FORUM STATISTICUM SLOVACUM Vol XVIII (2022) No 1 pp 59-68

Aplikace matematického programování při určení výběrové alokace v prostředí softwarového systému SAS
Application of mathematical programming in determining the sampling allocation in the SAS software system

Roman Pavelka
Štatistický úrad Slovenskej republiky, Odbor metód štatistických zisťovaní, Lamačská cesta 3, 840 05 Bratislava, Slovenská republika
Statistical Office of the Slovak Republic, Statistical Surveys and Methodology Department, Lamačská cesta 3, 840 05 Bratislava, Slovak Republic
roman.pavelka@statistics.sk

Abstrakt: *Matematické programování, zejména minimalizace nebo maximalizace funkcí, je ve statistických metodách všudypřítomná. Mnoho inferenčních metod je postaveno na principu maximální věrohodnosti, ve kterém předpokládaná funkce hustoty pravděpodobnosti nebo pravděpodobnostní funkce je optimalizována s ohledem na její parametry vzhledem k pozorovaným realizacím náhodné proměnné, jejíž rozdělení je popsáno pomocí funkce. Jiné metody inference zahrnují přizpůsobení modelu pozorovaným datům takovým způsobem, že odchylky pozorování od modelu jsou minimální. Mnoho důležitých statistických metod se používá před provedením závěrů o datech. V praxi se v ideálním případě uvažuje o statistických metodách dříve, než jsou shromážděny údaje. Například metody výběru vzorků nebo návrhu experimentu se stanoví tak, aby se maximalizovala hodnota dat pro vyvozování závěrů nejčastěji minimalizací rozptylu odhadů. V předkládaném článku bude uveden příklad optimální alokace výběrového souboru v programovacím prostředí SAS.*

Abstract: *Mathematical programming, especially function minimization or maximization, is ubiquitous in statistical methods. Many inference methods are based on the principle of maximum likelihood, in which the assumed probability density function or probability function is optimized with respect to its parameters with respect to the observed implementations of a random variable, the distribution of which is described by the function. Other methods of inference include adapting the model to the observed data in such a way that deviations of the observations from the model are minimal. Many important statistical methods are used before drawing conclusions about data. In practice, statistical methods are ideally considered before data are collected. For example, sampling methods or experimental design are determined to maximize the value of the data for drawing conclusions, most often by minimizing the variance of estimates. The present article will give an example of optimal sample allocation in the SAS programming environment.*

Klíčové slova: *matematické programování, optimalizace, SAS, výběrový soubor.*

Key words: *mathematical programming, optimization, SAS, sample set.*

1 Úvod

Nejdůležitější částí výzkumu sledovaného jevu je návrh výběrového šetření. Jestliže je výběrové šetření navrženo nesprávně, žádná další analýza nepřinese požadované informace a závěry. Návrh výběrového šetření zahrnuje především

metody kontroly výběrových a nevýběrových chyb. Jednoduchý náhodný výběr zahrnuje pouze jednu vlastnost návrhu: velikost vzorku. Pro stratifikovaný náhodný výběr je nutné definovat straty (vrstvy) a poté rozhodnout, kolik pozorování bude v každé jednotlivé stratě potřebné vybrat (alokovat) do výběrového souboru. Návrh výběrového šetření obecně a zejména návrh alokace výběrového souboru do zvolených strat (vrstev) může být považováno jako formulace matematického programování. Formulovat alokaci výběrového souboru jako problém matematického programování je zejména výhodné, pokud jsou na velikost výběrového souboru nebo na počty vybíraných jednotek v jednotlivých stratách (vrstvách) kladena určitá omezení (například výběr je limitován rozpočtem a maximálními náklady) při současném požadavku na přesnost odhadovaných výstupů. Tento článek představuje úvod do tohoto přístupu a snaží se přiblížit konkrétní aplikaci matematického programování na návrh výběrového šetření v prostředí programového systému SAS.

2 Návrh efektivního výběrového souboru

Statistici zabývající se návrhy výběrových šetření se věnují především navrhováním efektivních výběrových souborů (vzorků). Efektivita vzorku se měří buď náklady, nebo předpokládanou přesností výsledných odhadů. Konečný návrh vzorku často odráží snahu o dosažení určitého kompromisu mezi oběma požadavky. Prakticky pro každý návrh vzorku je k dispozici jednoznačné řešení, přinejmenším v kontextu jedné zájmové proměnné (Cochran, 1997, Hansen et al., 1993, Kish, 1965). Tato řešení buď zohledňují velikost vzorku potřebnou pro určitou přesnost odhadů, nebo udávají přesnost, kterou nabízí určitá velikost vzorku. V případě zahrnutí i dalších omezení (např. na velikost vzorku specifickou pro jednotlivé vrstvy) nebo návrh výběrového šetření pro současné odhady více požadovaných proměnných mohou tato řešení být komplikovanější anebo složitější.

Přesnost návrhu vzorku se obvykle měří odhady variancí odhadovaného statistického údaje. Variance odhadů jsou obvykle ovlivňovány stratifikací a shlukováním jednotek a efekty uvedených technik na jejich velikost jsou měřitelné nebo odhadnutelné. Rovněž měřitelné jsou varianční dopady dalších prvků návrhu, jako je např. odchylka od proporcionální alokace nebo zavedení rozdílných váhových efektů (Kish, 1992).

Pokud jsou k dispozici příslušné údaje, lze náklady na navrhovaný výběrový soubor měřit pomocí velikosti vzorku nebo pomocí nákladové funkce. Pokud jsou takové údaje dostupné spolu s rozptyly na úrovni strat (vrstev), lze použít vzorce pro optimální rozdělení (Cochran, 1977). Nedokonalost tohoto přístupu představuje skutečnost, že tyto vzorce v podstatě předpokládají jedinou

zájmovou proměnnou. Vzhledem ke skutečně vícerozměrné povaze většiny výběrových šetření, která mají za cíl současné odhady více ukazatelů, jsou pro složitější a efektivnější návrhy výběrových souborů nutné jiné techniky.

3 Výběrová alokace jako problém matematického programování

K řešení problému optimalizace návrhu výběrových souborů bylo navrženo matematické programování (Bethel,1989, Leaver et. al. 1997, Valliant a Gentle, 1997), které je vhodné i pro celou řadu dalších problémů vyskytujících se při navrhování výběrových šetření. Současné odhady více ukazatelů na základě jediného výběrového šetření za existence různých (a mnohdy protikladných) omezení znamená, že problém rozdělení počtů jednotek do strat (výběrová alokace) je podstatně složitější. V zásadě lze tyto cíle a omezení zohlednit pomocí technik matematického programování[9].

Matematické programování je obecný pojem, který se vztahuje k výběru nejlepšího řešení nějakého optimalizačního problému z dostupných alternativ. Formulace problému výběrové alokace spočívá v minimalizaci (nebo maximalizaci) určité funkce, která podléhá omezením týkajícím se nákladů, minimální velikosti výběrového souboru ve stratě (vrstvě), minimální velikosti výběrového souboru v doménách (subpopulacích) a výběrovou chybou (vyjádřenou zpravidla variačním koeficientem) odhadů ve stratách nebo doménách. Obecně má optimalizační problém při určení výběrové alokace následující části:

- Účelová (kriteriální) funkce - funkce jedné nebo více optimalizovaných veličin.
- Rozhodovací proměnné - veličiny upravované za účelem nalezení řešení, např. velikosti vzorků ve stratách (vrstvách),
- Parametry - fixní vstupy, s nimiž se zachází jako s konstantami, např. počty obyvatel ve vrstvách a rozptyly.
- Omezující podmínky - omezení rozhodovacích proměnných nebo jejich kombinací, např. velikosti domén a nákladů[10].

[9] Matematické programování se zabývá řešením optimalizačních úloh, ve kterých jde o nalezení maxima resp. minima předem definovaného kritéria (např. zisk, náklady, objem výroby, atd.) na množině všech možných přípustných variant dané úlohy. Prakticky to znamená, že hledáme extrém zmíněného kritéria při platnosti jistých omezujících podmínek. Dané kritérium je vyjádřené funkční závislostí (je to funkce více proměnných) a omezující podmínky jsou vyjádřeny soustavou rovnic či nerovnic. Úlohy matematického programování dělíme na lineární a nelineární.

[10] Optimalizační algoritmus se zaměřuje pouze na variabilní náklady spojený s velikostí výběrového souboru a jeho rozdělením do strat (alokací). Proto rozpočtové omezení je určeno pouze variabilními náklady.

Při návrhu výběrového šetření je možné vyjádřit účelovou (kriteriální) funkci, rozhodovací proměnné i ostatní části optimalizačního problému lineárními příp. nelineárními výrazy (Green, 2000). Problém matematického programování při návrhu výběrového šetření lze obecně vyjádřit následujícími vztahy.

Cílem návrhu výběrového šetření při lineárním programování je nalezení množiny velikostí vzorků (četností vybíraných jednotek) $\{n_h\}_{h=1}^{H}$ v *h*-té stratě, pro $h = 1, \cdots, H$ tak, aby se minimalizovala účelová (kriteriální) funkce Φ.

$$min(\Phi) = min\left[\sum_{j=1}^{J} \omega_j \cdot relvar\left(\hat{\bar{y}}_j\right)\right], \tag{1}$$

kde $\{\omega_j\}_{j=1}^{J}$ jsou váhy důležitosti[11] přiřazené *j*-té odhadované veličině $j = 1, \cdots, J$ a symbol $relvar\left(\hat{\bar{y}}_j\right)$ představuje tzv. relativní varianci *j*-tého odhadu $\hat{\bar{y}}_j$, pro kterou platí:

$$relvar\left(\hat{\bar{y}}_j\right) = CV^2\left(\hat{\bar{y}}_j\right) = \frac{V\left(\hat{\bar{y}}_j\right)}{\bar{y}_{U_j}^2}, \tag{2}$$

Označení $V\left(\hat{\bar{y}}_j\right)$ představuje varianci *j*-tého odhadu $\hat{\bar{y}}_j$, $CV\left(\hat{\bar{y}}_j\right)$ je variační koeficient *j*-tého odhadu $\hat{\bar{y}}_j$ a $\bar{y}_{U_j}$ populační průměr *j*-té odhadované veličiny y_j.

Hledání minima účelové (kriteriální) funkce (1) se musí realizovat vzhledem k následujícím omezením:

- Pro všechny straty (vrstvy) musí platit $n_h \leq N_h$, kde n_h je symbolem pro počet jednotek vybraných do vzorku ve stratě *h*, N_h je počet jednotek v *h*-té stratě populace pro $h = 1, \cdots, H$.
- Pro všechny straty (vrstvy) musí platit $n_h \geq n_{min}$, což vyjadřuje minimální počet vybraných jednotek do vzorku ve stratě *h*, pro $h = 1, \cdots, H$.
- Pro předem vybrané straty a odhadované veličiny musí platit $\left[CV\left(\bar{y}_{j,sh}\right)\right]^2 \leq \left[CV_{0jh}\right]^2$, tj. hodnota relativní variance výběrového průměru je menší než předem stanovená požadovaná $\left[CV_{0jh}\right]^2$. Relativní variance výběrového průměru $\bar{y}_{j,sh}$ *j*-té odhadované veličiny v *h*-té stratě je definována ve tvaru

$$\left[CV\left(\bar{y}_{j,sh}\right)\right]^2 = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \cdot \frac{S_h^2}{\bar{y}_{j,sh}^2}. \tag{3}$$

[11] I když suma vah důležitosti nemusí být rovná 1, jejich normalizace je vhodná, aby byly relativní velikosti vah dobře interpretovatelné. Mohou obsahovat také i nulové hodnoty, které označují veličinu „uvolněnou“ ze statistické inference na základě výběrového šetření (poznámka autora).

Symbol S_h^2 odpovídá populačnímu rozptylu *j*-té veličiny ve stratě *h*. Hodnoty populačního rozptylu S_h^2 se odhadují z předešlých výběrových šetření stejného anebo podobného typu.

- Pro předem vybrané domény a odhadované veličiny musí platit $\left[CV\left(\hat{\bar{y}}_{jd}\right)\right]^2 \leq \left[CV_{0jd}\right]^2$, tj. hodnota relativní variance odhadu populačního průměru domény je menší než předem stanovená požadovaná $\left[CV_{0jd}\right]^2$. Odhad doménového průměru $\hat{\bar{y}}_{jd}$ *j*-té odhadované veličiny v doméně *d* je definován ve tvaru:

$$\hat{\bar{y}}_{jd} = \frac{\sum_h \frac{N_h}{n_h} \cdot p_{dh} \cdot \bar{y}_{jd,sh}}{\sum_h \frac{N_h}{n_h} \cdot p_{dh}}, \tag{4}$$

  kde $\bar{y}_{jd,sh} = \sum_{i \in s_{dh}} \frac{y_{jhi}}{n_{dh}}$ je výběrový průměr proměnné *j* v doméně *d* v rámci straty *h* a $p_{dh} = \frac{n_{dh}}{n_h}$ je podíl jednotek ze straty *h* v doméně *d*.

- Výběrové šetření je také omezeno náklady, které nemohou překročit plánovaný rozpočet. Celkové náklady výběrového šetření *C* jsou složeny z nákladů fixních C_0 a z nákladů c_h na jednotku populace, tj. $C = C_0 + \sum_{h=1}^{H} n_h \cdot c_h$, kde *H* označuje počet strat (vrstev) ve sledované populaci.

Rozhodovacími proměnnými, které je třeba optimalizovat za účelem nalezení řešení, jsou četnosti jednotek vybraných do jednotlivých strat $\{n_h\}_{h=1}^{H}$.

Výše uvedený problém matematického programování lze řešit namísto relativních variancí použitím variačních koeficientů. Variační koeficienty jsou však nelineární funkcí rozptylu (odmocninou), čímž by vzrostla míra nelinearity při hledání optimálního řešení. Následkem toho by se zvýšila výpočetní náročnost řešení, případně by se podstatně ztížilo nalezení optimálního řešení.

4 Příklad optimalizace návrhu výběrového souboru v prostředí systému SAS

Vícekriteriální optimalizaci v systému SAS lze provádět pomocí programové procedury NLP (nelineární programování) nebo novější procedury OPTMODEL z programového modulu SAS/OR[12] v prostředí systému SAS (SAS Institute, 2020).

Starší procedura NLP nabízí sadu optimalizačních technik pro minimalizaci nebo maximalizaci spojité nelineární funkce *f(x)* o *n* rozhodovacích proměnných s okrajovými, obecnými lineárními a nelineárními omezujícími podmínkami rovnosti a nerovnosti. Procedura NLP podporuje řadu algoritmů pro řešení

[12] SAS/OR ve verzi 15.2 obsahuje procedury pro různé optimalizační postupy, řešitele a numerické algoritmy spolu s rozhraním pro SAS Simulation Studio. Pro další informace viz http://docume ntation.sas.com/doc/cs/pgmsascdc/9.4_3.4/ormpug/ormpug_whatsnew_sect001.htm

tohoto problému, které využívají speciální struktury na funkci *f(x)* a funkcích omezujících podmínek. Mimo obecných nelineárních algoritmů procedura využívá dva algoritmy speciálně navržené pro kvadratické optimalizační úlohy a další dva algoritmy pro efektivní řešení nelineárních úloh metodou nejmenších čtverců.

Novější procedura OPTMODEL zahrnuje nejmodernější technologie řešení pro několik tříd problémů matematického programování. Procedura OPTMODEL řeší problémy programování s omezujícími podmínkami, lineárního programování, smíšeného celočíselného programování, síťových algoritmů, obecného nelineárního programování, kvadratického programování a další optimalizační techniky. Možnosti procedury OPTMODEL při řešení problémů matematického programování jsou podobné jako u předchozí procedury NLP. Na rozdíl od předešlé procedury NLP procedura OPTMODEL využívá specifický modelovací jazyk, který lze velmi jednoduše využít k úspěšnému modelování problémů lineárního i nelineárního programování, a robustnější řešení složitých nelineárních problémů. Příkazy procedury OPTMODEL lze rozdělit do tří kategorií: příkaz procedury, deklarace a programovací příkazy. Příkaz procedury vyvolává vlastní proceduru a nastavuje počáteční hodnoty voleb. Deklarační příkazy deklarují součásti optimalizačního modelu (příkaz CON). Programovací příkazy čtou a zapisují data, volají řešitele (příkaz SOLVE) a tisknou výsledky optimalizačního procesu (příkaz PRINT). Tento specifický modelovací jazyk procedury OPTMODEL umožňuje jednoduchou transformaci slovního vyjádření optimalizačního problému do spustitelného programového kódu. Zejména z uvedených důvodů se procedura OPTMODEL stává vhodným nástrojem matematického programování při řešení komplexních problému alokace jednotek v rámci návrhu výběrových šetření.

Příklad: *Optimalizace vzorku podnikových jednotek pro výběrové šetření*

Předpokládejme, že chceme najít takové rozdělení jednotek výběrového souboru do strat (vrstev) při náhodném výběru bez vracení, které bude minimalizovat relativní varianci odhadovaných celkových výnosů, $\hat{T}_{rev} = \sum_h N_h \cdot \bar{y}_{sh}$, (kde $\bar{y}_{sh}$ je výběrový průměr v *h*-té stratě) při dodržení těchto omezení:

- Rozpočet na variabilní náklady činí *300 000 eur.*
- *CV ≤ 0,05* na odhadovaný celkový počet zaměstnanců,.
- V každém odvětví je vybráno nejméně *100* podniků, tj. $n_h \geq 100$.
- Počet podniků vybraných do vzorku v každé stratě (vrstvě) je menší než počet podniků v populaci, tj. $n_h \leq N_h$.
- *CV ≤ 0,03* na odhadovaném celkovém počtu provozoven, které žádají o daňovou úlevu na výzkum.

- *CV ≤ 0,03* na odhadovaném celkovém počtu podniků s pobočkami v zahraničí.

Tab. 2 Průměry, směrodatné odchylky a podíly populace podniků *(Zdroj: Valliant et. al. 2013)*

Strata h	Podnikový sektor	Počet jednotek N_h	c_h	Průměr		Směrodatná odchylka		Podíl jednotek	
				Výnosy (mil. eur)	Zaměstnanci	Výnosy (mil. eur)	Zaměstnanci	Daňový kredit	Zahraniční pobočka
1	Výroba	6 221	120	85	511	170,0	255,50	0,80	0,06
2	Maloobchod	11 738	80	11	21	8,8	5,25	0,20	0,03
3	Velkoobchod	4 333	80	23	70	23,0	35,00	0,50	0,03
4	Služby	22 809	90	17	32	25,5	32,00	0,30	0,21
5	Finance	5 467	150	126	157	315,0	471,00	0,90	0,77
Populace CELKEM		**50 568**							

Z předcházejících výběrových šetření, z daňových a účetních výkazů jsou známy výběrové průměry, počty zaměstnanců i směrodatné odchylky výnosů ve všech stratech. Z minulosti jsou známy i zájmové podíly podniků (viz tabulka 1).

Relativní variance odhadu úhrnu výnosů, která slouží jako účelová (kriteriální) funkce v příkladu optimalizace výběrového souboru, je definována ve tvaru:

$$relvar\left(\hat{\bar{y}}_j\right) = T_{rev}^{-2} \cdot \sum_h N_h \cdot \left(\frac{N_h}{n_h} - 1\right) \cdot S_h^2 \qquad (5)$$

Cílem matematického programování je nalezení takového rozdělení počtu jednotek výběrového souboru do straty *1* až *5*, který zajistí reprezentativní odhad úhrnu výnosů, počtu zaměstnanců a jednotlivých podílů za celou populaci při zadaných omezujících podmínkách.

Návrh výběrového souboru podnikových jednotek procedurou OPTMODEL:

Optimalizace návrhu výběrového souboru podle výše uvedeného zadání pomocí procedury OPTMODEL se realizuje pomocí následující příkazové syntaxe:

```
Title1 "Informace o výchozích hodnotách";
*************************************************************;
DATA work.Example;
        LENGTH Stratum 3 Nh UnitCost Revenue Emplyees Revnu_SD Emply_SD RCredit OffShore 8;
        LABEL Stratum="Stratum ID" Nh="Počty populace ve stratě"
                UnitCost="Náklady šetření na jednotku"
                Revenue="Průměrné výnosy (mil)" Emplyees="Průměrný počet zaměstnanců"
                Revnu_SD="Standardní odchylka výnosů (mil)"
                Emply_SD="Standardní odchylka počtu zaměstnanců"
                RCredit="Podíl jednotek s daňovými úlevami"
                OffShore="Podíl jednotek se zahraniční pobočkou";
        INPUT Stratum Nh UnitCost Revenue Emplyees Revnu_SD Emply_SD RCredit OffShore;
        CARDS;
1 6221 120 85 511 170.0 255.50 0.8 0.06
2 11738 80 11 21 8.8 5.25 0.2 0.03
3 4333 80 23 70 23.0 35.00 0.5 0.03
4 22809 90 17 32 25.5 32.00 0.3 0.21
5 5467 150 126 157 315.0 471.00 0.9 0.77;
```

```
RUN;

DATA work.Example;
        SET work.Example;
        ARRAY p_s RCredit OffShore;
        ARRAY sd_s RCrdt_SD OffSh_SD;
        DO OVER p_s;
                sd_s=SQRT(p_s * (1 - p_s) * Nh / (Nh - 1));
        END;
RUN;

************************************************************;
Title2 "Výběrová alokace - Řešení pomocí procedury OPTMODEL";
************************************************************;
PROC OPTMODEL;
        *_____ NAČTENÍ PARAMETRŮ _____*;
        *Populační počty po stratách;
        NUMBER Nh{1..5};
        READ DATA work.Example INTO [_n_] Nh;
        *Jednotkové náklady;
        NUMBER UnitCost{1..5};
        READ DATA work.Example INTO [_n_] UnitCost;
        *Populační průměry & standardní odchylky;
        NUMBER Revenue{1..5}, Emplyees{1..5}, RCredit{1..5}, OffShore{1..5},
                Revnu_SD{1..5}, Emply_SD{1..5}, RCrdt_SD{1..5}, OffSh_SD{1..5};
        READ DATA work.Example INTO [_n_] Revenue Emplyees RCredit OffShore
                Revnu_SD Emply_SD RCrdt_SD OffSh_SD;
        *_____ ROZHODOVACÍ PROMĚNNÉ _____*;
        *Velikost vzorků po stratách s přiřazením počátečních hodnot;
        VAR NSamp{i in 1..5} init 500;
        *_____ OMEZUJÍCÍ PODMÍNKY _____*;
        *Velikost vzorků po stratách >= 100, <= Velikost populace po stratách;
        CON SampSize{i in 1..5}: 100 <=NSamp[i] <=Nh[i];
        *Rozpočtové omezení;
        CON Budget: (SUM{i in 1..5} UnitCost[i] * NSamp[i]) <=300000;
        *Relativní variance pro průměrný počet zaměstnanců;
        CON RelVar1: (SUM{i in 1..5} Nh[i]*(Nh[i]/NSamp[i]-1)*Emply_SD[i]**2) /
                ((SUM{i in 1..5} Nh[i] * Emplyees[i])**2) <=(0.05**2);
        *Relativní variance pro podíl jednotek s daňovými úlevami;
        CON RelVar2: (SUM{i in 1..5} Nh[i]*(Nh[i]/NSamp[i]-1)*RCrdt_SD[i]**2) /
                ((SUM{i in 1..5} Nh[i] * RCredit[i])**2) <=(0.03**2);
        *Relativní variance pro podíl jednotek se zahraničními pobočkami;
        CON RelVar3: (SUM{i in 1..5} Nh[i]*(Nh[i]/NSamp[i]-1)*OffSh_SD[i]**2) /
                ((SUM{i in 1..5} Nh[i] * OffShore[i])**2) <=(0.03**2);
        *_____ ÚČELOVÁ (KRITERIÁLNÍ) FUNKCE _____*;
```

```
        MIN f=(SUM{i in 1..5} Nh[i] * (Nh[i]/NSamp[i] - 1) * Revnu_SD[i]**2) /
                ((SUM{i in 1..5} Nh[i] * Revenue[i])**2);
        *_____ ŘEŠENÍ _____*;
        SOLVE;
        ** NLP **;
        PRINT NSamp;
        PRINT (SUM{i in 1..5} NSamp[i]);
QUIT;
```

Procedura OPTMODEL dokáže sama rozpoznat třídu řešeného problému a příkazem SOLVE implicitně použít odpovídající algoritmus řešení. Výstupem z procedury OPTOMODEL je celkové shrnutí problému PROBLEM SUMMARY, shrnutí řešení SOLUTION SUMMARY a hodnoty optimalizovaných rozhodovacích veličin-alokace jednotek do strat (vrstev). Výstupy procedury jsou ilustrovány na obrázku 1.

Problem Summary	
Objective Sense	Minimization
Objective Function	f
Objective Type	Nonlinear
Number of Variables	5
Bounded Above	0
Bounded Below	0
Bounded Below and Above	0
Free	5
Fixed	0
Number of Constraints	9
Linear LE (<=)	1
Linear EQ (=)	0
Linear GE (>=)	0
Linear Range	5
Nonlinear LE (<=)	3
Nonlinear EQ (=)	0
Nonlinear GE (>=)	0
Nonlinear Range	0

Solution Summary	
Solver	NLP
Algorithm	Interior Point
Objective Function	f
Solution Status	Optimal
Objective Value	0.0021689216
Optimality Error	4.1118285E-7
Infeasibility	4.1118285E-7
Iterations	7
Presolve Time	0.00
Solution Time	0.01

[1]	NSamp
1	413.15
2	317.77
3	123.58
4	1396.32
5	596.30

Obrázek 1 Výstupy procedury OPTMODEL *(Zdroj: vlastní zpracování)*

Výstup procedury označený jako PROBLEM SUMMARY (na obrázku 1 levý sloupec hodnot) shrnuje zadání řešeného problému. Jsou zde uvedeny informace o účelové funkci, způsobu její optimalizace, zda je řešený problém lineární či nelineární, o počtu rozhodovacích proměnných a o počtu a typu omezovacích podmínek. Sloupec hodnot SOLUTION SUMMARY obsahuje název použitého řešitele včetně aplikovaného algoritmu, finální stav úlohy a minimalizovanou hodnotu účelové funkce v bodě optimality. Současně je zde uveden i počet iterací a optimalizační chyba. Sloupec hodnot z obrázku 1 na

pravé straně (nejmenší sloupec) obsahuje vektor hodnot rozhodovacích veličin, které byly optimalizovány podle zadání. Pokud tyto optimalizované hodnoty nejsou celá čísla, potom se zaokrouhlí na nejbližší vyšší celočíselnou hodnotu.

5 Závěr

Většina výběrových šetření, a to zejména výběrových šetření na populaci podniků, má za cíl současné odhady více ukazatelů. Návrhy výběrových souborů, které mají zajistit současné reprezentativní odhady více ukazatelů za existence omezujících podmínek, se proto stávají mnohem komplexnější a vyžadují sofistikovanější metody návrhu.

Jednou z možností, jak dosáhnout efektivní výběrový soubor pro vícerozměrná výběrová šetření, je přistupovat k návrhu výběrového souboru-především při stratifikovaném náhodném výběru-jako k problému matematického programování. Softwarový systém SAS ve svém programovém modulu SAS/OR nabízí celou sadu procedur, které umožňují řešení celé škály tříd problémů matematického programování. Pro optimální návrh výběrového souboru vícerozměrných šetření jsou nejhodnější procedury NLP a novější procedura OPTMODEL, která je ve srovnání s procedurou NLP robustnější a dovoluje používat komplexnější nelineární modely při řešení optimalizačních úloh.

6 Literatura

Bethel, J. (1989). Sample allocation in multivariate surveys. In: Survey Methodology, 15(1), s. 47-57.

Cochran, W. G. (1977). Sampling techniques. 3. vyd. New York, Wiley, 428 s.

Green, J., L. (2000). Mathematical programming for sample design and allocation problems. In: Proceedings of the Survey Research Methods Section, ASA, s. 688-692.

Hansen, M. et al. (1993). Sample survey methods and theory. Vol.1. New York: Wiley, 1016 s.

Kish, L. (1992). Weighting for unequal pi. In: Journal of Official Statistics, 8(2), s. 183-200.

Leaver, S. G., Johnson, W. H., Shoemaker, O. J. and Benson, T. S. (1999). Sample redesign for the introduction of the telephone point of purchase survey frames in the commodities and services component of the U.S. consumer price index. In: Proceedings of the Section on Government Statistics and Section on Social Statistics, s. 292-297.

SAS Institute Inc. 2020. SAS/OR® 15.2 User's guide: Mathematical programming. Cary (USA, NC): SAS Institute Inc.

Valliant, R., Gentle, J. E. (1997). An application of mathematical programming to sample Allocation. In: Computational Statistics & Data Analysis, 25, s. 337-360.

Valliant, R., Dever, J., A., Kreuter, F. (2013). Practical tools for designing and weighting survey Ssamples. New York: Springer, 776 s.