Colloquium Biometricum 46
2016, 15−29

COMPARISON OF PARMESAN CHEESE DRYING COURSE BY MEANS OF REGRESSION METHODS

Małgorzata Szczepanik[1], Paweł Sobczak[2], Jacek Mazur[2], Kazimierz Zawiślak[2], Rafał Nadulski[2], Mirosława Wesołowska-Janczarek[3], Dorota Domagała[1], Józef Grochowicz[4], Wioletta Żukiewicz-Sobczak[3]

[1]Department of Applied Mathematics and Computer Science,
University of Life Sciences, Głęboka 28, 20-612 Lublin,
[2] Department of Food Engineering and Machines,
University of Life Sciences, Doświadczalna 44, 20-236 Lublin
[3]Pope John Paul II State School of Higher Education,
Sidorska 95/97, 21-500, Biała Podlaska
[4] Warsaw School of Tourism and Hospitality in Warsaw, Stokłosy 3, Warszawa
e-mail: malgorzata.szczepanik@up.lublin.pl

Summary

The aim of the article is to compare regression lines which describe relations between the inverse of water content in cheese and time of drying. Each line involves one of the combinations of Parmesan's type cheese piece sizes (4 mm, 6 mm) and drying temperatures (30°C and 50°C) and piece size of 8 mm dried in 50°C. Regression lines were compared for five combinations of piece sizes and drying temperatures. The duration of the drying process was ten hours. Regression coefficients in the equations were compared using partial F-test. Multiple pair comparisons were examined by means of SimReg procedure. The latter method allowed to identify pairs of significantly different regression equations as well as pairs of equations that do not differ significantly. A similar analysis was performed for the observations regarding the second drying period (from 5th to 10th hour of drying). In addition to the aforementioned methods also the procedure of determining groups of mutually parallel straight lines was used. The application of that method enabled the division of five equations into two groups.

Keywords and phrases: linear regression, hypothesis testing, multiple comparisons, confidence bands, cheese, convection drying

Classification AMS 2010: 62F03, 62F25, 62J05, 62J15

1. Introduction

Conventional hot air drying is the most widely used technique for production of dehydrated fruits and vegetables. However, it involves high energy costs. In processes such as drying or ripening, which induce water transfer and weight loss, the in-depth knowledge of the mass transfer mechanisms substantially contributes not only to reduction of the operating costs, but also to improvement of the quality of the final product. In the case of cheese industry, ripening is the longest step in the manufacture of cheese, where an considerable loss of water occurs (Bradley and Vanderwarn, 2001; Law, 1999). During the drying process, water vaporizes from a wet surface of raw material to a stream of air (Castell-Palou et al., 2011). Drying cheese is useful to preserve the dairy product while creating special ingredients necessary for recipes such as chicken parmesan. An understanding of the mass transfer mechanisms of drying would thus contribute to improving engineering design and the quality of the final product (Castell-Palou et al., 2012; Zhang et al., 2013). Mass transfer and volume changes are also observed in the process of air frying. The results of analysis (Andrés et al., 2013) showed that oil uptake is much lower in air frying although a much longer processing time is required. Also, water loss and thus the loss of volume were much higher in air frying compared to the conventional process.

Modelling is essentially a way of representing processes or phenomena to explain the observed data and to predict behavior under different conditions. Mathematical models describing drying processes are necessary for engineering design and optimization (Castell-Palou and Simal, 2011).

When several groups of experimental units are considered, e.g. varieties, fractions, or units exposed to different types of heating, the question often appears whether regression equations fitted for these groups are the same. Therefore, we are interested in statistical comparison between linear regression models.

There are many methods to compare the means of a number of observation groups, including multiple comparisons of means of several groups proposed by Tukey (1953), Scheffe (1953) and Dunnet (1955).

In contrast to the methods employing comparison of means the ones that we are going to use in this paper i.e. comparing several regression models are less known and used occasionally. The best known test of comparison of two or more regression models is the partial F-test (Seber, 1977, Kleinbaum et al., 1998, Draper and Smith, 1998). The results for several regression models were presented by Gujarati (1970) and Spurrier (1999, 2002). In recent years, interesting studies based on simultaneous confidence bands were published by Liu et al. (2004), Jamshidian et al. (2005) and Jamshidian et al. (2010).

Another issue related to the comparison of several regression models is to test whether for several groups the change in the dependent variable described by regression equation is of similar dynamic. This problem can be resolved by checking the parallelism in the case of regression lines or regression surfaces. In these cases the partial F-test is most often used, but it provides only general information whether the hypothesis of parallelism for all test groups is rejected or not. Wojnar and Zieliński (2004) proposed a procedure for division slopes into homogeneous groups (groups of parallel regression lines).

Several methods mentioned above will be further used for the comparison of regression lines, which correspond to the five combinations of Parmesan's type cheese piece sizes and drying temperatures.

2. Materials and methods

The aim of the experiment was to determine changes in the water content of the convective type Parmesan cheese, shredded, intended for the purpose seasoning. The amount of water in the cheese to a large extent determines its technological usefulness and durability. The experiments were conducted using cheese specimens shredded on a disc grinder with a mesh of 4 mm (small) and 6 mm (medium), which were dried at the temperatures of 30°C and 50°C, and the particles formed in the blender with an aperture of 8 mm (large) dried at 50°C. 100 gram samples of comminuted cheese were placed on a sieve convection dryer and every hour the samples were tested for water content in $kg \cdot kg^{-1}$ dry matter by means of drier method in accordance with DIN EN ISO 5534 (ISO 5534:2004). Drying air flow velocity was 3 $m \cdot s^{-1}$. Water content was measured every hour, beginning from the start of the process, for 10 hours with three replicates (for later calculations the means were taken). Specific value of water content in cheese means how many kilograms of water contained in cheese there are in one kilogram of cheese dry matter. The scheme of laboratory convection drier is presented in Fig. 1.

Five combinations of piece sizes and drying temperatures were taken into consideration and were further denoted as 30-s, 30-m, 50-s, 50-m, 50-l (temperature: 30°C, 50°C, sizes: small, medium, large).

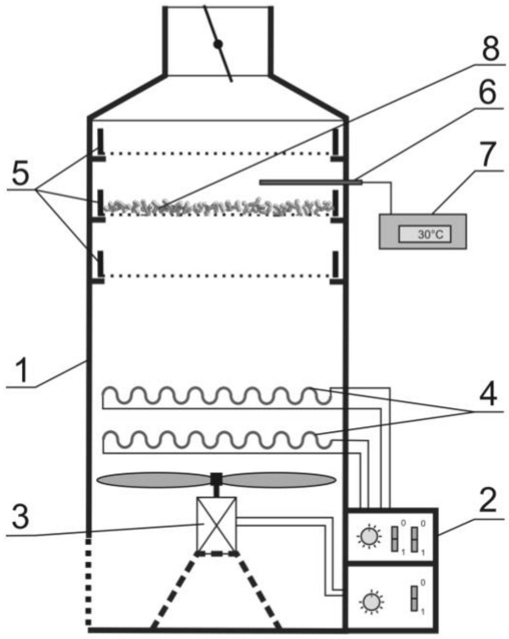

Fig. 1. Scheme of laboratory convection drier: 1-frame, 2-steering panel (fan and heater), 3-fan, 4-heaters, 5-sieves, 6-temperature sensor, 7-temperature recorder, 8-raw material

3. Models and results

3.1. Variable transformation and model

The changes of water content in cheese (*w*) in time (*t*) for the group 50-m are presented in Fig.2 (the upper chart). The data indicate that the relationship is curvilinear. It is possible to obtain linear shape using transformation of the dependent variable (water content: *w*), independent variable (time: *t*) or both variables simultaneously (Neter et al., 1996). In this case Box-Cox transformation of dependent variable was used: *w* to the form $y=w^{-1}$. The observations after transformation are presented in Fig.2 (the chart at the bottom). The points show the linear relationship between inverse of water content in cheese and drying time. Similar results were obtained for the other groups of observations. The change in the inverse of water content in cheese in time for five groups of observations was described by means of linear regression model:

$$y_{ij} = \alpha_i + \beta_i t_j + \varepsilon_{ij} \qquad (3.1)$$

where i is the number of group ($i=1,2,...,5$), j is the number of observation in group ($j=0,1,...10$). The parameters α_i, β_i are regression coefficients and t_j denotes time point ($t_j=j$). The error terms ε_{ij} are assumed to be independent and identically normally distributed with mean 0 and variance σ^2.

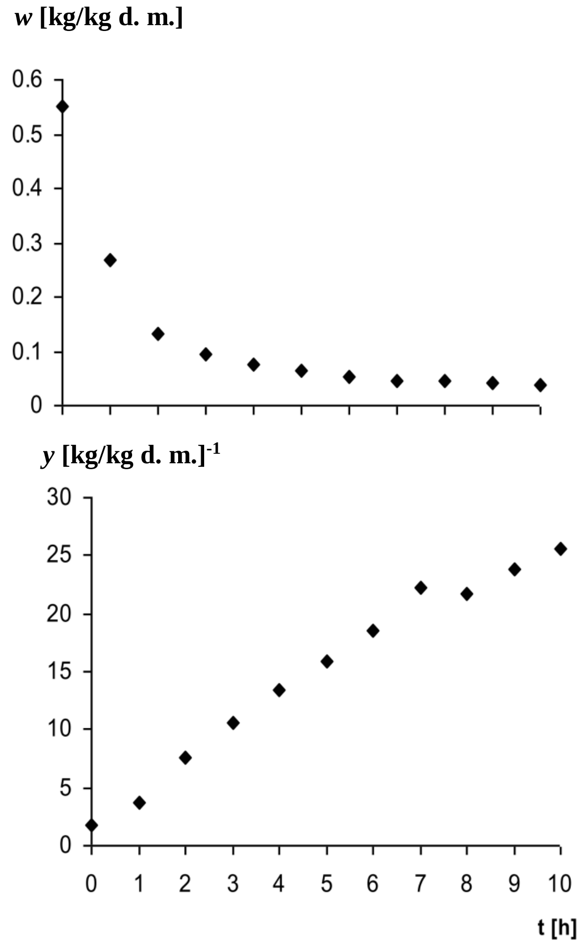

Fig. 2. Changes in water content in cheese (w) in time (t) and changes in inverse of water content in cheese (y=w-1) in time (t) for observations 50-m

Linear regression equations which coefficients were estimated by means of least square method and coefficients of determination for considered combinations of size and drying temperature are as follows:

30-s: $Y=4.55+1.95t$, $R^2=0.90$,

30-m: $Y=3.04+0.68t$, $R^2=0.94$,

50-s: $Y=6.95+2.86t$, $R^2=0.88$,

50-m: $Y=2.71+2.46t$, $R^2=0.98$,

50-l: $Y=2.24+1.06t$, $R^2=0.97$.

The charts of these regression lines are presented in Fig. 3.

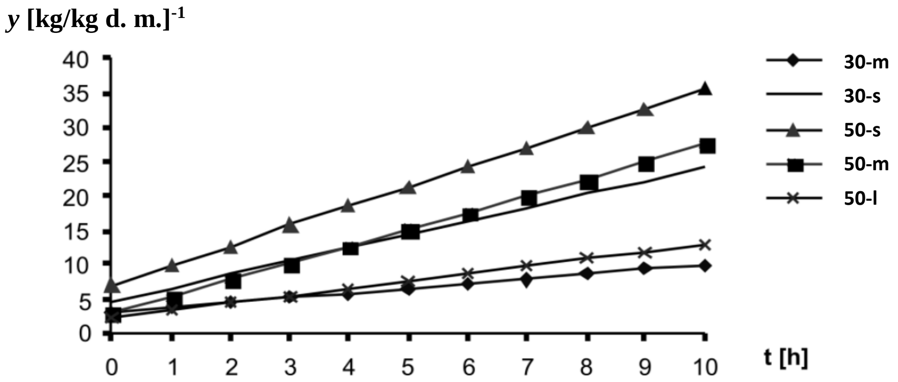

Fig. 3. The fitted regression lines describing the relationship between inverse of water content (*y*) and time of drying (*t*) during ten-hours process for five groups

3.2. Testing equality of regression coefficients (partial F-test)

The aim of further analysis is to check if combinations of drying temperature and size have similar influence on process of y change, or if regression lines fitted for these groups are significantly different. To compare regression lines described by (3.1), full regression model including four dummy variables (because five models are compared) should be built (Neter et al., 1996; Draper and Smith, 1998). The dummy variables are defined as follows:

$$z_1 = \begin{cases} 1 & \text{for group 30-m} \\ 0 & \text{for others,} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{for group 50-s} \\ 0 & \text{for others,} \end{cases}$$

$$z_3 = \begin{cases} 1 & \text{for group } 50-\text{m} \\ 0 & \text{for others,} \end{cases}$$

$$z_4 = \begin{cases} 1 & \text{for group } 50-\text{l} \\ 0 & \text{for others.} \end{cases}$$

Full model with dummy variables has the following form:

$$y = \gamma_0 + \gamma_1 t + z_1(\gamma_2 + \gamma_3 t) + z_2(\gamma_4 + \gamma_5 t) + z_3(\gamma_6 + \gamma_7 t) + \\ + z_4(\gamma_8 + \gamma_9 t) + \varepsilon \tag{3.2}$$

where y is the variable representing the inverse of water content in cheese, the variable t denotes time and

$$\gamma_0 = \alpha_1,\ \gamma_1 = \beta_1,\ \gamma_2 = \alpha_2 - \alpha_1,\ \gamma_3 = \beta_2 - \beta_1,\ \gamma_4 = \alpha_3 - \alpha_1,\ \gamma_5 = \beta_3 - \beta_1,$$

$$\gamma_6 = \alpha_4 - \alpha_1,\ \gamma_7 = \beta_4 - \beta_1,\ \gamma_8 = \alpha_5 - \alpha_1,\ \gamma_9 = \beta_5 - \beta_1$$

where $\alpha_i, \beta_i,\ i = 1,\ldots,5$ are regression coefficients of i-th group from (3.1). The error term ε is assumed to be identically and normally distributed with mean zero and variance σ^2 and observations are independent.

To check if regression equations estimated for the considered combinations of drying temperatures and sizes are the same, the following hypothesis should be formulated:

$H_0 : \gamma_2 = \gamma_3 = \ldots = \gamma_9 = 0$ against the hypothesis:

$H_1 :$ at least one $\gamma_k \neq 0$, $k = 2,3,\ldots,9$.

The relevant test statistics known also as partial F test (Seber, 1977; Neter et al., 1996; Draper and Smith, 1998; Kleinbaum et al., 1998) is:

$$F = \frac{(SSE_{H_0} - SSE)/q}{SSE/v}, \tag{3.3}$$

where SSE is the residual sum of squares in full model (3.2), SSE_{Ho} is the residual sum of squares in the model under true zero hypothesis, q is the difference in degrees of freedom for residuals in both models and ν denotes the degrees of freedom for residuals in the full model. The values can be taken from the analysis of variance tables for regression.

Using the values from Tables 1, 2 the value of statistics (3.3) is calculated F=59.58 for SSE=186.68, SSE_{Ho}=2163.87, q=53-45=8 and ν=45. The zero hypothesis is rejected at significance level 0.05 since $F>F_{0,05;8;45}$=2.15. In conclusion it is has been shown that at least one pair of coefficients in the compared lines is significantly different.

Table 1. The analysis of variance for full regression model (3.2)

	SS	df	MS	F	p-value
regression	3761.72	9	417.97	100.76	<0.000001
residual	186.68	45	4.15		
total	3948.39				

Table 2. The analysis of variance for the model under true zero hypothesis, where only one independent variable (t) is taken

	SS	df	MS	F	p-value
regression	1784.52	1	1784.52	43.71	<0.000001
residual	2163.87	53	40.83		
total	3948.39				

3.3 Multiple comparisons of regression lines

The SimReg procedure introduced in Jamshidian et al. (2005) was used to check which pairs of regression lines are significantly different. The results are presented in Table 3. It appears that only for pairs 30-s, 50-m and 30-m, 50-l there are no reasons to claim significant differences. In other comparisons, the pairs of regression lines differ significantly.

The graphic exemplary interpretation of the results can be seen in the charts of 95% confidence regions, which are presented in Fig. 4 and Fig. 5 (created by means of the SimReg procedure). If the horizontal line is contained in confidence region there is no reason to claim significant difference (Fig. 4). When the horizontal line crosses the bands of confidence region, the pair of regression lines is different (Fig. 5).

Table 3. Results of simultaneous pairwise comparisons with the use of SimReg procedure

pair	result	pair	result
30-s, 30-m	*	30-m, 50-s	*
30-s, 50-l	*	30-m, 50-m	*
30-s, 50-s	*	50-l, 50-s	*
30-s, 50-m	no significant difference	50-l, 50-m	*
30-m, 50-l	no significant difference	50-s, 50-m	*

* denotes significant difference at 0.05 significance level

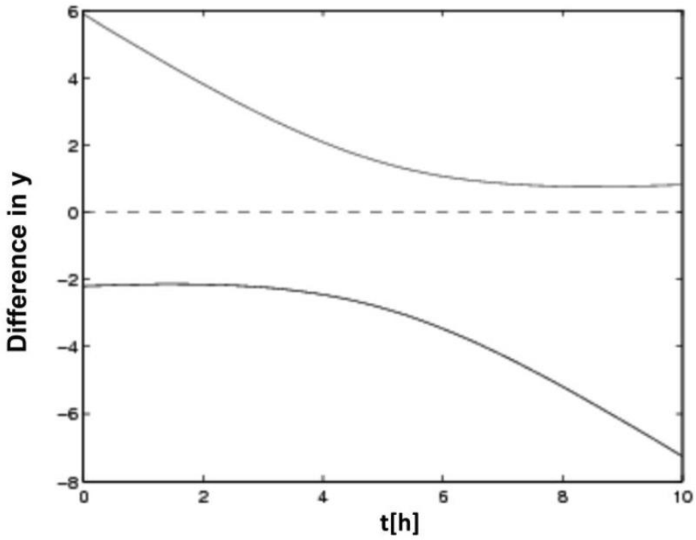

Fig. 4. Comparison plot for the pair 30-s and 50-m (no significant difference at 0.05 significance level)

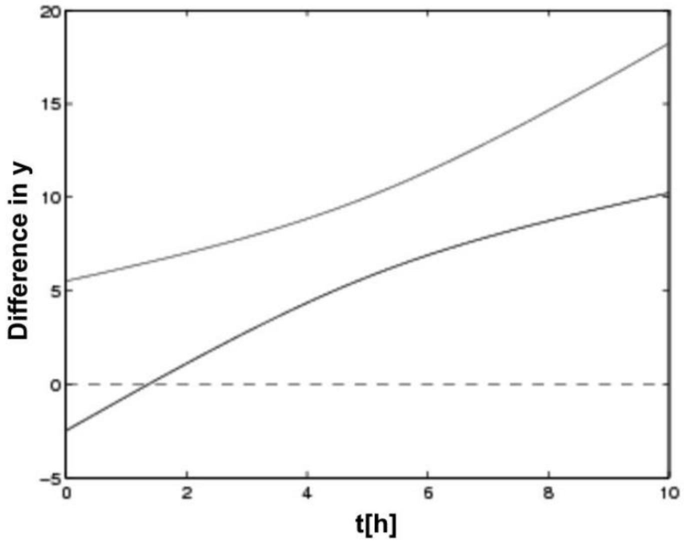

Fig. 5. Comparison plot for the pair 30-s and 30-m (significant difference)

3.4 Regression lines in the second drying period

The next step of the experiment's analysis is to compare the regression equations in the second drying period from 5^{th} to 10^{th} hour of drying or for $i=1,2,...,5$ and $j=4,5,...,10$ in (3.1). The second drying period is characterized by decreasing drying velocity. For the considered groups of observations the following regression lines parameters were estimated using least squares method and the coefficients of determination were calculated:

30-s: $Y=11.58+1.01t$, $R^2=0.95$,

30-m: $Y=4.57+0.47t$, $R^2=0.99$,

50-s: $Y=15.27+1.73t$, $R^2=0.85$,

50-m: $Y=6.16+2t$, $R^2=0.95$,

50-l: $Y=3.62+0.88t$, $R^2=0.88$.

The charts of regression lines are presented in Fig. 6.

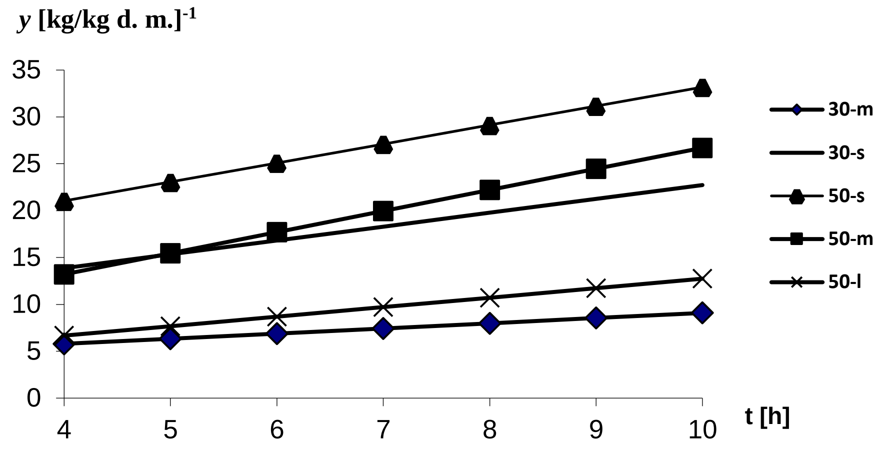

Fig. 6. Charts of fitted regression lines describing the relationship between inverse of water containing (*y*) and time of drying (*t*) for five groups during second drying period

3.5 Multiple comparisons of regression lines in the second drying period

The SimReg procedure was used to examine if the pairs of fitted regression lines are significantly different in the second drying period. The results are shown in Table 4. For all the considered comparisons there were revealed significant differences at significance level 0.05. It means that the combinations of drying temperatures and piece sizes differentiate the course of regression lines in the second drying period.

Table 4. The results of simultaneous pairwise comparisons in the second drying period using SimReg procedure

pair	result	pair	result
30-s, 30-m	*	30-m, 50-s	*
30-s, 50-l	*	30-m, 50-m	*
30-s,50-s	*	50-l, 50-s	*
30-s/50-m	*	50-l, 50-m	*
30-m/50-l	*	50-s, 50-m	*

* denotes significant difference at 0.05 significance level

3.6 Testing equality of slopes in the second drying period

Charts in Fig. 6 show that it is sensible to test hypothesis about the parallelism of regression lines. The parallelism would mean that changes in inverse water content in time proceed with similar dynamics. In order to determine the groups of mutually parallel lines the method from Wojnar and Zieliński (2004) is used. This method takes into consideration every possible division of the group of several equations into subgroups. In the first step the hypothesis about the parallelism of all lines is tested, in other words the hypothesis about the equality of all slopes from model (3.1) for $i=1,2,...,5$ and $j=4,5,...,10$:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_5 \qquad (3.4)$$

against the alternative at least one slope is distinct.

When the hypothesis (3.4) is not rejected the coefficients are considered to be equal and hence it is concluded that all regression lines are parallel. If the hypothesis (3.4) is rejected then in the second step all possible divisions of equations into two disjoint groups are considered and the equality of coefficients within these groups is tested. In the case of rejection of the hypotheses for the division into two groups the next step is to consider all possible divisions into three groups and so on.

In the Maple program the algorithm was implemented for the method. The results are shown in Table 5. In the first step the hypothesis (3.4) was rejected. In the second step the division of equations for homogenous subgroups was obtained: (30-s, 30-m, 50-l) and (50-s, 50-m). The value of the test statistics for this division (1.44) was smaller than the critical value (2.99). The hypothesis about the equality of the slopes in subgroups is not rejected. It means that changes of inverse water content in time proceed with similar dynamics (parallel) for observations from groups 30-s, 30-m, 50-l. Parallelism is also revealed for regression lines based on observations from groups 50-s, 50-m.

Table 5. Divisions, critical values (significance level 0.05) and test statistic values

Step	division	critical value	test statistic value
1	[30-s,30-m,50-s,50-m,50-l]	2.76	8.63
2	[30-s, 30-m, 50-s, 50-m], [50-l]	2.99	10.49
2	[30-s, 30-m, 50-s, 50-l], [50-m]	2.99	5.95
2	[30-s, 30-m, 50-m, 50-l], [50-s]	2.99	9.17
2	[30-s, 50-s, 50-m, 50-l], [30-m]	2.99	6.43
2	[30-m, 50-s, 50-m, 50-l], [30-s]	2.99	11.12
2	[30-s, 30-m, 50-s], [50-m, 50-l]	2.99	10.30
2	[30-s, 30-m, 50-m], [50-s, 50-l]	2.99	11.33
2	[30-s, 30-m, 50-l], [50-s, 50-m]	2.99	1.44
2	[50-s, 50-m, 50-l], [30-s, 30-m]	2.99	5.99
2	[30-s, 50-s, 50-m], [30-m, 50-l]	2.99	4.41
2	[30-s, 50-s, 50-l], [30-m, 50-m]	2.99	11.50
2	[30-m, 50-m, 50-l], [30-s, 50-s]	2.99	10.96
2	[30-s, 50-m, 50-l], [30-m, 50-s]	2.99	11.16
2	[30-m, 50-s, 50-l], [30-s, 50-m]	2.99	9.50
2	[30-m, 50-s, 50-m], [30-s, 50-l]	2.99	9.73

4. Conclusions

To sum up the results of considerations it can be claimed that the transformation of dependent variable enabled the description of relation inverse water content in cheese and time by the linear regression model. It was revealed that the considered combinations of drying temperatures and piece sizes differentiate the course of fitted regression lines during ten-hour drying process. The result of F-partial test (3.3) pointed out the differences in the course of regression lines. SimReg procedure enabled the examination of differences between the pairs of lines. At the significance level of 0.05 significant differences for 8 in 10 pairs of regression were discovered. There are no reasons to claim significant differences between lines fitted for 4mm pieces dried in the temperature of 30°C and 6mm pieces dried in the temperature of 50°C and between regression lines fitted for 6mm pieces dried in the temperature of 30°C and 8 mm pieces dried in the temperature of 50°C.

In the second drying period, significant differences in regression lines fitted for all pairs of five considered equations were revealed. Based on Wojnar and Zieliński (2004) method, the division of regression lines into two subgroups of parallel lines was obtained. This division implies that the changes of inverse water content in cheese in time proceed with similar dynamics for 4 mm and 6 mm pieces dried in the temperature of 30°C and for 8 mm pieces dried in the temperature of 50°C. Similar dynamics was also observed for regression lines fitted for 4 mm and 6 mm pieces dried in the temperature of 50°C.

The above conclusions may constitute a basis for further research on the choice of optimal time and temperature of drying. Normally slower drying enables to obtain better quality material, however such a process incurs higher energy costs. Higher drying velocity is more economical, but results in poor quality products.

References

Andrés A., Arguelles A., Castelló M.L., Heredia A. (2013). Mass transfer and volume changes in french fries during air frying. *Food and Bioprocess Technology* 6(2), 1917-1924.

Bradley R.L., Vanderwarn M.A. (2001). Determination of moisture in cheese and cheese products. *Journal of AOAC Internationa* 84(2), 570-592.

Castell-Palou A., Váquiro H.A., Cárcel J.A., Rosselló C., Femenia A., Simal S. (2012). Mathematical modeling of moisture distribution and kinetics in cheese drying. *Drying Technology* 30 (11-12), 1247-1255.

Castell-Palou A., Rosselló C., Femenia A., Bon J., Simal S. (2011). Moisture profiles in cheese drying determined by TD-NMR: Mathematical modeling of mass transfer. *Journal of Food Engineering* 104(4), 525–531.

Castell-Palou A., Simal S. (2011). Heat pump drying kinetics of a pressed type cheese. *Food Science and Technology* 44(2), 489–494.

Draper N.R., Smith H. (1998). *Applied Regression Analysis* (Third Edition). Wiley Series in Probability and Statistics.

Dunnett C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50, 1096-1121.

Gujarati D. (1970). Use of dummy variables in testing for equality between sets of coefficients in linear regressions: A generalization. *The American Statistician*, 24 (5), 18-22.

ISO 5534:2004. Cheese and processed cheese - Determination of the total solids content (Reference method).

Jamshidian M., Liu W., Bretz F. (2010). Simultaneous confidence bands for all contrasts of three or more simple linear regression models over an interval. *Computational Statistics & Data Analysis* 54(6), 1475-1483.

Jamshidian F., Zhang Y., Liu W., Jamshidian M. (2005). SimReg: A software including some new developments in multiple comparison and simultaneous confidence bands for linear regression models. *Journal of Statistical Software* 12(2), 1-23.

Kleinbaum D.G., Kupper L.L., Muller K.E., Nizam A. (1998). *Applied Regression Analysis and Other Multivariable Methods*, 3rd Edition. Duxbury Press.

Law B. (1999). *Technology of cheese making*. Sheffield Academic Press, Sheffield, UK.

Liu W., Jamshidian M., Zhang Y. (2004). Multiple comparison of several regression models. *Journal of the American Statistical Association*, 99, 395–403.

Neter J., Kutner M.H., Nachtsheim Ch.J., Wasserman W. (1996). *Applied Linear Statistical Models*. Chicago (4th ed.), IL: Irwin.

Scheffé H., (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* 40, 87–104.

Seber G.A.F. (1977). *Linear Regression Analysis*. John Wiley & Sons.

Spurrier J.D. (1999). Exact confidence bounds for all contrasts of three or more regression lines. *Journal of the American Statistical Association* 94, 483-488.

Spurrier J.D. (2002). Exact multiple comparisons of three or more regression lines: pair wise comparisons and comparisons with a control. *Biometrical Journal* 44, 801-812.

Tukey J.W., (1953). *The problem of multiple comparisons*. Unpublished manuscript. In The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948–1983 1–300.Chapman and Hall, New York.

Wojnar J., Zieliński W. (2004). Multiple comparisons of slopes of regression lines. *Biometrical Letters* 41(1), 25-35.

Zhang L., Nishizu T., Hayakawa S., Nakashima R., Goto K. (2013). Effects of different drying conditions on water absorption and gelatinization properties of pasta. *Food and Bioprocess Technology* 6(8), 2000-2009.