International Journal of Statistics and Probability; Vol. 13, No. 3; September 2024
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Approximation of Sample Mean and Standard Deviation from Percentiles and Application to Reference Model for Sprint Time of Elite Soccer Players

Kyra Batarse[1], Jessica Naranjo[1], Jakob Garcia[1] and Steven Kim[1*]

[1] Department of Mathematics and Statistics California State University, Monterey Bay

Correspondence: Steven Kim, Department of Mathematics and Statistics California State University, Monterey Bay

Received: June 5, 2024 Accepted: August 14, 2024 Online Published: August 31, 2024

doi:10.5539/ijsp.v13n3p25 URL: https://doi.org/10.5539/ijsp.v13n3p25

Abstract

The original goal of this project was to conduct meta-analysis and develop a reference model for 20-meter sprint time of elite soccer players with respect to age group. According to the likelihood function, we found that age-specific sample means and standard deviations are sufficient statistics. However, instead of the two sufficient statistics, the sample 10-th percentile, 90th-percentile, and quartiles were given to us. To deal with this unexpected challenge, we studied and compared methods of approximating the sample mean and standard deviation given various combinations of percentiles. Simulation studies showed that the likelihood method presented in this paper is superior when the sample 10-th percentile, 90-th percentile, and quartiles are given, and we applied the likelihood method to the meta-analysis and developed the reference model for 20-meter sprint time.

Keywords: order statistics, percentiles, sufficient statistics, piecewise regression, weighted least square estimation

1. Introduction

Among various factors of high soccer performance, sprint ability is an important physical factor (Haugen et al., 2014; Angius et al., 2013) as the most frequent action before a goal is sprinting (Faude et al., 2012; Haugen et al., 2014). The sprint ability is typically measured by how fast a 20-meter sprint is completed since most sprints are not longer than 20 meters during a match (Barnes et al., 2014). Like in most sports, the physical ability develops at young ages between 12 to 18 years (Le Gall et al. 2002; Mendez-Villanueva et al., 2011; Maly et al., 2015; Nikolaidis et al., 2016), and an age-specific reference model would be practically useful for soccer coaches and researchers. In this regard, the original objective of our statistical project was to develop an age-specific reference model for the 20-meter sprint time of young elite soccer players.

We found a valuable resource for our statistical project in literature. Nikolaidis et al. (2016) provided data of the 20-meter sprint performance of 474 elite soccer players aged 9 to 35 years. The researchers partitioned all observed players into 15 age groups (U10, U11, ..., U21, U25, U30, and U35) and provided age-specific reference values for 20-meter sprint time by the five percentile values including the 10-th percentile, 25-th percentile (first quartile), 50-th percentile (median), 75-th percentile (third quartile), and 90-th percentile (Table 4 of Nikolaidis et al., 2016). As shown in Figure 1, the reference values (time in seconds) are not smooth with respect to the age group, and we wanted to apply a regression model to provide more realistic reference values.

We had two statistical challenges. The first challenge we encountered was the absence of sufficient statistics. According to the likelihood function under our model assumptions, the sample mean and standard deviation of each age group are needed (Section 2). Nikolaidis et al. (2016) provided the five percentiles for each age group, instead of the two sufficient statistics. We then had to approximate the sample mean and standard deviation using the five percentiles. The second challenge was how to approximate the two sufficient statistics given the five percentiles.

The approximation of mean and standard deviation using the distance between percentiles has been proposed in the past (Pearson & Tukey, 1965), and many approximation methods have been developed given the typical 3- or 5-number summaries such as the sample minimum, first quartile, median, third quartile, and maximum (McGrath et al., 2020; Luo et al., 2018; Bland, 2015; Wan et al., 2014; Hozo et al., 2005). A Bayesian approximation method has been proposed, and it was shown to perform well under skewed or heavy-tailed distributions for the estimation of the standard deviation (Kwon & Reis, 2005; Kwon et al., 2021). However, these studies have not utilized the 10-th percentile and 90-th percentile with the quartiles for the approximation. Wan et al. (2014) addressed a similar problem based on the expected distance

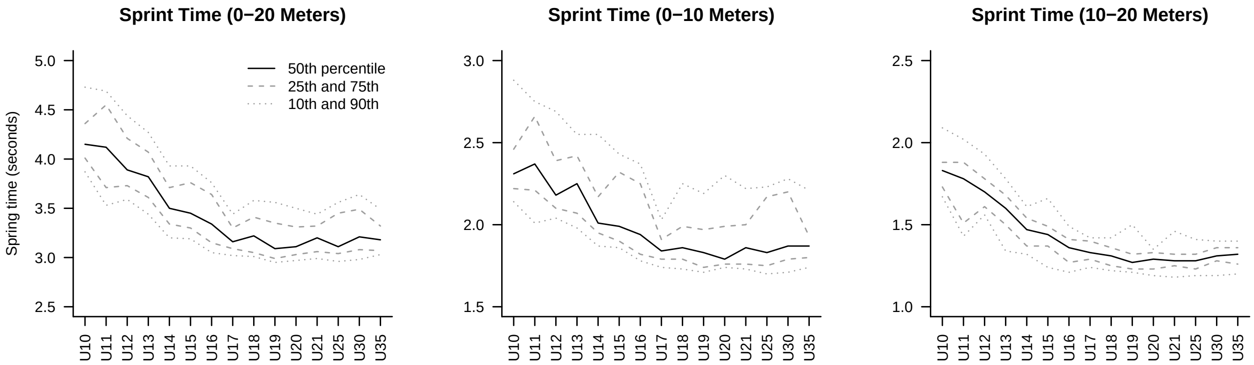

Figure 1. Age-specific reference sprint time (seconds) for 0-20 meters (left), 0-10 meters (middle), and 10-20 meters (right) among elite soccer players (Nikolaidis et al., 2016)

between two ordered statistics. We could see that their approach can be modified for our case (Section 3.1), and we considered another approach based on the likelihood function with the partial information (Section 3.2). We compared the two approaches using simulations (Section 4) and applied the better approximation method to our original objective (Section 5).

2. Model

2.1 Likelihood Function

We let y_{ij} be the sprint time in seconds observed from the j-th subject in the i-th age group for $j = 1, 2, \dots, n_i$ and $i = 1, 2, \dots, m$. Let a_i denote the i-th age group in years. Without loss of generality, we let $a_1 < a_2 < \cdots < a_m$. We assume all subjects are independent and $y_{ij} \sim N(\mu_i, \nu)$ where

$$\mu_i \equiv \mu_i(\beta_0, \beta_1, \delta) = \begin{cases} \beta_0 + \beta_1 a_i & \text{for } a_i < \delta\,, \\ \beta_0 + \beta_1 \delta & \text{for } a_i \geq \delta\,. \end{cases}$$

This is a special case of piecewise regression, which is also referred to as changepoint regression (Julious, 2001). We assume that the expected sprint time μ_i changes until the age of δ years. This assumption is based on a literature review that young players improve their sprint ability between 12 and 18 years (Le Gall et al. 2002; Mendez-Villanueva et al., 2011; Maly et al., 2015). In this model, we treat δ as an unknown integer-valued parameter.

For concise expression, let $\boldsymbol{\theta} = (\beta_0, \beta_1, \delta, \nu)^T$ be the vector of unknown model parameters, $t = \sum_{i=1}^{m} n_i$ be the total number of subjects combining all m age groups, $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ be the sample mean of the i-th age group, and $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ is the sample variance of the i-th age group. Using these notations, the likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(y_{ij} - \mu_i)^2}{2\nu}} = (2\pi\nu)^{-\frac{t}{2}} e^{-\frac{Q}{2\nu}}$$

where

$$\begin{aligned}
Q &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i+\bar{y}_i-\mu_i)^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[(y_{ij}-\bar{y}_i)^2+2(\bar{y}_i-\mu_i)(y_{ij}-\bar{y}_i)+(\bar{y}_i-\mu_i)^2\right] \\
&= \sum_{i=1}^{m}(n_i-1)s_i^2+2\sum_{i=1}^{m}(\bar{y}_i-\mu_i)\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)+\sum_{i=1}^{m}n_i(\bar{y}_i-\mu_i)^2 \\
&= \sum_{i=1}^{m}(n_i-1)s_i^2+\sum_{i=1}^{m}n_i(\bar{y}_i-\mu_i)^2 \\
&= \sum_{i=1}^{m}(n_i-1)s_i^2+Q^*(\boldsymbol{\beta},\delta)
\end{aligned}$$

where

$$Q^*(\boldsymbol{\beta},\delta)=\sum_{i=1}^{m}n_i(\bar{y}_i-\mu_i)^2\,.$$

Therefore, the log-likelihood function can be written as

$$l(\boldsymbol{\theta})=-\frac{t}{2}\cdot\ln(2\pi\nu)-\frac{\sum_{i=1}^{m}(n_i-1)s_i^2+Q^*(\boldsymbol{\beta},\delta)}{2\nu}\,.$$

2.2 Maximum Likelihood Estimation for $\boldsymbol{\theta}$

Given (δ,ν), the maximization of $l(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}=(\beta_0,\beta_1)^T$ is equivalent to the minimization of $Q^*(\boldsymbol{\beta},\delta)$ with respect to $\boldsymbol{\beta}$. Note that $Q^*(\boldsymbol{\beta},\delta)$ can be expressed using matrix notation as follows. Let

$$\mathbf{N}^{1/2}=\mathrm{diag}(\sqrt{n_1},\ \sqrt{n_2},\ldots,\ \sqrt{n_m})=\begin{pmatrix}\sqrt{n_1} & 0 & \ldots & 0\\ 0 & \sqrt{n_2} & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \sqrt{n_m}\end{pmatrix}$$

be the $m\times m$ diagonal matrix,

$$\bar{\mathbf{Y}}=\begin{pmatrix}\bar{y}_1\\ \bar{y}_2\\ \ldots\\ \bar{y}_m\end{pmatrix}$$

be the $m\times 1$ vector, and

$$\mathbf{X}_\delta=\begin{pmatrix}1 & a_1\\ 1 & a_2\\ \vdots & \vdots\\ 1 & a_{k-1}\\ 1 & \delta\\ \vdots & \vdots\\ 1 & \delta\end{pmatrix}$$

be the $m\times 2$ matrix, where a_{k-1} is the oldest age group younger than δ. Given δ, we can express

$$\begin{aligned}
Q^*(\boldsymbol{\beta},\delta) &= [\mathbf{N}^{1/2}(\bar{\mathbf{Y}}-\mathbf{X}_\delta\boldsymbol{\beta})]^T[\mathbf{N}^{1/2}(\bar{\mathbf{Y}}-\mathbf{X}_\delta\boldsymbol{\beta})] \\
&= (\bar{\mathbf{Y}}^*-\mathbf{X}_\delta^*\boldsymbol{\beta})^T(\bar{\mathbf{Y}}^*-\mathbf{X}_\delta^*\boldsymbol{\beta})
\end{aligned}$$

where $\bar{\mathbf{Y}}^* = \mathbf{N}^{1/2}\bar{\mathbf{Y}}$ and $\mathbf{X}^*_\delta = \mathbf{N}^{1/2}\mathbf{X}_\delta$. The minimization of $Q^*(\boldsymbol{\beta}, \delta)$ becomes the problem of the weighted least square estimation (WLSE; Kutner et al., 2004) with respect to $\boldsymbol{\beta}$ as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_\delta &= [(\mathbf{X}^*_\delta)^T\mathbf{X}^*_\delta]^{-1}[(\mathbf{X}^*_\delta)^T\bar{\mathbf{Y}}^*] \\ &= (\mathbf{X}^T_\delta\mathbf{N}\mathbf{X}_\delta)^{-1}\mathbf{X}^T_\delta\mathbf{N}\bar{\mathbf{Y}}\end{aligned}$$

where $\mathbf{N} = \text{diag}(n_1, n_2, \ldots, n_m)$, and $\hat{\boldsymbol{\beta}}_\delta$ is the maximum likelihood estimator (MLE) for $\boldsymbol{\beta}$ given δ for any value of $\nu > 0$. Hence, the MLE for $\boldsymbol{\theta} = (\beta_0, \beta_1, \delta, \nu)^T$ can be done as follows.

1. Find $\hat{\boldsymbol{\beta}}_\delta$ given δ for all values of δ under consideration.
2. Evaluate $Q^*(\boldsymbol{\beta}, \delta)$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\delta$ with respect to all values of δ. When $Q^*(\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\delta, \delta)$ is minimized at $\delta = \hat{\delta}$, the MLE for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\hat{\delta}}$.
3. The previous step determines that $\hat{\delta}$ is the MLE for δ.
4. Let $\hat{Q}^*$ be the minimum value of $Q^*(\boldsymbol{\beta}, \delta)$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\delta = \hat{\delta}$. Let $\hat{\nu}$ be the MLE for ν which maximizes $l(\boldsymbol{\theta})$ with respect to ν at $Q^*(\boldsymbol{\beta}, \delta) = \hat{Q}^*$. The closed form solution is

$$\hat{\nu} = \frac{\sum_{i=1}^m (n_i - 1)s_i^2 + \hat{Q}^*}{t}$$

by solving $\frac{\partial l(\boldsymbol{\theta})}{\partial \nu} = 0$ given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\delta = \hat{\delta}$.

2.3 Our Problem

For each age a_i and the sample size n_i in the i-th age group, the MLE for $\boldsymbol{\theta}$ is possible with the sufficient statistics, $\bar{y}_i$ and s_i, for each age group $i = 1, 2, \ldots, m$. Unfortunately, our problem becomes complicated because $\bar{y}_i$ and s_i are not available for $i = 1, 2, \ldots, m$ or we cannot calculate $\bar{y}_i$ and s_i from available information.

In meta-analysis, we often see that researchers report different statistics than the sample mean and standard deviation. Commonly reported statistics are the sample minimum, first quartile, median, third quartile, and maximum without the sample mean and standard deviation. Wan et al. (2014) addressed how to approximate the sample mean and standard deviation when the 5-number statistics are given. However, Nikolaidis et al. (2016) reported the 10-th percentile, first quartile (i.e. 25-th percentile), median (i.e. 50-th percentile), third quartile (i.e. 75-th percentile), and 90-th percentile, and we could not find in literature how to approximate the sample mean and standard deviation when these five percentiles are given.

3. Approximations of Sample Mean and Standard Deviation Using Ordered Statistics

3.1 Distance-based Approximation

We use the following notations similar to the notations used by Wan et al. (2014). Given a sample $(y_1, y_2, \ldots, y_n)$ of size n from a single population, let a be the sample minimum, q_1 be the sample first quartile (i.e. 25-th percentile), m be the sample median (i.e. 50-th percentile), q_3 be the sample third quartile (i.e. 75-th percentile), b be the sample maximum, $\bar{y}$ be the sample mean, and s be the sample standard deviation where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i\,,$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2}\,.$$

Assume $\bar{y}$ and s are not reported. Wan et al. (2014) approximated $\bar{y}$ and s in the following three cases. The first case is when $C_1 = \{a, m, b, n\}$ is known (Section 3.1.1), the second case is when $C_2 = \{a, q_1, m, q_3, b, n\}$ is known (Section 3.1.2), and the third case is when $C_3 = \{q_1, m, q_3, n\}$ is known (Section 3.1.3). In particular, Wan et al. (2014) made significant contribution of approximating s by calculating the expected distance between two ordered statistics more accurately than the past methods (Hozo et al., 2005; bland, 2015), and the general idea is briefly explained in this section (Section 3.1.4). The idea is applied to the fourth case when $C_4 = \{c, q_1, m, q_3, d, n\}$ is given, where c and d are the 10-th percentile and 90-th percentile, respectively (Section 3.1.5).

3.1.1 Approximations under Case 1

Given $C_1 = \{a, m, b, n\}$, Hozo et al. (2005) proposed to approximate $\bar{y}$ and s by

$$\bar{y}_h^* = \frac{a + 2m + b}{4},$$
$$s_h^* = \begin{cases} \frac{1}{\sqrt{12}}\left[(b-a)^2 + \frac{(a-2m+b)^2}{4}\right]^{1/2} & \text{for } n \le 15, \\ \frac{b-a}{4} & \text{for } 15 < n \le 70, \\ \frac{b-a}{6} & \text{for } n > 70. \end{cases} \tag{1}$$

Wan et al. (2014) attempted to improve s_h by

$$s_w^* = \frac{b-a}{2\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)} \tag{2}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard normal distribution, N(0,1), and it can be calculated by the `qnorm` function in R (R Core Team, 2023).

3.1.2 Approximations under Case 2

Given $C_2 = \{a, q_1, m, q_3, b, n\}$, Bland (2015) proposed to approximate $\bar{y}$ and s by

$$\bar{y}_b^{**} = \frac{a + 2q_1 + 2m + 2q_3 + b}{8},$$
$$s_b^{**} = \frac{1}{2}\left(\frac{b-a}{\xi(n)} + \frac{q_3 - q_1}{\eta(n)}\right) \tag{3}$$

where values of $\xi(n)$ and $\eta(n)$ are provided in Wan et al. (2014) for $n = 5, 9, 13, \ldots, 201$ (i.e. for $Q = 1, 2, \ldots, 50$ where $n = 4Q + 1$). Wan et al. (2014) attempted to improve s_b by

$$s_w^{**} = \frac{1}{2}\left[\frac{b-a}{2\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)} + \frac{q_3 - q_1}{2\Phi^{-1}\left(\frac{0.75n-0.125}{n+0.25}\right)}\right]. \tag{4}$$

3.1.3 Approximations Under Case 3

Given $C_3 = \{q_1, m, q_3, n\}$, Wan et al. (2014) proposed to approximate $\bar{y}$ and s by

$$\bar{y}_w^{***} = \frac{q_1 + m + q_3}{3},$$
$$s_w^{***} = \frac{q_3 - q_1}{2\Phi^{-1}\left(\frac{0.75n-0.125}{n+0.25}\right)}. \tag{5}$$

3.1.4 Expected Distance between Two Ordered Statistics

Wan et al. (2014) considered the expected distance between two ordered statistics to derive s_w^* in Equation (2), s_w^{**} in Equation (4), and s_w^{***} in Equation (5). Assume $(Y_1, Y_2, \ldots, Y_n)$ is an i.i.d. sample of size n from a normal distribution with mean μ and standard deviation σ. If we standardize Y_i as

$$Z_i = \frac{Y_i - \mu}{\sigma}, \quad i = 1, 2, \ldots, n,$$

then $(Z_1, Z_2, \ldots, Z_n)$ is an i.i.d. sample of size n from the standard normal distribution, $N(0, 1)$. Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the probability density function (PDF) and cumulative distribution function (CDF) of $N(0, 1)$. For any r-th order statistic $Z_{(r)}$, the PDF is given by

$$f_{Z_{(r)}}(z) = \frac{n!}{(r-1)!\,(n-r)!}[\Phi(z)]^{r-1}[1-\Phi(z)]^{n-r}\phi(z), \quad -\infty < z < \infty,$$

and the expectation of $Z_{(r)}$ is given by

$$\mathcal{E}(Z_{(r)}) = \frac{n!}{(r-1)!\,(n-r)!}\int_{-\infty}^{\infty} z\,[\Phi(z)]^{r-1}[1-\Phi(z)]^{n-r}\phi(z)\,dz \tag{6}$$

(David & Nagaraja, 2003). Wan et al. (2014) considered a sample size of $n = 4Q+1$ for $Q = 1, 2, \ldots, 50$ for computational convenience so that $Q+1$ and $3Q+1$ are also integers, and $Z_{(Q+1)}$ and $Z_{(3Q+1)}$ represent the first quartile and third quartile, respectively. Since $\mathcal{E}(Z_{(Q+1)}) = -\mathcal{E}(Z_{(3Q+1)})$, the expected distance between the first quartile and third quartile can be approximated by $2\mathcal{E}(Z_{(3Q+1)})$ standard deviations, and Blom (1958) suggested the following approximation:

$$\mathcal{E}(Z_{(r)}) \approx \Phi^{-1}\left(\frac{r-0.375}{n+0.25}\right), \quad r = 1, 2, \ldots, n\,. \tag{7}$$

Since $Q = \frac{n-1}{4}$, we can replace r by $3Q+1 = \frac{3(n-1)}{4} + 1 = 0.75n + 0.25$ in Equation (7). Using the similar approach, the expected distance between the maximum and minimum can be approximated by $2\mathcal{E}(Z_{(n)})$ standard deviations.

3.1.5 Approximation Under Case 4

Let c and d be the sample 10th percentile and 90th percentile, respectively. We now consider the fourth case when $C_4 = \{c, q_1, m, q_3, d, n\}$ is given. The idea of Wan et al. (2014) is also applicable to this case as follows:

$$s_w^{****} = \frac{1}{2}\left[\frac{q_3 - q_1}{2\mathcal{E}(Z_{(0.75n)})} + \frac{d-c}{2\mathcal{E}(Z_{(0.9n)})}\right]. \tag{8}$$

This is similar to s_w^{**} in Equation (4), and we can numerically calculate $\mathcal{E}(Z_{(0.75n)})$ and $\mathcal{E}(Z_{(0.9n)})$, which are special cases of $\mathcal{E}(Z_{(r)})$ in Equation (6), using the `integrate` function with `qnorm` function for $\Phi(\cdot)$ and `dnorm` function for $\phi(\cdot)$ in R (R Core Team, 2023).

3.2 Likelihood-based Approximation

3.2.1 Case 1

Let $\mathcal{F}(y; \mu, \sigma)$ be the CDF of $N(\mu, \sigma^2)$ for unknown model mean μ and model standard deviation σ,

$$\mathcal{F}(y; \mu, \sigma) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\,.$$

For the first case $C_1 = \{a, m, b, n\}$, we can guess that one data point is at a or below, one data point is at b or above, $\frac{n-2}{2} = 0.5n - 1$ data points are between a and m, and $0.5n - 1$ data points are between m and b. Using these clues, we can express the likelihood of observing such a sample as follows.

$$\begin{aligned}\mathcal{L}_1(\mu, \sigma) = \mathcal{F}(a; \mu, \sigma)[\mathcal{F}(m; \mu, \sigma) - \mathcal{F}(a; \mu, \sigma)]^{0.5n-1} \\ [\mathcal{F}(b; \mu, \sigma) - \mathcal{F}(m; \mu, \sigma)]^{0.5n-1}[1 - \mathcal{F}(b; \mu, \sigma)]\,.\end{aligned}$$

Suppose $\mathcal{L}_1(\mu, \sigma)$ is maximized at $\mu = \tilde{\mu}_{[1]}$ and $\sigma = \tilde{\sigma}_{[1]}$ using numerical search. Note that

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,,$$

$$s\sqrt{\frac{n-1}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

are the MLE for μ and σ, respectively. We propose to approximate $\bar{y}$ and s by

$$\begin{aligned}\bar{y}_l^* &= \tilde{\mu}_{[1]}\,, \\ s_l^* &= \tilde{\sigma}_{[1]}\sqrt{\frac{n}{n-1}}\end{aligned} \tag{9}$$

for the first case given $C_1 = \{a, m, b, n\}$.

3.2.2 Extensions to Other Cases

For the second case $C_2 = \{a, q_1, m, q_3, b, n\}$, third case $C_3 = \{q_1, m, q_3, n\}$, and the fourth case $C_4 = \{c, q_1, m, q_3, d, n\}$, we can express the respective likelihood functions as follows:

$$\begin{aligned}\mathcal{L}_2(\mu,\sigma) = {} & \mathcal{F}(a;\mu,\sigma)[\mathcal{F}(q_1;\mu,\sigma) - \mathcal{F}(a;\mu,\sigma)]^{0.25n-0.5} \\ & [\mathcal{F}(m;\mu,\sigma) - \mathcal{F}(q_1;\mu,\sigma)]^{0.25n-0.5}[\mathcal{F}(q_3;\mu,\sigma) - \mathcal{F}(m;\mu,\sigma)]^{0.25n-0.5} \\ & [\mathcal{F}(b;\mu,\sigma) - \mathcal{F}(q_3;\mu,\sigma)]^{0.25n-0.5}[1 - \mathcal{F}(b;\mu,\sigma)]\,,\end{aligned}$$

$$\begin{aligned}\mathcal{L}_3(\mu,\sigma) = {} & [\mathcal{F}(q_1;\mu,\sigma)]^{0.25n}[\mathcal{F}(m;\mu,\sigma) - \mathcal{F}(q1;\mu,\sigma)]^{0.25n} \\ & [\mathcal{F}(q_3;\mu,\sigma) - \mathcal{F}(m;\mu,\sigma)]^{0.25n}[1 - \mathcal{F}(q_3;\mu,\sigma)]^{0.25n}\,,\end{aligned}$$

$$\begin{aligned}\mathcal{L}_4(\mu,\sigma) = {} & [\mathcal{F}(c;\mu,\sigma)]^{0.1n}[\mathcal{F}(q_1;\mu,\sigma) - \mathcal{F}(c;\mu,\sigma)]^{0.15n} \\ & [\mathcal{F}(m;\mu,\sigma) - \mathcal{F}(q_1;\mu,\sigma)]^{0.25n} \\ & [\mathcal{F}(d;\mu,\sigma) - \mathcal{F}(q_3;\mu,\sigma)]^{0.15n}[1 - \mathcal{F}(d;\mu,\sigma)]^{0.1n}\,.\end{aligned}$$

Suppose $\mathcal{L}_k(\mu,\sigma)$ is maximized at $\mu = \tilde{\mu}_{[k]}$ and $\sigma = \tilde{\sigma}_{[k]}$ for $k = 2, 3, 4$. Similar to the first case, we propose to approximate $\bar{y}$ and s by

$$\begin{aligned}\bar{y}_l^{**} &= \tilde{\mu}_{[2]}\,, \\ s_l^{**} &= \tilde{\sigma}_{[2]}\sqrt{\frac{n}{n-1}}\,.\end{aligned} \tag{10}$$

for the second case given $C_2 = \{a, q_1, m, q_3, b, n\}$,

$$\begin{aligned}\bar{y}_l^{***} &= \tilde{\mu}_{[3]}\,, \\ s_l^{***} &= \tilde{\sigma}_{[3]}\sqrt{\frac{n}{n-1}}\,.\end{aligned} \tag{11}$$

for the third case given $C_3 = \{q_1, m, q_3, n\}$, and

$$\begin{aligned}\bar{y}_l^{****} &= \tilde{\mu}_{[4]}\,, \\ s_l^{****} &= \tilde{\sigma}_{[4]}\sqrt{\frac{n}{n-1}}\,.\end{aligned} \tag{12}$$

for the fourth case given $C_4 = \{c, q_1, m, q_3, d, n\}$. All approximation methods presented in Section 3 are summarized in Table 1.

4. Simulations

4.1 Simulation Scenarios

We conducted simulation studies to compare the approximation methods presented in Table 1. Given a simulation scenario, we generated an i.i.d. sample of size n from $N(\mu,\sigma)$ for each of the four cases. For each case, we considered the following parameter values: $\mu = 3$, $\sigma = 0.1, 0.3, 0.5$, and $n = 10, 20, 30, 40, 50, 100, 150$. Only one value of $\mu = 3$ was considered because it is a location parameter, the center of a normal distribution, which does not affect the quality of the approximation methods for $\bar{y}$ and s. The values of σ were chosen to cover the range of likely values of σ after observing the data shown in Figure 1. The values of n were chosen because the sample size per age group ranged between 10 and 51 in Nikolaidis et al. (2016), and we wanted to further investigate the large-sample property at $n = 100$ and $n = 150$. Each simulation scenario was repeated 50,000 times.

4.2 Evaluation Criteria

Let $\tilde{s}$ be an approximation method for s. The relative error (RE) of $\tilde{s}$ to s was calculated by

$$\mathrm{RE}(\tilde{s}, s) = \frac{\tilde{s} - s}{s}$$

following Wan et al. (2014). The relative error of an approximation method for $\bar{y}$ was calculated similarly. In addition, we evaluated each approximation method by the root mean square error (RMSE) with respect to the true parameter value. We note that the RE measures the closeness to the random sample mean $\bar{y}$ and standard deviation s, whereas the RMSE measures the quality of estimating the unknown fixed mean parameter μ and standard deviation parameter σ.

4.3 Simulation Results

Table 2 presents the relative error (RE) and root mean square error (RMSE) of the approximation methods for s. In Cases 1 and 2, when the sample extrema (a and b) are given, the likelihood-based approaches (s_l^* and s_l^{**}) underperform the distance-based approach (s_w^* and s_w^{**}). In Case 3, when the three quartiles (q_1, m, and q_3) are given, the likelihood-based approach (s_l^{***}) is slightly less accurate than the distance-based approach (s_w^{***}) in terms of the RE (closeness to s), but s_l^{***} has a slightly lower RMSE than s_w^{***} in most scenarios in terms of the RMSE (a better estimator for σ). In Case 4, when the 10-th percentile and 90-th percentile with the quartiles (c, q_1, m, q_3, and d) are given, the likelihood-based approach (s_l^{****}) is consistently closer to s than the distance-based approach (s_w^{****}) in terms of the RE, and it is also a better estimator for σ in terms of the RMSE.

When we compare between the likelihood-based approaches of s_l^{***} (Case 3) and s_l^{****} (Case 4), it is interesting that knowing c and d in addition to q_1, m, and q_3 is a distractor in terms of being close to s, but the additional information of c and d helps estimating σ. It appears to be clear that the likelihood-based approaches of s_l^* (Case 1) and s_l^{**} (Case 2) need improvement when a and b are given.

Table 3 presents the RE and RMSE of the approximation methods for $\bar{y}$. All methods considered in the simulations are fairly close to $\bar{y}$ in terms of RE (the absolute relative error < 0.001). In terms of RMSE, however, the likelihood-based approaches $\bar{y}_l^*$ in Case 1 and $\bar{y}_l^{**}$ in Case 2 are better than their counterparts $\bar{y}_h^*$ and $\bar{y}_b^{**}$, respectively, and $\bar{y}_l^{***}$ and $\bar{y}_w^{***}$ are very close in Case 3.

5. Application

As aforementioned in Section 1, Nikolaidis et al. (2016) reported $C_4 = \{c, q_1, m, q_3, d, n\}$ per age group, where c and d are the sample 10-th percentile and 90-th percentile. For this case, the simulation results suggest the likelihood-based approach (Section 3.2), especially for approximating s. Using the approximated values of $\bar{y}_i$ and s_i, we used the MLE to estimate $\boldsymbol{\theta} = (\beta_0, \beta_1, \delta, \nu)$ (Section 2.2).

Figure 2 shows the resulting piecewise regression models. They follow the patterns observed in Figure 1, but they are smoother by the regression assumption. For all three sprint segments (0-20 meters, 0-10 meters, and 10-20 meters), the resulting models suggest that the sprint performance improves until U17. The uncertainty is bigger for the first half segment of 0-20 meters (0-10 meters) than for the second half (10-20 meters). In other words, there is higher individual variability during the acceleration than after the acceleration. Under the normality assumption, the age-specific estimates are as follows for each segment.

- 0-20 meters: `time` $\sim N(\hat{\mu}, \hat{\nu})$ where $\sqrt{\hat{\nu}} = 0.2720$ (standard deviation) and

$$\hat{\mu} = \begin{cases} 5.6872 - 0.1455 \times \texttt{age} & \text{for } \texttt{age} < \texttt{U17}, \\ 3.2137 & \text{for } \texttt{age} \geq \texttt{U17}. \end{cases}$$

- 0-10 meters: `time` $\sim N(\hat{\mu}, \hat{\nu})$ where $\sqrt{\hat{\nu}} = 0.2209$ and

$$\hat{\mu} = \begin{cases} 3.1930 - 0.0754 \times \texttt{age} & \text{for } \texttt{age} < \texttt{U17}, \\ 1.9112 & \text{for } \texttt{age} \geq \texttt{U17}. \end{cases}$$

- 10-20 meters: `time` $\sim N(\hat{\mu}, \hat{\nu}^2)$ where $\sqrt{\hat{\nu}} = 0.1151$ and

$$\hat{\mu} = \begin{cases} 2.5410 - 0.0732 \times \texttt{age} & \text{for } \texttt{age} < \texttt{U17}, \\ 1.2966 & \text{for } \texttt{age} \geq \texttt{U17}. \end{cases}$$

In the above representations, `U17` means an age group of between 16 and 17, and it is treated as a numeric value of 17.

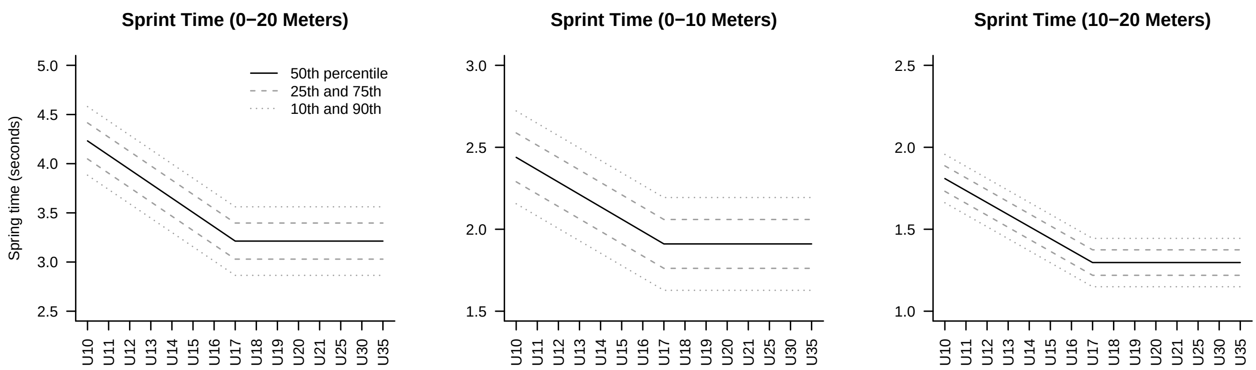

Figure 2. Age-specific reference sprint time (seconds) for 0-20 meters (left), 0-10 meters (middle), and 10-20 meters (right) among elite soccer players (the piecewise regression model estimated by the data of Nikolaidis et al. (2016))

Table 1. Approximation methods presented in Section 3

	Approximation for $\bar{y}$	Approximation for s
Case 1: $C_1 = \{a, m, b, n\}$	$\bar{y}_h^*$, Eq. (1) $\bar{y}_l^*$, Eq. (9)	s_h^*, Eq. (1) s_w^*, Eq. (2) s_l^*, Eq. (9)
Case 2: $C_2 = \{a, q_1, m, q_3, b, n\}$	$\bar{y}_b^{**}$, Eq. (3) $\bar{y}_l^{**}$, Eq. (10)	s_b^{**}, Eq. (3) s_w^{**}, Eq. (4) s_l^{**}, Eq. (10)
Case 3: $C_3 = \{q_1, m, q_3, n\}$	$\bar{y}_w^{***}$, Eq. (5) $\bar{y}_l^{***}$, Eq. (11)	s_w^{***}, Eq. (5) s_l^{***}, Eq. (11)
Case 4: $C_4 = \{c, q_1, m, q_3, d, n\}$	 $\bar{y}_l^{****}$, Eq. (12)	s_w^{****}, Eq. (8) s_l^{****}, Eq. (12)

Table 2. Evaluation of the approximation methods for s

		Case 1 $C_1 = \{a, m, b, n\}$			Case 2 $C_2 = \{a, q_1, m, q_3, b, n\}$			Case 3 $C_3 = \{q_1, m, q_3, n\}$		Case 4 $C_4 = \{c, q_1, m, q_3, d, n\}$		
RE	n	s_h^*	s_w^*	s_l^*	s_b^{**}	s_w^{**}	s_l^{**}	s_w^{***}	s_l^{***}	s_w^{****}	s_l^{****}	s
	10	–0.078	+0.023	+0.310	+0.040	+0.029	+0.300	+0.035	–0.053	+0.130	–0.085	–
	20	–0.053	+0.014	+0.181	+0.011	+0.014	+0.161	+0.013	–0.030	+0.057	–0.047	–
	30	+0.031	+0.011	+0.139	+0.011	+0.011	+0.116	+0.012	–0.016	+0.038	–0.031	–
$\sigma = 0.1$	40	+0.087	+0.008	+0.115	+0.007	+0.009	+0.092	+0.009	–0.012	+0.028	–0.024	–
	50	+0.130	+0.008	+0.102	+0.006	+0.008	+0.077	+0.008	–0.009	+0.022	–0.019	–
	100	–0.162	+0.006	+0.071	+0.002	+0.004	+0.044	+0.002	–0.006	+0.010	–0.010	–
	150	–0.115	+0.006	+0.058	+0.002	+0.004	+0.033	+0.002	–0.003	+0.007	–0.007	–
	10	–0.077	+0.024	+0.309	+0.041	+0.030	+0.301	+0.036	–0.052	+0.131	–0.085	–
	20	–0.054	+0.013	+0.179	+0.011	+0.014	+0.160	+0.015	–0.028	+0.058	–0.046	–
	30	+0.032	+0.012	+0.134	+0.011	+0.011	+0.116	+0.010	–0.018	+0.036	–0.033	–
$\sigma = 0.3$	40	+0.088	+0.009	+0.122	+0.006	+0.008	+0.091	+0.007	–0.014	+0.028	–0.024	–
	50	+0.130	+0.008	+0.104	+0.005	+0.006	+0.076	+0.005	–0.012	+0.021	–0.020	–
	100	–0.163	+0.005	+0.079	+0.003	+0.005	+0.045	+0.004	–0.004	+0.011	–0.010	–
	150	–0.117	+0.003	+0.059	+0.001	+0.003	+0.032	+0.002	–0.002	+0.008	–0.005	–
	10	–0.077	+0.024	+0.310	+0.039	+0.028	+0.301	+0.032	–0.055	+0.129	–0.085	–
	20	–0.052	+0.015	+0.188	+0.011	+0.013	+0.161	+0.011	–0.031	+0.055	–0.048	–
	30	+0.030	+0.009	+0.144	+0.010	+0.010	+0.114	+0.011	–0.017	+0.038	–0.031	–
$\sigma = 0.5$	40	+0.087	+0.008	+0.118	+0.006	+0.008	+0.091	+0.007	–0.013	+0.027	–0.024	–
	50	+0.132	+0.009	+0.109	+0.006	+0.007	+0.077	+0.005	–0.011	+0.021	–0.020	–
	100	–0.161	+0.007	+0.075	+0.002	+0.004	+0.044	+0.002	–0.005	+0.010	–0.010	–
	150	–0.116	+0.005	+0.068	+0.001	+0.003	+0.032	+0.002	–0.001	+0.007	–0.006	–
RMSE	n	s_h^*	s_w^*	s_l^*	s_b^{**}	s_w^{**}	s_l^{**}	s_w^{**}	s_l^{***}	s_w^{****}	s_l^{****}	s
	10	0.025	0.026	0.043	0.026	0.026	0.041	0.036	0.034	0.034	0.027	0.023
	20	0.019	0.019	0.028	0.018	0.018	0.025	0.026	0.025	0.022	0.019	0.016
	30	0.018	0.017	0.023	0.015	0.015	0.020	0.021	0.021	0.017	0.016	0.013
$\sigma = 0.1$	40	0.018	0.015	0.020	0.013	0.013	0.016	0.018	0.018	0.015	0.013	0.011
	50	0.020	0.015	0.018	0.012	0.012	0.014	0.016	0.016	0.013	0.012	0.010
	100	0.019	0.012	0.015	0.009	0.009	0.010	0.012	0.011	0.009	0.009	0.007
	150	0.015	0.011	0.013	0.007	0.008	0.008	0.009	0.009	0.007	0.007	0.006
	10	0.077	0.078	0.128	0.080	0.079	0.123	0.110	0.103	0.102	0.081	0.071
	20	0.058	0.059	0.087	0.055	0.055	0.075	0.078	0.075	0.065	0.057	0.049
	30	0.052	0.051	0.069	0.045	0.045	0.058	0.064	0.063	0.052	0.047	0.040
$\sigma = 0.3$	40	0.055	0.046	0.062	0.039	0.039	0.049	0.055	0.054	0.044	0.041	0.034
	50	0.061	0.043	0.057	0.035	0.035	0.043	0.049	0.049	0.039	0.036	0.030
	100	0.058	0.036	0.045	0.027	0.027	0.031	0.035	0.035	0.027	0.026	0.021
	150	0.046	0.033	0.040	0.023	0.023	0.025	0.029	0.029	0.022	0.021	0.017
	10	0.127	0.128	0.214	0.132	0.130	0.202	0.183	0.172	0.168	0.134	0.116
	20	0.096	0.097	0.141	0.090	0.090	0.124	0.128	0.125	0.107	0.094	0.080
	30	0.087	0.085	0.118	0.074	0.074	0.096	0.105	0.102	0.085	0.077	0.065
$\sigma = 0.5$	40	0.093	0.078	0.102	0.065	0.066	0.082	0.092	0.090	0.074	0.067	0.057
	50	0.103	0.072	0.095	0.059	0.059	0.073	0.083	0.082	0.065	0.060	0.050
	100	0.096	0.061	0.073	0.044	0.044	0.051	0.058	0.058	0.045	0.043	0.036
	150	0.076	0.054	0.066	0.038	0.038	0.042	0.047	0.048	0.037	0.035	0.029

Table 3. Evaluation of the approximation methods for $\bar{y}$

		Case 1 $C_1 = \{a, m, b, n\}$		Case 2 $C_2 = \{a, q_1, m, q_3, b, n\}$		Case 3 $C_3 = \{q_1, m, q_3, n\}$		Case 4 $C_4 = \{c, q_1, m, q_3, d, n\}$	
RE	n	$\bar{y}_h^*$	$\bar{y}_l^*$	$\bar{y}_b^{**}$	$\bar{y}_l^{**}$	$\bar{y}_w^{***}$	$\bar{y}_l^{***}$	$\bar{y}_l^{****}$	$\bar{y}$
$\sigma = 0.1$	10	–0.000	–0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	20	+0.000	+0.000	+0.000	–0.000	–0.000	–0.000	–0.000	–
	30	–0.000	–0.000	–0.000	–0.000	–0.000	–0.000	+0.000	–
	40	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	50	+0.000	+0.000	+0.000	–0.000	+0.000	–0.000	+0.000	–
	100	–0.000	–0.000	–0.000	–0.000	+0.000	+0.000	–0.000	–
	150	–0.000	+0.000	–0.000	–0.000	–0.000	–0.000	–0.000	–
$\sigma = 0.3$	10	+0.000	+0.000	+0.000	+0.000	–0.000	–0.000	–0.000	–
	20	–0.000	–0.000	–0.000	–0.000	+0.000	+0.000	–0.000	–
	30	–0.000	–0.000	–0.000	–0.000	–0.000	–0.000	–0.000	–
	40	+0.000	–0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	50	+0.000	+0.000	–0.000	+0.000	+0.000	+0.000	–0.000	–
	100	+0.000	+0.000	+0.000	+0.000	–0.000	–0.000	+0.000	–
	150	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–0.000	–
$\sigma = 0.5$	10	+0.000	+0.000	+0.000	+0.000	–0.000	–0.000	–0.000	–
	20	–0.000	+0.000	–0.000	–0.000	+0.000	+0.000	+0.000	–
	30	–0.000	–0.000	–0.000	+0.000	+0.000	+0.000	+0.000	–
	40	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	50	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	100	+0.000	–0.000	+0.000	+0.000	+0.000	+0.000	+0.000	–
	150	–0.000	–0.000	+0.000	–0.000	–0.000	+0.000	+0.000	–
RMSE	n	$\bar{y}_h^*$	$\bar{y}_l^*$	$\bar{y}_b^{**}$	$\bar{y}_l^{**}$	$\bar{y}_w^{***}$	$\bar{y}_l^{***}$	$\bar{y}_l^{****}$	$\bar{y}$
$\sigma = 0.1$	10	0.033	0.033	0.032	0.032	0.033	0.033	0.032	0.031
	20	0.026	0.025	0.023	0.023	0.024	0.024	0.023	0.022
	30	0.023	0.021	0.019	0.019	0.019	0.019	0.019	0.018
	40	0.021	0.018	0.017	0.016	0.017	0.017	0.016	0.016
	50	0.020	0.016	0.015	0.015	0.015	0.015	0.014	0.014
	100	0.017	0.012	0.012	0.011	0.011	0.011	0.010	0.010
	150	0.016	0.010	0.010	0.009	0.009	0.009	0.008	0.008
$\sigma = 0.3$	10	0.100	0.100	0.095	0.096	0.100	0.100	0.096	0.094
	20	0.078	0.075	0.069	0.069	0.072	0.072	0.069	0.067
	30	0.069	0.062	0.058	0.057	0.059	0.059	0.057	0.055
	40	0.064	0.055	0.051	0.050	0.051	0.051	0.049	0.048
	50	0.060	0.050	0.047	0.045	0.046	0.046	0.044	0.043
	100	0.051	0.036	0.035	0.032	0.032	0.032	0.031	0.030
	150	0.048	0.030	0.031	0.027	0.027	0.027	0.025	0.025
$\sigma = 0.5$	10	0.169	0.169	0.160	0.161	0.169	0.169	0.162	0.159
	20	0.129	0.123	0.114	0.114	0.119	0.119	0.114	0.111
	30	0.116	0.105	0.096	0.095	0.097	0.097	0.093	0.091
	40	0.106	0.091	0.085	0.082	0.084	0.084	0.081	0.079
	50	0.099	0.082	0.077	0.073	0.075	0.075	0.072	0.070
	100	0.086	0.060	0.060	0.053	0.054	0.054	0.051	0.050
	150	0.079	0.050	0.052	0.044	0.044	0.044	0.042	0.041

6. Summary

Under the normality assumption, the sample mean and standard deviation are sufficient statistics for the two model parameters, whereas percentiles are not sufficient. In literature, there are various methods of utilizing the sample minimum and maximum and/or the quartiles for approximating the unreported sample mean and standard deviation in meta-analysis. One of the popular methods is based on the expected distance between order statistics, and this distance-based approach can be generalized to cases when other percentiles are available. We generalized to the case when the sample 10-th percentile and 90-th percentile are given instead of the sample minimum and maximum. In addition, we considered the likelihood-based approach and investigated its operating characteristics via simulations. We learned the following: (1) The likelihood-based approach presented in this paper needs substantial improvement when the sample extrema (minimum and maximum) are given; (2) the distance-based approach is closer to random s when only the three quartiles are available, but the likelihood-based approach is a better estimator for fixed σ in this case; (3) the likelihood-based approach outperforms the distance-based approach in terms of approximating s and estimating σ when the 10-th percentile and 90-th percentile are given in addition to the quartiles. Since it was our case in the applied example, we implemented the likelihood-based approach to provide the age-specific reference model for 20-meter sprint time using the piecewise regression model.

Acknowledgement

Not applicable.

Authors contributions

Kyra Batarse and Jessica Naranjo were responsible for literature review, data collection, and application and drafted the relevant sections. Jakob Garcia was responsible for simulations and drafted. Dr. Steven Kim was responsible for leading the students, compiling the drafts, and writing and revising the manuscript.

Funding

The article processing fee was supported by the Department of Mathematics and Statistics at Cal State Monterey Bay through its 2015 California Award for Innovation in Higher Education.

Competing interests

The authors declare no known competing financial interests or personal relationships that influenced the work reported in this paper.

References

Angius, L., Olla, S., Piras, F., Mura, R., Ibba, G., Todde, F., ... Crisafulli, A. (2013). Indexes of physical capacity and repeated sprint ability of young soccer players. *Sport Sciences for Health, 9*, 1-6. https://doi.org/10.1007/s11332-013-0137-x

Barnes, C., Archer, D. T., Hogg, B., Bush, M., & Bradley, P. S. (2014). The evolution of physical and technical performance parameters in the English Premier League. *International Journal of Sports Medicine, 35*(13), 1095-1100. https://doi.org/10.1055/s-0034-1375695

Bland, M. (2015). Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. *International Journal of Statistics in Medical Research, 4*(1), 57-64. https://doi.org/10.6000/1929-6029.2015.04.01.6

Blom, G. (1958). *Statistical estimates and transformed beta-variables.* New York, NY, John Wiley & Sons.

David, H. A., & Nagaraja, H. N. (2003). Order statistics. (3rd ed.). New York, NY, John Wiley & Sons.

Faude, O., Koch, T., & Meyer, T. (2012). Straight sprinting is the most frequent action in goal situations in professional football. *Journal of Sports Sciences, 30*(7), 625-631. https://doi.org/10.1080/02640414.2012.665940

Haugen, T., T?nnessen, E., Hisdal, J., & Seiler, S. (2014). The role and development of sprinting speed in soccer. *International Journal of Sports Physiology and Performance, 9*(3), 432-441. https://doi:10.1123/ijspp.2013-0121

Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology, 5*, 13. https://doi.org/10.1186/1471-2288-5-13

Julious, S. A. (2001). Inference and estimation in a changepoint regression problem. *The Statistician, 50*(1), 51-61. https://doi.org/10.1111/1467-9884.00260

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied linear statistical models.* (5th ed.). New York, NY, McGraw-Hill/Irwin.

Kwon, D., Reddy, R. R. S., & Reis, I. M. (2021). ABCMETAapp: R shiny application for simulation-based estimation of mean and standard deviation for meta-analysis via approximate Bayesian computation. *Research Synthesis Methods, 12*(6), 842-848. https://doi.org/10.1002/jrsm.1505

Kwon, D., & Reis, I. M. (2015). Simulation-based estimation of mean and standard deviation for meta-analysis via Approximate Bayesian Computation (ABC). *BMC Medical Research Methodology, 15*, 61. https://doi.org/10.1186/s12874-015-0055-5

Le Gall, F., Beillot, J., & Rochcongar, P. (2002). The improvement in maximal anaerobic power from soccer player during growth. *cience & Sports, 17*(4), 177-188. https://doi.org/10.1016/S0765-1597(02)00155-7

Luo, D., Wan, X., Liu, J., & Tong, T. (2018). Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Statistical Methods in Medical Research, 27*(6), 1785-1805. https://doi.org/10.1177/0962280216669183

Maly, T., Zah¢lka, F., Hr¢sky, P., Mala, L., I?ovsk¢, J., Bujnovsky, D., Dragijsky, M., & Mihal, J. (2015). Age-related differences in linear sprint and power characteristics in youth elite soccer players. *Journal of Physical Education and Sport, 15*(4), 857-863. https://doi.org/10.7752/jpes.2015.04132

McGrath, S., Zhao, X., Steele, R., Thombs, B. D., Benedetti, A., & DEPRESsion Screening Data (DEPRESSD) Collaboration. (2020). Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. *Statistical Methods in Medical Research, 29*(9), 2520-2537. https://doi.org/10.1177/0962280219889080

Mendez-Villanueva, A., Buchheit, M., Kuitunen, S., Douglas, A., Peltola, E., & Bourdon., P. (2011). Age-related differences in acceleration, maximum running speed, and repeated-sprint performance in young soccer players. *Journal of Sports Sciences, 29*(5), 477-484. https://doi.org/10.1080/02640414.2010.536248

Nikolaidis, P. T., Knechtle, B., Clemente, F. M., & Torres-Luque, G. (2016). Reference values for the sprint performance in male football players aged from 9C35 years. *Biomedical Human Kinetics, 8*, 103-112. https://doi.org/10.1515/bhk-2016-0015

Pearson, E. S., & Tukey, J. W. (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika, 52*, 533-546. https://doi.org/10.2307/2333703

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC medical research methodology, 14*, 135. https://doi.org/10.1186/1471-2288-14-135

Copyrights