Computing the Average Body Mass Index: A Study with Systematic Sampling Using Auxiliary Information

Gajendra K. Vishwakarma
Indian Institute of Technology Dhanbad
Neha Singh
Indian Institute of Technology Dhanbad
Surendra Pal Singh (surendra.geomatics@gmail.com)
Wollega University https://orcid.org/0000-0002-6949-4614

Research article

Keywords: Systematic sampling, Body mass index, Body circumference, Observational error

Posted Date: August 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-50197/v1

Version of Record: A version of this preprint was published at Mathematical Problems in Engineering on January 19th, 2022. See the published version at https://doi.org/10.1155/2022/9291313.

Computing the average body mass index: A study with systematic sampling using auxiliary information

Gajendra K. Vishwakarma[1], Neha Singh[1] and Surendra Pal Singh[2*]

Abstract

Background: The use of body mass index (BMI) could lead to over/under estimation of fat mass percentage. Systematic sampling is to be applied only if the given population is logically homogeneous, because systematic sample units are uniformly distributed over the population. The method of estimation for mean of the study variable under systematic sampling using auxiliary information has been proposed to estimate the body mass index (BMI).

Methods: The measures of different body parts are taken as auxiliary variables. The observation available on different body parts are assumed to be recorded with observational error. Thus we also propose method of estimation for mean in the presence of observational error. Numerical study has been done to reveal the efficacy of the proposed procedure for estimation of mean. Simulation study has also been done to demonstrate the effect of observational error on the estimation of body mass index.

Results: The properties of the proposed estimation method have been derived under large sampling approximation and obtained the conditions under which proposed method are more efficient.

Conclusions: The study provides an easy approach and simplest way to obtain the BMI estimate with and without observational error. Thus the suggested method may be used by statistician for this problem and for many others similar problem in the estimation of mean.

Keywords: Systematic sampling, Body mass index**,** Body circumference, Observational error.

[*] Correspondence: surendra.geomatics@gmail.com

[2]College of Engineering & Technology, Wollega University, Ethiopia

Full list of author information is available at the end of the article

1. Background

In the survey, it may often happen that the data are observed with some error and it is termed as measurement error or observational error. It is defined as the discrepancy between observed value and true value of the sample. [1, 2] have discussed real-life situations when data are obtained with errors. [3, 4, 5, 6] have discussed the observational error in the context of linear and non-linear regression models. [7, 8, 9, 10, 11] have studied the observational error in the estimation of ratio, product and regression methods of estimation.

A study was done by [12] to derive a prediction equation for body fat percentage in men (n = 252, age 22-81 years) from simple body measurements. Body density determined by underwater weighing and body fat percentage was determined from [13] equation. The data set includes the following variables [14, pp. 45-48], for observational techniques: density determined from underwater weighing, percent body-fat from Siri's equation, age in years, weight in lbs, height in inches and circumference of neck, chest, abdomen, hip, thigh, knee, ankle, bicep, arm, wrist in centimetre. In this article, we approach a different method rather than multiple regression method by [15] and method by [12]. [16] estimate optimal sample size by using body mass index for a dietetic supplement. We attempt to estimate for the body mass index in place of body-fat by using one of the auxiliary variables. The hip, thigh, knee, ankle, bicep, arm, wrist circumference can be taken as single auxiliary variable to estimate the body mass index. The correlation coefficient for each auxiliary variable has been obtained. If the data are systematically distributed the systematic sampling has nice features of selecting every kth element by choosing first element arbitrary. Many authors such as [17, 18, 19, 20, 21] have done pioneered work using systematic sampling at the estimation stage. [22] and [23] found for certain natural population like forest areas, the estimation of volume of timber is convenient by using systematic sampling. Use of auxiliary variable is prevalent as ratio, product and regression estimator. In the case of estimating the volume of timber, [24] proposed ratio estimator under systematic sampling suggested the leaf area or the girth of the tree may be taken as the auxiliary variable and. [25] proposed product estimators in the context of systematic sampling.

Suppose, the population consists of N units $u=(u_1,u_2,...,u_N)$ from a finite population. The population size is divided into k intervals such that $N=nk$. To select a sample, the first unit is selected at random from the first k units. This sampling method is similar to that of selecting a cluster at random out of k cluster (each cluster containing n units), made such that ith cluster contains serially numbered units $i, i+k, i+2k,..., i+(n-1)k$. After sampling of n units, we observe both the study and auxiliary variables. In this article we consider a situation where each data value may be observed with error. In order to compute the effect of observational error, it is assumed that (x_{ij}, y_{ij}) are observed values instead of their true values (X_{ij}, Y_{ij}) for every ij^{th} $(i=1,2,...,k, j=1,2,...,n)$ unit. In such a way, these values are expressible in additive form as, $x_{ij}=X_{ij}+V_{ij}$ and $y_{ij}=Y_{ij}+U_{ij}$. We consider that the errors

(U, V) are normally distributed with mean zero and variance (σ_U^2, σ_V^2). We assume that the error variables U and V are uncorrelated to each other as well as uncorrelated to all combinations with X and Y, respectively. This implies $Cov(X,U) = Cov(X, V) = Cov(Y, V) = Cov(Y, U) = Cov(U, V) = 0$ and $Cov(X,Y) \neq 0$. Let μ_{Ysy}, μ_{Xsy} be the population mean and $\sigma_{Ysy}^2, \sigma_{Xsy}^2$ be the population variance of the study and the auxiliary variables respectively. ρ is the correlation coefficient between study and auxiliary variable. Further, the sample mean of the observed data are unbiased estimators of the population mean μ_{Xsy} and μ_{Ysy} respectively.

The systematic sample means are

$$\mu_{Ysy} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} \quad \text{and} \quad \mu_{X\,sy} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}$$

The sample means are unbiased estimators of population means $\bar{y}$ and $\bar{x}$ of μ_{Ysy} and $\mu_{X\,sy}$ respectively.

$$\bar{y}_{sy} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}, (i = 1, 2, ..., k), \tag{1}$$

$$\bar{x}_{sy} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, (i = 1, 2, ..., k). \tag{2}$$

For determining variance, it is expressed by means of error terms e_0 and e_1, which are defined as

$$\bar{y}_{sy} = \mu_{Ysy}(1+e_0) \quad \text{and} \quad \bar{x}_{sy} = \mu_{Xsy}(1+e_1).$$

We can write $E(e_0) = E(e_1) = 0$,

and $E(e_1^2) = \frac{k}{\mu_{Xsy}^2}\left\{\sigma_{Xsy}^2 + \sigma_{Vsy}^2\right\}$, $E(e_0^2) = \frac{k}{\mu_{Ysy}^2}\left\{\sigma_{Ysy}^2 + \sigma_{Usy}^2\right\}$, $E(e_0 e_1) = \frac{k\rho\sigma_{Xsy}\sigma_{Ysy}}{\mu_{Ysy}\mu_{Xsy}}$, $R = \frac{\mu_{Xsy}}{\mu_{Ysy}}$,

$$\sigma_{Ysy}^2 = \frac{1}{k}\sum_{i=1}^{n}\left(\bar{y}_{sy} - \mu_{Ysy}\right)^2, \quad \sigma_{Xsy}^2 = \frac{1}{k}\sum_{i=1}^{n}\left(\bar{x}_{sy} - \mu_{Xsy}\right)^2, \quad \sigma_{Usy}^2 = \frac{1}{k}\sum_{i=1}^{k}\left(\bar{U}_{i.}\right)^2, \quad \sigma_{Vsy}^2 = \frac{1}{k}\sum_{i=1}^{k}\left(\bar{V}_{i.}\right)^2.$$

2. Methods

Three well-known form of the estimator has been proposed to estimate body mass index .We use ratio estimator by [24], product estimator by [25] and difference estimators under systematic sampling.

$$\bar{y}_{Rsy} = \bar{Y}_{sy}\frac{\mu_{Xsy}}{\bar{X}_{sy}}, \tag{3}$$

$$\bar{y}_{Psy} = \bar{Y}_{sy} \frac{\bar{X}_{sy}}{\mu_{Xsy}} \tag{4}$$

$$\bar{y}_{dsy} = \bar{Y}_{sy} + b\left(\mu_{Xsy} - \bar{X}_{sy}\right) \tag{5}$$

The mean square error of the ratio estimator is given as

$$MSE\left(\bar{y}_{Rsy}\right) = k\left(\sigma^2_{Ysy} + R^2\sigma^2_{Xsy} - 2\rho R\sigma_{Ysy}\sigma_{Xsy}\right) \tag{6}$$

The mean square error of the estimator is obtained by Shukla (1971) as

$$MSE\left(\bar{y}_{Psy}\right) = k\left(\sigma^2_{Ysy} + R^2\sigma^2_{Xsy} + 2\rho R\sigma_{Ysy}\sigma_{Xsy}\right). \tag{7}$$

The variance of difference estimator is given as

$$V\left(\bar{y}_{dsy}\right) = k\left[\sigma^2_{Ysy}\left(1-\rho^2\right)\right]. \tag{8}$$

2.1 The proposed estimation under observational error

Since we are taking account that observation recorded during data collection are obtained with some error. We consider the severity of misleading inference based on data obtained with observational error. In this section, we propose ratio, product, difference and mean estimators when the data are recorded with observational errors. In previous section, we have used well-known methods of estimation but in this section we derive the expression for mean square error and variance for all estimators when the data are observed with error.

Considering that the observations are recorded with observational error, then the variance in the presence of observational error is

$$\sigma^2_{Ysym} = \frac{1}{k}\sum_{i=1}^{n}\left(\bar{y}_{sy} - \mu_{Ysy}\right)^2 \tag{9}$$

$$\sigma^2_{Ysym} = \sigma^2_{Ysy} + \sigma^2_{Usy} = V\left(\bar{y}_{sym}\right), \tag{10}$$

where the term σ^2_{Usy} is the variance due to observational error.

We consider the situations when both the study variables and the auxiliary variables are observed with observational error. In that case, we propose the ratio estimator as

$$\bar{y}_{Rsym} = \bar{y}_{sy} \frac{\mu_{Xsy}}{\bar{x}_{sy}}. \tag{11}$$

In order to obtain the bias and mean square error we can write (11) as

$$\bar{y}_{Rsym} = \mu_{Ysy}\left(1+e_0\right)\frac{\mu_{Xsy}}{\mu_{Xsy}\left(1+e_1\right)}, \tag{12}$$

$$\bar{y}_{Rsym} = \mu_{Ysy}\left(1+e_0\right)\left(1+e_1\right)^{-1}, \tag{13}$$

For bias of the estimator, we obtained from (13)

$$\bar{y}_{Rsym} = \mu_{Ysy}\left(e_1^2 - e_0e_1\right). \tag{14}$$

Taking expectation of (14), we get the bias of the estimator as

$$bias\left(\overline{y}_{Rsym}\right)=\frac{k}{\mu_{Ysy}}\left\{R^2\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)-\rho R\sigma_{Xsy}\sigma_{Ysy}\right\}. \tag{15}$$

For mean square error, we can write from (13)

$$\left(\overline{y}_{Rsym}-\mu_{Ysy}\right)^2=\mu_{Ysy}^2\left(e_0^2+e_1^2-2e_0e_1\right), \tag{16}$$

Taking expectation of (16), we get the mean square error as

$$MSE\left(\overline{y}_{Rsym}\right)=k\left[\sigma_{Ysy}^2+\sigma_{Usy}^2+R^2\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)-2\rho R\sigma_{Ysy}\sigma_{Xsy}\right]. \tag{17}$$

We can obtain the result under no observational error by putting σ_{Usy}^2 and σ_{Vsy}^2 equal to zero. This will give the same result as obtained by [24]. From (6) and (17) we can write that MSE in the presence of observational error is always high.
The product estimator is proposed under the consideration of observational error as

$$\overline{y}_{Psym}=\overline{y}_{sy}\frac{\overline{x}_{sy}}{\mu_{Xsy}}, \tag{18}$$

To obtain the bias and mean square error, we can write (18) as

$$\overline{y}_{Psym}=\mu_{Ysy}\left(1+e_0\right)\left(1+e_1\right). \tag{19}$$

For the bias, by taking the expectation of (19) we get

$$bias\left(\overline{y}_{Psym}\right)=2\rho R\sigma_{Ysy}\sigma_{Xsy}. \tag{20}$$

For the mean square error, we can write from (19)

$$\left(\overline{y}_{Psym}-\mu_{Ysy}\right)^2=\mu_{Ysy}^2\left(e_0^2+e_1^2+2e_0e_1\right), \tag{21}$$

Taking the expectation of (21), we get the mean square error as

$$MSE\left(\overline{y}_{Psym}\right)=k\left[\sigma_{Ysy}^2+\sigma_{Usy}^2+R^2\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)+2\rho R\sigma_{Ysy}\sigma_{Xsy}\right]. \tag{22}$$

By substituting the value σ_{Usy}^2 and σ_{Vsy}^2 equal to zero we can obtain the MSE without observational error which is the same as obtained by [25]. From (7) and (22) we can conclude that MSE is always high in the presence of observational error.
The difference type estimator as proposed under the influence of observational error

$$\overline{y}_{dsym}=\overline{y}_{sy}+b\left(\mu_{Xsy}-\overline{x}_{sy}\right) \tag{23}$$

In order to obtain variance, we can write (24) as

$$\overline{y}_{dsym}=\mu_{Ysy}\left(1+e_0\right)+b\left\{\mu_{Xsy}-\mu_{Xsy}\left(1+e_1\right)\right\} \tag{24}$$

$$\left(\overline{y}_{dsym}-\mu_{Ysy}\right)=\left(\mu_{Ysy}e_0-be_1\mu_{Xsy}\right). \tag{25}$$

Squaring both sides of (25) and taking expectation

$$E\left(\bar{y}_{dsym}-\mu_{Ysy}\right)^2=E\left[\mu_{Ysy}^2e_0^2+b^2e_1^2\mu_{Xsy}^2-2b\mu_{Ysy}\mu_{Xsy}e_0e_1\right] \tag{26}$$

From (26) we can get the variance of the estimator as

$$V\left(\bar{y}_{dsym}\right)=k\left[\left(\sigma_{Ysy}^2+\sigma_{Usy}^2\right)+b^2\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)-2b\rho\sigma_{Ysy}\sigma_{Xsy}\right] \tag{27}$$

To obtain minimum variance differentiate (27) with respect to b and equate it to zero we get

$$b=\frac{\rho\sigma_{Ysy}\sigma_{Xsy}}{\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)}. \tag{28}$$

By substituting the value of b in (27), we get the minimum variance of the estimator as

$$V\left(\bar{y}_{dsym}\right)=k\left[\left(\sigma_{Ysy}^2+\sigma_{Usy}^2\right)-\frac{\rho^2\sigma_{Ysy}^2\sigma_{Xsy}^2}{\left(\sigma_{Xsy}^2+\sigma_{Vsy}^2\right)}\right]. \tag{29}$$

From (8) and (29), we can write that MSE in the presence of observational error is always high. By putting σ_{Usy}^2 and σ_{Vsy}^2 equal to zero, we can obtain the MSE under no observational error which is the same as given in (8).

3. Results

Numerical study has been carried out to show the efficacy of the proposed methods. We have taken the data from http://lib.stat.cmu.edu/datasets/bodyfat, this is a comprehensive dataset that lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. Earlier study by [15] is a linear regression model fitting. The present study is another approach to estimate the body mass index. The population is taken of size 252. In systematic sampling every kth sample is chosen such that $N=nk$. With this population, two sample population for $k=10,25$ has been chosen using systematic sampling. The correlation coefficient between BMI and circumference of different body parts has been obtained. The measures of different body parts are taken as single auxiliary variable. The mean square error has been obtained when different measures of body parts are used as auxiliary variable.

In this manuscript we also consider the presence of observational error in sample data. For the study of observational error, we have done simulation study. A hypothetical population has been generated by using mean and variance of original data under study. A population of size 5000 units with mean vector and a covariance matrix has been generated. The data matrices on X, Y, u and v have been generated using multivariate normal distribution for four variables with mean vector $\left(\mu_Y \quad \mu_X \quad 0 \quad 0\right)$ and covariance matrix

$$\begin{pmatrix} \sigma_Y^2 & \rho\sigma_X\sigma_Y & 0 & 0 \\ \rho\sigma_X\sigma_Y & \sigma_X^2 & 0 & 0 \\ 0 & 0 & \sigma_U^2 & 0 \\ 0 & 0 & 0 & \sigma_V^2 \end{pmatrix}$$

Two sets for $k=10,25$ has been chosen by using systematic sampling. The mean and variances have been computed for all the auxiliary variable. The mean square error and the variance have been computed. The above process has been replicated 5000 times and corresponding grand mean has been obtained. The percent relative efficiency of an estimator $\phi(=\bar{y}_{Rsym}, \bar{y}_{Psym}, \bar{y}_{dsym})$ with respect to usual unbiased estimator $\bar{y}_{sym}$ is calculated by

$$PRE(\phi,\ \bar{y})=\frac{V\left(\bar{y}_{sym}\right)}{MSE\left(\phi\right)}\times 100. \tag{30}$$

The results of the numerical and simulation study are given in table 1 and 2. Table 1 shows the MSE and PRE of the data linked in the abstract. From the table we can see for all the measures of body parts, ratio and regression estimators perform better than usual estimator. In all cases, the use of body measures of hip has maximum efficiency over other body measures as it has maximum correlation coefficient with body mass index. After hip, the use of body measures of thigh has more efficiency in the estimation. The body measures of abdomen has also better correlation with body mass index so it has better efficiency. The body measures of ankle and forearm have less correlation coefficient to the body mass index resultant have less efficiency in the estimation. The circumference of wrist has minimum correlation coefficient with body mass index. The mean square error for wrist is maximum thus it is better not to use the wrist circumference in the estimation of body mass index. Table 2, shows the results of the data when the error variance (σ_U^2, σ_V^2 = 0.5, 0.1). The MSE in the presence of observational errors is always high for all the estimators. The above results of different body measures follow the same trends in the presence of observational error. Hence the properties of estimators do not change in the presence of observational error but the value of mean square error is large. Study with related to sample size, the value of mean square error is less when the size of sample is large i.e. k is small. When k is large, the size of the sample is small and MSE is high for all the proposed estimators for all the body measures. This result can be seen from table 1 and 2.

Table 1: *MSE* and *PRE* of various estimators of $\bar{y}_{sym}$ for BMI estimation for $(k=25, 10)$ when no error variance

Body Part	Estimator	$k=25$		$k=10$	
		MSE	PRE	*MSE*	PRE
Circumference of abdomen	$\bar{y}_{dsym}$	173.7477	121.42	25.1320	121.42
	$\bar{y}_{Rsym}$	174.3850	120.97	26.0398	117.19

	$\bar{y}_{Psym}$	303.7588	69.45	38.7288	78.79
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of neck	$\bar{y}_{dsym}$	196.0010	107.63	28.3509	107.63
	$\bar{y}_{Rsym}$	198.0751	106.51	28.4656	107.20
	$\bar{y}_{Psym}$	235.6348	89.53	35.1281	86.87
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of chest	$\bar{y}_{dsym}$	179.9560	117.23	26.0300	117.23
	$\bar{y}_{Rsym}$	184.5389	114.32	27.1705	112.31
	$\bar{y}_{Psym}$	260.8786	80.87	36.0632	84.61
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of hip	$\bar{y}_{dsym}$	165.9304	127.14	24.0013	127.14
	$\bar{y}_{Rsym}$	180.6452	116.78	26.9497	113.23
	$\bar{y}_{Psym}$	257.8028	81.83	35.4747	86.02
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of thigh	$\bar{y}_{dsym}$	171.4538	123.04	24.8002	123.04
	$\bar{y}_{Rsym}$	176.8878	119.26	26.1525	116.68
	$\bar{y}_{Psym}$	276.3093	76.35	37.8913	80.53
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of knee	$\bar{y}_{dsym}$	182.9926	115.28	26.4693	115.28
	$\bar{y}_{Rsym}$	190.3175	110.85	27.5349	110.82
	$\bar{y}_{Psym}$	244.9394	86.13	35.4117	86.17
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of ankle	$\bar{y}_{dsym}$	201.6285	104.63	29.1649	104.63
	$\bar{y}_{Rsym}$	201.6853	104.60	29.4117	103.75
	$\bar{y}_{Psym}$	236.1036	89.35	32.5023	93.88
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of biceps	$\bar{y}_{dsym}$	190.5693	110.70	27.5652	110.70
	$\bar{y}_{Rsym}$	190.6847	110.63	27.6162	110.50
	$\bar{y}_{Psym}$	266.1175	79.27	37.8633	80.59
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of forearm	$\bar{y}_{dsym}$	201.1914	104.86	29.1016	104.86
	$\bar{y}_{Rsym}$	201.2036	104.85	29.1353	104.73

	$\bar{y}_{Psym}$	238.9017	88.30	33.9158	89.97
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00
Circumference of wrist	$\bar{y}_{dsym}$	203.3305	103.75	29.4111	103.75
	$\bar{y}_{Rsym}$	203.5102	103.66	29.5349	103.32
	$\bar{y}_{Psym}$	229.3498	91.98	32.4712	93.98
	$\bar{y}_{sym}$	210.9612	100.00	30.5148	100.00

Table 2: *MSE* and *PRE* of various estimators of $\bar{y}_{sym}$ for BMI estimation for $(k = 25, 10)$ when the error variance $(\sigma_U^2, \sigma_V^2 = 0.5, 0.1)$

Body Part	Estimator	$k = 25$		$k = 10$	
		MSE	PRE	*MSE*	PRE
Circumference of abdomen	$\bar{y}_{dsym}$	203.3305	103.75	30.5490	116.26
	$\bar{y}_{Rsym}$	203.5102	103.66	31.2247	113.74
	$\bar{y}_{Psym}$	229.3498	91.98	43.7609	81.16
	$\bar{y}_{sym}$	210.9612	100.00	35.5148	100.00
Circumference of neck	$\bar{y}_{dsym}$	210.7290	106.04	34.0874	104.19
	$\bar{y}_{Rsym}$	211.6064	105.60	34.1271	104.07
	$\bar{y}_{Psym}$	249.1661	89.68	40.7895	87.07
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of chest	$\bar{y}_{dsym}$	192.8664	115.86	31.3796	113.18
	$\bar{y}_{Rsym}$	197.1965	113.32	32.2637	110.08
	$\bar{y}_{Psym}$	273.5362	81.69	41.1564	86.29
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of hip	$\bar{y}_{dsym}$	179.2454	124.67	29.7633	119.32
	$\bar{y}_{Rsym}$	193.2975	115.60	32.0421	110.84
	$\bar{y}_{Psym}$	270.4551	82.62	40.5670	87.55
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of thigh	$\bar{y}_{dsym}$	184.9911	120.80	30.6509	115.87
	$\bar{y}_{Rsym}$	189.8094	117.73	31.4161	113.05
	$\bar{y}_{Psym}$	289.2310	77.26	43.1549	82.30
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of knee	$\bar{y}_{dsym}$	199.1622	112.20	33.0655	107.41
	$\bar{y}_{Rsym}$	203.8243	109.63	33.1596	107.10

	$\bar{y}_{Psym}$	258.4462	86.46	41.0364	86.54
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of ankle	$\bar{y}_{dsym}$	216.5653	103.18	35.2467	100.76
	$\bar{y}_{Rsym}$	216.9886	102.98	36.1959	98.12
	$\bar{y}_{Psym}$	251.4069	88.88	39.2865	90.40
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of biceps	$\bar{y}_{dsym}$	204.6725	109.18	33.433	106.23
	$\bar{y}_{Rsym}$	204.6728	109.18	33.5436	105.88
	$\bar{y}_{Psym}$	280.1056	79.78	43.7907	81.10
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of forearm	$\bar{y}_{dsym}$	215.2909	103.80	34.8574	101.89
	$\bar{y}_{Rsym}$	215.4835	103.70	35.2974	100.62
	$\bar{y}_{Psym}$	253.1816	88.26	40.0779	88.61
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00
Circumference of wrist	$\bar{y}_{dsym}$	219.3158	101.89	35.3544	100.45
	$\bar{y}_{Rsym}$	220.6081	101.29	37.4055	94.95
	$\bar{y}_{Psym}$	246.4477	90.67	40.3418	88.03
	$\bar{y}_{sym}$	223.4612	100.00	35.5148	100.00

4. Conclusions

We give a different approach from [12] to estimate BMI rather than body-fat. This study is used about systematic sampling by using auxiliary variables. Body mass index is calculated using the ratio, product, regression and unbiased mean estimator. The different measures of body are used as auxiliary variables. From the study, we may conclude that regression estimator under systematic sampling has maximum efficiency in the estimation of body mass index. The efficacy of the methods are depends on the correlation between body mass index and circumference of the different measures of the body. The correlation coefficient for the body measurement of hip, abdomen, and thigh is good, so these variable provide better estimate for body mass index when the circumferences of these parts are used as auxiliary variable. The circumferences of body parts wrist, forearm and ankle have least correlation coefficient with body mass index thus may not be used in estimation of BMI. From the tables, we can also conclude that the ratio estimator and regression estimator are always more efficient than unbiased mean estimator. So it is better to use ratio and regression methods of estimation by using the different measures of the body as auxiliary variables. Since in this article, we are assuming the presence of observational error in the study of 252 men. The efficiency of the regression estimators is better in the presence of observational error. Also the presence of observational error does not change the properties of the estimators. From table 1 and 2, we can conclude the effect of observational error on mean square error. The above study is provide an easy approach and simplest way to obtain the BMI estimate with and without observational error. Thus the suggested method may be used by statistician for this problem and for many others similar problem in the estimation of mean.

Abbreviations: BMI: body mass index, MSE: mean square error, PRE: percent relative efficiency.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: Not applicable.

Competing interests: The authors declare that there are no financial and competing interests.

Funding: Not Available.

Authors' contributions GKV conceptualized the study, NS developed the data analysis methodology. NS analyzed the data and drafted this manuscript. SPS reviewed the draft manuscript and provided technical inputs to GKV to finalize the manuscript. All authors read and approved the final manuscript.

Author information: [1]Department of Mathematics & Computing, Indian Institute of Technology (ISM), Dhanbad-826004, India. [2]College of Engineering & Technology, Wollega University, Ethiopia.

Acknowledgements: NA

References

[1] Cochran, W.G., Relative accuracy of systematic and stratified random samples for a certain class of populations. *The Annals of Mathematical Statistics*, 1946; 17(2):164-177.

[2] Murthy, M.N., *Sampling theory and methods.* Statistical Publishing Society, Calcutta, India, 1967.

[3] Biemer, P. Stokes, S.L., Approaches to the modeling of measurement error; Measurement error in Surveys. *Wiley*, 1991; 485-516

[4] Fuller, W.A., *Measurement error models.* John Wiley & Sons, Inc., 1987.

[5] Cheng, C.L. and Vanness, J.W., On estimating linear relationships when both variables are subject to error. *Journal of the Royal Statistical Society. Series B (Methodological),* 1994; 56(1): 167-183.

[6] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M., *Measurement error in nonlinear models: a modern perspective,* CRC Press, 2006.

[7] Shalabh, Ratio method of estimation in the presence of measurement error. *Journal of the Indian Society of Agricultural Statistics*, 1997; 50(2): 150-155.

[8] Maneesha and Singh, R.K., Role of regression estimator involving measurement errors. *Brazilian Journal of Probability and Statistics*, 2002; 16, 39-46.

[9] Sahoo, L.N., Sahoo, R.K., Senapati, S.C., An empirical study on the accuracy of ratio and regression estimators in the presence of measurement error. *Monte Carlo Methods and Applications,* 2006; 12(5): 495-501.

[10] Singh, N., Vishwakarma, G.K. and Kim, J.M., Computing the effect of measurement error on efficient variant of ratio and product estimator using auxiliary variable. *Communications in Statistics-Simulation and Computation*. 2019; DOI:10.1080/03610918. 2019.1656742.

[11] Singh, N. and Vishwakarma, G.K., A generalised class of estimator of population mean with the combined effect of measurement errors and non-response in sample survey. *Revista Investigación Operacional*, 2019; 40(2): 275-285.

[12] Penrose, K.W., Nelson, A.G. and Fisher, A.G., Generalized body composition prediction equation for men using simple measurement techniques. *Medicine & Science in Sports & Exercise,* 1985; 17(2): 189.

[13] Siri, W.E., In advances in biological and medical physics. *London and New York, Academic press Inc.*, 1956; 4: 239-280.

[14] Behnke, A.R. and Wilmore, J.H., *Evaluation and regulation of body build and composition*. Prentice Hall, 1974.

[15] Johnson, R.W., Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education*, 1996; 4(1): 1-8.

[16] Bouza, C.N., Allende-Alonso, S.M., Vishwakarma, G.K. and Singh, N., Estimation of optimum sample size allocation: An illustration with body mass index for evaluating the effect of a dietetic supplement. *International Journal of Biomathematics*, 2019; 12(8): 1950086.

[17] Cochran, W.G., Errors of measurement in statistics. *Technometrics*, 1968; 10(4): 637-666.

[18] Gautschi, W., Some remarks on systematic sampling. *The Annals of Mathematical Statistics*, 1957; 28(2): 385-394.

[19] Madow, W.G., On the theory of systematic sampling, II. *The Annals of Mathematical Statistics,* 1949; 20(3): 333-354.

[20] Madow, W.G., On the theory of systematic sampling, III. Comparison of centered and random start systematic sampling. *The Annals of Mathematical Statistics*, 1953; 24: 101-106.

[21] Williams, R.M., The variance of the mean of systematic samples. *Biometrika,* 1956; 43: 137-148.

[22] Hasel, H.A., Estimation of volume of timber stands by strip sampling. *Annals of Mathematical Statistics*, 1942; 13: 179-206.

[23] Griffth, A.L., The efficiency of enumerations. Indian Forest Leaflets, 83-93. *Forest- Research Institute*: Dehra Dun, 1945.

[24] Swain, A.K.P.C., The use of systematic sampling in ratio estimate, *Journal of Indian Statistical Association,* 1964; 2(213): 160-164.

[25] Shukla, N.D., Systematic sampling and product method of estimation. *In Proceeding of all India Seminar on Demography and Statistics*. BHU, Varanasi, India, 1971.