DHM2020
L. Hanson et al. (Eds.)

doi:10.3233/ATDE200005

Skewed Boundary Confidence Ellipses for Anthropometric Data

Erik BROLIN [a,1], Dan HÖGBERG [a] and Lars HANSON [a,b]
[a] *University of Skövde, School of Engineering Science, 541 28 Skövde, Sweden*
[b] *Scania CV, Södertälje, Sweden*

Abstract. Some anthropometric measurements, such as body weight often show a positively skewed distribution. Different types of transformations can be applied when handling skewed data in order to make the data more normally distributed. This paper presents and visualises how square root, log normal and, multiplicative inverse transformations can affect the data when creating boundary confidence ellipses. The paper also shows the difference of created manikin families, i.e. groups of manikin cases, when using transformed distributions or not, for three populations with different skewness. The results from the study show that transforming skewed distributions when generating confidence ellipses and boundary cases is appropriate to more accurately consider this type of diversity and correctly describe the shape of the actual skewed distribution. Transforming the data to create accurate boundary confidence regions is thought to be advantageous, as this would create digital manikins with enhanced accuracy that would produce more realistic and accurate simulations and evaluations when using DHM tools for the design of products and workplaces.

Keywords. Anthropometry, Skewness, Boundary Cases, Confidence Ellipses

1. Introduction

Digital human modelling (DHM) tools enable simulations and analyses of ergonomics in virtual environments. Functionality for consideration of anthropometric diversity and methods for ergonomics evaluations are central features when using DHM tools for product and production development to ensure that the design fits the intended proportion of the targeted population from a physical perspective. Working with anthropometric data, using mathematical and statistical treatment, it is possible to create boundary confidence regions in the form of ellipses or ellipsoids [1]. This is done under the assumption that the measurement distribution can be approximated with a normal distribution. However, body weight, width and circumference measurements as well as muscular strength often show a positively skewed distribution [2-3]. Comparing older data of a relatively fit population, e.g. ANSUR with military data from 1989 [4], to a more recent civilian population from 3D body scan studies, e.g. CAESAR data from 2002 [5], and to an even more recent data with a bigger sample, e.g. NHANES from 2007 [6], shows clear differences in skewness between both fitness level of different populations and the year when the data was measured [7]. Different types of transformations can be applied when handling skewed data in order to make the data

[1] Corresponding Author, Email: erik.brolin@his.se.

more normally distributed [8-9]. The transformed data could potentially be more representative when creating boundary confidence ellipsoids and selecting subsequent manikin cases, i.e. virtual human models with a number of key anthropometric dimensions defined. This paper visualises how square root, log normal and, multiplicative inverse transformations can affect the data when creating boundary confidence ellipses, i.e. make the shape of the created confidence ellipses more similar to and accurately describing the original data. The paper also shows the difference of created manikin families, i.e. groups of manikin cases, when using transformed distributions or not, for populations with different skewness.

2. Method

The applied methodology for consideration of skewness when defining boundary case manikins includes two parts:

- the first part handles transformation of skewed data to make it more normally distributed as well as transforming the generated boundary case data back to real values for visualisation and input for subsequent DHM simulation
- the second part handles the generation of boundary confidence region and definition of cases on that region

2.1. Transformation of skewed data

Skewness is a measure that describes the asymmetry of the distribution where a positive skew indicates that there is a number of persons that have values relatively far from the median value, thus forming a tail on the right side of the distribution. Skewness is here defined as

$$\frac{n\sum_{i=1}^{n}(x_i-\mu)^3}{\left((n-1)(n-2)\right)\sigma^3}, \tag{1}$$

where n is the sample size, μ the sample mean and σ the standard deviation. Different methods can be used to consider positively skewed anthropometric data, e.g. using body mass index (BMI) instead of body weight or using the positively skewed log normal distribution instead of the symmetrical normal distribution [7]. Another general method for transforming data, box-cox transformation [8], can also be used. In this study, three different methods for transforming body weight data, *w*, have been evaluated:

- square-root $w^{1/2}$ (2)
- log normal $\ln(w)$ (3)
- reciprocal or multiplicative inverse w^{-1} (4)

The three transformation methods have in this study been applied only on original body weight data and not using BMI due to space limitations. After the transformed data have been used in statistical methods, e.g. for the generation of boundary confidence region and cases, the data can be transformed back into real values. The three methods are transformed back as:

- square-root $(w^{1/2})^2$ (5)

- log normal $e^{\ln(w)}$ (6)
- reciprocal or multiplicative inverse $(w^{-1})^{-1}$ (7)

2.2. Generation of boundary region and definition of cases

The transformed data is used, together with stature data, to form boundary regions in the shape of two-dimensional confidence ellipses, and then boundary cases manikins are defined on edges of these ellipses [1]. The mathematical process for calculating boundary case data based on the correlation matrix is described in Table 1.

Table 1. Mathematical process for calculating boundary case data based on the correlation matrix.

Description:		Mathematical definition:
1.	Correlation matrix	$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$
2.	Eigenvalue matrix	$\lambda = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$
3.	Eigenvector matrix	$\boldsymbol{v} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$
4.	Two dimensional scale factor k (P=95%)	$k = \sqrt{x_2^2(1-0.95)} = 2.45$
5.	Matrix of scaled axes	$A = \begin{bmatrix} 2.45 \times \sqrt{1+\rho} & 0 \\ 0 & 2.45 \times \sqrt{1-\rho} \end{bmatrix}$
6.	Experimental design plan	$E = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$
7.	Boundary cases in standardised space	$C_Z = \boldsymbol{v}(EA)^T$
8.	Boundary cases in real or transformed space	$C = [C_{Z-x} \times \sigma_x + \mu_x, C_{Z-y} \times \sigma_y + \mu_y]$

To give each distribution the same significance in the calculations the data is, in addition to the previous transformation due to skewness, transformed into standard normal distributions in which the mean values are 0 and standard deviation are 1 [10]. The two-dimensional confidence ellipses are defined by the length and direction of the axes, which are given by the eigenvalues and eigenvectors of the correlation matrix. In a two-dimensional standard normal distribution the eigenvalues and eigenvectors are relatively easy to calculate (Table 1). To get the final length of the ellipses axes the square root of the eigenvalues are multiplied with the scale factor k. The scale factor k is calculated from the chi-squared distribution, in this case with two degrees of freedom since we have two dimensions and with a sought accommodation level of 95 %, i.e. the confidence ellipses are supposed to cover 95 % of the data points. The boundary cases are calculated by using an experimental design matrix that defines four axis cases on the edges of the two axes of the ellipses and four box cases at the corners of a rectangle that spans the biggest area inside the ellipses [11]. The boundary cases are then calculated by multiplying the eigenvector matrix with the transpose of the experimental design plan multiplied with the matrix of scaled axes. The values for the boundary cases are in the end transformed back from standard normal distribution to the real space or transformed space due to skewness.

3. Results

The suggested method was applied on three different populations: 1. ANSUR with military data from 1989 [4], 2. CAESAR with civilian data from 2002 [5], and 3. NHANES with more recent data and a bigger sample from 2011-2014 [12]. The study was, due to space limitations, limited to female data but the three population show a range of skewness and correlation for stature and body weight (Table 2).

Table 2. Skewness and correlation of stature and body weight for three different populations [4,5,12].

Data source	Skewness		Correlation between stature and body weight
	Stature	Body weight	
ANSUR [4]	0.139	0.536	0.529
CAESAR [5]	0.129	1.748	0.296
NHANES [12]	0.020	1.230	0.329

The different transformation methods affect both the skewness and the correlation to stature (Table 3). This can also be visualised using quantile-quantile plots (Q-Q plot) (Figure 1-3). The resulting boundary ellipses with eight boundary cases for each transformation as well as the original weight data are visualised in Figure 4-6 and the corresponding measurement and percentile values are presented in Table 4-6.

Table 3. Skewness and correlation to stature for the three transformation methods.

Data source	Skewness			Correlation to stature		
	Square-root, $w^{½}$	*Log normal,* $\ln(w)$	*Reciprocal,* w^{-1}	Square-root, $w^{½}$	*Log normal,* $\ln(w)$	*Reciprocal,* w^{-1}
ANSUR [4]	0.325	0.119	0.287	0.531	0.533	-0.533
CAESAR [5]	1.285	0.882	-0.203	0.314	0.331	-0.359
NHANES [12]	0.759	0.352	0.388	0.339	0.347	-0.354

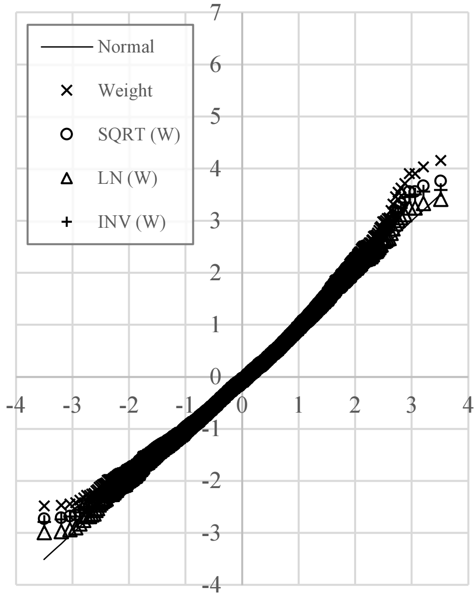

Figure 1. Q-Q plot of the positively skewed distribution of body weight as well as the three transformation methods, data from ANSUR [4].

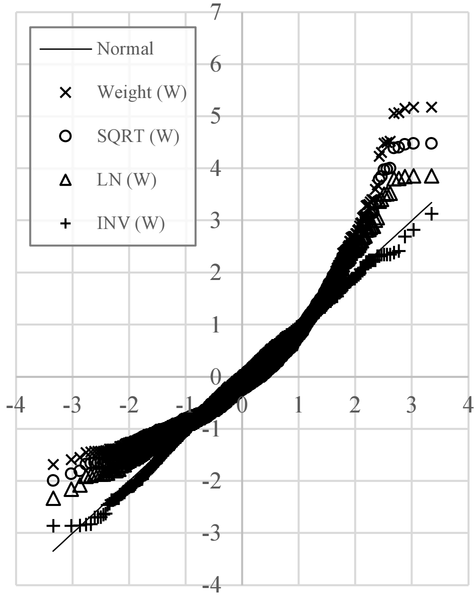

Figure 2. Q-Q plot of the positively skewed distribution of body weight as well as the three transformation methods, data from CAESAR [5].

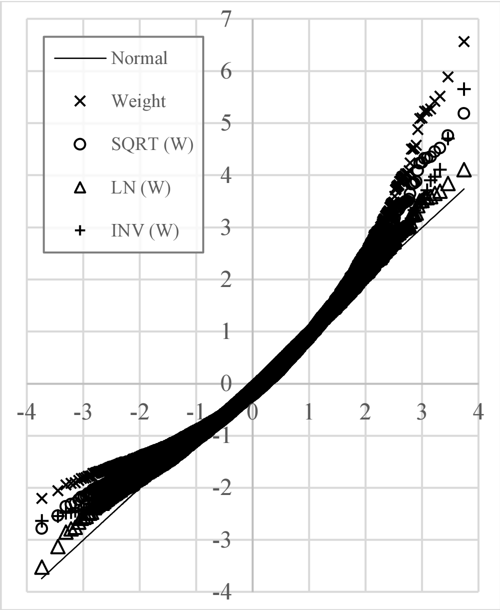

Figure 3. Q-Q plot of the positively skewed distribution of body weight as well as the three transformation methods, data from NHANES [12].

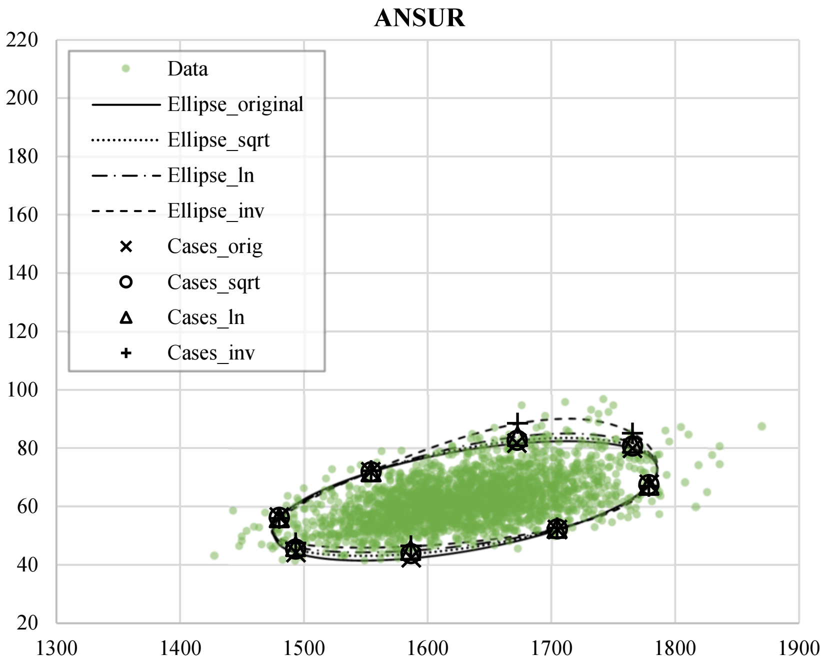

Figure 4. Confidence ellipses and boundary cases for the original body weight data and for the three transformation methods caption, data from ANSUR [4]. Stature (mm) on x-axis and weight (kg) on y-axis.

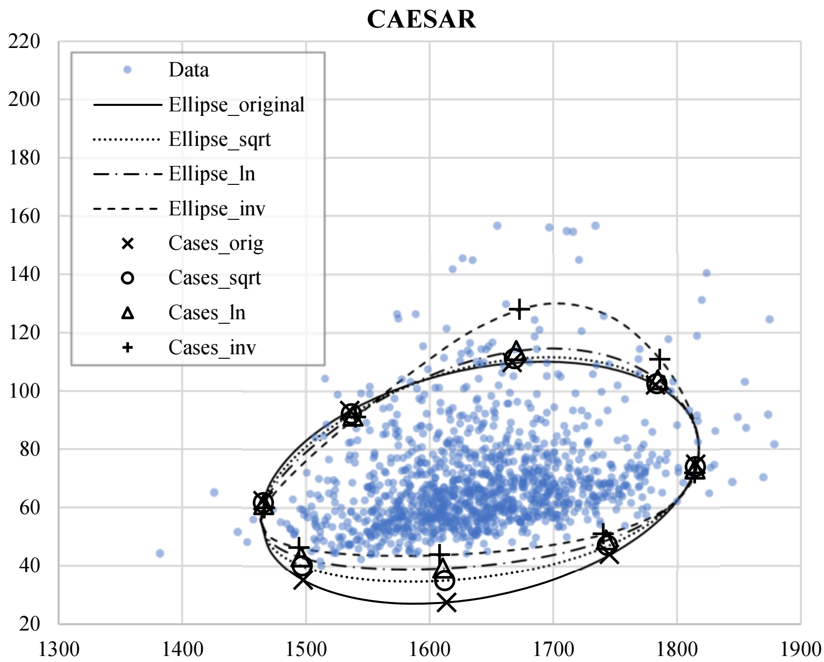

Figure 5. Confidence ellipses and boundary cases for the original body weight data and for the three transformation methods caption, data from CAESAR [5]. Stature (mm) on x-axis and weight (kg) on y-axis.

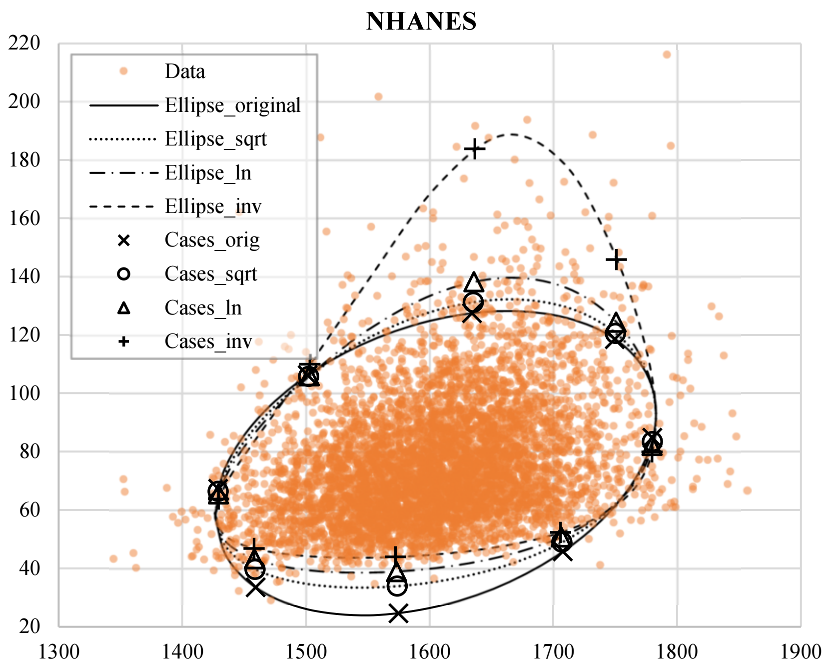

Figure 6. Confidence ellipses and boundary cases for the original body weight data and for the three transformation methods caption, data from NHANES [12]. Stature (mm) on x-axis and weight (kg) on y-axis.

Table 4. Boundary cases for each transformation as well as the original weight data and the corresponding measurement and percentile values, data from ANSUR [4].

ANSUR	**Stature** [mm,%-ile]	**Body weight (w)** [kg,%-ile]	**Square-root** $w^{½}$ [kg,%-ile]	**Log normal ln(w)** [kg,%-ile]	**Reciprocal** w^{-1} [kg,%-ile]
Case 1	1493 (1.6)	44 (0.6)	45 (0.9)	46 (1.4)	47 (2.6)
Case 2	1766 (98.1)	80 (97.1)	81 (97.5)	82 (97.9)	85 (99.1)
Case 3	1705 (87.6)	52 (10.5)	52 (11.1)	53 (11.9)	53 (12.2)
Case 4	1554 (11.1)	72 (88.0)	72 (88.0)	72 (88.0)	72 (88.7)
Case 5	1586 (25.7)	42 (0.2)	44 (0.5)	45 (0.8)	46 (1.5)
Case 6	1779 (98.9)	68 (77.3)	67 (76.5)	67 (75.9)	67 (74.9)
Case 7	1480 (0.7)	56 (26.0)	56 (25.5)	56 (25.1)	56 (24.2)
Case 8	1672 (75.0)	82 (97.9)	83 (98.3)	84 (98.9)	89 (99.5)

Table 5. Boundary cases for each transformation as well as the original weight data and the corresponding measurement and percentile values, data from CAESAR [5].

CAESAR	**Stature** [mm,%-ile]	**Body weight (w)** [kg,%-ile]	**Square-root** $w^{½}$ [kg,%-ile]	**Log normal ln(w)** [kg,%-ile]	**Reciprocal** w^{-1} [kg,%-ile]
Case 1	1497 (1.3)	35 (0.0)	40 (0.1)	43 (0.3)	46 (1.9)
Case 2	1784 (97.2)	102 (95.0)	103 (95.0)	104 (95.3)	111 (96.9)
Case 3	1744 (91.7)	44 (0.4)	47 (2.3)	49 (4.1)	51 (6.9)
Case 4	1537 (6.5)	93 (92.0)	92 (91.2)	91 (90.5)	91 (90.5)
Case 5	1612 (36.6)	27 (0.0)	35 (0.0)	39 (0.0)	44 (0.3)
Case 6	1815 (98.5)	75 (74.5)	74 (73.6)	73 (71.8)	72 (69.0)
Case 7	1466 (0.5)	62 (42.1)	62 (39.8)	61 (38.0)	60 (34.2)
Case 8	1668 (67.7)	110 (96.5)	111 (96.9)	114 (97.2)	128 (99.0)

Table 6. Boundary cases for each transformation as well as the original weight data and the corresponding measurement and percentile values, data from NHANES [12].

NHANES	**Stature** [mm,%-ile]	**Body weight (w)** [kg,%-ile]	**Square-root** $w^{½}$ [kg,%-ile]	**Log normal ln(w)** [kg,%-ile]	**Reciprocal** w^{-1} [kg,%-ile]
Case 1	1459 (2.3)	34 (0.0)	40 (0.3)	43 (1.0)	47 (2.8)
Case 2	1750 (97.6)	119 (95.7)	121 (96.3)	124 (96.9)	146 (99.1)
Case 3	1707 (92.4)	46 (2.2)	49 (4.9)	51 (7.1)	52 (9.2)
Case 4	1502 (8.3)	106 (91.0)	106 (90.7)	106 (90.9)	110 (92.9)
Case 5	1574 (33.6)	25 (0.0)	34 (0.1)	39 (0.2)	44 (1.4)
Case 6	1780 (99.1)	85 (71.8)	83 (70.1)	82 (68.6)	80 (65.1)
Case 7	1429 (0.6)	67 (39.7)	66 (37.7)	66 (35.7)	64 (31.9)
Case 8	1635 (66.3)	128 (97.4)	131 (97.9)	138 (98.7)	184 (99.8)

4. Discussion

The results from the study show that transforming skewed distributions when generating confidence ellipses and boundary cases is possible, suitable and often even necessary to more accurately consider this type of diversity. The shape of the created confidence ellipses are more similar to and accurately describes the original skewed data for the CAESAR and NHANES female population. For the ANSUR female population, which have a less skewed weight distribution, transforming the data does not affect the shape of the ellipses to any great extent. But the tested transformation methods does not either create any subsequent issues or inaccuracies when generating confidence ellipses.

When looking at the measurement and percentile values of the generated cases it is evident that not transforming the data of a skewed weight distribution will generate cases that are relatively far outside or far inside the actual distribution. Case 5 for both CAESAR and NHANES when not transforming the data have an extremely low weight of 25 and 27, respectively. Not transforming the data will also lead to an underestimation of the higher percentile values. Case 8 for all three populations have the highest weight values but at the same time relatively low values when not transforming the data. When transforming the data, the values increases from square-root to log normal and from log normal to multiplicative inverse. For the NHANES data the multiplicative inverse transformation method leads to a boundary case with a weight of 184 kg which can seem extremely high, however that case can still be found within the actual distribution. Future research will also include BMI and other non-normal distributed anthropometric variables as well testing additional transformation methods. To have accurate boundary confidence regions is thought to be advantageous, whether the manikin cases are selected as boundary cases located towards the edges or as distributed cases spread throughout a region, randomly or by some systematic approach. This would give digital human models with anthropometry that better resembles the variance and diversity that exist within human populations. This would in turn produce more realistic and accurate simulations and evaluations and thus give better assistance to engineers and designers using DHM tools when developing products and workplaces.

Acknowledgement

This work has been made possible with support from the Knowledge Foundation and the associated INFINIT research environment at the University of Skövde (projects: Synergy Virtual Ergonomics and ADOPTIVE), and with support from Vinnova in the VIVA project, and SAFER - Vehicle and Traffic Safety Centre at Chalmers, Sweden, and by the participating organizations. This support is gratefully acknowledged.

References

[1] Brolin E, Högberg D, Hanson L. Description of boundary case methodology for anthropometric diversity consideration, International Journal of Human Factors Modelling and Simulation, 2012, 3(2), 204–223. [online] https://doi.org/10.1504/IJHFMS.2012.051097.

[2] Vasu M, Mital A. Evaluation of the validity of anthropometric design assumptions, Industrial Ergonomics, 2000, 26, 19–37.

[3] Pheasant S, Haslegrave CM. Bodyspace: Anthropometry, Ergonomics and the Design of Work. CRC Press, Boca Raton, 2006.

[4] Gordon CC, Churchill T, Clauser CE, Bradtmiller B, Mcconville JT, Tebbetts I, Walker RA. 1988 Anthropometric Survey of US Army Personnel: Methods and Summary Statistics. U.S. Army Natick Research, Development and Engineering Center, Natick, MA, 1989.
[5] Robinette KM, Blackwell S, Daanen H, Boehmer M, Fleming S, Brill T, Hoeferlin D, Burnsides D. Civilian American and European Surface Anthropometry Resource (CAESAR). Final report: Air Force Research laboratory, Wright-Patterson AFB, OH, and Society of Automotive Engineers International, Warrendale, PA, 2002.
[6] U.S. Centers for Disease Control and Prevention. National Health and Nutrition. Examination Survey (NHANES). National Center for Health Statistics, 2008. http://www.cdc.gov/nchs/nhanes.htm.
[7] Parkinson MB, Reed MP. Creating virtual user populations by analysis of anthropometric data. Int. J. Industrial Ergonomics, 2010, 40(1), 106-111.
[8] Box GE, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 1964, 26(2), 211-243.
[9] Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr, 1990, 44(1), 45-60.
[10] Glenberg A, Andrzejewski M. Learning from Data: An Introduction to Statistical Reasoning, 3rd ed. Routledge Academic, New York, 2007.
[11] Bertilsson E, Högberg D, Hanson L. Using experimental design to define boundary manikins. Work: A Journal of Prevention, Assessment and Rehabilitation, 2012, 41, 4598–4605 [online] https://doi.org/10.3233/WOR-2012-0075-4598.
[12] Fryar CD, Gu Q, Ogden CL, Flegal KM. Anthropometric reference data for children and adults: United States, 2011–2014, National Center for Health Statistics. Vital Health Stat, 2016, 3(39).