

Research Article

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

Gajendra K. Vishwakarma ¹, Neha Singh ¹ and Surendra P. Singh ²

¹Department of Mathematics & Computing, Indian Institute of Technology (ISM) Dhanbad, Dhanbad-826004, Jharkhand, India

²Institute of Engineering & Technology, Wollega University, Nekemte, Ethiopia

Correspondence should be addressed to Surendra P. Singh; surendra.geomatics@gmail.com

Received 21 September 2021; Accepted 16 December 2021; Published 19 January 2022

Academic Editor: Peter Dabnichki

Copyright © 2022 Gajendra K. Vishwakarma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. The use of body mass index (BMI) is prevalent, to measure the fat in the body. Sometimes, during a clinical survey, different measures of body parts of people may be available, but the actual weight and height are not available. In this article, we have shown a method to estimate the body mass index using the measures of different body parts. Systematic sampling is to be applied only if the given population is logically homogeneous because systematic sample units are uniformly distributed over the population. **Methods.** The method of estimation for the mean of the study variable under systematic sampling using auxiliary information has been used to estimate the body mass index (BMI). We also have shown the effect of observational error in the estimation. The measures of different body parts are taken as auxiliary variables. The correlation coefficient between BMI and the circumference of different body parts has been obtained. The efficacy of methods in terms of mean square error has been obtained in the estimation of BMI. Also, the observations available on different body parts are assumed to be recorded with observational error. Thus, we propose a method of estimation of BMI in the presence of observational error. A simulation study has been conducted to demonstrate the effect of the 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 the conditions under which the proposed method is more efficient are found. We assume the presence of observational error in the study of 252 men. The efficiency of the difference estimators is better in the presence of observational error. Also, the presence of observational error does not change the properties of the estimators. **Conclusions.** The study provides an easy approach and the simplest way to obtain the BMI estimation with and without observational error. Thus, the suggested method may be used by statisticians for this problem and for many other similar problems in the estimation of mean.

1. Background

In a 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 the observed value and the true value of the sample. There are several examples of real-life situations when data are obtained with errors [1, 2]. The observational error in the context of linear and nonlinear regression models also has been thoroughly discussed in the literature [3–6]. Several research studies have been performed on the observational errors in the estimation of ratio, product, and regression methods of estimation, which are available in [7–11]. If the data are

systematically distributed, the systematic sampling has nice features of selecting every k^{th} element by choosing the first element arbitrary. Many authors have done pioneering work using systematic sampling at the estimation stage (see [12–16]). The estimation of parameters for certain natural population is convenient using systematic sampling [17, 18]. The use of auxiliary variables is prevalent as ratio, product, and regression estimator. In case of estimating the volume of timber, the proposed ratio estimator under systematic sampling suggested that the leaf area or the girth of the tree may be taken as the auxiliary variable [19]. The product estimators in the context of systematic sampling have been discussed in [20]. Some pioneer works in systematic sampling have been introduced in [21, 22].

A study was conducted to derive a prediction equation for body fat percentage in men ($n=252$, age 22–81 years) from simple body measurements [23]. Body density determined by underwater weighing and body fat percentage was determined. The dataset includes the following variables, given in [24], pp. 45–48), for observational techniques: density determined from underwater weighing, percent body-fat from Siri's equation [25], age in years, weight in lbs, height in inches, and circumference of the neck, chest, abdomen, hip, thigh, knee, ankle, bicep, arm, and wrist in centimetre. In this article, we propose a different method to estimate the body mass index rather than the already established multiple regression method in [23, 26]. The body mass index is highly correlated with the body parts, so in case if BMI is not known for a large population, we can estimate using sampling methods ratio, product, and difference estimator. We attempt to estimate the body mass index in place of body fat by using one of the auxiliary variables. The circumference of hip, thigh, knee, ankle, bicep, arm, and wrist can be taken as a single auxiliary variable to estimate the body mass index. The correlation coefficient for each auxiliary variable has been obtained. Since the data are natural, we used systematic sampling. An estimated optimal sample size by using the body mass index for a dietetic supplement has been calculated [27].

The method of estimation for the mean of the study variable under systematic sampling using auxiliary information has been used to estimate the body mass index (BMI). We also have shown the effect of observational error in the estimation. The measures of different body parts are taken as auxiliary variables. The correlation coefficient between BMI and the circumference of different body parts has been obtained. The efficacy of methods in terms of mean square error has been obtained in the estimation of BMI. Also, the observations available on different body parts are assumed to be recorded with observational error. Thus, we also propose a method of estimation of BMI in the presence of observational error. A simulation study has been conducted to demonstrate the effect of the observational error on the estimation of body mass index.

Suppose, the population consists of N units $u = (u_1, u_2, \dots, u_N)$ 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 i^{th} cluster contains serially numbered units $\{i, (i+k), (i+2k), \dots, (i+(n-1)k)\}$. After the 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 the observed values instead of their true values (X_{ij}, Y_{ij}) for every i^{th} ($i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$) 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 $\text{Cov}(X, U) = \text{Cov}(X, V) = \text{Cov}(Y, U) = \text{Cov}(Y, V) = 0$, and $\text{Cov}(X, Y) \neq 0$. Let μ_{Ysy} , μ_{Xsy} be the population mean and σ_{Ysy}^2 , σ_{Xsy}^2 be the population variance of the study and the auxiliary variables, respectively. ρ is the correlation coefficient between the study and auxiliary variable. Furthermore, the sample means of the observed data are the unbiased estimators of the population means μ_{Xsy} and μ_{Ysy} , respectively.

The population means are $\mu_{Ysy} = 1/nk \sum_{i=1}^k \sum_{j=1}^n y_{ij}$ and $\mu_{Xsy} = 1/nk \sum_{i=1}^k \sum_{j=1}^n x_{ij}$.

The sample means are the unbiased estimators of the population means μ_{Ysy} and μ_{Xsy} , respectively.

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

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

For determining variance, it is expressed by means of error terms e_0 and e_1 , which are defined as $\bar{y}_{sy} = \mu_{Ysy}(1 + e_0)$ and $\bar{x}_{sy} = \mu_{Xsy}(1 + e_1)$.

We can write

$$\begin{aligned} E(e_0) &= E(e_1) \\ E(e_1^2) &= \frac{1}{k\mu_{Xsy}^2} \{ \sigma_{Xsy}^2 + \sigma_{Vsy}^2 \}, \\ E(e_0^2) &= \frac{1}{k\mu_{Ysy}^2} \{ \sigma_{Ysy}^2 + \sigma_{U_{sy}}^2 \}, \\ E(e_0e_1) &= \frac{1}{k} \frac{\rho\sigma_{Xsy}\sigma_{Ysy}}{\mu_{Ysy}\mu_{Xsy}}, \\ R &= \frac{\mu_{Xsy}}{\mu_{Ysy}}, \end{aligned} \quad (2)$$

$$\sigma_{Ysy}^2 = \frac{1}{k} \sum_{i=1}^k (\bar{y}_{sy} - \mu_{Ysy})^2,$$

$$\sigma_{Xsy}^2 = \frac{1}{k} \sum_{i=1}^k (\bar{x}_{sy} - \mu_{Xsy})^2,$$

$$\sigma_{U_{sy}}^2 = \frac{1}{k} \sum_{i=1}^k (\bar{U}_i)^2,$$

$$\sigma_{Vsy}^2 = \frac{1}{k} \sum_{i=1}^k (\bar{V}_i)^2.$$

2. Methods

Three well-known forms of the estimator have been proposed to estimate the body mass index. We use ratio estimator [19], product estimator [20], and difference estimator under systematic sampling.

$$\begin{aligned} \bar{y}_{Rsy} &= \bar{Y}_{sy} \frac{\mu_{Xsy}}{\bar{X}_{sy}}, \\ \bar{y}_{Psy} &= \bar{Y}_{sy} \frac{\bar{X}_{sy}}{\mu_{Xsy}}, \\ \bar{y}_{dsy} &= \bar{Y}_{sy} + b(\mu_{Xsy} - \bar{X}_{sy}). \end{aligned} \quad (3)$$

The mean square error of the ratio estimator is given as

$$\text{MSE}(\bar{y}_{Rsy}) = \frac{1}{k} (\sigma_{Ysy}^2 + R^2 \sigma_{Xsy}^2 - 2\rho R \sigma_{Ysy} \sigma_{Xsy}). \quad (4)$$

The mean square error of the product estimator is obtained in [20] as

$$\text{MSE}(\bar{y}_{Psy}) = \frac{1}{k} (\sigma_{Ysy}^2 + R^2 \sigma_{Xsy}^2 + 2\rho R \sigma_{Ysy} \sigma_{Xsy}). \quad (5)$$

The variance of the difference estimator is given as

$$V(\bar{y}_{dsy}) = \frac{1}{k} [\sigma_{Ysy}^2 (1 - \rho^2)]. \quad (6)$$

2.1. The Proposed Estimation under Observational Error.

The observation recorded during data collection is 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 the 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 is

$$\sigma_{Ysym}^2 = \frac{1}{k} \sum_{i=1}^n (\bar{y}_{sy} - \mu_{Ysy})^2, \quad (7)$$

$$\sigma_{Ysym}^2 = \sigma_{Ysy}^2 + \sigma_{U_{sy}}^2 = V(\bar{y}_{sym}),$$

where the term $\sigma_{U_{sy}}^2$ is the variance due to observational error.

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

$$\bar{y}_{Rsym} = \bar{y}_{sy} \frac{\mu_{Xsy}}{\bar{x}_{sy}}. \quad (8)$$

In order to obtain the bias and mean square error, we can write equation (8) as

$$\bar{y}_{Rsym} = \mu_{Ysy} (1 + e_0) \frac{\mu_{Xsy}}{\mu_{Xsy} (1 + e_1)}, \quad (9)$$

$$\bar{y}_{Rsym} = \mu_{Ysy} (1 + e_0) (1 + e_1)^{-1}. \quad (10)$$

For the bias of the estimator, we obtained from equation (10) as

$$\bar{y}_{Rsym} = \mu_{Ysy} (e_1^2 - e_0 e_1). \quad (11)$$

Taking the expectation of equation (11), we get the bias of the estimator as

$$\text{bias}(\bar{y}_{Rsym}) = \frac{1}{k \mu_{Ysy}} \{R^2 (\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2) - \rho R \sigma_{Xsy} \sigma_{Ysy}\}. \quad (12)$$

For the mean square error, we can write from equation (10) as

$$(\bar{y}_{Rsym} - \mu_{Ysy})^2 = \mu_{Ysy}^2 (e_0^2 + e_1^2 - 2e_0 e_1). \quad (13)$$

Taking the expectation of equation (13), we get the mean square error as

$$\text{MSE}(\bar{y}_{Rsym}) = \frac{1}{k} [\sigma_{Ysy}^2 + \sigma_{U_{sy}}^2 + R^2 (\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2) - 2\rho R \sigma_{Ysy} \sigma_{Xsy}]. \quad (14)$$

We can obtain the result under no observational error by putting $\sigma_{U_{sy}}^2$ and $\sigma_{V_{sy}}^2$ equal to zero. This will give the same result as obtained in [19]. From equations (4) and (14), we can write that MSE in the presence of the observational error is always high.

The product estimator is proposed under the consideration of observational error as

$$\bar{y}_{Psy} = \bar{y}_{sy} \frac{\bar{x}_{sy}}{\mu_{Xsy}}. \quad (15)$$

To obtain the bias and mean square error, we can write equation (15) as

$$\bar{y}_{Psy} = \mu_{Ysy} (1 + e_0) (1 + e_1). \quad (16)$$

For the bias, by taking the expectation of equation (16), we get

$$\text{bias}(\bar{y}_{Psy}) = 2\rho R \sigma_{Ysy} \sigma_{Xsy}. \quad (17)$$

For the mean square error, we can write from equation (16) as

$$(\bar{y}_{Psy} - \mu_{Ysy})^2 = \mu_{Ysy}^2 (e_0^2 + e_1^2 + 2e_0 e_1). \quad (18)$$

Taking the expectation of equation (18), we get the mean square error as

$$\text{MSE}(\bar{y}_{Psy}) = \frac{1}{k} [\sigma_{Ysy}^2 + \sigma_{U_{sy}}^2 + R^2 (\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2) + 2\rho R \sigma_{Ysy} \sigma_{Xsy}]. \quad (19)$$

By substituting the value $\sigma_{U_{sy}}^2$ and $\sigma_{V_{sy}}^2$ equal to zero, we can obtain the MSE without observational error which is the same as obtained in [20]. From equations (4) and (19), 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 is

$$\bar{y}_{dsym} = \bar{y}_{sy} + b(\mu_{Xsy} - \bar{x}_{sy}). \quad (20)$$

In order to obtain variance, we can write equation (20) as

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 | Expected values | $k = 25$ | | $k = 10$ | |
|------------------------------|------------------|-----------------|----------|--------|----------|--------|
| | | | MSE | PRE | MSE | PRE |
| Circumference of the abdomen | \bar{y}_{dsym} | 27.72 | 173.7477 | 121.42 | 25.1320 | 121.42 |
| | \bar{y}_{Rsym} | 28.14 | 174.3850 | 120.97 | 26.0398 | 117.19 |
| | \bar{y}_{Psym} | 24.24 | 303.7588 | 69.45 | 38.7288 | 78.79 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the neck | \bar{y}_{dsym} | 24.98 | 196.0010 | 107.63 | 28.3509 | 107.63 |
| | \bar{y}_{Rsym} | 25.89 | 198.0751 | 106.51 | 28.4656 | 107.20 |
| | \bar{y}_{Psym} | 26.10 | 235.6348 | 89.53 | 35.1281 | 86.87 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the chest | \bar{y}_{dsym} | 25.12 | 179.9560 | 117.23 | 26.0300 | 117.23 |
| | \bar{y}_{Rsym} | 25.81 | 184.5389 | 114.32 | 27.1705 | 112.31 |
| | \bar{y}_{Psym} | 26.25 | 260.8786 | 80.87 | 36.0632 | 84.61 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the hip | \bar{y}_{dsym} | 24.15 | 165.9304 | 127.14 | 24.0013 | 127.14 |
| | \bar{y}_{Rsym} | 25.77 | 180.6452 | 116.78 | 26.9497 | 113.23 |
| | \bar{y}_{Psym} | 26.26 | 257.8028 | 81.83 | 35.4747 | 86.02 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the thigh | \bar{y}_{dsym} | 25.01 | 171.4538 | 123.04 | 24.8002 | 123.04 |
| | \bar{y}_{Rsym} | 25.80 | 176.8878 | 119.26 | 26.1525 | 116.68 |
| | \bar{y}_{Psym} | 26.30 | 276.3093 | 76.35 | 37.8913 | 80.53 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the knee | \bar{y}_{dsym} | 24.68 | 182.9926 | 115.28 | 26.4693 | 115.28 |
| | \bar{y}_{Rsym} | 25.83 | 190.3175 | 110.85 | 27.5349 | 110.82 |
| | \bar{y}_{Psym} | 26.16 | 244.9394 | 86.13 | 35.4117 | 86.17 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the ankle | \bar{y}_{dsym} | 25.23 | 201.6285 | 104.63 | 29.1649 | 104.63 |
| | \bar{y}_{Rsym} | 25.92 | 201.6853 | 104.60 | 29.4117 | 103.75 |
| | \bar{y}_{Psym} | 26.09 | 236.1036 | 89.35 | 32.5023 | 93.88 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of biceps | \bar{y}_{dsym} | 24.99 | 190.5693 | 110.70 | 27.5652 | 110.70 |
| | \bar{y}_{Rsym} | 25.79 | 190.6847 | 110.63 | 27.6162 | 110.50 |
| | \bar{y}_{Psym} | 26.32 | 266.1175 | 79.27 | 37.8633 | 80.59 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the forearm | \bar{y}_{dsym} | 25.02 | 201.1914 | 104.86 | 29.1016 | 104.86 |
| | \bar{y}_{Rsym} | 25.93 | 201.2036 | 104.85 | 29.1353 | 104.73 |
| | \bar{y}_{Psym} | 26.08 | 238.9017 | 88.30 | 33.9158 | 89.97 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |
| Circumference of the wrist | \bar{y}_{dsym} | 25.00 | 203.3305 | 103.75 | 29.4111 | 103.75 |
| | \bar{y}_{Rsym} | 25.92 | 203.5102 | 103.66 | 29.5349 | 103.32 |
| | \bar{y}_{Psym} | 26.03 | 229.3498 | 91.98 | 32.4712 | 93.98 |
| | \bar{y}_{sym} | 25.94 | 210.9612 | 100.00 | 30.5148 | 100.00 |

$$\bar{y}_{dsym} = \mu_{Ysy}(1 + e_0) + b\{\mu_{Xsy} - \mu_{Xsy}(1 + e_1)\}, \quad (21)$$

$$(\bar{y}_{dsym} - \mu_{Ysy}) = (\mu_{Ysy}e_0 - be_1\mu_{Xsy}). \quad (22)$$

By squaring both sides of equation (22) and taking expectation,

$$E(\bar{y}_{dsym} - \mu_{Ysy})^2 = E[\mu_{Ysy}^2 e_0^2 + b^2 e_1^2 \mu_{Xsy}^2 - 2b\mu_{Ysy}\mu_{Xsy}e_0e_1]. \quad (23)$$

From equation (23), we can get the variance of the estimator as

$$V(\bar{y}_{dsym}) = \frac{1}{k} [(\sigma_{Ysy}^2 + \sigma_{U_{sy}}^2) + b^2(\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2) - 2b\rho\sigma_{Ysy}\sigma_{Xsy}]. \quad (24)$$

To obtain minimum variance differentiate from equation (24) with respect to b and equate it to zero, we get

$$b = \frac{\rho\sigma_{Ysy}\sigma_{Xsy}}{(\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2)}. \quad (25)$$

By substituting the value of b in equation (24), we get the minimum variance of the estimator as

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 | Expected values | $k = 25$ | | $k = 10$ | |
|------------------------------|------------------|-----------------|----------|--------|----------|--------|
| | | | MSE | PRE | MSE | PRE |
| Circumference of the abdomen | \bar{y}_{dsym} | 27.72 | 203.3305 | 103.75 | 30.5490 | 116.26 |
| | \bar{y}_{Rsym} | 28.14 | 203.5102 | 103.66 | 31.2247 | 113.74 |
| | \bar{y}_{Psym} | 24.24 | 239.3498 | 91.98 | 43.7609 | 81.16 |
| | \bar{y}_{sym} | 25.94 | 218.9612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the neck | \bar{y}_{dsym} | 24.98 | 210.7290 | 106.04 | 34.0874 | 104.19 |
| | \bar{y}_{Rsym} | 25.89 | 211.6064 | 105.60 | 34.1271 | 104.07 |
| | \bar{y}_{Psym} | 26.10 | 249.1661 | 89.68 | 40.7895 | 87.07 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the chest | \bar{y}_{dsym} | 25.12 | 192.8664 | 115.86 | 31.3796 | 113.18 |
| | \bar{y}_{Rsym} | 25.81 | 197.1965 | 113.32 | 32.2637 | 110.08 |
| | \bar{y}_{Psym} | 26.25 | 273.5362 | 81.69 | 41.1564 | 86.29 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the hip | \bar{y}_{dsym} | 24.15 | 179.2454 | 124.67 | 29.7633 | 119.32 |
| | \bar{y}_{Rsym} | 25.77 | 193.2975 | 115.60 | 32.0421 | 110.84 |
| | \bar{y}_{Psym} | 26.26 | 270.4551 | 82.62 | 40.5670 | 87.55 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the thigh | \bar{y}_{dsym} | 25.01 | 184.9911 | 120.80 | 30.6509 | 115.87 |
| | \bar{y}_{Rsym} | 25.80 | 189.8094 | 117.73 | 31.4161 | 113.05 |
| | \bar{y}_{Psym} | 26.30 | 289.2310 | 77.26 | 43.1549 | 82.30 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the knee | \bar{y}_{dsym} | 24.68 | 199.1622 | 112.20 | 33.0655 | 107.41 |
| | \bar{y}_{Rsym} | 25.83 | 203.8243 | 109.63 | 33.1596 | 107.10 |
| | \bar{y}_{Psym} | 26.16 | 258.4462 | 86.46 | 41.0364 | 86.54 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the ankle | \bar{y}_{dsym} | 25.23 | 216.5653 | 103.18 | 35.2467 | 100.76 |
| | \bar{y}_{Rsym} | 25.92 | 216.9886 | 102.98 | 36.1959 | 98.12 |
| | \bar{y}_{Psym} | 26.09 | 251.4069 | 88.88 | 39.2865 | 90.40 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of biceps | \bar{y}_{dsym} | 24.99 | 204.6725 | 109.18 | 33.433 | 106.23 |
| | \bar{y}_{Rsym} | 25.79 | 204.6728 | 109.18 | 33.5436 | 105.88 |
| | \bar{y}_{Psym} | 26.32 | 280.1056 | 79.78 | 43.7907 | 81.10 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the forearm | \bar{y}_{dsym} | 25.02 | 215.2909 | 103.80 | 34.8574 | 101.89 |
| | \bar{y}_{Rsym} | 25.93 | 215.4835 | 103.70 | 35.2974 | 100.62 |
| | \bar{y}_{Psym} | 26.08 | 253.1816 | 88.26 | 40.0779 | 88.61 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |
| Circumference of the wrist | \bar{y}_{dsym} | 25.00 | 219.3158 | 101.89 | 35.3544 | 100.45 |
| | \bar{y}_{Rsym} | 25.92 | 220.6081 | 101.29 | 37.4055 | 94.95 |
| | \bar{y}_{Psym} | 26.03 | 246.4477 | 90.67 | 40.3418 | 88.03 |
| | \bar{y}_{sym} | 25.94 | 223.4612 | 100.00 | 35.5148 | 100.00 |

$$V(\bar{y}_{dsym}) = \frac{1}{k} \left[(\sigma_{Ysy}^2 + \sigma_{U_{sy}}^2) - \frac{\rho^2 \sigma_{Ysy}^2 \sigma_{Xsy}^2}{(\sigma_{Xsy}^2 + \sigma_{V_{sy}}^2)} \right]. \quad (26)$$

From equations (6) and (25), we can write that MSE in the presence of observational error is always high. By putting $\sigma_{U_{sy}}^2$ and $\sigma_{V_{sy}}^2$ equal to zero, we can obtain the MSE under no observational error which is the same as given in equation (6).

3. Results

A numerical study has been carried out to show the efficacy of the proposed methods. We have taken the data from

<https://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. With this population, two sample populations for $k = 10, 25$ have been chosen using systematic sampling.

In this manuscript, we also consider the presence of observational error in sample data. For the study of observational error, we have conducted a simulation study. A hypothetical population has been generated by using the 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 the mean vector $(\mu_Y \ \mu_X \ 0 \ 0)$ and covariance matrix

$$\begin{pmatrix} \sigma_Y^2 & \rho\sigma_X\sigma_Y & 0 & 0 \\ \rho\sigma_X\sigma_Y & \sigma_X^2 & 0 & 0 \\ 0 & 0 & \sigma_U^2 & 0 \\ 0 & 0 & 0 & \sigma_V^2 \end{pmatrix}. \quad (27)$$

Two sets for $k = 10, 25$ have been chosen by using systematic sampling. The mean and variances have been computed for all the auxiliary variables. The mean square error and the variance have been computed. The above process has been replicated 5000 times, and the 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 the usual unbiased estimator \bar{y}_{sym} is calculated by

$$PRE(\phi, \bar{y}) = \frac{V(\bar{y}_{sym})}{MSE(\phi)} \times 100. \quad (28)$$

The results of the numerical and simulation studies are given in Tables 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 difference estimators perform better than the usual estimator. In all cases, the use of body measures of the hip has maximum efficiency over other body measures as it has maximum correlation coefficient with the body mass index. After the hip, the use of body measures of the thigh has more efficiency in the estimation. The body measures of the abdomen have also better correlation with the body mass index, so it has better efficiency. The body measures of the ankle and the forearm have less correlation coefficient with the body mass index, and the resultant has less efficiency in the estimation. The circumference of the wrist has minimum correlation coefficient with the body mass index. The mean square error for the wrist is maximum; thus, it is better not to use the wrist circumference in the estimation of the body mass index. Table 2 shows the results of the data with error variance ($\sigma_U^2, \sigma_V^2 = 0.5, 0.1$). 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. In a study related to the 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 Tables 1 and 2.

4. Conclusions

We have given a different approach to estimate BMI rather than the available method [21]. This study is used for systematic sampling by using auxiliary variables. The different

measures of the body are used as auxiliary variables. From the study, we may conclude that a difference estimator under systematic sampling has maximum efficiency in the estimation of the body mass index. The efficacy of the methods depends on the correlation between the body mass index and the circumference of the different measures of the body. The correlation coefficient for the body measurement of the hip, abdomen, and thigh is good, so these variables provide better estimation for the body mass index when the circumferences of these parts are used as auxiliary variables. The circumferences of body parts the wrist, forearm, and ankle have the least correlation coefficient with the body mass index and thus may not be used in the estimation of BMI. From the tables, we can also conclude that the ratio estimator and difference estimator are always more efficient than the unbiased mean estimator. So it is better to use ratio and difference 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 difference estimators is better in the presence of observational error. Also, the presence of observational error does not change the properties of the estimators. From Tables 1 and 2, we can conclude the effect of the observational error on mean square error. The above study provided an easy approach and the simplest way to obtain the BMI estimation with and without observational error. Thus, the suggested method may be used by statisticians for this problem and many other similar problems in the estimation of parameters of a natural population.

5. Limitations of Study

The present study proposes a simple method to estimate BMI. Although, the current methodology is confined to the homogeneous population or natural population or population for close geographical areas. The strengths contain the fact that BMI is cheap and relatively easy to use. The weaknesses include the fact that BMI percentiles are not extensively used, and the classification of BMI percentiles may not satisfactorily define the risk of comorbid conditions. In addition, for stratifying children and adolescents with a very high BMI, percentiles are not optimal. In spite of limitations, BMI and BMI percentiles have immense utility in the clinical setting, and the impending to be even more constructive as BMI is used more frequently and more suitably by primary care providers.

Abbreviations

BMI: Body mass index
MSE: Mean square error
PRE: Percent relative efficiency.

Data Availability

All the relevant data are included in the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

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 the authors read and approved the final manuscript.

Acknowledgments

The authors thank the editor Professor Peter Dabnichki and learned referees for their constructive suggestions leading to improving the quality of contents and presentation of the original manuscript.

References

- [1] W. G. Cochran, "Relative accuracy of systematic and stratified random samples for a certain class of populations," *The Annals of Mathematical Statistics*, vol. 17, no. 2, pp. 164–177, 1946.
- [2] M. N. Murthy, *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India, 1967.
- [3] P. Biemer and S. L. Stokes, "Approaches to the modeling of measurement error," *Measurement Error in Surveys*, pp. 485–516, Wiley, Hoboken, NJ, USA, 1991.
- [4] W. A. Fuller, *Measurement Error Models*, John Wiley & Sons, Hoboken, NJ, USA, 1987.
- [5] C.-L. Cheng and J. W. Van Ness, "On estimating linear relationships when both variables are subject to errors," *Journal of the Royal Statistical Society: Series B*, vol. 56, no. 1, pp. 167–183, 1994.
- [6] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC Press, Boca Raton, FL, USA, 2006.
- [7] Shalabh, "Ratio method of estimation in the presence of measurement error," *Journal of the Indian Society of Agricultural Statistics*, vol. 50, no. 2, pp. 150–155, 1997.
- [8] S. Maneesha and R. K. Singh, "Role of regression estimator involving measurement errors," *Brazilian Journal of Probability and Statistics*, vol. 16, pp. 39–46, 2002.
- [9] L. N. Sahoo, R. K. Sahoo, and S. C. Senapati, "An empirical study on the accuracy of ratio and regression estimators in the presence of measurement error," *Monte Carlo Methods and Applications*, vol. 12, no. 5, pp. 495–501, 2006.
- [10] N. Singh, G. K. Vishwakarma, and J. M. Kim, "Computing the effect of measurement error on efficient variant of ratio and product estimator using auxiliary variable," *Communications in Statistics-Simulation and Computation*, 2019.
- [11] N. Singh and G. K. Vishwakarma, "A generalised class of estimator of population mean with the combined effect of measurement errors and non-response in sample survey," *Revista Investigación Operacional*, vol. 40, no. 2, pp. 275–285, 2019.
- [12] W. G. Cochran, "Errors of measurement in statistics," *Technometrics*, vol. 10, no. 4, pp. 637–666, 1968.
- [13] W. Gautschi, "Some remarks on systematic sampling," *The Annals of Mathematical Statistics*, vol. 28, no. 2, pp. 385–394, 1957.
- [14] W. G. Madow, "On the theory of systematic sampling, II," *The Annals of Mathematical Statistics*, vol. 20, no. 3, pp. 333–354, 1949.
- [15] W. G. Madow, "On the theory of systematic sampling, III. Comparison of centered and random start systematic sampling," *The Annals of Mathematical Statistics*, vol. 24, no. 1, pp. 101–106, 1953.
- [16] R. M. Williams, "The variance of the mean of systematic samples," *Biometrika*, vol. 43, no. 1-2, pp. 137–148, 1956.
- [17] A. A. Hasel, "Estimation of volume in timber stands by strip sampling," *The Annals of Mathematical Statistics*, vol. 13, no. 2, pp. 179–206, 1942.
- [18] A. L. Griffith, "The efficiency of enumerations," *Indian Forest Leaflets*, Forest- Research Institute, Dehradun, India, 1945.
- [19] A. K. P. C. Swain, "The use of systematic sampling in ratio estimate," *Journal of Indian Statistical Association*, vol. 2, no. 213, pp. 160–164, 1964.
- [20] N. D. Shukla, "Systematic sampling and product method of estimation," in *Proceedings of the all India Seminar on Demography and Statistics*, BHU, Varanasi, India, 1971.
- [21] Z. Khan, J. Shabbir, S. Gupta, and A. Shamim, "An optimal systematic sampling scheme," *Journal of Statistical Computation and Simulation*, vol. 90, no. 11, Article ID 29023, 2020.
- [22] Z. Khan, J. Shabbir, and S. Gupta, "Generalized systematic sampling," *Communications in Statistics - Simulation and Computation*, vol. 44, no. 9, pp. 2240–2250, 2015.
- [23] K. W. Penrose, A. G. Nelson, and A. G. Fisher, "Generalized body composition prediction equation for men using simple measurement techniques," *Medicine & Science in Sports & Exercise*, vol. 17, no. 2, p. 189, 1985.
- [24] A. R. Behnke and J. H. Wilmore, *Evaluation and Regulation of Body Build and Composition*, Prentice-Hall, Hoboken, NJ, USA, 1974.
- [25] W. E. Siri, "The gross composition of the body," *Advances in Biological and Medical Physics*, vol. 4, pp. 239–280, 1956.
- [26] R. W. Johnson, "Fitting percentage of body fat to simple body measurements," *Journal of Statistics Education*, vol. 4, no. 1, pp. 1–8, 1996.
- [27] C. N. Bouza, S. M. Allende-Alonso, G. K. Vishwakarma, and N. Singh, "Estimation of optimum sample size allocation: an illustration with body mass index for evaluating the effect of a dietetic supplement," *International Journal of Biomathematics*, vol. 12, no. 8, Article ID 1950086, 2019.