

## HODNOCENÍ REPREZENTATIVNOSTI VÝBĚROVÉHO ŠETŘENÍ

Marie Prášilová, Rudolf Zeipelt

katedra statistiky, PEF, Vysoká škola zemědělská,  
165 00 Praha 6 - Suchbátka, tel.(02) 3382235

### **Anotace :**

Příspěvek je výsledkem spolupráce katedry statistiky s VÚZE v Praze. Zabývá se ověřením reprezentativnosti výběrového šetření, které proběhlo v roce 1994 s cílem zjistit výsledky hospodaření v roce 1992 a 1993 ve všech formách právních subjektů v zemědělství. Ve výběrovém šetření bylo zahrnuto 581 respondentů, z toho 357 fyzických osob. K ověření reprezentativnosti byly použity testovací statistické postupy a ke zvýšení homogenity souboru byla k vyloučení extrémních hodnot použita explorační analýza. Ověřování prokázalo, že výběrové šetření je reprezentativní.

### **Summary :**

An article is a result of the co-operation between the Department of Statistics and the Research Institute of Agricultural Economics. There are verified the representation of the sampling survey with an aim to obtain the economical results in the years 1992 and 1993 for all types of legal subjects in agriculture. The sampling survey has 581 respondents ( 357 family farms ). Statistical testing methods were used for verifying of representation and exploratory data analysis were used for increasing of data homogeneity. The procedure indicated the sampling survey as a representative survey.

### **Klíčová slova :**

Výběrové šetření, reprezentativnost, explorační analýza.

### **Key words :**

Sampling survey, representation, exploratory data analysis.

V letošním roce bylo ve VÚZE v Praze ve spolupráci s územními odbory ministerstva zemědělství provedeno výběrové šetření u 581 respondentů .Výsledky tohoto šetření byly použity v tzv." zelené zprávě ", která analyzovala současný stav zemědělské výroby v České republice. Ve výběrovém šetření byly zahrnuty výsledky hospodaření za rok 1992 a 1993. Před využitím získaných výsledků bylo nutné ověřit, zda na jejich základě je možno činit obecně platné závěry.

## **Reprezentativnost výběrového souboru**

Ve výběrovém šetření bylo zahrnuto 162 zemědělských družstev, což činí 14,8 % zemědělských družstev v ČR po transformaci. Tento počet lze považovat za dostatečný; to bylo ověřeno též výpočtem minimálního nutného rozsahu výběrového souboru. Za poněkud problematickou lze považovat reprezentativnost zemědělských družstev dle výrobních oblastí (vzhledem k malému zastoupení výrobní oblasti kukuřičné a horské). Nedostatkem výběrového šetření je malé zastoupení řady produkčních okresů, což ovlivňuje vypovídací schopnost výběrového souboru z hlediska regionálního třídění.

Státní statky jsou ve výběrovém šetření zastoupeny pouze dvanácti respondenty. Vzhledem k tomu, že tento sektor prochází procesem privatizace, lze považovat tento počet za dostatečný z hlediska současného celkového počtu. Sektor státní statky není možné na základě tohoto šetření samostatně hodnotit. Ve výběrovém šetření bylo zahrnuto 50 respondentů zastupujících obchodní společnosti, to je 8,8 % základního souboru. Tento počet je na hranici minimálního počtu nutného rozsahu výběrového souboru. Vypovídací schopnost má výběrové šetření z hlediska obchodních společností celkem, ale ne již podle jednotlivých kategorií.

Fyzické osoby jsou ve výběrovém šetření zastoupeny 357 respondenty, což je pouhé jedno % ze základního souboru. Poněkud jinak situace vypadá, hodnotíme-li zvláště jednotlivé velikostní skupiny. Ve velikostní skupině 10 - 50 ha je 245 respondentů, což činí 8,8 %. Ve velikostní skupině 50 - 100 ha je 33 respondentů; tento počet odpovídá téměř 26 % ze základního souboru. Ve velikostní skupině nad 100 ha tvoří 33 respondentů více než 50 % základního souboru. Na základě uvedených údajů lze konstatovat, že výběrové šetření v sektoru fyzických osob je plně reprezentativní pro velikostní skupiny podniků nad 10 ha. Tyto velikostní skupiny zastupují více než 50 % výměry obhospodařované půdy fyzickými osobami. Z předchozího vyplývá, že výběrové šetření provedené v tomto rozsahu lze plně využít k formulaci závěrů zobecňujících výsledky na celý základní soubor. Dílčí nedostatky, které byly z hlediska vypovídací schopnosti výběrového souboru, lze překlenout podrobnějšími odhady provedenými na základě dalších známých údajů.

## **Statistická analýza výsledků výběrového šetření**

Pro další statistická šetření, založená na výběrových souborech, je užitečné získat představu o počtech a velikostech extrémních jednotek (zemědělských podniků), a to jak z hlediska výměry, produkce, tržeb a dalších objemových ukazatelů, tak i z hlediska hektarových výnosů, nákladů na 1 ha zem. půdy a tržeb na 1 ha, tedy ukazatelů intenzivních.

Explorační (průzkumová) analýza je název pro souhrn metod, sloužících k rychlému orientačnímu poznání charakteru variačních řad, m.j. i k detekci tzv. odlehlých (extrémních) pozorování. Základním nástrojem této analýzy je zpracování vstupních dat do tzv. krabičkového grafu (box-and-whisker plot), který jednoduše a velmi názorně uvádí rozložení pěti nejdůležitějších číselných hodnot variační řady : minima, maxima, mediánu a obou kvartilů. Námí použitý statistický programový paket STATGRAPHICS uvádí navíc i zmíněná odlehlá pozorování (extrémní hodnoty), a to v podobě tečky, popř. hvězdičky (jde-li o velmi odlehlé hodnoty).

Všeobecně platí, že ve větších souborech je odlehlých hodnot více , a to i relativně. Je tomu tak proto, že hodnoty, které se při malém počtu pozorování nedají (s danou pravděpodobností) označit jako odlehlé (extrémní), mohou se jako takové jevit (se stejnou pravděpodobností) v souborech rozsáhlejších.

Pomocí uvedené grafické metody byly stanoveny odlehlé hodnoty, které vlastně do souboru nepatří. Tyto hodnoty byly bez nebezpečí snížení vypovídací hodnoty ze souboru vyloučeny. Tím se zvýšila homogenita souboru a snížila se variabilita sledovaných znaků. Teprve takto upravené údaje byly zpracovány. Výsledky zpracování za některé vybrané ukazatele jsou uvedeny v tabulce č. 1 .

Pro další ověření reprezentativnosti výběr.šetření byly v případech, kde jsou k dispozici údaje za základní soubor (právnícké osoby), provedeny testy, pomocí kterých bylo zjištěno, zda existuje rozdíl mezi hodnotou zjištěnou z výběrového souboru a skutečnou hodnotou ze základního souboru (údaje z oficiálního statistického výkaznictví). K řešení byl použit test o hodnotě parametru  $\mu$ :

$$U = \frac{|\bar{x} - \mu_0|}{s} \cdot \sqrt{n}$$

- kde: U ... hodnota testovacího kriteriia  
 $\bar{x}$  ... aritmetický průměr hodnoceného ukazatele  
 $\mu_0$  ... údaj zjištěný se statistického výkaznictví  
s ... směrodatná odchylka  
n ... rozsah souboru

U většiny sledovaných ukazatelů test potvrdil nevýznamnost rozdílu výběrové charakteristiky s charakteristikou ze základního souboru. Tam kde byl prokázán statisticky významný rozdíl, byl soubor znovu podroben explorační analýze dat a ve většině případů po vyřazení jedné či dvou extrémních hodnot se variabilita snížila natolik, že se projevila opět statistická nevýznamnost rozdílu.

**Tabulka č. 1: Vybrané ukazatele výběrového šetření**

Komodita	ukazatel	jednotky	rozsah	průměr	maximum	minimum	standardní chyba	variační koef(%)
<b>Pšenice 92</b>	výnos/os.pl.	t/ha	222	4,66	7,09	2,45	0,0672	21,48
	prodej/os.pl.	t/ha	215	3,964	9,506	0,052	0,105	38,88
	tržby/os.pl.	Kčs/ha	218	9945,32	26796,70	145,214	287,289	42,65
<b>Brambory 92</b>	výn./os.pl.	t/ha	193	17,82	34,600	6,589	0,3782	29,49
	prodej/os.pl.	t/ha	181	14,185	38,025	0,423	0,459	43,53
	tržby/os.pl.	Kčs/ha	182	28044,0	74,330	46,059	1154,19	55,52
<b>Dojnice 92</b>	produkce/KD	l/KD	205	10,52	13,12	5,40	0,0945	12,87
	prodej/KD	l/KD	203	9,49	13,56	4,74	0,104	15,61
	tržby/KD	Ksč/KD	216	53,65	96,438	12,982	0,730	19,99
<b>Výkrm prasat 92</b>	produkce/KD	kg/KD	194	0,566	0,924	0,245	0,0076	18,64
	prodej/KD	kg/KD	195	0,654	1,350	0,0002	0,018	39,22
	tržby/KD	Kčs/KD	192	17,902	38,638	0,987	0,492	38,07
<b>Výkrm skotu 92</b>	produkce/KD	kg/KD	203	0,706	1,111	0,246	0,0094	19,08
	prodej/KD	kg/KD	197	1,285	3,506	0,0002	0,043	46,86
	tržba/KD	Kčs/KD	199	33,54	87,494	0,036	1,023	43,04
<b>Hosp. výsl. 92</b>	hosp.výsl.na obhosp.půdu	tis.Kč/ha	133	1,584	6,641	-9,734	0,211	-153,90

### Výsledky provedeného ověřování

Provedené testy potvrdily předpoklad, že na základě uskutečněného výběrového šetření je možno konstruovat kvalitní odhady pro zemědělství České republiky. Tyto odhady je možno chápat buď jako odhady bodové, kdy se k odhadu průměrné hodnoty použije přímo průměr získaný z výběrového šetření, nebo jako odhady intervalové. Intervalové odhady stanoví interval, ve kterém požadovaná střední hodnota leží s předem určenou (zpravidla 95 %) pravděpodobností. Jako příklad je uvedeno stanovení intervalu spolehlivosti pro denní dojivost dojnic. Byl vypočten 95 % interval spolehlivosti v rozmezí (10,33 - 10,70) l/KD. Tento výsledek lze považovat za velmi dobrý, neboť skutečně známá hodnota ze základního souboru je 10,38 l/KD.

Provedená statistická analýza výsledků výběrového šetření potvrzuje závěry, které byly stanoveny na základě počtu respondentů. Je tedy možné na základě provedeného výběrového šetření provést kvalitní odhady s tím, že při jejich konstrukci je nutno přihlídnout k dílčím nedostatkům, které toto výběrové šetření má.