ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume LXI 295 Number 7, 2013
http://dx.doi.org/10.11118/actaun201361072639

REPRESENTATIVENESS OF THE FADN CZ SAMPLE OF AGRICULTURAL ENTERPRISES AND WAYS OF ITS VERIFICATION

Marie Prášilová, Radka Procházková, Martina Harvilíková

Received: August 30, 2013

Abstract

PRÁŠILOVÁ MARIE, PROCHÁZKOVÁ RADKA, HARVILÍKOVÁ MARTINA: *Representativeness of the FADN CZ sample of agricultural enterprises and ways of its verification.* Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2013, LXI, No. 7, pp. 2639–2648

Following accession to the European Union, Czech Republic joined the EU FADN system. The Farm Accountancy Data Network (FADN) is the main EC information source on real economic situation of agricultural enterprises. It is a sample survey of individual farms and legal entities from the primary farm production domain and it supplies very detailed information on economic results. Representativeness, i.e., the degree of correspondence of sample information with population information in all relevant parameters except for size, is the priority requirement on sample surveys. The aim of testing the FADN CZ sample survey representativeness was to verify whether it was possible to draw conclusions on the population from the sample outcomes, in order to assess the whole of agriculture. Czech Statistical Office (CSO) data bases supplied data on the population.
The paper deals with statistical techniques and methods suitable for the assessment of representativeness and the extent of sample needed. The actual results of representativeness verification are offered here, based on FADN CZ 2011 accounting year sample survey results. These were obtained in 2012. STATISTICA 10 special package was employed in the representativeness testing exercise. The analysis also offers methodology proposals in order to increase quality of the sample.

representativeness, agriculture, agricultural enterprise, agricultural holding, farm accountancy data network FADN CZ, sample survey, testing

The Farm Accountancy Data Network was launched 1965 in the EU, after a legal basis for its organisation was established by the EC Regulation 79/65. FADN system is the only EC information source on the real economic situation of agricultural enterprises.

The Czech Republic joined the FADN starting 1994. Based on a network of test units (farms) the then Institute of Agricultural Economics and Information provided sample surveys of economic results in agriculture already before. In 2003 the FADN CZ Liaison Agency was established based on ministerial decision. Czech Republic accounting and advisory offices have been providing collection of FADN data since then. These are also responsible for processing and forwarding FADN farm data to the Liaison Agency. Data have been then forwarded annualy from the Czech Republic to the EU Directorate-General for Agriculture and Rural Development from 2004 on.

A special classification system of agricultural enterprises has been established in EU statistics for the purposes of sample surveys. This system makes it possible to order the enterprises according to their economic sizes and types of farming. It makes it possible to categorize all the farms involved in the Agrocensus into the farming classes by size and type of farming, and to design selection plans.

The number of farms in the FADN sample is higher (up to 20% more) in the almost all of the EU member states in order to ensure forwarding of required number to the EC. The EU Member States

employ additional criteria for selection of farms into the FADN system. In the Czech Republic the EU methodology of selection is satisfied but moreover, other factors are taken into account for selection of farms into the FADN CZ sample. First of all it is the regional location of farms, materialized in the administrative and natural regions (production areas, less favoured areas). The aim here, is to ensure as high a representativeness of the FADN CZ sample, as possible.

Sample survey outcomes concerning economic results of agricultural farms included in the FADN CZ sample are published in the Appendix to the Report on the State of Czech Agriculture. The sample surveys are organized by the Institute of Agricultural Economics and Information in Prague. The aim of this network is to collect accountancy data from farms and to provide income and business analysis of agricultural enterprises of all legal forms, economic size classes and farming types. Hence, data from the FADN data base are employed intensively for many different purposes. Apart from the regular outcomes, (Report on the State of Czech Agriculture and some special tasks), the data are used by the Czech Ministry of Agriculture in managing actual problems, analyses and prognoses of agricultural policies. Furthermore, the FADN outcomes are supplied to the Czech Statistical Office and other users of agricultural statistics (Palát *et al*., 2012).

The sample of 1, 422 agricultural enterprises from the FADN CZ data base was forwarded to the EU data base for the accounting year 2011 and the same sample was employed in the statistical representativeness verification exercise. The whole area of agricultural lands of the farms included in the sample was 938,095 hectares, representing 26.5% of the total agricultural lands utilized in the Czech Republic in 2011 (Kol., 2012).

MATERIAL AND METHODS

Statistical testing of the FADN CZ sample representativeness was held with the aim of verifying the possibility to draw conclusions about the population from which the sample was taken. The quality of results of statistical analysis methods, which were applied on the concrete samples of data, depends on quality of these samples. In case of unsuitable characteristics of data sets, is using of some of methods incorrect and acquired results does not correspond with the reality. The data sets characteristics depend on the way of the collecting.

The basic assumption of correct results is that parameters of a random variable, which was acquired as estimations of parameters by analysing of the sample, differs as little as possible from the real parameters, which would be acquired by analysing of the population (Čermák, Vrabec, 1999). The population is presented by the CSO data in the paper.

A sample survey could not only replace the census but also could supplement it properly. Within a sample survey we are able to pay a much higher attention to the actual data collection and therefore to achieve higher reliability of data or else to collect more variables i.e. to extend the survey (Čermák, Vrabec, 1999). This possibility results from the fact that the researched variables are surveyed only for selected units and therefore the apparatus of these who are engaged is considerably reduced and only persons who are highly qualified remain.

An important requirement of sample surveys is to ensure the representativeness of the sampling. Blahuš (2000) says: *the representativeness is the basic characteristic of the studied sample. It is an extent to which the studied sample corresponds to the whole population in all relevant parameters (attributes, characteristics) except the size. The purpose of the studied sample is to represent the whole population. The representativeness is the extent to which it is successful.* The representativeness is moreover influenced by a chosen method of the sample selection and, to a lesser extent, by the sample size. The representativeness is a continuum and the unattainable ideal maximum. Therefore it is necessary to optimize it for purposes of the research (Blahuš, 2000*)*.

It means that the sample must contain substantial and characteristic features of the population. Only then it is possible to generalize the results.

Statistical procedures for verifying the representativeness

Estimations from the sample survey results are not fully identical with results of the exhaustive census, since the sample survey is always influenced by a sampling error. However when making the selection properly it is possible to eliminate this error. The representativeness is not guaranteed by the sample size, but by the selection procedure (Carver, 1994).

For other statistical surveys, which are based on samples, it is useful to get the idea of numbers and size of units (agricultural holdings) form the point of view of area, production, revenues and other volume variables and also intensive variables such as hectare yields, costs per hectare and revenues per hectare. For this purpose an explorative analysis is used. This analysis provides also many possibilities for improving the sample resulting in better results of statistical analyses.

The task of the explorative data analysis is to make the first evaluation of data quality and to establish assumptions for the following statistical processing. Primarily the aim is to find a curiosity in statistical behaviour of data. As explorative analysis methods are used only procedures, which are not influenced by the type of the surveyed random variable. The explorative analysis is a complex of procedures used for a quick recognition of variation rows and also for detection of outlying (extreme) observations. The basic tool of this analysis is box-and-whisker plot, which in a simple and illustrative way graphically depicts the layout of minimum, maximum, median and both of the quartiles. It is generally considered

that there are more outliers in large samples. Values, which cannot be (with given probability), due to the small number of observations, determined as outlying (extreme), may seem (with the same probability) as outlying in larger samples (Hebák *at al*., 2007).

Outlying values can be detected using a graphic method. These values can be excluded from the sample without decreasing informative value. Therefore homogeneity of the sample increases and variability of surveyed variables decreases. Data were adjusted in this way and then processed.

First of all it was necessary to verify, if representation of individual farms and legal entities is identical in all of the size classes. For this purpose the Pearson's chi-squared test was used. This test checks if there is the equal frequency distribution in the sample and in the population.

The following test criterion was used for calculation:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}, \quad (1)$$

where
k ..... number of cells,
n_j ..... observed frequency,
np_j ..... theoretical frequency (in the population).

If the value of the test criterion χ^2 is not higher than the critical value χ^2_α, the fit of a frequency distribution according to the categories in the sample and the population may be confirmed.

Furthermore differences in proportional representation of subjects in the sample and in the population were checked using the relative frequency (empirical probability) test.

$$u = \frac{f_i - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}, \quad (2)$$

where
u ..... value of test criterion,
f_i ..... relative frequency in the sample,
π_0 ..... anticipated value of relative frequency in the population,
n ..... sample size.

If the value of the test criterion is higher than the critical value of the normal distribution, statistically significant difference is proved. In that case it is possible to subject the sample to the explorative analysis again and after excluding one or two of the outliers the variability of a variable may decrease and insignificance of the difference will be confirmed (Čermák, Vrabec, 1999).

It is possible to test the sample (relative frequency of the variable) of the accounting year 2011 and 2012.

For this purpose was used a test on a difference between relative frequency distributions with the test criterion u:

$$u = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{\bar{p} \times \bar{q}}{n}}}, \ \bar{p} = \frac{m_1 + m_2}{n_1 + n_2}, \ \bar{q} = 1 - \bar{p}, \ n = \frac{n_1 \times n_2}{n_1 + n_2}, \quad (3)$$

where
u ..... test criterion,
n_1 ..... extent of the first sample,
n_2 ..... extent of the second sample,
m_1 ..... number of elements with occurrence of a certain feature sin the first sample,
m_2 ..... number of elements with occurrence of a certain feature sin the second sample.

If the value of the test criterion is lower than the critical value of the normal distribution, the difference between proportions in both of the samples is insignificant. Apart from the representativeness of number of sampled subjects it was also necessary to evaluate the representativeness of established variables.

In cases where data from the population were available (especially within legal entities) tests of differences between the value from the sample and the real value from the population (data from census) were performed. A test on a value of the μ_0 parameter was used.

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}, \quad (4)$$

where
t ..... test criterion,
$\bar{x}$ ..... arithmetic mean of the examined variable,
μ_0 ..... entry from the census,
s^2 ..... variance of the sample,
n ..... extent of the sample.

If the calculated value of the test criterion t is higher than the critical value of the Student's t-distribution, the difference between the value from the sample and the value from the population is statistically significant. The purpose of the tests was to confirm, that on the basis of the sample survey it would be possible to draw quality estimates for the Czech agriculture. It is possible to understand these estimates as point estimation, in that case sample survey average is used for average value estimation, or interval estimation with a predetermined probability (usually 95%).

Methodical proposals for increasing estimates accuracy

In case of a simple random selection is preparation of the selecting scheme reduced to determining the sample extent. The sample extent is limited by several factors, especially by time and financial capacities, their influence on the sample survey in preparation might be crucial during the sample extent defining (Čermák, Vrabec, 1998).

At first we consider point estimation of the average value and we presume, that a request for its accuracy is pre-determined as a standard deviation D($\bar{y}$). For selection without re-placement it is stated:

$$n = \frac{NS'^2}{(N-1)D^2(\bar{y}) + S'^2} = \frac{NS^2}{ND^2(\bar{y}) + S^2}. \quad (5)$$

For interval estimation of the average is the request for estimation accuracy defined as a permissible error $\Delta = u_{1-\alpha/2} D(\bar{y})$, which should not be exceeded (with reliability 1 – α). The extent of the sample is thus calculated as:

$$n = \frac{NS'^2 u^2_{1-\alpha/2}}{(N-1)\Delta^2 + S'^2 u^2_{1-\alpha/2}} = \frac{NS^2 u^2_{1-\alpha/2}}{N\Delta^2 + S^2 u^2_{1-\alpha/2}}. \quad (6)$$

Sometimes during determining of the sample extent dimensionless variables are used, it means that calculation is based on the required relative standard error i.e. on the variation coefficient V($\bar{y}$) or on the set relative permissible error δ regarding interval estimations:

$$n = \frac{NV'^2 u^2_{1-\alpha/2}}{(N-1)\delta^2 + V'^2 u^2_{1-\alpha/2}} = \frac{NV^2 u^2_{1-\alpha/2}}{N\delta^2 + V^2 u^2_{1-\alpha/2}}. \quad (7)$$

It is obvious in all the above mentioned cases that for calculation of the sample extent n it is necessary to know the variance of the population S^2 or its i mpartial estimate S^2 or eventually the variation coefficient V' or its estimate V.

Determination of some of the variability characteristics (even if only approximate) is, in practice, one of the most difficult tasks before the actual survey. If these information sources are not available, a pre-sampling is performed before the actual sample survey and the investigated characteristic (variance, variation coefficient) is estimated (Čermák, Vrabec, 1998).

In our case sample surveys from the previous years may be considered as the pre-sampling.

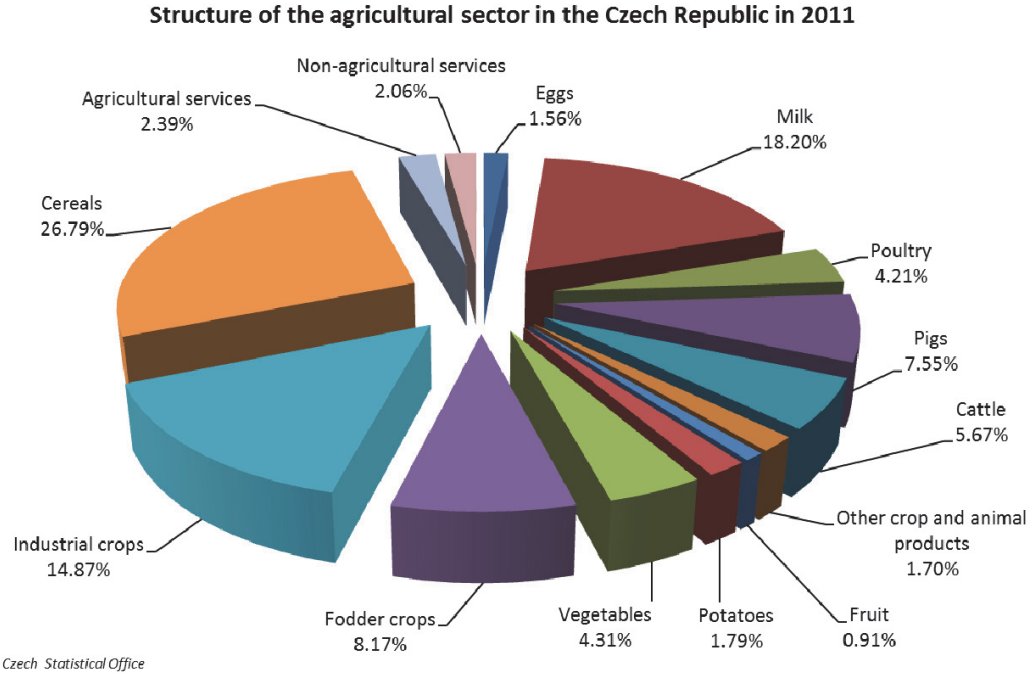
RESULTS

A specialized statistical system Statistica version 10 was used for the representativeness testing. Data from the Report on the state of the Czech agriculture, data from the sample survey of the economic results in agriculture and data from the periodical survey of the CSO was used for the sample survey representativeness verification within the sample of agricultural holdings. The verification of the sample survey of agricultural holdings was performed for crop yields and livestock efficiency. Specifically, the following variables were chosen: hectare yields of wheat, barley, potatoes and rape; milk yield, daily weight gain of pigs and daily weight gain of cattle for fattening.

The reason for choosing these variables was the prevailing share in the agriculture in the CR (Fig. 1).

It was surveyed, if there are differences between the value surveyed within the sample of the agricultural holdings and the real value of the population (the official CSO results of 2011, which are available on the www.czso.cz).

If the calculated value of the test criterion is higher than the value of the normal distribution, the difference between the value from the sample and from the population is statistically significant.



Source: Czech Statistical Office

1: *Production structure of the agricultural sector in 2011 in the Czech Republic*

I: *Statistical characteristics of samples for selected crops – year 2011 (the Czech Republic)*

Crops	Per hectare yield t/ha				
	the total CR	Individual farms	Number of subjects	Legal entities	Number of subjects
Wheat	5.474	5.1728	467	5.765	483
Barley	4.76	4.5223	368	4.8695	368
Potatoes	25.4934	25.0140	103	26.0188	94
Rape	2.9523	2.9646	276	2.9439	407

Source: Calculations on the basis of the CSO data and FADN CZ data

The critical value of the normal distribution was determined as $u_{0,05} = 1.96$.

The statistical system Statistica version 10 which was used for the exploratory analysis of the sample of data and the representativeness testing offers the *p-value*, which states the possible lowest significance level for contradiction of the null hypothesis. This value if offered by all of statistical software systems in all outputs.

If within the study the tests were performed with the reliability level of 95%, i.e. the significance level $\alpha = 0.05$, then it would be considered, if $p > \alpha$, that the null hypothesis cannot be contradicted and the difference between the value from the sample FADN CZ 2011 and from the population (the CSO outputs) is statistically insignificant.

On the contrary, if $p \leq \alpha$, the null hypothesis is contradicted and the difference between the value from the sample FADN CZ 2011 and from the population (the CSO outputs) is statistically significant

The valuation was carried out separately for individual holdings and for legal entities with double-entry bookkeeping. These samples of agricultural holdings are from the whole CR and are classified according to the FADN EU methodology. The following Tab. I and Tab. II summarize results from the whole CR and use relations 1–4.

II: *Values of test criteria for selected crops – year 2011 (the Czech Republic)*

Crop	Test criterion t	
	Individual farms	Legal entities
Wheat	–10.7940	1.1927
Barley	–6.83219	1.9587
Potatoes	–6.36729	–5.18255
Rape	3.89800	3.53445

Source: Calculations on the basis of the CSO data and FADN CZ data

Within the CR there was concordance between hectare yields of crucial agricultural crops calculated within the FADN sample of legal entities and hectare yields presented by the CSO. Although in some cases the parametric tests proved differences between values calculated within the FADN CZ 2011 sample and the reference constant, it is necessary to realize, that in many cases of the expected values the differences were small (in tenths). Results of the exploratory analysis of wheat in structure of individual farms (J) and legal entities (P) are presented in Annex 1.

There is a significant difference between results for potatoes for individual farms and legal entities. The difference was caused by high variability of hectare yields due to seasonal fluctuations. The variation coefficient for potatoes is high (34.64% for individual farms, 31.86% for legal entities).

The exploratory analysis and the representativeness testing of the agricultural holdings sample were also performed for the chosen variables of the livestock production. It was surveyed if there was a difference between the found out value from the sample and the value from the population (the CSO database).

If the calculated value of the test criterion is higher than the value of the normal distribution, the difference between the sample value and the population value is statistically significant. The critical value of the normal distribution was determined as $u_{0.05} = 1.96$. Calculated values of test criteria concerning milk yield are presented in Tab. III.

While the average weight gains of pigs and the weight gains variability differences between individual farms and legal entities are not statistically provable, weight gains of cattle for fattening differences are statistically significant (value of the variation coefficient for individual farms is 19.85%, for legal entities 25.05 %).

It is obvious from the statistical analyses results, that the average milk yield variability (measured by

III: *Values of test criteria for dairy cows' milk yield – year 2011 (litres/day) according to the type of farm*

Milk yield l/day					Test criterion t	
the total CR	Individual farms	Number of subjects	Legal entities	Number of subjects	Individual farms	Legal entities
17.81	15.73	143	18.7	324	–8.70185	–3.22154

Source: Calculations on the basis of the CSO data and FADN CZ data

the variation coefficient) of the individual farms is 10% higher than of legal entities.

Some of the results of the primary analyses of the milk yield variable are presented in Annex 2.

CONCLUSION

The analysis proved that the quality of the FADN CZ 2011 sample of the agricultural holdings is sufficient. The results confirmed that the sample of the agricultural holdings is representative. For many surveyed variables the values from the analysis of the FADN CZ 2011 sample did not statistically differ from the real values, i.e. values which would be calculated from the whole population. Differences, mostly on the regional level, appeared only between some of the variables in the sample, especially within the individual farms; this was due to higher variability of some variables in this sample.

Although in some cases the parametric tests proved differences between the variables, it is necessary to realize, that in many cases of the expected values the differences were very small. The sample of the agricultural holdings is sufficiently large, therefore confidence intervals are narrow, i.e. the FADN CZ sample enables accurate and highly reliable estimates. It is not necessary to widen the sample. However it would be possible to decrease the variability of some variables by excluding outliers from the sample using special statistical diagnostic procedures. It is a topic for discussion, if this recommendation is not in conflict with the client's methodology.

Annex 1: Exploratory data analysis for the total wheat (part of the output) *

Descriptive statistics parsing table (FADN wheat sta.)
N=950

Legal form	area Average	area N	area Minimum	area Maximum	area Stand. deviation	area Coefficient var.	area confid. interval -95.000%	area confid. interval +95.000%
P	358.2330	483	3.000000	2225.450	318.9778	89.0420	329.7145	386.7516
J	52.7308	467	0.800000	81.5839	81.5839	154.7176	45.3122	60.1494
Total	208.0546	950	0.800000	279.8156	279.8156	134.4914	190.2385	225.8707

Rozkladová tabulka popisných statistik (FADN-pšenice.sta)
N=950(V seznamu záv. prom. nejsou ChD)

ÚČET	plocha Průměr	plocha N	plocha Minimum	plocha Maximum	plocha Sm.odch.	plocha Koef.prom.	plocha Int. spolehl. -95.000%	plocha Int. spolehl. +95.000%
P	358,2330	483	3,000000	2225,450	318,9778	89,0420	329,7145	386,7516
J	52,7308	467	0,800000	791,000	81,5839	154,7176	45,3122	60,1494
Vš.skup.	208,0546	950	0,800000	2225,450	279,8156	134,4914	190,2385	225,8707

Descriptive statistics parsing table (FADN wheat sta.)
N=950

Legal form	harvest tons Average	yield N	harvest tons Minimum	harvest tons Maximum	harvest tons Stand. deviation	harvest tons Coefficient var.	harvest tons confid. interval -95.000%	harvest tons confid. interval +95.000%
P	2167.797	483	18.00000	15502.00	2106.134	97.1555	1979.496	2356.098
J	281.534	467	3.00000	4524.00	448.846	159.4288	240.719	322.348
Total	1240.550	950	3.00000	15502.00	1800.574	145.1432	1125.906	1355.194

Rozkladová tabulka popisných statistik (FADN-pšenice.sta)
N=950(V seznamu záv. prom. nejsou ChD)

ÚČET	sklizeň v t Průměr	sklizeň v t N	sklizeň v t Minimum	sklizeň v t Maximum	sklizeň v t Sm.odch.	sklizeň v t Koef.prom.	sklizeň v t Int. spolehl. -95.000%	sklizeň v t Int. spolehl. +95.000%
P	2167,797	483	18,00000	15502,00	2106,134	97,1555	1979,496	2356,098
J	281,534	467	3,00000	4525,00	448,846	159,4288	240,719	322,348
Vš.skup.	1240,550	950	3,00000	15502,00	1800,574	145,1432	1125,906	1355,194

Descriptive statistics parsing table (FADN wheat sta.)
N=950

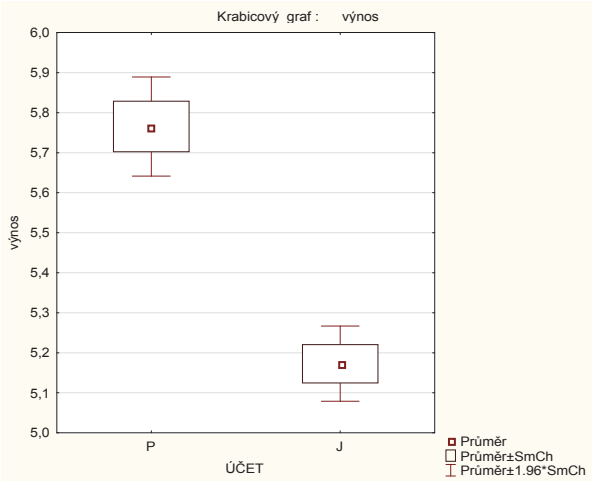
Legal form	yield Average	yield N	yield Minimum	yield Maximum	yield Stand. deviation	yield Coefficient var.	yield confid. interval -95.000%	yield confid. interval +95.000%
P	5.765284	483	1.485149	9.757370	1.387263	24.06236	5.641254	5.889314
J	5.172767	467	1.125000	9.000000	1.036632	20.01892	5.078604	5.266931
Total	5.474015	950	1.125000	9.757370	1.261687	23.04866	5.393682	6.664348

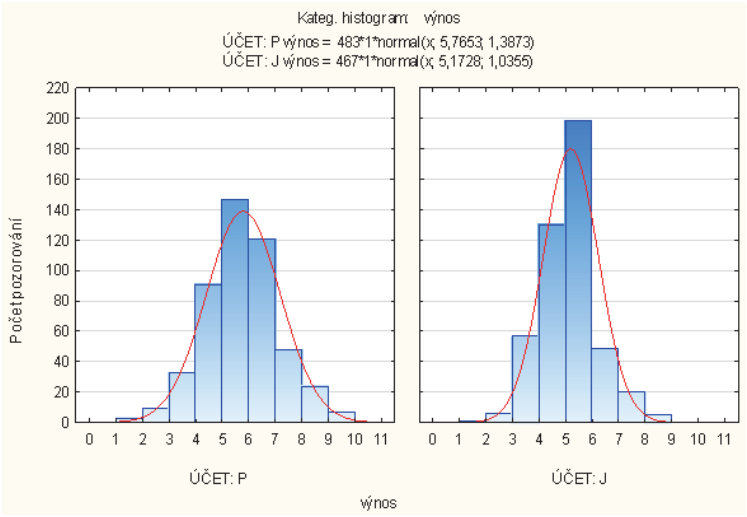
Rozkladová tabulka popisných statistik (FADN-pšenice.sta) N=950(V seznamu záv. prom. nejsou ChD)								
ÚČET	výnos Průměr	výnos N	výnos Minimum	výnos Maximum	výnos Sm.odch.	výnos Koef.prom.	výnos Int. spolehl. -95.000%	výnos Int. spolehl. +95.000%
P	5,765284	483	1,485149	9,757370	1,387263	24,06236	5,641254	5,889314
J	5,172767	467	1,125000	9,000000	1,035532	20,01892	5,078604	5,266931
Vš.skup.	5,474015	950	1,125000	9,757370	1,261687	23,04866	5,393682	5,554348

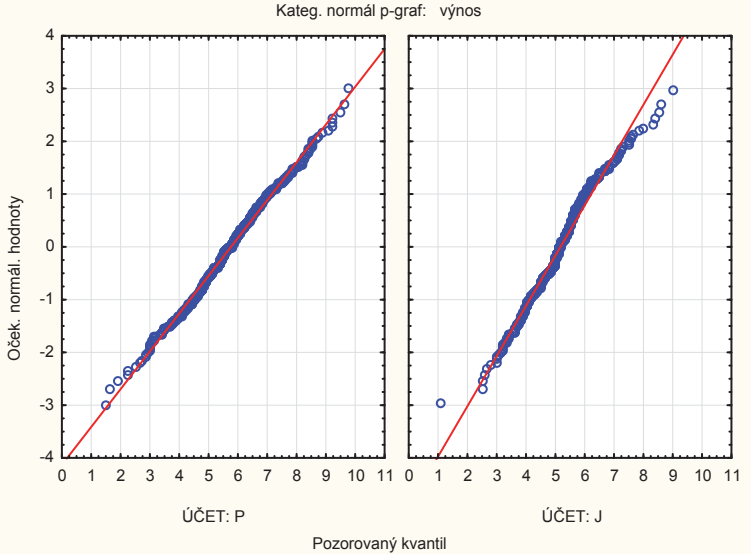
	t-tests; grouped: legal form (FADN-wheat sta.) Group 1:P Group2: J										
Variable	Average P	Average J	t	sv	p	Number P	Number J	Standard deviation P	Standard deviation J	F-ratio Variance	p Variance
yield	5.765284	5.172767	7.4407229	948	0.000000	483	467	1.387263	1.035532	1.794696	0.000000

	t-testy; grupováno: ÚČET (FADN-pšenice.sta) Skup. 1: P Skup. 2: J										
Proměnná	Průměr P	Průměr J	t	sv	p	Poč.plat P	Poč.plat. J	Sm.odch. P	Sm.odch. J	F-poměr Rozptyly	p Rozptyly
výnos	5,765284	5,172767	7,440729	948	0,000000	483	467	1,387263	1,035532	1,794696	0,000000

*individual farms (J) a legal entities (P)







*Annex 2: Exploratory data analysis of the milk yield (l/feeding day), part of the output**

Descriptive statistics parsing table (FADN milk yield sta.)
N=467

Legal form	daily milk yield Average	daily milk yield N	daily milk yield Minimum	daily milk yield Maximum	daily milk yield Stand. deviation	daily milk yield Coefficient var.	daily milk yield confid. interval -95.000%	daily milk yield confid. interval +95.000%
P	18.73458	324	7.104284	29.08944	4.276687	22.82777	18.26716	19.20201
J	15.72954	143	5.964286	29.07895	5.181383	32.94046	14.87301	16.58607
Total	17.81441	467	5.964286	29.08944	4.772919	26.79246	17.38040	18.24842

Rozkladová tabulka popisných statistik (FADN-dojivost.sta)
N=467(V seznamu záv. prom. nejsou ChD)

ÚČET	denní dojivost Průměr	denní dojivost N	denní dojivost Minimum	denní dojivost Maximum	denní dojivost Sm.odch.	denní dojivost Koef.prom.	denní dojivost Int. spolehl. -95.000%	denní dojivost Int. spolehl. +95.000%
P	18,73458	324	7,104284	29,08944	4,276687	22,82777	18,26716	19,20201
J	15,72954	143	5,964286	29,07895	5,181383	32,94046	14,87301	16,58607
Vš.skup.	17,81441	467	5,964286	29,08944	4,772919	26,79246	17,38040	18,24842

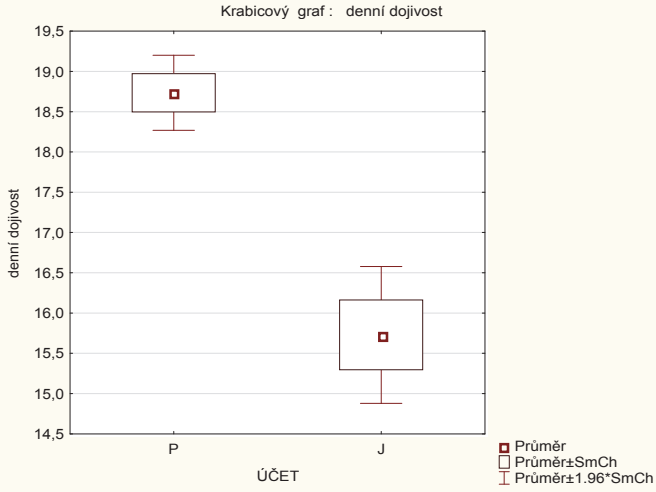
t-tests; grouped: legal form (FADN-milk yield sta.)
Group 1:P
Group2: J

Variable	Average P	Average J	t	sv	p	Number P	Number J	Standard deviation P	Standard deviation J	F-ratio Variance	p Variance
daily milk yield	18.73458	15.72954	6.546778	465	0.000000	324	143	4.276687	5.181383	1.467833	0.005577

t-testy; grupováno: ÚČET (FADN-dojivost.sta)
Skup. 1: P
Skup. 2: J

Proměnná	Průměr P	Průměr J	t	sv	p	Poč.plat P	Poč.plat. J	Sm.odch. P	Sm.odch. J	F-poměr Rozptyly	p Rozptyly
denní dojivost	18,73458	15,72954	6,546778	465	0,000000	324	143	4,276687	5,181383	1,467833	0,005577

* individual farms (J) a legal entities (P)



SUMMARY

Statistical significance of the results of analyses is being subjected recently to criticism on the part of methodologists, statisticians, as well as researchers themselves, in spite of their own remaining at a thoughtless way of its usage. Its purposeful application, anyway, is limited to representative samples obtained by random sampling methods. An important requirement of sampling surveys is just securing of sample representativeness. Representativeness as such is not a graded property, a sample cannot be „more representative" or „less representative".

The paper clarifies the concept of representativeness and it introduces statistical methods applied in the verification of representativeness of the random sample of agricultural enterprises in FADN CZ data network. The agricultural accountancy data network (FADN) plays the role of the principal EU information source on real economic situation of agricultural enterprises in all the EU member countries. It is a random sample of both physical and corporate bodies doing business in agricultural primary production, and very detailed data on economic results can be obtained using questionnaire surveys based on it. Also suggestions concerning methodology aimed at improving quality of the given random sample assessment are included into the analyses done.

The object of sample representativeness verification were the results of husbandry of agricultural enterprises from the FADN CZ sample in 2011, obtained in 2012. Testing of the FADN CZ sample survey representativeness was done in order to verify whether it was possible to draw generalized conclusions for assessment of the whole of agriculture on its basis. Data representing the whole population of farms were taken from Czech Statistical Office data bases and from the Report on the State of Czech Agriculture.The test farms sample survey representativeness verification was done for the data on hectare yields of wheat, barley, potatoes and rape, and also daily milk yields, daily weight increments in pigs and in beef cattle.The reason for selection of these data was their decisive share in the production structure of the whole of Czech agriculture. It was examined whether a difference exists between the value found in the test farms sample and the actual value in the whole population of farms.

The analysis confirmed sufficient quality of the FADN CZ 2011 test farms sample. The results obtained confirm that, the sample of test farms is representative. In many of the results tested the parameters obtained by the FADN CZ 2011 sample do not differ significantly from the actual parameters that could be obtained by an analysis of the whole population of farms. Some differences, mostly at regional level, were shown in some data only, first of all at physical entrepreneurial subjects; this is caused by a higher level of variability recorded in the data from this sample.

It is not needed to expand the random sample in order to further raise its quality. Anyway, it could be possible to reduce variability of some data removing their extreme values from the sample using special statistical diagnostic methods. But it is a sake of discussion whether this recommendation is not at odds with the researchers' intentions.

REFERENCES

BLAHUŠ, P., 2000: Statistická významnost proti vědecké průkaznosti výsledků výzkumu. *Česká kinantropologie.* 4, 2: 53–72. ISSN 1211-9261.

CARVER, R., 1994. The case against statistical significance testing, revisited. *Harvard Educational Review,* 61: 287–292. ISSN 0017-8055.

ČERMÁK, V., VRABEC, M., 1999: *Teorie výběrových šetření – část I.* Praha: VŠE, 177 p. ISBN 80-7079-191-8.

ČERMÁK, V., VRABEC, M., 1998: *Teorie výběrových šetření* – část II. Praha: VŠE, 79 p. ISBN 80-7079-609-X.

HEBÁK, P., HUSTOPECKÝ, J., PECÁKOVÁ, I., PRŮŠA, M., ŘEZANKOVÁ, H., SVOBODOVÁ, A., VLACH, P., 2007: *Vícerozměrné statistické metody [3]*. Praha: Informatorium, 271 p. ISBN 80-7333-039-3.

Kol., 2012: *Zemědělská účetní datová síť FADN CZ – Výběrové šetření hospodářských výsledků zemědělských podniků v síti FADN CZ za rok 2011* – Samostatná příloha ke Zprávě o stavu zemědělství ČR za rok 2011. Praha: ÚZEI Praha. ISBN 978-80-86671-96-3.

PALÁT, M., DVOŘÁKOVÁ, Š., KUPKOVÁ, N., 2012: Consumption of beef in the Czech Republic. *Agricultural Economics-Zemědělská ekonomika,* 58, 7: 308–314. ISSN 0139-570X.

Address

doc. Ing. Marie Prášilová, CSc., Ing. Radka Procházková, Ph.D, Ing. Martina Harvilíková, Department of Statistics,, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Praha 6 – Suchdol, Institute of Agricultural Economics and Information, Mánesova 1453/75, 120 00 Praha-Vinohrady, Czech Republic, e-mail: prasilova@pef.czu.cz, prochazkova@pef.czu.cz, harvilikova@fadn.cz