critical thresholds W_α are tabulated by Shapiro & Wilk (1965), at the chosen significance threshold. The test also generates an attached probability (SPSS software) and the hypothesis of normality is accepted if the probability is greater than 0.05 at the chosen significance threshold $\alpha = 5\%$.

The Jarque-Bera test is a statistical test for verifying the assumption of normality of a distribution. It is based on measuring the skewness S and the kurtosis K of a distribution. The tested variable is

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right)$$

which follows a χ -square distribution law, where n is the number of observations or degrees of freedom. The null hypothesis of the Jarque-Bera test is H_0 : the data are normally distributed and is evaluated against the alternative hypothesis H_1 : the data are not normally distributed and follows another distribution from the family of Pearson distributions. Using the t-Student test and the standard error, confidence intervals were calculated for the mean at a chosen significance level of 5%, for normal distributions, given by

$$\left[\bar{x}_s - t \cdot \frac{\sigma_s}{\sqrt{n}}, \bar{x}_s + t \cdot \frac{\sigma_s}{\sqrt{n}} \right].$$

In order to evaluate the parameters in regression models, we used the method of least squares, where the coefficient of determination R^2 shows us the percentage by which the influence of the independent variables on the dependent variable is explained. Student's t-test was used to test the null hypothesis H_0 (coefficients are not significantly different from 0) and the alternative hypothesis H_1 (coefficients are significantly different from 0). The F-test verifies the null hypothesis H_0 (all coefficients are not significantly different from 0) and the alternative hypothesis H_1 (there is at least one coefficient different from 0).

Furthermore, we used the Durbin-Watson test to check first-order autocorrelation or independence of model errors, the Breusch-Godfrey test to check higher-order autocorrelation of errors, the Kolmogorov-Smirnov, Shapiro-Wilk and Jarque-Bera tests to see if the model errors follow a normal distribution and White's test to check the homoscedasticity or heteroscedasticity of the regression model (see Georgescu, 2005, Maddala, 2001, Wooldridge, 1999).

4. The main results

4.1. Statistical analysis

In this section we will statistically analyze the qualitative and quantitative indicators that characterize the group of cattle (females) chosen for the study and the quality of the milk collected from them. The numerical indicators of the tendency of the distribution series were calculated, such as the selection mean, median, minimum and maximum value, but also variation indicators such as the selection dispersion, the standard deviation and the coefficient of variation (Table 1). The Kolmogorov-Smirnov, Shapiro-Wilk (Table 3) and Jarque-Bera (Table 2) tests were also used to verify whether or not the data series follow a normal distribution. Finally, using the t-Student test and the standard error, the confidence interval for the mean was calculated in the case of series with normal distributions at a significance threshold of 5% (Table 4).

Table 1. Descriptive statistics of zootechnical indicators

	Mean of selection	Standard error	Median	Standard deviation	Coef. of variation	Dispersion of selection	Minimum value	Maximum value
Duration of lactation (days)	298,493	2,117	305	11,787	3,94%	138,951	254	305
Total amount of milk (litres)	9652,771	210.680	9898	1173,022	12,15%	1375980,66	6437,33	11248,66
Daily quantity of milk (litres)	32,355	0,700	32,872	3,898	12,04%	15,196	21,846	37,655
Weight (kg)	604,838	5.,72	600	33,253	5,49%	1105,806	560	680
Nutritional units vital functions (kg)	3,024	0,0298	3	0,166	5,49%	0,027	2,8	3,4
Lactation nutritional units (kg)	16,177	0,350	16,436	1,949	12,04%	3,799	10,923	18,827
Digestible protein vital functions (gr)	423,387	4,180	420	23,277	5,49%	541,845	392	476
Digestible milk protein (gr)	2103,078	45,509	2136,6	253,386	12,04 %	64204,572	1420	2447,61

Fat percentage	5,066	0,057	5,036	0,320	6,31%	0,103	4,24	5,68
Protein percentage	3,319	0,016	3,320	0,094	2,83%	0,008	3,104	3,525
Lactose percentage	4,714	0,015	4,710	0,085	1,80%	0,007	4,520	4,862

Source: Developed by author

Table 2. Jarque-Bera Test test for normality for zootechnical indicators

	Value Jarque-Bera test	Probability	Asymmetry	Kurtosis
Duration of lactation	59.10458	0.000000	-2.259477	8.033630
Total amount of milk	3.598778	0.165400	-0.829908	3.176517
Daily amount of milk	1.905396	0.385699	-0.604684	2.887860
Weight	4.281751	0.117552	0.910343	2.997998
Nutrient units vital functions	4.281751	0.117552	0.910343	2.997998
Lactation nutritional units	1.905396	0.385699	-0.604684	2.887860
Digestible protein vital functions	4.281751	0.117552	0.910343	2.997998
Digestible milk protein	1.905396	0.385699	-0.604684	2.887860
Fat percentage	0.128277	0.937875	-0.154128	2.934504
Protein percentage	0.015024	0.992516	0.003634	2.892397
Lactose percentage	0.829881	0.660379	-0.221434	2.331901

Source: Developed by author with Eviews software

Table 3. Kolmogorov-Smirnov and Shapiro-Wilk tests for zootechnical indicators

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Duration of lactation	,292	31	,000	,631	31	,000
Total amount of milk	,128	31	,200	,937	31	,066
Daily amount of milk	,120	31	,200	,949	31	,145
Weight	,203	31	,002	,900	31	,007
Nutrient units vital functions	,203	31	,002	,900	31	,007
Lactation nutritional units	,120	31	,200	,949	31	,144
Digestible protein vital functions	,203	31	,002	,900	31	,007
Digestible milk protein	,120	31	,200	,949	31	,144
Fat percentage	,083	31	,200	,979	31	,796
Protein percentage	,068	31	,200	,993	31	,999
Lactose percentage	,096	31	,200	,968	31	,458

Source: Developed by author with SPSS software

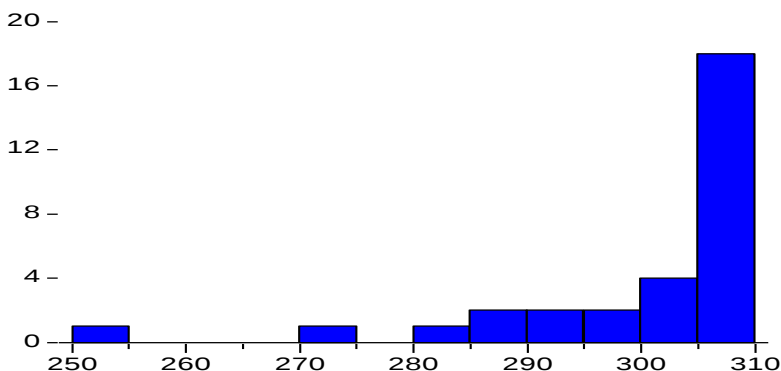
Table 4. The confidence interval for the mean of zootechnical indicators

	Mean	$t \cdot \frac{\sigma_s}{\sqrt{n}}$	confidence interval
Total amount of milk	9652,771	430,267	[9222.503; 10083.038]
Daily amount of milk	32,355	1,429	[30,926; 33,784]
Lactation nutritional units	16,177	0,714	[15,463; 16,891]
Digestible milk protein	2103.078	92.942	[2010,136; 2196,02]
Fat percentage	5.066	0.117	[4,949; 5,183]
Protein percentage	3.319	0,034	[3.285; 3.353]
Lactose percentage	4.714	0.031	[4,683; 4,745]

Source: Developed by author

From the analysis of the results for the normality of the distributions, it is found that the duration of lactation does not follow a normal distribution, according to the Kolmogorov-Smirnov (prob.=0.00), Shapiro-Wilk (prob.=0.00) and Jarque-Bera (prob.=0.00). We also consider that the weight does not follow a normal distribution according to the Kolmogorov-Smirnov (prob.=0.002), Shapiro-Wilk (prob.=0.007) tests, although the Jarque-Bera test has an attached probability of 0.11. Due to the number of 31 observations ($n < 50$), the Shapiro-Wilk test is considered the most representative. It is found that the nutritional units vital functions and the digestible protein vital functions, which are proportional to the weight, do not follow a normal distribution. For the other indicators, the probabilities attached to the three normality tests are greater than 0.05, that is, they tend towards normal distributions. Next, we will do an analysis for each indicator separately

The duration of lactation for the chosen sample has a mean of 298.4 days, with a standard deviation of 11.7 days. The coefficient of variation has the value of 3.94%, which shows that the series is homogeneous and the mean representative. The values are in the range of 254-305 days, most of them being at the upper edge and the series does not have a normal distribution (stopping lactation in deliberately, for a new insemination).

**Figure 1. Lactation duration histogram**

The total amount of milk collected during the lactation period is a normally distributed variable, according to the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.066) and Jarque-Bera tests with the value 3.598 and the attached probability 0.165 (asymmetry =-0.829, kurtosis=3.176). The sample mean is 9652.7 liters with a standard deviation of 1173.02 liters. The coefficient of variation has the value of 12.15% indicating a homogeneous data series, the recorded values are located around the mean, which is significant. The confidence interval for the mean, at a significance threshold of 5%, is [9,222.5; 10,083.03].

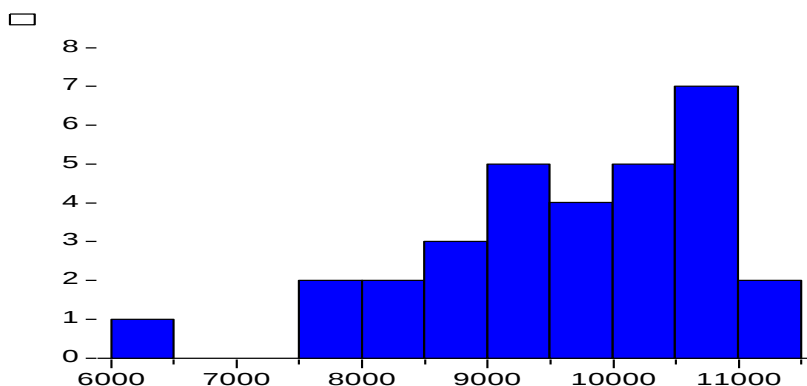


Figure 2. Total amount of milk histogram

The average daily amount of milk is obtained by dividing the total amount by the duration of lactation and the mean obtained for this sample is 32.355 liters with a standard deviation of 3.898 liters. The data are normally distributed, according to the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.145) and Jarque-Bera tests with the value 1.905 and the attached probability 0.385 (asymmetry=-0.604, kurtosis=2.887). The coefficient of variation has the value of 12.04%, indicating a homogeneous data series, with a significant mean. The confidence interval for the mean, at a significance threshold of 5%, is [30.926; 33.784].

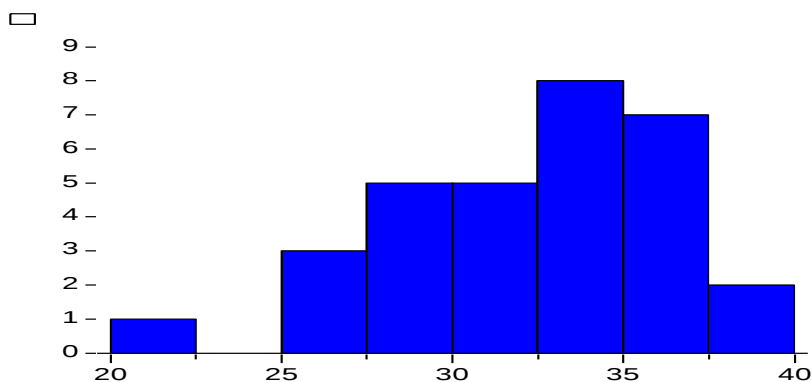


Figure 3. Daily quantity of milk histogram

The weight of the selected cattle is another indicator taken into account and the mean of the sample is found to be 604.838 kg, with a standard deviation of 33.253 kg. The sample data are not normally distributed according to the Kolmogorov-Smirnov (prob.=0.002), Shapiro-Wilk (prob.=0.007) tests. However, the coefficient of variation is 5.49%, meaning a small variation in the data, which is around the mean, which is representative.

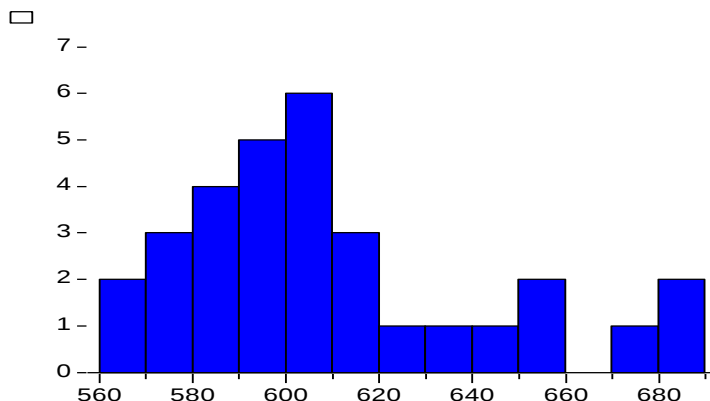


Figure 4. Weight histogram

The vital functions nutritional unit indicator is calculated by multiplying the weight by 0.05, having an identical distribution to the weight indicator. The lactation nutritional unit indicator is obtained from the daily amount of milk multiplied by 0.5, so that the two indicators have the same distributions. The vital functions digestible protein indicator (grams) is calculated by multiplying the weight by 0.7 and consequently they have identical distributions. Also, the indicator of digestible milk protein (grams) is obtained by multiplying the daily amount by 65 and they have the same distributions

Analyzing the milk fat percentage, it is found that this indicator has a mean of 5.066 and a standard deviation of 0.32 for the chosen sample, the data following a normal distribution, according to the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk tests (prob.=0.796) and Jarque-Bera of value 0.128 with the probability of 0.937 (asymmetry=-0.154; kurtosis =2.934). The coefficient of variation has a value of 6.31%, which indicates a homogeneous series, with values located around the mean, which is representative. The confidence interval for the mean is [4.949; 5.183] for a significance threshold of 5%.

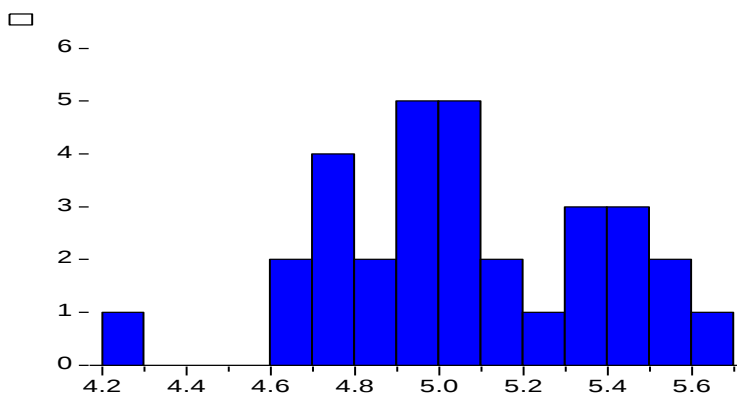


Figure 5. Milk fat percentage histogram

The *milk protein percentage* is another studied indicator, having a mean of 3.319 for our sample, with a standard deviation of 0.094. The coefficient of variation has a value of 2.83%, indicating a homogeneous series with a representative mean. The data are normal distributed according to the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.999) and Jarque-Bera tests of value 0.015 with the probability of 0.992 (asymmetry=0.003; kurtosis=2.892). The confidence interval for the mean at the 5% significance level is [3.285; 3.353].

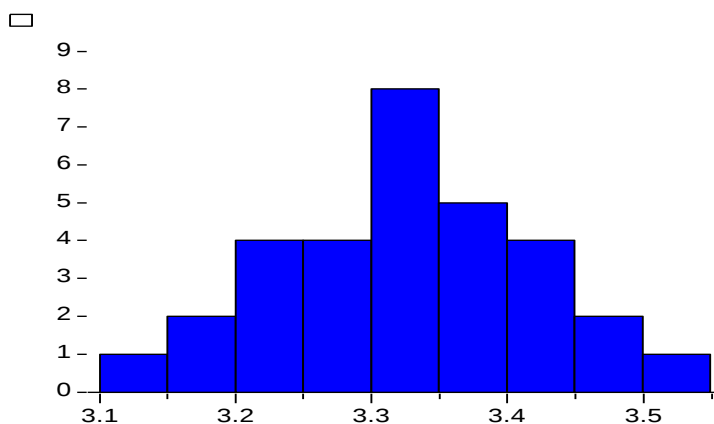


Figure 6. Milk protein percentage histogram

The *lactose percentage* in milk has a mean of 4.714 for this sample, with a standard deviation of 0.085. The coefficient of variation has a value of 1.8%, which means that the series is homogeneous, the values are around the mean, which is representative. The data series follows a normal distribution, a result confirmed by the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.458) and Jarque-Bera tests with a value of 0.829 and an attached probability of 0.66, (asymmetry=-0.221; kurtosis=2.331). The confidence interval in which the mean for the entire population is located with a probability of 95% is [4.683; 4.745].

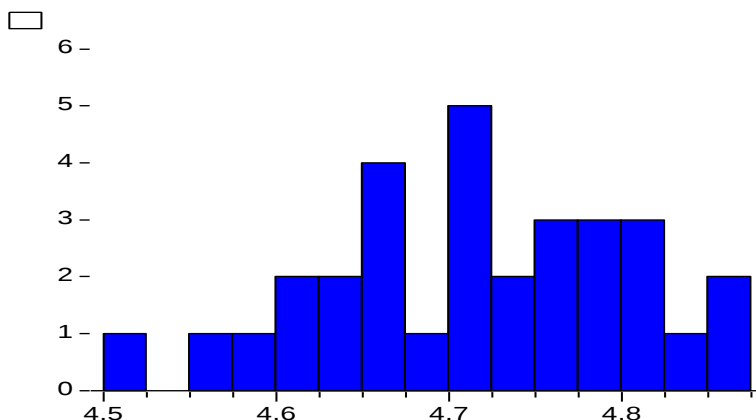


Figura 7. Milk lactose percentage histogram

4.2 The correlation matrix

From the analysis of the correlation matrix between the indicators (Table 5) it can be seen that we have small relative values of the correlation coefficients. The highest value, which is singular, is 0.94, but this is due to the way of calculating the daily amount of milk, which is obtained by dividing the total amount by the duration of lactation.

Thus, the total amount of milk collected is positively correlated with the duration of lactation, with a coefficient of only 0.228, meaning that a longer duration of lactation leads to a higher total amount of milk. Also, the total amount is negatively correlated with the percentages of fat (-0.247), protein (-0.159) and lactose (-0.074). This fact can lead to the conclusion that an increase in the total amount of milk collected generates a decrease in the percentages of fat, protein and lactose in this product.

The daily amount of milk is negatively correlated with the duration of lactation with a small coefficient of -0.112, which means that an increase in the duration of lactation can lead to a lower average daily amount. Also, the daily amount is negatively correlated with the percentages of fat (-0.359), protein (-0.175), lactose (-0.165) in milk, so an increase in the daily amount causes a decrease in these indicators.

The weight is positively correlated with the total quantity (0.146) and the daily quantity (0.221) and negatively correlated with the percentages of fat (-0.178), protein (-0.250), lactose (-0.210) in milk, but with small values. This fact suggests that an increase in weight can lead to a greater amount of milk collected, but also to a decrease in the percentages of fat, protein and lactose.

The duration of lactation is positively correlated with the percentages of fat (0.267), protein (0.048), lactose (0.265), which may suggest that a longer duration of lactation, which also means a lower daily amount, leads to an increase in these indicators of the milk. The percentages of fat, protein, lactose have positive correlation coefficients between them (0.384), (0.223), respectively (0.552), which suggests that an increase in one of these indicators is correlated with the increase in the other two.

It is important to point out that correlation coefficients only indicate a linear relationship between variables, and not direct causality. Other factors and unobserved variables can influence these indicators, such as the genetic potential of the animal, feeding diet, age, number of calvings and calving season.

Table 5. Correlation matrix between indicators

	Duration of lactation	Total quantity	Daily amount	Weight	Fat %	Protein %	Lactose %
Duration of lactation	1						
Total quantity	0.228	1					
Daily amount	-0.112	0.940	1				
Weight	-0.214	0.146	0.221	1			
Fat percentage	0.267	-0.247	-0.359	-0.178	1		
Protein percentage	0.048	-0.159	-0.175	-0.250	0.384	1	
Lactose percentage	0.265	-0.074	-0.165	-0.210	0.223	0.552	1

Source: Made by the author

4.3 Linear regression models

In what follows, we will present some simple linear regression models. It should be mentioned that several regression models were checked, but only those that passed all the validation tests for simple or multiple linear regression were kept. The first regression model contains the milk fat percentage as the dependent variable and the daily amount of milk as the independent variable.

Dependent Variable: Fat percentage

Method: Least Squares

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.020513	0.465549	12.93207	0.0000
Daily amount	-0.029497	0.014289	-2.064367	0.0480
R-squared	0.128124	Mean dependent var		5.066129
Adjusted R-squared	0.098059	S.D. dependent var		0.321244
S.E. of regression	0.305087	Akaike info criterion		0.525904
Sum squared resid	2.699272	Schwarz criterion		0.618420
Log likelihood	-6.151515	F-statistic		4.261609
Durbin-Watson stat	2.334137	Prob(F-statistic)		0.048029
Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	0.530048	Probability		0.594576
Obs*R-squared	1.171163	Probability		0.556782
White Heteroskedasticity Test:				
F-statistic	0.142734	Probability		0.867611
Obs*R-squared	0.312865	Probability		0.855189

Source: Developed by author with Eviews software

The regression equation is:

$$\text{Fat_percentage} = 6,02 - 0,029 \times \text{Daily_amount} + \alpha \quad (1)$$

For the independent variable the coefficient is -0.029. This means that with an increase of one unit in the daily amount of milk, we can have a decrease of 0.029 units in the percentage of fat. The probabilities associated with the coefficients indicate their level of statistical significance. In this case, the coefficients are significantly different from zero at the 5% confidence level, with associated probabilities less than 0.05. R-squared measures the proportion of the variation in the dependent variable that is explained by the independent variable in the model. In this case, the R-squared is 0.1281, which means that only about 12.81% of the variation in fat percentage can be explained by the variation in the daily amount of milk obtained. The rest of 87.19% of the variation in fat percentage is due to other factors, which can be listed as follows: the animal's genetic potential, feeding diet, age, number of litters and season. Each of them affects the fat content by decreasing or increasing it.

The F statistic is used to assess the overall significance of the regression model. In this case, the F-statistic value is 4.261 and the associated probability is 0.048. The associated probability is the chance of obtaining an F-statistic as extreme as that observed if the regression coefficients were indeed zero. As the probability is lower than the chosen significance level of 0.05, then we can reject the null hypothesis that all coefficients are zero and accept the alternative hypothesis that at least one coefficient is significantly different from zero. In this case we can consider that the model has a global statistical significance.

The Durbin-Watson statistic is used to test for the presence of first-order autocorrelation (correlation between adjacent residuals) in the residuals of the regression model. In this case, the value of the Durbin-Watson statistic is 2.311. The reference value for this statistic is around 2, the range in which we have no autocorrelation is [1.50; 2.50] suggesting no autocorrelation.

The Breusch-Godfrey serial autocorrelation test (or Breusch-Godfrey serial autocorrelation LM test) has two associated statistics as well as the corresponding probability values. The F statistic is used to test whether there is higher-order serial autocorrelation in the residuals of the regression model. In this case, the value of the F statistic is 0.530. A smaller value of it indicates that there is insufficient evidence to support the existence of serial autocorrelation in the residuals. The probability associated with the F statistic is 0.594. This probability represents the chance of obtaining an F statistic as extreme as that observed if there was no serial autocorrelation in the model residuals. In this case, the probability is high, above 0.05 which suggests that there is insufficient evidence to reject the null hypothesis that there is no serial autocorrelation. Obs*R-squared is another statistic used in the Breusch-Godfrey test for serial autocorrelation. This statistic is calculated by multiplying the number of observations (Obs) by the R-squared from the regression model. In this case, the Obs*R-squared value is 1.171 and the associated probability is 0.556. This probability represents the chance of obtaining an Obs*R-squared statistic as extreme as that observed when there is no serial autocorrelation in the model residuals. Similar to the F-statistic, a probability greater than 0.05 suggests that there is insufficient evidence to reject the null hypothesis that there is no serial autocorrelation.

In conclusion, the results of the Breusch-Godfrey test for serial autocorrelation indicate that there is insufficient evidence to support the hypothesis that there is serial autocorrelation in the residuals of the regression model. The probability values associated with the F-statistic and Obs*R-squared are high, suggesting that the residuals are not serially correlated with each other.

The White heteroscedasticity test has two associated statistics as well as the corresponding probability values. The F statistic is used to test whether there is a relationship between the variability of the residuals and the independent variables in the regression model. In this case, the value of the F-statistic is 0.142, which indicates a relatively small value of the statistic. The lower the value of the F statistic, the less likely heteroscedasticity exists in the data. The probability associated with the F-statistic is 0.867 and this probability represents the chance of obtaining an F-statistic as extreme as that observed if there is no heteroscedasticity in the data. In this case, the probability is high, above 0.05, suggesting that there is insufficient evidence to reject the null hypothesis that there is no heteroscedasticity in the data. Obs*R-squared is another statistic used in White's test for heteroscedasticity. This statistic is calculated by multiplying the number of observations (Obs) by the R-squared (coefficient of determination) from the regression model. In this case, the Obs*R-squared value is 0.312, with the associated probability of 0.855. This probability represents the chance of obtaining an Obs*R-squared statistic as extreme as that observed when there is no heteroscedasticity in the data. Similar to the F statistic, a high probability suggests that there is insufficient evidence to reject the null hypothesis that there is no heteroscedasticity in the data. In conclusion, the results of White's test for heteroscedasticity indicate that there is insufficient evidence to support the hypothesis that there is heteroscedasticity in the data. Thus, the model is homoscedastic, with constant variation of residuals.

The residuals follow a normal distribution according to the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.757) and Jarque-Bera tests with a value of 0.935 and the associated probability of 0.626 (Figure 7).

□

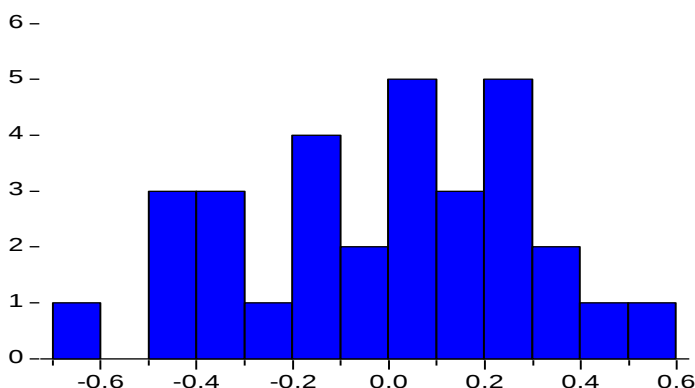


Figure 7. Histogram of residuals in regression model 1

Table 6. Tests of normality of residuals in regression model 1

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	0,122	31	0,200	0,978	31	0,757

Source: Made by the author with SPSS software

The second regression model contains the milk fat percentage as the dependent variable and the protein percentage as the independent variable. The regression equation is:

$$\text{Fat_percentage} = 0,769 + 1,294 \times \text{Protein_percentage} + a \quad (2)$$

The value of the t-Student test for the coefficient of the independent variable is 2.23 with the attached probability of 0.033, which indicates that the independent variable has a coefficient significantly different from zero. The F-statistic test has a value of 5.01 with the attached probability 0.033 which leads to the conclusion that the regression model is valid. The value of the Durbin-Watson test is 2.27, which is in the range [1.5; 2.5] in which the residuals are independent, that is, we have no first-order autocorrelation. Also, the model passes the test of lack of higher-order autocorrelation, a fact confirmed by the Breusch-Godfrey test whose associated probability is 0.75. The lack of heteroscedasticity is found by applying the White test with an associated probability of 0.58, the model being homoscedastic, that is, the variation of the residuals is constant. Also, the residuals follow a normal distribution, a fact confirmed by the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.616) and Jarque-Bera tests of value 0.953 and the attached probability 0.62. (Figure 8).

Dependent Variable: Fat Percentage

Method: Least Squares

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.769110	1.920393	0.400496	0.6917
Protein Percentage	1.294031	0.578089	2.238463	0.0330
R-squared	0.147328	Mean dependent var		5.066129
Adjusted R-squared	0.117925	S.D. dependent var		0.321244
S.E. of regression	0.301709	Akaike info criterion		0.503632
Sum squared resid	2.639819	Schwarz criterion		0.596148
Log likelihood	-5.806301	F-statistic		5.010718
Durbin-Watson stat	2.276494	Prob(F-statistic)		0.033033
Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	0.283596	Probability		0.755286
Obs*R-squared	0.637822	Probability		0.726940
White Heteroskedasticity Test:				
F-statistic	0.553057	Probability		0.581344
Obs*R-squared	1.178088	Probability		0.554858

Source: Developed by author with Eviews software

Thus, from the regression equation (2) it results that for an increase of one unit in the milk protein percentage, we have an increase of 1.29 units in the fat percentage. However, the coefficient of determination R-square has the value of 0.1473, so that only a percentage of 14.73% of the variation of the milk fat percentage is explained by the variation of the protein percentage.

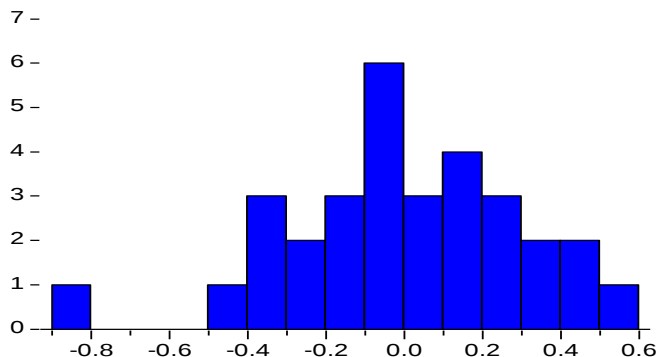


Figure 8. Histogram of residuals in regression model 2

Table 7. Tests of normality of residuals in regression model 2

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	0,100	31	0,200	0,973	31	0,616

Source: Made by the author with SPSS software

The third regression model contains the percentage of milk protein as the dependent variable and the percentage of lactose as the independent variable.

Dependent Variable: Protein Percentage

Method: Least Squares

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.552435	0.822971	0.671269	0.5074
Lactose Percentage	0.587208	0.174545	3.364219	0.0022
R-squared	0.280718	Mean dependent var		3.320645
Adjusted R-squared	0.255915	S.D. dependent var		0.095287
S.E. of regression	0.082195	Akaike info criterion		-2.097110
Sum squared resid	0.195923	Schwarz criterion		-2.004595
Log likelihood	34.50521	F-statistic		11.31797
Durbin-Watson stat	1.960422	Prob(F-statistic)		0.002174
Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	0.345946	Probability		0.710641
Obs*R-squared	0.774545	Probability		0.678906
White Heteroskedasticity Test:				
F-statistic	1.162411	Probability		0.327368
Obs*R-squared	2.376584	Probability		0.304741

Source: Developed by author with Eviews software

The regression equation is:

$$\text{Protein_percentage} = 0,552 + 0,587 \times \text{Lactose_percentage} + a \quad (3)$$

The probability attached to the t-Student test has a value of 0.0022 and is lower than the chosen significance threshold of 0.05, which indicates that the independent variable has a coefficient significantly different from zero. The F-statistic test has the value of 11.31 with the attached probability 0.0021 and indicates that the regression model is valid. The Durbin-Watson test has a value of 1.96, which is within the range [1.5; 2.5] in which the residuals are independent, which means that we have no first-order autocorrelation. Also, the linear regression model passes the test of the absence of higher-order autocorrelation, a fact that is confirmed by applying the Breusch-Godfrey test whose associated probability is 0.71. The lack of heteroscedasticity is found by applying the White test with an associated probability of 0.32, the model being homoscedastic, which means constant residual variation. The residuals follow a normal distribution, this hypothesis being confirmed by the Kolmogorov-Smirnov (prob.=0.2), Shapiro-Wilk (prob.=0.775) and Jarque-Bera tests with a value of 1.784 and the attached probability 0.409 (Figure 9)

From the regression equation (3) it is found that for an increase of one unit in the percentage of lactose in milk, we have an increase of 0.58 units in the percentage of protein. Finally, the coefficient of determination R-square has the value of 0.28, so that only a percentage of 28% of the variation of the milk protein percentage is explained by the variation of the lactose percentage.

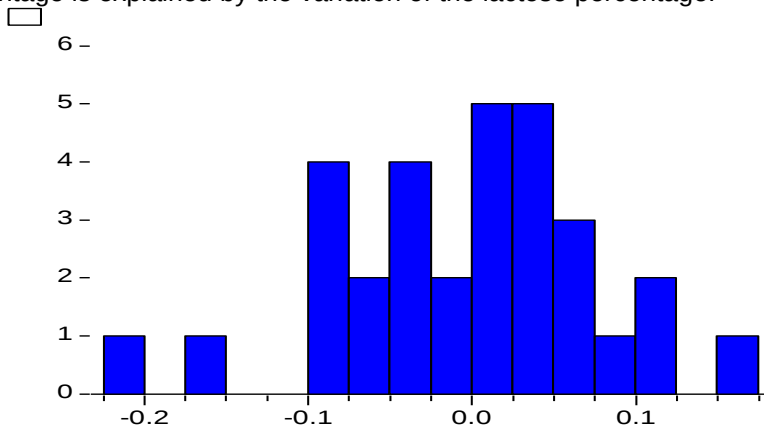


Figure 9. Histogram of residuals in regression model 3

Table 8. Tests of normality of residuals in regression model 3

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	0,102	31	0,200	0,979	31	0,775

Source: Made by the author with SPSS software

5. Conclusions

It was found that indicators such as total amount and daily amount of milk, nutritional units of lactation, digestible protein of lactation, as well as the percentages of fat, protein and lactose follow a normal distribution. For these indicators, the confidence interval for the mean was calculated at a significance threshold of 5% (Table 4). It was found that the confidence intervals for the daily amount of milk and the percentages of fat, protein, lactose are narrow, indicating a stable and uniform production of milk. The confidence interval for the mean of livestock indicators in milk thus provides crucial information for decision-making in the dairy industry, helping to manage quality, optimize processes and ensure compliance with quality standards. Also, from the correlation matrix (Table 5) it can be seen that the value of the correlation coefficients is relatively small, they are included in the range [-0.359; 0.552]. From the regression equation (1) it resulted that with an increase of one unit in the daily amount of milk, we can have a decrease of 0.029 units in the percentage of fat. From the regression equation (2) it follows that for an increase of one unit in the milk protein percentage, we can have an increase of 1.29 units in the fat percentage. From the regression equation (3) it was found that for an increase of one unit in the percentage of lactose in milk, we can have an increase of 0.58 units in the percentage of protein. However, from the validated linear regression models it is found that the values of the determination coefficients are relatively small with values between [12.81%, 28%] which leads to the conclusion that there are other causes that can explain the evolution of the dependent variables as fat percentage and protein percentage studied. Other quantitative and qualitative indicators that can explain the evolution of the dependent variables could be the genetic potential of the animal, the feeding diet, the age, the number of calvings and the calving season.

REFERENCES

- Anghel, M.G., Anghelache, C. Panait, M., 2017. Evolution of agricultural activity in the European Union. *Romanian Statistical Review, Supplement*, 6, 63-74.
- Anghelache, C., Dumitrescu, D., 2015. The Production Indices in Agriculture. *Romanian Statistical Review, Supplement*, 1, 67-71.
- Anghelache C. et al., 2014. Analysis of the Evolution of the Agricultural Production in Romania. *Romanian Statistical Review, Supplement*, 4, 78-86.
- Cucu, I., Maciuc, V., 2004. Elemente de biostatistică și folosirea lor în cercetarea științifică zootehnică. *Rev. Cercetări Agronomice în Moldova*, Iași.
- Grosu, H., Lungu, S., Soysal Ihsan, M., 2003. *Estimation of the genetic parameters and variance components in the production records of farm animals*. Trakya University, Tubitak, Tekirdag, Turcia.
- Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S., Rufino, M.C., 2013. The roles of livestock in developing countries. *Animal*, Volume 7, Supplement 1, 3-18. <https://doi.org/10.1017/S1751731112001954>.
- Georgescu, V., 2005. *Statistica descriptiva si inferentiala*, Ed. Universitaria.

- FAO, 2009. *The state of food and agriculture: Livestock in the balance*. FAO, Rome.
- Farkas, J.Z., Kőszegi, I.R., Hoyk, E. Szalai, Á., 2023. Challenges and Future Visions of the Hungarian Livestock Sector from a Rural Development Viewpoint. *Agriculture*, 13, 1206. <https://doi.org/10.3390/agriculture13061206>
- Herath, D., Weersink, A., Carpentier, C. L., 2005. Spatial Dynamics of the Livestock Sector in the United States: Do Environmental Regulations Matter? *Journal of Agricultural and Resource Economics*, vol. 30, no. 1 pp. 45-68.
- Korkmaz, M., Güney, S., Yiğiter, Ş., 2012. The importance of logistic regression implementations in the turkish livestock sector. *Harran Journal of Agricultural and Food Science*, 16 (2), 25-36.
- Maciuc, V., 2006. *Managementul creşterii bovinelor*. Editura Alfa, Iaşi.
- Maddala, G. S., 2001. *Introduction to Econometrics*. John Wiley & Sons.
- Robinson, T.P., Wint G.R.W., Conchedda, G., Van Boeckel, T.P. Ercoli, V., Palamara, E. et al., 2014. Mapping the Global Distribution of Livestock. *PLoS ONE* 9(5): e96084. <https://doi.org/10.1371/journal.pone.0096084>
- Sandu, M., 2015. *Assessment of Romanian Livestock Sector Available from the Perspective of Animal Food Products*. In: Impactul transformărilor socio-economice şi tehnologice la nivel national, european si mondial; Nr. 7, Vol.7, Available at SSRN: <https://ssrn.com/abstract=2672839>
- Shapiro, S., Wilk, M., 1965. An analysis of variance test for normality (complete samples). *Biometrika*. 52 (3–4): 591–611. [doi:10.1093/biomet/52.3-4.591](https://doi.org/10.1093/biomet/52.3-4.591)
- Wooldridge, J., 1999. *Introductory Econometrics: A modern Approach*. South-Western College Pub.

Annex 1. Statistical indicators

No.	Term lactation	Total amount of milk (litres)	Daily quantity of milk (litres)	Weight (Kg)	Nutrient units vital functions	Lactation nutritional units	Digestible protein vital functions	Digestible milk protein	Fat percentage	Protein percentage	Lactose percentage
1	285	10381	36,40	650	3,25	18,200	455	2365,936	5,07	3,36	4,66
2	305	10843	35,55	600	3	17,775	420	2310,732	5,09	3,22	4,70
3	285	9465	33,24	580	2,9	16,619	406	2160,522	4,73	3,29	4,65
4	301	11076	36,84	680	3,4	18,419	476	2394,479	5,09	3,35	4,80
5	302	10635	35,18	560	2,8	17,588	392	2286,395	4,82	3,18	4,57
6	305	10092,5	33,09	650	3,25	16,545	455	2150,861	5,41	3,22	4,59
7	304	10770	35,47	580	2,9	17,734	406	2305,395	5,04	3,24	4,64
8	300	11249	37,45	640	3,2	18,727	448	2434,506	5,18	3,24	4,62
9	291	10970	37,66	600	3	18,828	420	2447,614	4,98	3,31	4,66
10	291	10970	37,66	620	3,1	18,828	434	2447,614	4,98	3,31	4,66
11	299	8895	29,78	670	3,35	14,892	469	1935,926	4,68	3,10	4,72
12	295	6437	21,85	590	2,95	10,923	413	1420,000	5,47	3,35	4,71
13	284	7594	26,74	680	3,4	13,370	476	1738,140	5,2	3,33	4,64
14	272	7625	28,00	560	2,8	13,999	392	1819,841	5,57	3,34	4,68
15	305	10026	32,87	590	2,95	16,436	413	2136,689	4,75	3,38	4,7
16	305	8364	27,42	570	2,85	13,711	399	1782,492	5,48	3,53	4,80
17	305	9067,5	29,73	600	3	14,865	420	1932,418	5,35	3,41	4,79
18	305	10361	33,97	610	3,05	16,984	427	2207,975	5,54	3,44	4,71
19	305	9518	31,21	590	2,95	15,603	413	2028,426	4,98	3,25	4,85
20	254	9423	37,10	630	3,15	18,549	441	2411,398	4,24	3,31	4,62
21	305	9925	32,54	570	2,85	16,270	399	2115,164	5,31	3,33	4,86
22	305	8283	27,16	580	2,9	13,579	406	1765,230	4,78	3,30	4,77
23	305	10622	34,83	600	3	17,413	420	2263,705	4,97	3,38	4,77
24	305	10550	34,59	610	3,05	17,295	427	2248,361	4,63	3,25	4,74
25	305	10429	34,19	590	2,95	17,097	413	2222,574	5,31	3,32	4,73

26	305	9176	30,09	600	3	15,043	420	1955,541	5,03	3,27	4,76
27	305	8611	28,23	610	3,05	14,116	427	1835,131	4,78	3,18	4,52
28	305	9496	31,13	600	3	15,567	420	2023,738	4,86	3,49	4,79
29	305	9898	32,45	580	2,9	16,226	406	2109,410	4,93	3,45	4,78
30	305	8595	28,18	570	2,85	14,090	399	1831,721	5,68	3,40	4,81
31	305	9889	32,42	590	2,95	16,211	413	2107,492	5,12	3,41	4,84

Source: Data taken from SCDA Simnic/ www.pedigriu.ro