Forensic Science International 289 (2018) 304–309

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Technical Note

TDStats—A fast standardized capability for facial soft tissue thickness analysis in R☆

Carl N. Stephan

Laboratory for Human Craniofacial and Skeletal Identification (HuCS-ID Lab), School of Biomedical Sciences, The University of Queensland, Brisbane, 4072, Australia

ARTICLE INFO

Article history:
Received 26 August 2017
Received in revised form 3 April 2018
Accepted 24 May 2018
Available online 31 May 2018

Keywords:
Forensic science
Craniofacial identification
Facial soft tissue thickness
Facial soft tissue depth
Shorth
Shormax

ABSTRACT

The popular and repeated calculation of facial soft tissue thickness in different samples by different practitioners makes the development of a free, standard, and open-source tool that calculates all necessary statistics in an automated fashion useful. This tool should enable data analysis for single or multiple studies in both independent study-specific format and as a pooled aggregate. The tool should be fast, should facilitate exploratory data analysis, and should provide robust central tendency statistics (shorth and shormaxes). The tool should facilitate effortless analysis, so that once the raw data are entered, little additional input is required of the analyst to maximize user-friendliness. This paper describes in detail such a capability (*TDStats v2017.1*) specifically formulated for soft tissue thickness analyses and demonstrates its utility by analysing 14,201 publically available data points at 24 landmarks from 1086 subjects in under 45 s. Outputs include 120 plots, 15 tables, and 5 summary pdfs. A step-by-step user guide to *TDStats* is provided as well as additional comments and clarity on the utility of robust central tendency statistics.

1. Introduction

Facial soft tissue thickness research was first initiated by Welcker to assist craniofacial identification in 1883 [1], some 135 years ago at the time of this paper's writing. Since then more than 100 FSTT papers have been published with more than one third entering the literature in the last 10 years (Fig. 1). Across the last 30 years, the number of published FSTT studies has risen substantially. Now, FSTT studies easily represent the most frequently published topic within the craniofacial identification domain [2] and without any sign of a decrease, the number of FSTT studies to be published each year in the future can only be expected to increase (Fig. 1). This will create an ever expanding set of data adding to the already abundant aggregate: >160,500 measurements of >10,300 adults [3]; and for sub-adults: >66,900 measurements of >8500 individuals [3].

With the formulation of new FSTT raw data repositories [4] and employment of new and different descriptive statistics [2], new tools that meet new data analysis demands and facilitate increased analytical speed are required. Necessary new functions include the ability to analyse multiple independent and pooled datasets in automated and quick fashion with the production of a comprehensive suite of exploratory data plots and summary statistic outputs to help the researcher formulate informed scientific inferences and produce the best quality summary statistics for casework use.

This paper describes in detail a purpose built R [6] capability to standardize and expedite FSTT analysis in craniofacial identification, including all pertinent summary statistics and emphasizing exploratory data analysis functions. The first version of this capability was released in 2012 (v2012.1), but has since gone through a number of updates—the latest being v2017.1 as described herein. It is called *TDStats*—short for Tissue Depth Statistics. In addition to its in-depth description to facilitate user operation, this technical note aims to determine the processing time required for the script to conduct a representative facial soft tissue depth analysis, which includes multiple independent tissue thickness studies.

2. Materials and methods

2.1. Background to capability development and intent

TDStats was initiated to address a need, identified through a series of FSTT studies [2,5,7], for a program that facilitated

☆ Portions of this work have been presented at the 59th Annual Meeting of the American Academy of Forensic Sciences February 2007, in San Antonio, USA, and the 17th Biennial Meeting of the International Association of Craniofacial Identification July 2017, in Brisbane, Australia.

E-mail address: c.stephan@uq.edu.au (C.N. Stephan).

https://doi.org/10.1016/j.forsciint.2018.05.039
0379-0738/

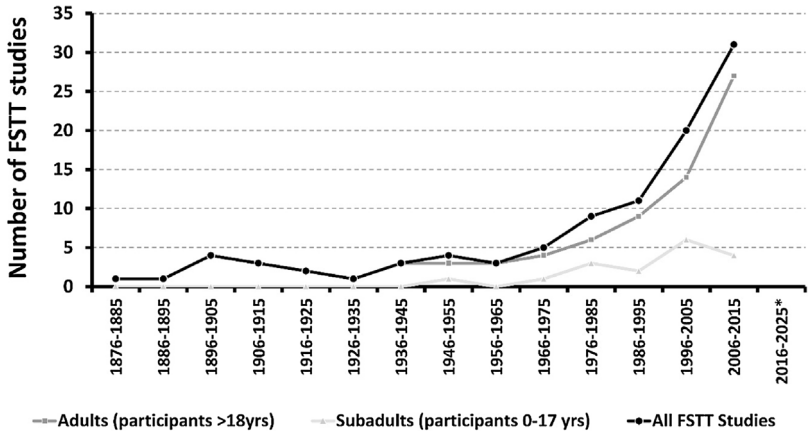

Fig. 1. Numbers of FSTT publications by year since 1883. Plot created using study numbers drawn from the literature in Ref. [5] and further updated by Ref. [3]. *Active time bracket.

exploratory data analysis [8–11] and provided robust estimators of location [12]. At the time, this was contrary to the more popular and mainstream practice of statistical significance testing largely with disregard for measurement errors, effect sizes, and potential problems associated with sampling procedures [13].

The principle underpinning exploratory data analysis (EDA) is to interrogate the data visually for its underlying patterns to draw scientific inferences (Fig. 2), rather than performing statistical significance tests in isolation [8–11,14]. The value of such data visualisation is widely recognised and saliently demonstrated by Anscombe's famed data quartet of data [11,15]: data trends are difficult to read from numbers alone, especially when summary statistics provide near identical results (Table 1), but underlying differences in data patterns are immediately obvious and salient when the data are plotted or visualised (Fig. 2).

One goal of *TDStats* was to harness these principles and subsequently no statistical significance test functions are explicitly provided within the *TDStats* tool. These other tests are left to the user to conduct on their own risk/accord (see e.g., Refs. [3,5,13,16]) and instead the focus is given to summary statistics and plots (Fig. 3).

2.2. Capability development

2.2.1. Why R?

Inspiration for exploratory data analysis of FSTT and *TDStats* was originally provided when undertaking FSTT analysis [7] in the *JMP*® program, which is also EDA orientated. *JMP*® did not, however, provide shorth [12] or 75-shormax calculations [2,17] and functionality was not tailored to full pipeline FSTT analysis in streamlined fashion. This could, however, be achieved in the fully customisable statistical programming environment of R [6]. The R environment also held the advantage of being open-source and free, making it easily accessible to any craniofacial identification researchers or practitioners [6,18]. The open-source nature of R and *TDStats* additionally provides the option that the code can be audited and/or extended by any practitioner truly making it a community resource [18].

TDStats integrates a number of graphical user interfaces (GUIs) into its workflow to offer users point-and-click functionality, such that no entries to the R console window are required. This permits *TDStats* to be easily used by those with limited coding skills, R language fluency, or familiarity with R.

Table 1
Ancombe's Quartet [11] as called in R [6] with summary statistics calculated.

	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	x1	y1	x2	y2	x3	y3	x4	y4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.1	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.1	4	5.39	19	12.5
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
SD	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
Variance	11.00	4.13	11.00	4.13	11.00	4.12	11.00	4.12
Correlation	0.82		0.82		0.82		0.82	

2.2.2. Licensing

TDStats code is open source and has been released under the General Public License, version 2 (GPLv2), meaning that the code can be freely distributed and modified for other purposes, provided it too is released under the GPL v2 license and contains acknowledgements back to [2] and/or this paper.

2.2.3. Workflow

The program is designed to undertake the FSTT analysis in two steps so far as the end-user is concerned: (1) load the facial soft tissue thickness data via an open-dialogue box, and (2) automatically, but serially conduct statistical evaluations of the data and output the results to a single working directory ready for user examination. See Supplementary File 1 for the *TDStats* User Manual and Supplementary File 2 for a .pdf of the source code (NB. original .r source code is available at CRANIOFACIALidentification.com).

2.2.4. Data input

The FSTT data should be entered as a .csv file in the standard format of the C-Table data repository [4]. Note here the .csv file can easily be viewed, modified and changed in Microsoft® Excel® with very similar format/appearance to .xls files (but should always be saved with a .csv extension).

The number of landmarks used in the analysis is set by the user at an entry dialogue box upon running the *TDStats* program, with defaults set to the standard pre-set 27 landmarks of the C-Table. A successful load of the data is recognised by the program and communicated to the user by an onscreen message.

TDStats is designed to handle multiple FSTT datasets in the one .csv file, and it will produce summary data and boxplots for