International Journal of Industrial Ergonomics 69 (2019) 23–28

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

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

Sample size calculations for a functional human motion analysis: Application to vehicle ingress discomfort prediction

Hadi Ibrahim Masoud[a,*], Matthew P. Reed[b,c], Jionghua (Judy) Jin[c]

[a] Department of Industrial Engineering, University of Jeddah, Jeddah, Saudi Arabia
[b] University of Michigan Transportation Research Institute, Ann Arbor, USA
[c] Industrial and Operations Engineering, University of Michigan, Ann Arbor, USA

ARTICLE INFO

Keywords:
Sample size
Ingress
Human motion
Fraction disaccommodated
Functional data analysis

ABSTRACT

The ease of entering a vehicle, known as ingress, is one of the important ergonomic factors that car manufacturers consider during the process of vehicle design. Manufacturers frequently conduct human subject tests to assess ingress discomfort for different vehicle designs. Using subject tests, manufacturers are able to estimate the proportion of participants that report that they are discomfortable entering a vehicle, referred to in this paper as *fraction disaccommodated* (FD). Manufacturers then conduct statistical tests in order to determine if the FD of two vehicle designs are significantly different, and to determine the required sample size in testing the FD difference between two vehicle designs under pre-specified testing power. Since conducting human subject tests is often expensive and time consuming, another alternative is to estimate the FD using simulated human motion data. Determining the number of simulations that is required is an important statistical question that is dependent on the prediction performance of the simulation analysis. In this paper, a dual bootstrap approach is proposed to obtain the standard deviation of the estimated FD based on functional predictors. This standard deviation is then used to calculate the power in testing the difference between two estimated FDs.

1. Introduction

The ease of getting into a vehicle, known as ingress, is an important consideration for customer satisfaction in the automotive industry (Morgans and Thorness, 2013). This has motivated vehicle manufacturers to focus on assessing and improving ingress discomfort. The most straightforward way to assess ingress discomfort is to build prototypes or mockups and have human participants test these potential vehicle designs. Participants rate the ease of getting into the vehicle using a Likert scale. For example, using a 10-point scale, participants might rate a design 1 out of 10 if it is very difficult to get into the vehicle and 10 out of 10 if the ingress motion is exceptionally comfortable. These ingress ratings can also be transformed into binary responses using a cutpoint. Using cutpoint 5, for example, ratings below or equal to 5 are transformed to 0 (or "uncomfortable") and ratings above 5 are transformed to 1 (or "comfortable"). One metric of interest is the proportion of participants who rated the ingress discomfort of a design above a defined cutpoint, referred to as *fraction disaccommodated* (FD). As the population *FD* (true *FD*) for a certain vehicle design is unknown, the participants responses are usually considered as a sample for estimating the population ingress fraction disaccommodated, which is denoted as $\widetilde{FD}$ in this research.

As it is generally expensive and time-consuming to conduct tests with participants to assess ingress discomfort, manufacturers seek more efficient ways to assess ingress discomfort, including computer simulation (Wegner et al., 2007). Advances in digital human modeling technologies have provided the ability to simulate the ingress motion of people with a wide range of anthropometric features (Reed et al., 2006; Reed and Huang, 2008). However, even if accurate methods for simulating ingress motions are available, it is still necessary to predict the subjective responses from the motion data. Masoud et al. (2016) developed a systematic framework that used human motion trajectories to predict subjective ingress discomfort responses using a machine-learning approach based on support vector machines (SVM). By using this framework, the FD of a vehicle design can be predicted by conducting simulations for a wide range of drivers (e.g., tall and short, young and old) and predicting subjective responses from the simulated motion data. This simulation-based approach can expedite the vehicle design validation process and reduce the cost of testing participants in physical mockups. To differentiate between the estimated FD obtained using participant responses ($\widetilde{\mathrm{FD}}$) and the predicted FD obtained using actual or simulated human motion data, we denote the latter as $\widehat{\mathrm{FD}}$.

* Corresponding author.
E-mail addresses: masoud@uj.edu.sa (H.I. Masoud), mreed@umich.edu (M.P. Reed), jhjin@umich.edu (J.J. Jin).

https://doi.org/10.1016/j.ergon.2018.09.010
Received 31 October 2017; Received in revised form 16 May 2018; Accepted 26 September 2018

In many cases, manufacturers are interested in knowing whether the ingress discomfort of one design is better than that of another. For this purpose, manufacturers may conduct a statistical hypothesis test to examine whether the FD of one design is significantly higher than that of another. Moreover, after a design change has been made, manufacturers seek to determine the minimum sample size that can provide a definitive assessment of the difference between two designs in terms of their FD values. In literature, many methods have been developed to test whether there is a significant difference between two proportions (Newcombe, 1998). Power calculations and sample size determination for testing the difference between proportions have also been studied (Faul et al., 2007; Cohen, 2013). In these methods, the responses used to estimate the proportions are assumed to be i.i.d (independent and identically distributed) and to follow a binomial distribution, i.e., each response has an equal probability of success (p) and the standard deviation of the sample proportions is equal to $\sigma_p = \sqrt{\frac{p\,(1-p)}{n}}$. Although this assumption is appropriate for $\widetilde{FD}$, which is estimated using subjective responses, it is not immediately apparent that this relationship can be used to estimate the standard deviation of $\widehat{FD}$ due to the complex relationship between the motion model parameterization and the predicted subjective responses.

The objective of this paper is to develop a method for conducting power calculations in comparing two $\widehat{FD}$s in which the response proportion $\widehat{FD}$ are predicted from functional data obtained either from physical or virtual experiments. To conduct the power calculations, we must estimate the standard deviation of $\widehat{FD}$, referred to as σ_{FD} in this research. We developed a dual-bootstrapping approach that enables us to consider the two sources of variation in σ_{FD}. One is the modeling variation, which is due to the uncertainty of the estimated prediction model (σ_m) under different training datasets, and the other is the sampling variation due to the randomness of selecting test participants from the population (σ_s).

2. Methods

2.1. Data source

The data in this study was obtained from a vehicle ingress experiment that was conducted to study and reduce discomfort during ingress (Masoud et al., 2016). In brief, the experiment captured human motion data from 32 participants during vehicle ingress trials. Participants evaluated 17 vehicle designs that differed widely in the layout of the driver entry area. During each ingress test, reflective markers were used to record the location over time (trajectories) of 20 different joints. The trajectories of each joint were modeled by 27 B-spline coefficients. After participants completed an ingress trial (sample), they rated the ease of getting into the car on a 10-point scale, where 1 represents an unacceptable ingress experience and 10, an exceptionally comfortable ingress experience. The ingress discomfort rating was then transformed into a binary response using the cutpoint equal to 5, i.e., ratings below or equal to 5 were set as 0, and those above 5 were set as 1. In this research, the Cartesian trajectories of the 5 joints (left hip, right shoulder, right elbow, S1L5, and head) are used. The coordinates of these kinematic joints were identified by Masoud et al. (2016) as the most informative kinematic data for predicting ingress discomfort.

2.2. Method overview

A dual bootstrap or resampling approach was developed to estimate σ_{FD}, which includes two types of variation, σ_m and σ_s. A bootstrap approach is necessary because the complex relationship between the motion model and the predicted subjective responses precludes the use of the binomial distribution for estimating the standard deviation for the response proportion $\widehat{FD}$. As shown in Fig. 1, the first step is to generate a set of "bootstrap training datasets" by randomly resampling from the original dataset (X_t, Y_t) obtained from physical participant-tests described in the previous section, where X_t is the human motion data and Y_t, the corresponding participant ingress discomfort response. Each of the generated bootstrap training datasets (X_t^{*b}, Y_t^{*b}) ($b = 1,\ldots, B$) is used to train a prediction model using an SVM classifier (Masoud et al., 2016). With B bootstrap training datasets, we can obtain a set of prediction models, i.e., B different SVM classifiers, as shown in Fig. 1. The second step is to generate "bootstrap prediction datasets" for the two designs to be compared. As shown in Fig. 1, the bootstrap prediction datasets are generated by randomly resampling from X_p to generate J bootstrap prediction datasets X_p^{*1}, X_p^{*2}, ...,X_p^{*j}, ..., X_p^{*J}. These bootstrap prediction datasets are then used along with one trained SVM model to predict J $\widehat{FD}$ for the design of interest (i.e., one $\widehat{FD}$ for each bootstrap prediction dataset). These predicted $\widehat{FD}$ are used to predict the sampling variance (σ_s) that arises due to the randomness in the prediction dataset. By repeating this process B times, through each of the SVM models, we can estimate the modeling variance (σ_m) induced by the uncertainty in the estimated prediction models. The details of each step are discussed in the following subsections.

2.3. Generate bootstrap training datasets

Assume that $X_t = (x_1^t, x_2^t, \ldots, x_{n_0}^t)$ represents the original training dataset obtained from the human participant-tests, where x_i^t represents the vector of human motion data of one ingress sample, represented as B-spline coefficients; n_0, the number of samples; and $Y_t = (y_1^t, y_2^t, \ldots, y_{n_0}^t)$, the participant's binary ingress discomfort responses, where y_i^t is the discomfort rating corresponding to the motion data sample x_i^t. A bootstrap training dataset $X_t^{*b} = (x_1^{*b}, x_2^{*b}, \ldots, x_{n0}^{*b})$, $Y_t^{*b} = (y_1^{*b}, y_2^{*b}, \ldots, y_{n_0}^{*b})$ is generated by randomly resampling, with replacement, n_0 times from the original dataset X_t and Y_t, where * represents a bootstrap sample and b, the bootstrap replication index. This replication process is performed B times to generate a large number of bootstrap training datasets X_t^{*1}, X_t^{*2}, ..., X_t^{*B} and Y_t^{*1}, Y_t^{*2}, ..., Y_t^{*B}. In this analysis, the number of bootstrap datasets, denoted as B, was set to 100.

2.4. Train SVM prediction models

In this step, the bootstrap training datasets are used to train SVM prediction models (Cortes and Vapnik, 1995). SVM is a supervised learning classifier that has gained popularity in recent years as it can handle nonlinear classification and is robust to outliers (Cherkassky and Ma, 2004; Pal and Foody, 2010).

In this work, each set of bootstrap training datasets, X_t^{*b} and Y_t^{*b}, was used to train a separate SVM classifier, thus generating B different SVM models (${}^{(1)}SVM$, ${}^{(2)}SVM$, ..., ${}^{(b)}SVM$, ...,${}^{(B)}SVM$). The SVM models were trained using a Gaussian RBF kernel. The parameters of the RBF kernel were optimized for the original datasets X_t and Y_t using grid search to minimize the bias between the FD estimated from the prediction model ($\widehat{FD}$) and that estimated from participant responses ($\widetilde{FD}$); i.e., $\sum_{d=1}^{D}\left(\widehat{FD}_d - \widetilde{FD}_d\right)^2$ is minimized, where d is the index of different vehicle designs. Details of training an SVM model for classifying functional data can be found in Masoud et al. (2016).

2.5. Generate bootstrap prediction datasets

Assume that $X_p = (x_1^p, x_2^p, \ldots, x_n^p)$ *and* $X_{p'} = (x_1^{p'}, x_2^{p'}, \ldots, x_n^{p'})$ represent the human motion data corresponding to two different designs indicated by subscripts p and p' respectively, where n represents the number of motion data samples obtained though visual experimental tests or computer simulations. The participants tested in Design p can be either different from those in Design p', referred to as independent