ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 228, 2009

*Janusz L. Wywiał**

ESTIMATION OF MEAN IN DOMAIN WHEN DISTRIBUTION OF VARIABLE IS SKEWED

Abstract. The problem of estimation the expected value in the case when a random variable has skewed probability distribution was considered e.g. by Carroll and Ruppert (1988), Chandra and Chambers (2006), Chen and Chen (1996), Karlberg (2000). Their results are based on transformation of skewed data. In the paper another approach is presented. The proposed estimators are constructed on the rather well known following property. Kendall and Stuart (1967) showed that the covariance between sample variance and sample mean is proportional to the third central moment of a variable. This property is applied to construction of several estimators of mean in a domain. The estimators are useful in the case when the variable under study has asymmetrical distribution because under some additional assumption they are more accurate than the sample mean. The results of the paper can be applied in survey sampling of economic populations.

Key words: small area sampling, skewnees coefficient, regression estimator, mean domain, relative efficiency.

I. RELATIONSHIP BETWEEN SAMPLE MEAN AND SAMPLE VARIANCE

Let $s = [Y_1, Y_2, ..., Y_n]$ be the simple sample from the distribution of a random variable Y. We assume that this random variable has at least six central moments which are denoted by $v_r(Y) = v_r = E(Y-E(Y))^r$, $r = 1, 2,...$ The mean value of Y will be denoted by $\mu = E(Y)$.

Let us consider the relationship between the sample mean $\overline{Y}_s$ and the sample variance denoted by $V_{2s} = \frac{1}{n-1}\sum_{i\in s}(Y_i - \overline{Y}_s)^2$ where $\overline{Y}_s = \frac{1}{n}\sum_{i\in s} Y_i$. Moreover, let us suppose that a distribution function of the variable Y is right skewed (positive asymmetric). Hence, the dominant of the distribution of the random variable Y is on the left from its expected value. So, we can expect, that the observation of sample spread around dominant are more frequent and they give rather smaller sample variance and sample mean than those in the sample spread about expected

*Professor, Department of Statistics, Katowice University of Economics.

value. So, in this case the positive relationship between the sample mean and the sample variance should be expected. In the case of left skewed distribution the expected value is on the left of the dominant. So, the small sample variance is related to rather large sample mean. The covariance between the sample mean and the sample variance is: $\text{Cov}(\overline{Y}_s, V_{2,s}) = \frac{1}{n} v_3 + O(n^{-2})$, see e.g. Kendall and Stuart (1967). Hence, we can consider the following linear relationship:

$$\overline{Y}_s \approx \alpha V_{2,s} + \beta \tag{1}$$

The criterion $E(\overline{Y}_s - \alpha V_{2,s} + \beta)^2 = \min$ leads to the following parameters α and β:

$$\alpha = \frac{\text{Cov}(\overline{Y}_s, V_{2,s})}{D^2(V_{2,s})}, \quad \beta = \mu - \alpha E(V_{2,s}) \tag{2}$$

It is well known (see, e.g. Cramér (1945) or Kendall and Stuard (1958)) that $E(V_{2,s}) = v_2 + O(n^{-1})$, $D^2(V_{2,s}) = \frac{1}{n}(v_4 - v_2^2) + O(n^{-2})$, So, the equations (2) can be rewritten in the following way:

$$\alpha = \frac{v_3}{v_4 - v_2^2} = \kappa \sqrt{\frac{v_4 - v_2^2}{v_2}}, \quad \beta = \mu - \alpha v_2 \tag{3}$$

where

$$\kappa = \frac{v_3}{\sqrt{v_2(v_4 - v_2^2)}}, \; -1 \le \kappa \le 1 \tag{4}$$

is the coefficient of skewnees (asymmetry) of a random variable distribution. In the case of symmetric distribution the normalized coefficient of skewnees $\kappa = 0$ in the case of right (left) asymmetric distribution $\kappa \ge 0$ $(\kappa \le 0)$. Moreover, let us note that the coefficient κ can be evaluated as the correlation coefficient of the random variables Y and $(Y - E(Y))^2$, see Wywiał (1981, 1983).

Example 1. Let us consider the sample drawn from the exponential distribution with the density function $f(y) = \lambda \exp(-\lambda y)$ for $y \ge 0$ and $f(y) = 0$ for $y < 0$. The moments of this distributions are: $E(Y) = \lambda^{-1}$, $v_2 = \lambda^{-2}$, $v_3 = 2\lambda^{-3}$,

$v_4 = 9\lambda^{-4}$, So, the given by the expression (4) coefficient κ takes the value $\kappa = \frac{1}{\sqrt{2}} \approx 0.7071$.

Example 2. Let us consider the sample drawn from the Pareto distribution with the density function $f(y) = \beta\alpha^{\beta} y^{-\beta-1}$ for $y \geq \alpha$ and $f(y) = 0$ $y < \alpha$. The shape parameter is denoted by β and the parameter of location and scale is denoted by α. The moments of this distributions are: $E(Y) = \frac{\alpha\beta}{\beta-1}$ for $\beta > 1$, $v_2 = \frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$ for $\beta > 2$, $v_3 = \frac{2\alpha^3\beta(\beta+1)}{(\beta-1)^3(\beta-2)(\beta-3)}$ for $\beta > 3$ and $v_3 = \frac{6\alpha^4\beta(\beta^3+\beta^2-6\beta-2)}{(\beta-1)^4(\beta-2)^2(\beta-3)(\beta-4)}$ for $\beta > 4$. The given by the expression (4) coefficient κ takes the value $\kappa = \sqrt{\frac{(\beta-2)(\beta-4)}{(\beta-3)(5\beta^3+13\beta^2-48\beta-12)}}$. So, the coefficient κ do not depend on the location and scale parameter denoted by α. If $\beta = 5$, $\kappa \approx 0.5563$. When $\beta \to \infty$, $\kappa = \frac{2}{\sqrt{5}} \approx 0.8944$.

II. ESTIMATION OF MEAN IN THE CASE OF KNOWN VARIANCE

As it is well known the sample mean $\overline{Y}_s = \frac{1}{n}\sum_{i\in s} Y_i$ is unbiased estimator of the expected value $\mu = E(Y)$ and its variance is: $D^2(\overline{Y}_s) = v_2 = \frac{v_2}{n}$. Below we consider estimators involving the above properties of relationship between sample mean and sample variance.

The sample moments defines the following expression:

$$V_{r,s} = \frac{1}{n-1}\sum_{i\in s}(Y_i - \overline{Y}_s)^r,\ r = 2, 3, \ldots \tag{5}$$

Under the assumption that the variance v_2 is know the following estimator based on the egression relationship, given by the expression (1), can be proposed:

$$\hat{Y}_{1s} = \overline{Y}_s + \alpha_s(v_2 - V_{2,s}) \tag{6}$$

where;

$$\alpha_s = \frac{V_{3,s}}{V_{4,s} - V_{2,s}^2} \tag{7}$$

It is the regression type estimator which is well known in survey sampling, see e.g. Cochran (1963).

On the basis of well known approximation method of deriving the variance of moment functions (see e.g. Cramér (1945) or Kendall and Stuart (1958)) we have the following.

$$D^2(\hat{Y}_{1,s}) \approx D^2(\overline{Y}_s) + \frac{v_3^2}{(v_4 - v_2^2)^2} D^2(V_{2,s}) - \frac{2v_3}{v_4 - v_2^2} Cov(\hat{Y}_{1,s}, V_{2,s}) = \frac{v_2}{n} - \frac{v_3^2}{(v_4 - v_2^2)n}$$

Result 1. If $n \to \infty$, then $\hat{Y}_{1,s} \sim N(\mu, D^2(\hat{Y}_{1,s}))$ where

$$D^2(\hat{Y}_{1,s}) \approx \frac{v_2}{n}(1 - \kappa^2). \tag{8}$$

If $\kappa \neq 0$ then $D^2(\overline{Y}_s) - D^2(\hat{Y}_{1,s}) > 0$.

The Result 1 was derived on the basis of the well known theorems about asymptotic probability distribution of a moment function (see e.g. Cramér (1945) or Rao (1965)).

Hence, in the case of asymmetric distribution and sufficiently large sample size the statistic $\hat{Y}_{1,s}$ is asymptotically unbiased estimator of the mean μ and it is not less precise than the sample mean $\overline{Y}_s$. The estimator $\hat{Y}_{1,s}$ is better than $\overline{Y}_s$ because the former is the function of auxiliary information which results from relationship between the sample mean and the sample variance.

The well known interpretation of the regression type estimator lead to conclusion that if the sample variance $V_{2,s}$ is not equal to the population variance v_2 then the estimator $\hat{Y}_{1,s}$ add the correction $\alpha_s(v_2 - V_{2,s})$ to the sample mean $\overline{Y}_s$.

Example 3. Similarly like in the example 1 we consider the sample drawn from exponential distribution. It is well known that in this case the best estimator of the parameter $E(Y)$ is the sample mean $\overline{Y}_s$. The results of the example 1 and expressions (8) and (4) lead to the conclusion that for large sample size $D^2(\hat{Y}_{1,s}) \approx \frac{1}{2n\lambda^2} = \frac{1}{2} D^2(\overline{Y}_s)$. Hence, in this case the estimator $\hat{Y}_{1,s}$ is two times more precise than the sample mean $\overline{Y}_s$ because the probability distribution of $\hat{Y}_{1,s}$ depend on the variance v_2 which is the auxiliary information.

Example 4. In the case of the Pareto probability distribution considered in the Example 2 we have that $D^2(\hat{Y}_{1,s}) = 0.208 D^2(\overline{Y}_s)$. So, when the large sample is drawn from Pareto population the estimator $\hat{Y}_{1,s}$ is more accurate than the sample mean $\overline{Y}_s$.

The estimator $\hat{Y}_{1,s}$ is not useful in practice because usually the variance v_2 is not known. But its modified version proposed in the next chapter can be valuable in practical statistical research.

III. ESTIMATION OF MEAN VALUE IN DOMAIN

Let a population U consists of D such non-empty and disjoint domains U_k, $k = 1,\ldots, D$. We will consider the following model. Random variables Y_i, $i = 1,\ldots, N$, have finite moments of at least sixth order. Let $E(Y_i) = \mu_k$ if $i \in U_k$. We assume that the variance of the all variables are the same, so $D^2(Y_i) = v_2$, $I = 1,\ldots,N$. The central moments of the variables are denoted by $v_{r,k} = E(Y_i - \mu_k)^2$ for all $i \in U_k$, $k = 1,\ldots, D$ and $r = 3, 4,\ldots$

Let s_k be simple sample drawn from k-th domain. The sample s of size n consists of all subsamples s_k, $k = 1,\ldots, D$. The size of s_k is denoted by n_k, so $n = \sum_{k=1}^{D} n_k$. We assume that $n_k > 1$ for $k = 1,\ldots, D$. The variance v_2 can be estimated by following statistics.

$$V_{2,s_k} = \frac{1}{n_k - 1}\sum_{i\in s_k}(Y_i - \overline{Y}_{s_k})^2, \quad \overline{Y}_{s_k} = \frac{1}{n_k}\sum_{i\in s_k} Y_i, \quad k=1,\ldots,D, \tag{9}$$

$$\overline{V}_{2,\sim s_d} = \frac{1}{D-1}\sum_{i=}^{D} V_{2,s_k}, \quad \overline{V}_{2,s} = \frac{1}{D}\sum_{k=1}^{D} V_{2,s_k}, \tag{10}$$

Under the stated assumptions all of them are unbiased estimators of the variance v_2.

Let us suppose that the distribution function of a variable in a k-the domain is skewed. In order to estimate its expected value we consider the following regression type estimator:

$$\hat{Y}_{2,s_d} = \overline{Y}_{s_d} + \alpha_{s_d}(\overline{V}_{2,\sim s_d} - V_{2,s_d}) \tag{11}$$

where the statistic $\overline{V}_{2,\sim s_d}$ is given by the expression (10) and

$$\alpha_{s_d} = \frac{V_{3,s_d}}{V_{4,s_d} - V_{2,s_d}^2}. \tag{12}$$

Similarly like the Result 1 we can derive the following.

Result 2. If $n \to \infty$ for k = 1,..., D then $\hat{Y}_{2,s_d} \sim N\left(\mu_d, D^2\left(\hat{Y}_{2,s_d}\right)\right)$ where

$$D^2\left(\hat{Y}_{2,s_d}\right) \approx D^2\left(\hat{Y}_{1,s_d}\right) + \frac{\nu_2 \kappa_d^2}{(D-1)\overline{n}_{H,D-1}} \geq D^2\left(\hat{Y}_{1,s_d}\right) + \frac{\nu_2 \kappa_d^2}{n - n_d} \tag{13}$$

where

$$D^2\left(\hat{Y}_{1,s_d}\right) \approx \frac{\nu_2}{n}\left(1 - \kappa_d^2\right), \tag{14}$$

$$\overline{n}_{H,D-1} = \frac{D-1}{\sum_{k=1,k\neq d}^{D} \frac{1}{n_k}} \leq \overline{n}_{D-1} = \frac{1}{D-1} \sum_{k=1,k\neq d}^{D} n_k = \frac{n - n_d}{D-1}$$

is the harmonic mean of the subsample sizes selected from domains.

The next estimator is as follows

$$\hat{Y}_{3,s_d} = \overline{Y}_{s_d} + \alpha_{s_d}(\overline{V}_{2,s} - V_{2,s_d}) \tag{15}$$

where the statistic $\overline{V}_{2,s}$ is defined in the expression (10).

Result 3. If $n \to \infty$ for k=1,...,D then $\hat{Y}_{3,s_d} \sim N\left(\mu_d, D^2\left(\hat{Y}_{3,s_d}\right)\right)$ where

$$D^2\left(\hat{Y}_{3,s_d}\right) \approx D^2\left(\hat{Y}_{1,s_d}\right) + \frac{\nu_2 \kappa_d^2}{D\overline{n}_{H,D}} \geq D^2\left(\hat{Y}_{1,s_d}\right) + \frac{\nu_2 \kappa_d^2}{n} \tag{13}$$

where

$$\bar{n}_{H,D} = \frac{D-1}{\sum_{k=1}^{D}\frac{1}{n_k}} \le \bar{n}_D = \frac{1}{D}\sum_{k=1}^{D} n_k = \frac{n}{D}.$$

is the harmonic mean of the subsample sizes selected from domains.

IV. SIMULATION ANALYSIS

The variances of the defined estimators of the mean value in a domain are derived approximately and they are valid only in the case of large sample. In order to study their accuracy in the case of small or moderate sample, a simulation analysis is developed.

We assume that in a d-th domain random variables have the two-parameter exponential density function $f_d(y_d)=\lambda\exp\{\lambda(y_d-\gamma_d)\}$ and $E(Y_d)=1/\lambda+\gamma_d=\mu_d$, $D^2(Y_d)=v_d=1/\lambda^2$. In the d-th domain the variable has appropriately shifted exponential distribution. The variances of those distributions are the same. This and the properties of the considered estimators lead to the conclusion that it is sufficiently to consider only the one-parameter exponential distribution (without the shift parameter) in order to study the accuracy of the estimation of the mean in a particular domain. Similarly we are going to consider the introduced earlier the Pareto probability distribution.

The appropriate samples s_k were generated 10000 times according to exponential probability distribution function. Such samples are denoted by $s_{d,h}$. Next the values of the considered estimators $\hat{Y}_{i,s_d}$, $i=1, 2$, were calculated. The variance $D^2(\hat{Y}_{i,s_d})$ was assessed by means of $\hat{D}^2(\hat{Y}_{i,s_d})=0.0001\sum_{h=1}^{10000}\left(\hat{Y}_{i,s_{d,h}}-\hat{E}(\hat{Y}_{i,s_d})\right)^2$ where $E(\hat{Y}_{i,s_d})=0.0001\sum_{h=1}^{10000}\hat{Y}_{i,s_{d,h}}$. The accuracy of the estimators are compared with the ordinary extension estimator by means of the coefficient $deff(\hat{Y}_{i,s_d})=100\hat{D}^2(\hat{Y}_{i,s_d})/D^2(\bar{Y}_{s_d})$ where $D^2(\bar{Y}_{s_d})=v_2/n_d$. Moreover, the relative bias is calculated on the basis of the expression $rb(\hat{Y}_{i,s_d})=100(\hat{E}(\hat{Y}_{i,s_d})/E(Y)-1)$.

Table 1. Values of the coefficient $def(\hat{Y}_{1,s})$ and $rb(\hat{Y}_{1,s})$.
The sample drawn from the exponential distribution

λ	0.01		0.05		0.10		1.00		10.00		20.00		100.00	
n	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*Rb*	*def*	*rb*	*def*	*rb*
20	79	9	7	8	78	8	77	8	79	8	77	8	77	9
30	68	6	70	6	69	6	70	6	68	6	70	6	67	6
50	62	4	61	4	63	4	62	4	63	4	64	4	64	4
100	57	2	57	2	56	2	57	2	54	2	57	2	58	2
300	54	1	53	1	54	1	53	1	53	1	53	1	52	1
500	52	0	51	0	53	0	51	0	51	0	52	0	50	0

In general, the analysis of Table 1 leads to two following conclusion. The accuracy of the estimator $\hat{Y}_{1,s}$ increases rather slowly with an increasing sample size, although for sample size 50 its mean square error is about 63 % of the variance of the simple sample mean $\bar{Y}_s$. Let us remind (see the Example 2) that $def(\hat{Y}_{1,s}) \geq 50\%$. The relative bias is rather large for small sample sizes but it decreases when the sample size increases. The relative bias is not large than 10%. The next conclusion is that under the fixed sample size the mean square error of the estimator $\hat{Y}_{1,s}$ do not change significantly with changing value of a parameter λ. So, in the next simulation studies we consider only the exponential distribution for the fixed parameter λ.

Table 2 leads to the similar conclusions like above ones, but in this case the accuracy of the estimator increases when the parameter θ of the Pareto distribution increases. The relative bias of the estimator is not larger than 1%.

Table 2. Values of the coefficient $def(\hat{Y}_{1,s})$.
The sample drawn from the Pareto distribution

n	θ=5	θ=10	θ=15	θ=20	θ=30	θ=50	θ=100
20	237	125	106	95	89	84	81
30	188	104	87	82	78	75	72
50	149	84	77	72	68	64	64
100	109	71	65	62	58	58	58
300	81	59	56	56	54	53	52
500	75	57	54	53	53	52	51

Table 3. Values of the coefficient *def* and rb (relative bias) for the estimator $\hat{Y}_{2,s_d}$ for equal-size samples drawn from domains. The exponential distribution with $\lambda = 1$

n_i	15		20		30		50		100		200	
D	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*Rb*
3	203	7	172	7	150	7	125	5	106	4	94	2
5	136	9	117	10	104	8	94	7	83	4	73	3
10	103	12	89	13	82	13	75	8	66	6	63	3
20	87	15	78	15	73	14	68	10	61	6	58	4
30	84	15	74	16	71	15	65	11	60	6	56	4

In the case when the sample (drawn from domains) sizes are the same the Table 3 leads to conclusion that the estimator $\hat{Y}_{2,s_d}$ is at least 20 % accurate as $\overline{Y}_{s_d}$ for number of domains at least equal 20 and size of the sample equal or grater 20 or for D≥10 and $n_i \geq 50$. The relative bias increases when number of domains increases but it decreases if sample size increases. In the case of rater large sample size and number of domains the bias can be neglected. Hence, it seems that the proposed estimator is quite good for not too large size of the samples and number of domains.

Table 4. Values of the coefficient *def* for the estimator $\hat{Y}_{2,s_d}$ for equal-size samples drawn from domains. The Pareto distribution

n_i:	30			50			100			200		
$D\backslash\theta$	20	50	100	20	50	100	20	50	100	20	50	100
5	143	117	110	116	101	96	93	86	82	79	75	74
10	103	88	86	90	78	78	75	69	69	67	65	65
20	87	78	73	73	71	68	68	63	62	63	59	58
30	84	74	74	76	68	66	67	66	63	58	58	58

The relative bias of the estimator $\hat{Y}_{2,s_d}$ was not greater than 1% in the case of the Pareto distribution for the considered parameters in the Table 4. Similarly like in the case of the Table 2 the accuracy of the estimator increases when the value of the parameter θ increases. Moreover the Tables 3 and 4 lead to conclusions that for the same sizes of domains and samples the accuracy of the estimation based on the sample drawn from exponential distribution is better than in the casee of Pareto distribution for $20 \leq \theta \leq 100$.

Table 5. Values of the coefficient *def* and *rb* (relative bias) for the estimator $\hat{Y}_{2,s_d}$ for samples drawn from domain. The case of the exponential distribution with λ=1

	$n_d=n_1$				$n_d=n_{(D+1)/2}$				$n_d=n_D$			
	n_k=50*k*		n_k=100*k*		n_k=50*k*		n_k=100*k*		n_k=50*k*		n_k=100*k*	
D	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*	*def*	*rb*
3	90	3.9	77	1.7	125	1.8	106	1.0	151	1.2	133	1.0
5	73	3.6	66	2.1	84	1.2	82	0.7	97	0.6	93	0.3
11	65	3.6	60	2.1	62	0.7	61	0.4	71	0.3	69	0.1

Table 6. Values of the coefficient *def* for the estimator $\hat{Y}_{2,s_d}$ for samples drawn from domain. The case of the Pareto distribution

	$n_d=n_1$				$n_d=n_{(D+1)/2}$				$n_d=n_D$			
	n_k=50*k*		n_k=100*k*		n_k=50*k*		n_k=100*k*		n_k=50*k*		n_k=100*k*	
$D\backslash\theta$	20	100	20	100	20	100	20	100	20	100	20	100
3	109	94	90	78	139	128	121	109	186	155	154	138
5	86	77	74	66	97	90	87	81	115	102	103	95
11	76	70	65	60	70	66	66	65	75	71	73	70

An analysis of the Table 5 leads to main conclusion that the accuracy of the estimator $\hat{Y}_{2,s_d}$ is highest when the sample size n_d of the domain U_d is the smallest one among the domains sample sizes. The bias of the estimator is small because its relative level is less than 4%. Moreover, the mean square error decreases when the number of domains or the sample sizes of all domains increase.

In the case of the Pareto distribution Table 6 leads to the similar conclussions like those evaluated during analysis of the Table 5. Let us not that in this case the relative bias of the estimator $\hat{Y}_{2,s_d}$ was not greater than 1% for the considered parameters.

V. CONCLUSIONS

The proposed estimator $\hat{Y}_{2,s_d}$ of mean value of the skewed random variable is more accurate than the sample mean but rather not useful in practice. But sometimes in special cases the value of the population variance can be assessed on the basis of census survey. For instance in the case of some economical variables we can expect that such assessed value of the variance can be almost the same in some period after the census survey.

The properties of the proposed estimator $\hat{Y}_{2,s_d}$ under some additional assumptions stated on value of variance of a variable under study let improve precision of estimation of mean value in a domain. The proposed estimator deal with situation when distribution of the variable in the domain is skewed. The simulation analysis leads to the conclusion that the estimator can be useful in the case of rather not too large sizes of the samples and quite large number of domains. We can expect that the estimator $\hat{Y}_{2,s_d}$ has the similar properties as estimator $\hat{Y}_{2,s_d}$.

The accuracy studies should be continued in the next papers. Some other theoretical distributions should be taken into account as well as e.g. a problem of stability of the variance in domains.

REFERENCES

Carroll R., Ruppert D. (1988), *Transformation and Weighting in Regression.* Chapman and Hall, New York.

Chandra H., Chambers R. (2006), Small area estimation with skewed data. *Southampton Statistical Sciences Research Institute Methodology Working Papers,* M06/05, University of Southampton, U.K.

Chen G., Chen J. (1996), A transformation method for finite population sampling calibrated with empirical likelihood. *Survey Methodology*, 22, pp.139–146.

Cochran W.G. (1963), *Sampling Techniques*. John Wiley & Sons, New York.

Cramér, H. (1945), *Mathematical Methods of Statistics*. Uppsala: Almqvist and Wiksells.

Karlberg F. (2000), Population total prediction under a lognormal superpopulation model. *Metron*, pp. 53–80.

Kendall, M. G. Stuart, A. (1958), *The Advanced Theory of Statistics. Vol.1 Distribution Theory*. Charles Griffin and Company Limited, London.

Kendall, M. G. Stuart, A. (1967), *The Advanced Theory of Statistics. Vol. 2 Inference and Relationship*. Charles Griffin and Company Limited, London.

Rao, C.R. (1965), *Linear Statistical Inference and Its Applications*. John Wiley and Sons, New York, London, Sydney, Toronto.

Wywiał, J. L. (1981), On some normalized coefficients of asymmetry and kurtosis of the random variable distribution (in Polish) . *Przegląd Statystyczny*, 28, 263–269.

Wywiał, J. L. (1983), Normalized coefficients of deviation from multi-normal distribution (in Polish) . *Przegląd Statystyczny*, 30, 77–86.

Janusz L. Wywiał

ESTYMACJA ŚREDNIEJ ZMIENNEJ O ROZKŁADZIE ASYMETRYCZNYM W DOMENIE

Rozważana jest nadpopulacja w której wyróżniono domeny badań. Celem wnioskowania jest estymacja wartości średniej w wyróżnionej domenie. Zakłada się, że rozkład prawdopodobieństwa zmiennych w domenach może być nawet silnie asymetryczny, jednocześnie przyjmując, że wszystkie zmienne tworzące model nadpopulacji mają tę samą wariancję. Pozwala to na konstrukcję specyficznego estymatora typu regresyjnego średniej w wyróżnionej domenie. Korzysta się przy tym ze znanego faktu, że kowariancja średniej z próby i wariancji z próby jest proporcjonalna do trzeciego momentu centralnego zmiennej. Okazuje się, że proponowany estymator może dawać dokładniejsze oceny średniej w domenie, gdy właśnie rozkład zmiennej jest asymetryczny. Wykazano to na podstawie odpowiednio zaprojektowanych i przeprowadzonych badań symulacyjnych.