Energy & Buildings 281 (2023) 112755

Contents lists available at ScienceDirect

Energy & Buildings

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

Analysis of subjective thermal comfort data: A statistical point of view

Matteo Favero [a], Antonio Luparelli [b], Salvatore Carlucci [c,*]

[a] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway
[b] Automation and Information Systems Area, CETMA – European Research Centre for Technologies Design and Materials, Brindisi, Italy
[c] Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 9 May 2022
Revised 17 November 2022
Accepted 27 December 2022
Available online 2 January 2023

Keywords:
Subjective thermal comfort data
Rating scales
Level of measurement
Ordinal regression
Bayesian analysis
Statistical thinking

ABSTRACT

Thermal comfort research aims to determine the relationship between the thermal environment and the human sense of warmth. This is usually achieved by measuring the subjective human thermal response to different thermal environments. However, it is common practice to use simple linear regression to analyse data collected using ordinal scales. This practice may lead to severe errors in inference. This study first set the methodological foundations to analyse subjective thermal comfort data from a statistical perspective. Subsequently, we show the practical consequences of fallacious assumptions by utilising a Bayesian approach and show, through an illustrative example, that a linear regression model applied to ordinal data suggests results different from those obtained using ordinal regression. Specifically, linear regression found no difference in means and effect size between genders, while the ordinal regression model led to the opposite conclusion. In addition, the linear regression model distorts the estimated regression coefficient for air temperature compared to the ordinal model. Finally, the ordinal model shows that the distance between adjacent response categories of the ASHRAE 7-point thermal sensation scale is not equidistant. Given the abovementioned issues, we advocate utilising ordinal models instead of metric models to analyse ordinal data.

1. Introduction

Thermal comfort is defined as 'the condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation' [1]. Subjective evaluation is usually obtained using rating scales, the most adopted of which is the ASHRAE 7-point thermal sensation scale, which consists of seven verbal anchors: 'cold', 'cool', 'slightly cool', 'neutral', 'slightly warm', 'warm', and 'hot'. This is a perceptual judgement scale [2] and is utilised to measure thermal sensation. Other rating scales are also employed in thermal comfort studies: the most common ones being thermal evaluation, preference, and acceptability. ISO 10551:2019 [2], beyond those already mentioned, also introduces a 'tolerance scale', which is rarely used in the scientific literature. Each one of these scales can be presented in different formats (e.g., discontinuous *versus* continuous format) and methods (e.g., paper- *versus* computer-based). Independently of the format and method used, it is common practice to assign a numerical value to each level (i.e., the verbal anchors) of a scale. For instance, the ASHRAE 7-point thermal sensation scale generally varies from −3 ('cold') to +3 ('hot'). However, different values can be assigned, such as 1 for 'cold' and 7 for 'hot'. This interchangeability is possible because these numbers are merely placeholders without an underlying meaning. Nevertheless, it is common practice to calculate the mean of the thermal sensation votes of a group of people (e.g., Refs [3,4]). The reasoning behind this method is that, while the variable is ordinal in nature, a vote created by averaging different responses is continuous. Furthermore, the averaged votes will result in a more normal-looking distribution and, therefore, statistical methods that assume normality (e.g., linear regression and analysis of variance) can be applied. The origin of this approach can be found in early works to measure attitudes, such as in Thurstone [5] and Likert [6]. However, there are two problems with this approach. Firstly, it is not appropriate to calculate the mean of an ordinal variable because its linearity (i.e., equally spaced divisions) is an arbitrary assumption imposed on the original scale values. This assumption was also recently questioned by Schweiker et al. [7,8]. Secondly, this approach conflates the problem of the

Abbreviations: ANOVA, Analysis of variance; CI, Credible interval; CDF, Cumulative distribution function; CLMs, Cumulative link models; GLM, General linear model; HDI, Highest density interval; K–S test, Kolmogorov–Smirnov test; LOOCV, Leave-one-out cross-validation; LOOIC, Leave-one-out information criterion; MCMC, Markov Chain Monte Carlo; PDF, Probability density function; SD, Standard deviation; TSV, Thermal sensation votes.

* Corresponding author at: Konstantinou Kavafi 20, 2121 Aglantzia, Cyprus.
E-mail address: s.carlucci@cyi.ac.cy (S. Carlucci).

https://doi.org/10.1016/j.enbuild.2022.112755
0378-7788/

level of measurement with that of the distribution of a variable. Averaging ordinal data may improve the degree to which the distribution of votes resembles a normal distribution, but it does not change the nature of the observations from ordinal to interval.

Concerning the analyses of subjective thermal comfort data, ISO 10551:2019 [2] gives guidance to the analysis of ordinal data. Unfortunately, it uses disputable arguments, based on McIntyre's work [9], to legitimise treating ordinal data from the ASHRAE 7-point thermal sensation scale as a continuous variable. In his paper published in 1978, McIntyre clearly stated that the 7-point warmth scale is ordinal and that, therefore, non-parametric statistics are the appropriate method. However, McIntyre also said that non-parametric statistics are generally related to hypothesis testing and are quite limiting for thermal comfort analysis. Therefore, utilising the method of graded dichotomies, he investigates whether these scales can be treated as intervals (i.e., if the psychological width of the categories can be approximated to be of equal spacing). McIntyre concluded that there is 'no reason to suppose that we are not dealing with an equal interval scale', even if nothing can be said to the extreme categories, that is 'cold' and 'hot' [9]. In addition, performing a Kolmogorov–Smirnov test[1] (K–S test), he found no significant deviation from normality and deduced that it is appropriate to use statistical methods that presuppose normality. However, checking whether an ordinal variable can be assumed to be interval for analytic purposes may work in some cases, but it does not constitute general proof. While this practice seemed reasonable at the time, considering that, until the 1960s, there was relatively little development of models for categorical responses (see page 1 of Ref [10]), nowadays, it is not. Advances in statistics and statistical software have provided many options for appropriate models of ordinal response variables, and many parametric ordinal models can be found in the literature. Furthermore, the K–S test can be applied only to continuous distributions, which is not the case analysed by McIntyre. In addition, the distribution used to compare the sample must be fully specified, that is, the location and scale parameters (i.e., mean and standard deviation) of the normal distribution must be known *a priori* and not estimated from the data. If these parameters are calculated from the data, as in the case of McIntyre, the critical region of the K–S test is no longer valid and should be determined by simulation.

In the following two sections, the notion of 'level of measurement' is introduced (Section 1.1), and the issue of analysing ordinal data as metric is discussed (Section 1.3). Discussion regarding the different types of scales employed (e.g., categorical scale, visual analogue scale, and graphic categorical scale), the number of anchors utilised, and the assumptions underlying their usage are outside the scope of this study. The interested reader is referred to previous studies such as Refs [7,8,11,12] for further discussions of these topics.

1.1. Level of measurement

A level of measurement is a classification that represents the nature of the information contained in the values assigned to the variables [13]. A widespread measurement classification is Stevens's typology [14], which is divided into four classes: nominal, ordinal, interval, and ratio. The nominal scale identifies or categorises the values of the variables but cannot order the categories; the ordinal scale, in which the values of the variables are ranked or ordered, is used for this purpose. For the interval scale, the intervals between the values of the variables are equally spaced, and the zero on the scale is arbitrary (i.e., the zero on the scale is a matter of convention or convenience). Conversely, the ratio scale has a true zero point, which defines the absence of the quantity being measured. As a consequence, ratios of magnitudes can be defined.

In Stevens's view, it is important to know which kind of scale one is dealing with because 'to each of these types of scales certain statistics are appropriate and others are not' [15], and a scale that retains meaning under a certain class of transformations should be limited to statistics whose meaning would not change if those transformations were applied to the data. Table 1 shows the different types of scales with their empirical operations, invariant mathematical transformations, and (permissible) measures of central tendency.

Stevens went beyond his simple typology and classified not just simple operations but also statistical procedures according to the scales for which they were permissible. The idea that a particular level of measurement prescribes or proscribes statistical methods has been strongly criticised by statisticians [16–18], and alternative taxonomies have been proposed. Mosteller and Tukey's typology [19] and Chrisman's typology [20] introduced an expanded list of levels of measurement to account for various measurements that do not fit well into Stevens's framework. The difference is that they do not prescribe statistical methods nor even suggest that statistical methods should depend on the levels of measurement. Statistical analyses make assumptions about the distributions of variables and/or errors, not about measurement levels. Of course, it is necessary to verify that these assumptions comply with the data at hand. However, to conclude that there is no value in the data types would be inaccurate. The notion of scale type is important, and Stevens's nomenclature is frequently used. For example, any designed experiment must distinguish between categorical factors (usually nominal or ordinal in Stevens's terminology) and metric/continuous covariates (usually intervals or ratios) [16]. However, these scale types derive from how the data were measured rather than being fundamental characteristics of the data themselves.

1.2. Statistical methods: A brief overview

As stated previously, one of the goals of thermal comfort research is to establish a relationship between the thermal environment and the human response. In a statistical modelling framework, this is generally achieved through regression analysis. Regression analysis is 'the blanket name for a family of data analysis techniques that examine relationships between variables' [21], which are categorised into a dependent variable ('outcome' or 'response' variable), Y, and one or more independent variables ('explanatory variables', 'predictors', 'covariates' or 'features'), X.

The most common approach utilised in thermal comfort research for the analysis of subjective thermal comfort data is linear regression. Another approach, even if far less common, is ordinal regression (e.g., Ref [22]). The main difference between the two is that linear regression requires the dependent variable to be continuous, while ordinal regression requires it to be ordinal. Even though different regression models have different mathematical underpinnings, they share a general form that can be expressed as the function of a random component, $g(\cdot)$, which refers to the conditional probability distribution of the response variable, and a systematic component, $h(\cdot)$, which refers to the explanatory variables. The systematic component is used as the predicted tendency of Y given the predictors. Nevertheless, Y is not predicted to be exactly $h(\cdot)$, but near $h(\cdot)$. That is, the best that can be done is to predict the probability that Y will have any particular value, given x. This probability density function (PDF) is the random component $g(\cdot)$. This is a more general notation that encompasses different models (i.e., it extends more easily to other models by focusing

[1] The K–S test was used to test if a sample comes from a population with a specific distribution.

Table 1
Types of measurement scales (from Ref [15]).

Scale	Empirical operations	Permissible transformations	Permissible measures of central tendency
Nominal	Determination of equality	Any one-to-one substitution	Mode
Ordinal	Determination of greater or lesser (rank-order)	Any increasing monotonic transform	Median
Interval	Determination of the equality of intervals or of differences	Multiplication by and addition of a constant	Arithmetic mean
Ratio	Determinations of the equality of ratios	Multiplication by a constant	Geometric mean Harmonic mean

on the conditional distribution of the response rather than the distribution of the error term [23]). The following sections briefly describe two regression-type models utilised to model subjective thermal comfort data.

1.2.1. General linear model

The most common approach utilised in thermal comfort research for the analysis of subjective thermal comfort data is the general linear model (GLM), which usually refers to the linear regression model. In this model, the continuous response variable is modelled given some predictors, generally assuming a conditional normal distribution of the response:

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2)$$

$$\mu_i = \eta_i \tag{1}$$

$$\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$$

where μ_i is the mean, σ is the standard deviation, and η_i is the predictor term function of some predictors $\mathbf{x}_i^{\mathrm{T}}$. The subscript i is to stress the dependency on the *ith* observation.

1.2.2. Cumulative link model

As mentioned previously, if the response variable is assumed to be ordinal (and therefore measured as ordinal), it is proper to analyse it with ordinal models. Cumulative link models (CLMs) belong to a broad class of models known as ordinal regression models. Following the categorisation of Bürkner and Vuorre [24], in addition to the cumulative models, other two distinct model classes belong to the ordinal regression models: sequential and adjacent-category models. Each of these models has a different rationale behind it and, consequently, a different application.

The rationale behind choosing a CLM lies in the fact that this model has a latent variable representation, which is in line with the general assumption underlying the rating scales. The idea is that the dependent variable Y is the categorisation of a latent (not observable) continuous variable $\tilde{\tilde{Y}}$. Fig. 1 illustrates this concept. The categorical outcome, Y (Fig. 1.a and .b) is a categorised version of an unobservable (latent) continuous variable, $\tilde{\tilde{Y}}$ (Fig. 1. c and .d). The dotted lines in the bottom figures divide the continuous latent variable into $K+1$ bins according to the threshold parameters $\{\tau_k\}$, with $k \in \{1, \cdots, K\}$. Consequently, the area under the curve in each bin represents the probability of the corresponding observed ordinal response (Fig. 1.a and .b). In Fig. 1, the thresholds are shown as not equidistant and equidistant (Fig. 1.c and .d, respectively) for illustrative purposes only. In practice, the thresholds are determined by nature; they are parameters to be estimated.

The conditional distribution of the response variable Y is assumed to follow a multinomial distribution where its probability vector is $\boldsymbol{\pi} = \{\pi_1, ..., \pi_k\}$ with $\pi_k = \Pr(Y = k)$. The cumulative probability corresponding to π_k is $\gamma_k = \Pr(Y \leq k)$ so that $\gamma_k = \pi_1 + ... + \pi_k$. The cumulative probabilities are then mapped to the real numbers through a link function. In this study, the probit function was chosen as the link function. The reason is that the probit link assumes the latent variable to be normally distributed[2] around the predicted central tendency (i.e., the mean of the latent scale) and is therefore comparable with linear regression. The mathematical form of the model can be written as:

$$Y_i \sim \text{Multinomial}(n, \boldsymbol{\pi}_i)$$

$$\text{Probit}(\gamma_{ik}) = \tau_k - \eta_i \tag{2}$$

$$\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$$

where τ_k are the thresholds parameters and η_i is the linear predictor term without an intercept.[3] The subscript i is to stress the dependency on the i^{th} observation.

For more explanations and practical guidelines for using this and other methods (i.e., sequential and adjacent-category models), along with detailed mathematical derivations and discussions, the reader is referred to Bürkner and Vuorre [24].

1.3. Ordinal-as-metric

While it is generally recognised that ordinal data are not metric, it is commonplace to analyse them with methods that assume metric responses. This is inappropriate for the following reasons. First and foremost, the ordinal variable's categories may not be equidistant since it is unknown the psychological distance between adjacent categories and whether these distances are the same across subjects. In a survey respondent's thinking, the difference between 'neutral' and 'slightly warm', for example, may be considerably smaller than the difference between 'warm' and 'hot', as demonstrated by Schweiker et al. [7,8]. Second, the distribution of ordinal categories can be nonnormal, especially if low (e.g., 'cold') or high (e.g., 'hot') values are commonly chosen. Third, the variances of the unobserved variables underlying the observed ordinal categories can vary, for example, between periods (e.g., seasons) and groups (e.g., gender). The ordinal-as-metric method cannot account for such uneven variances.

The issue of examining ordinal data as metrics was analysed in great detail by Liddell and Kruschke [25], whose arguments are summarised hereafter. To facilitate their understanding and explanation, Fig. 3 and Fig. 4 in Ref [25] have been adapted and reproduced here as Fig. 2. In this figure, the mean of the ordinal values (i.e., when the ordinal values are treated as metric) is plotted as a function of the latent mean, μ, and standard deviation (SD), σ. The four letter-labelled points represent a specific combination of μ and σ on the underlying latent scale that, if used as parameters in a cumulative probit model, would generate a particular pattern in the ordinal data. For instance, the point indicated by Ⓑ (i.e., group B) has a latent mean and standard deviation of $\mu = 2$ and $\sigma = 1$, respectively (Fig. 2.c) and an ordinal mean of 1.93 (Fig. 2.b).

From Fig. 2.a, four different 'effects' can be observed:

[2] Technically, the distributional assumption should be made on the error term, not the response variable. However, in linear regression, to assume the error as normally distributed around zero is equivalent to assuming the response to be normally distributed around the regression line.

[3] Omitting the intercept term allows the full set of thresholds $\tau_1, \cdots, \tau_k$ to be identified.

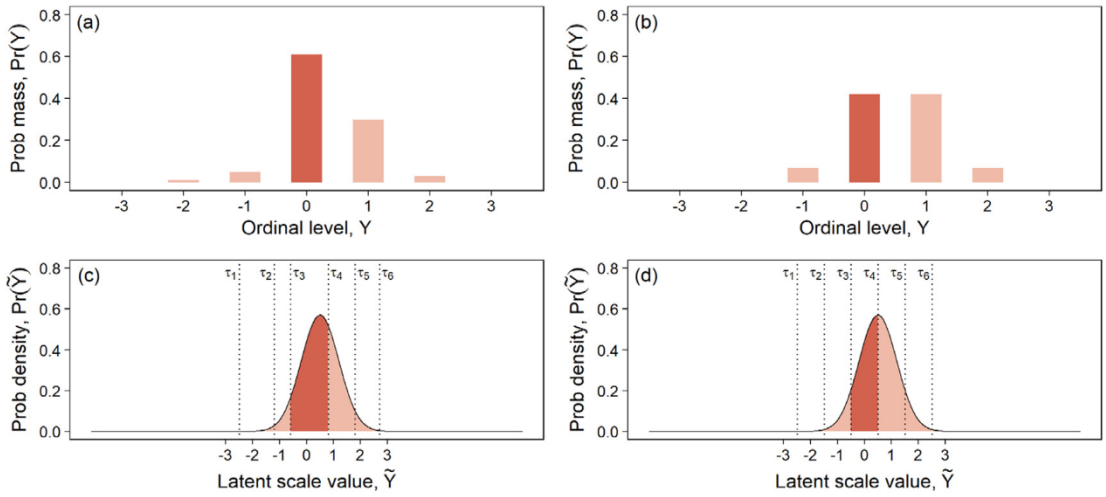

Fig. 1. Representation of the latent variable interpretation: observed values ((a) and (b)) and underlying latent distribution ((c) and (d)). Note. The thresholds τ_k (the dotted lines in (c) and (d)) are defined here as being not equidistant (c) and equidistant (d).

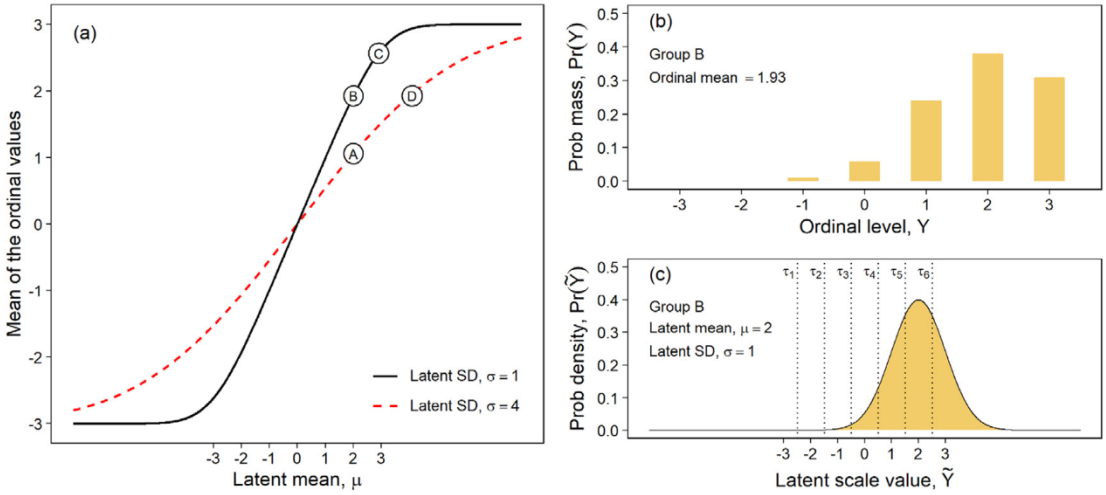

Fig. 2. Mean of the ordinal values as a function of latent mean, μ, and SD, σ (a); ordinal level (b) and latent scale value (c) for group Ⓑ (adapted from Ref [25]). Note. The thresholds τ_k (the dotted lines in (c)) are defined here as being equidistant.

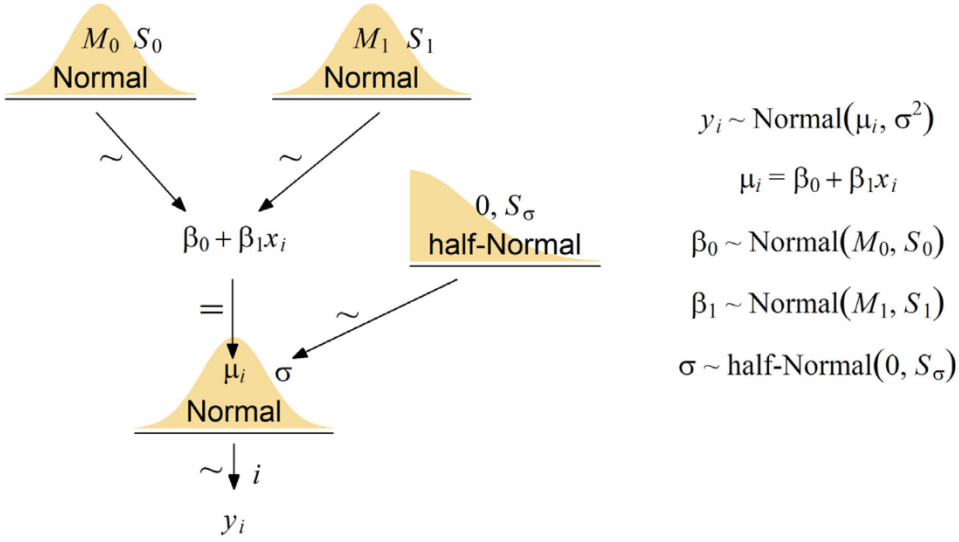

Fig. 3. Dependency diagram for a simple linear regression model (adapted from Ref [47]).

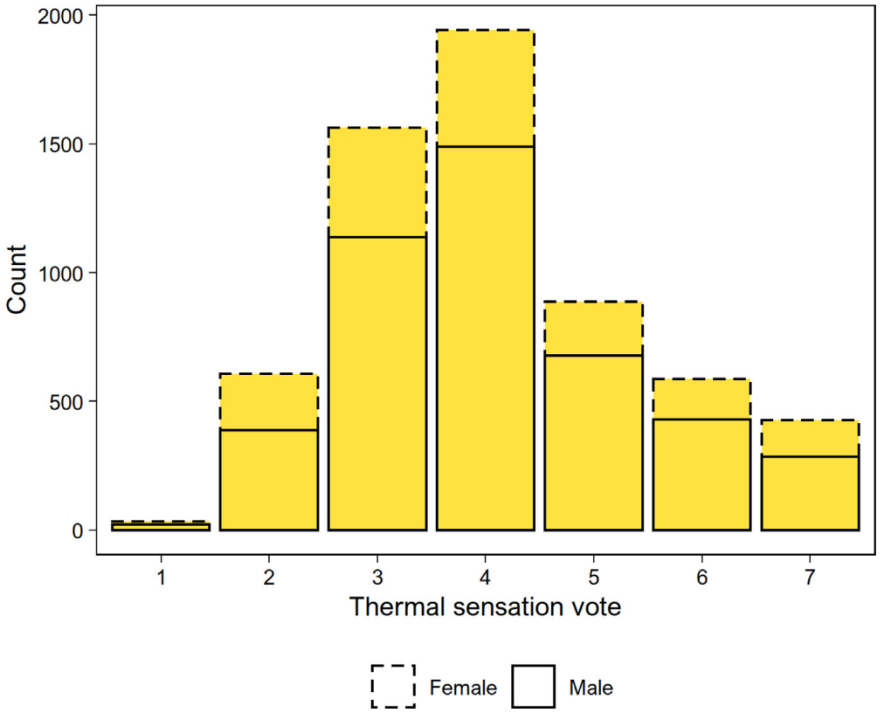

Fig. 4. Distribution of the thermal sensation vote.

1. Points Ⓐ and Ⓑ illustrate a false-alarm rate (Type I 'error'): these two groups have the same latent means, but the ordinal means are estimated as very different.
2. Points Ⓑ and Ⓓ illustrate a low correct-detection rate (Type II 'error'): for these two groups, the latent means are quite different, but the ordinal means are estimated as equal.
3. Points Ⓐ and Ⓓ illustrate a distorted effect-size estimate: here, the two groups have identical latent variances, but the difference in means on the horizontal axis (i.e., on the underlying latent scale) is larger than the corresponding difference on the vertical axis (i.e., on the ordinal-as-metric scale).
4. Points Ⓒ and Ⓓ illustrate a reversed effect-size estimate: here, the latent mean of group Ⓓ is greater than that of group Ⓒ, but the ordinal means for group Ⓒ are incorrectly estimated to be greater than those of group Ⓓ.

Liddell and Kruschke [25] posited that there are infinite combinations of underlying parameter values (μ and σ) that lead to inflated false-alarm rates, or low rates of correct detection, or distorted effect-size estimates, or inversions of differences between groups. Consequently, analysing ordinal data with metric methods (i.e., methods that assume continuous response variables), such as *t*-test, analysis of variance (ANOVA) and linear regression, could lead to the aforementioned issues. Furthermore, linear regression applied directly to ordinal values can misestimate regression coefficients, leading to incorrect inferences about differences or non-differences in slopes across conditions, as well as the existence or absence of non-linear trends. For further discussion and examples, the reader is referred to Liddell and Kruschke [25].

1.4. Objective and relevance of this study

Establishing the link between the thermal environment and the human sense of warmth is one of the goals of thermal comfort research. This is usually achieved by measuring the subjective human thermal response to different thermal environments. In this field, it is common practice to analyse subjective human thermal responses independently of how they have been measured. That is, the statistical analysis is unrelated to the modalities of the data that have been acquired. For example, Zhang and de Dear [26] state that thermal sensation vote 'although it is essentially an ordinal variable, the thermal comfort research community has usually regarded it as a continuous variable'. From this statement, the authors (i) highlight that there is a difference between ordinal and continuous variables but (ii) specify that, within the thermal comfort research community, there is the tendency to consider it as continuous. In other words, linear regression is widely used to analyse thermal sensation votes (TSV) measured on an ordinal scale. Liddell and Kruschke [25] showed that analysing ordinal data as if they were continuous could lead to misleading results. This is particularly relevant for thermal comfort research, since thermal comfort models are mainly based on ordinal data analysed as if they were continuous (e.g., Refs [26–39]). This might be a concurrent factor to explain why conflicting results were found in previous research where, for example, gender was shown to be or not an influential factor in determining human responses to the thermal environment. Furthermore, these models are included in international standards, such as EN 15251:2007 [40], replaced by EN 16798–1:2019 [41], and ASHRAE 55:2020 [1], which are used in the design and operation of buildings all around the world.

This paper focuses on analysing the data once they have been collected and not on the correctness of the level of measurement utilised to measure them (see Refs [7;8] for further discussions of this topic). For this purpose, this study leverages the largest global thermal comfort database to date. The aim of the paper is twofold. The first aim is to overview the methods commonly used to analyse subjective thermal comfort data from a statistical perspective. The second aim is to highlight the ordinal-as-metric issue that is often not considered and to spur researchers to analyse these kinds of data more critically. It is essential to emphasise that we are not advocating that the specific approach hereafter presented as the best way to analyse these kinds of data: the approach presented is merely-one of the possible ways to do so. Statistics is a field that is an art as much as it is a science. Although statistical theory is founded on exact assumptions and conditions, the real world is seldom that straightforward. Consequently, the practice of statistics

involves a tremendous number of choices, and the challenge is how to make those choices.

2. Methodology

2.1. Bayesian approach to regression

In this study, a Bayesian approach is used to analyse the data. This approach is not entirely new in thermal comfort studies (e.g., Refs [42–44]); however, it is not an established practice either. Since statistical knowledge in this field generally tends towards 'frequentist' principles, it is essential to explain the Bayesian approach and compare it with the frequentist one. Nevertheless, the aim of this paper is neither to go into details about their differences, nor to be a full introduction to either approach. For a more complete treatment, see, for example, Refs [45] and [46].

Essentially, the divide between frequentists and Bayesians is in the interpretation of probability. For frequentists, probabilities are associated with frequencies of events. For Bayesian, probabilities are related to their own understanding (i.e., certainty or uncertainty) of events. This difference has important implications in the analysis of data. For instance, in a frequentist view, the parameter θ is considered a fixed (i.e., constant) but unknown quantity and only the information from the sampling data is relevant for the inference. On the contrary, Bayesian statistics estimate the full (joint) posterior distribution of the parameters (i.e., the probability of the parameters given the observed set of data), which is generally calculated as:

$$\Pr(\theta|Y) = \frac{\Pr(Y|\theta)\Pr(\theta)}{\Pr(Y)} \tag{3}$$

where $\Pr(Y|\theta)$ is the likelihood, $\Pr(\theta)$ is the prior distribution, and $\Pr(Y)$ is the marginal likelihood. Here the parameters are considered random variables and not constant, as in the frequentist approach.

In Eq. (3), the $\Pr(\theta)$ represent the prior 'belief' about the distribution of the parameters, and such a belief must be specified. Since there is no single method for choosing a prior (i.e., prior probability distribution), different priors can be introduced, leading to potentially different posterior distributions and conclusions. This subjectivity is the main criticism of Bayesian inference. Furthermore, obtaining the posterior distribution analytically is rarely possible. Consequently, Bayesian statistics relies on Markov Chain Monte Carlo (MCMC) methods to estimate the posterior distributions of the parameters of interest. MCMC methods have a higher computational cost and fitting a model with Bayesian statistics is generally slower than the frequentist approach. However, Bayesian methods are usually more flexible and have more informative results (e.g., estimating a full posterior distribution, rather than a single point with a measure of uncertainty). Such advantages are often worth the increase in computational cost. Bayesian estimation does not have specific assumptions but relies on the model's assumptions, since those are the assumptions about the likelihood function. The fundamental assumption is that the likelihood function chosen is a reasonable representation of the data.

Generally, the assumptions behind a Bayesian model are not directly mentioned because they are stated when defining likelihood and priors. For example, Fig. 3 illustrates the formulation of a Bayesian model for simple linear regression. The corresponding mathematical formulations are added to the side for clarity.

This figure shows the assumptions about the random component (i.e., the conditional distribution assumptions for y) and the functional form of the systematic component (i.e., the expression for μ). The distributions of the parameters β_0, β_1, and σ are the priors. Since the standard deviation cannot be less than zero, a half-normal distribution was selected as its prior (however, other distributions could have been chosen, such as exponential and uniform). For an introduction to Bayesian analysis or more advanced treatment, see Refs [48] and [47], respectively.

2.2. Data preparation and software

As mentioned in Section 1.4, this study leverages the largest global thermal comfort database to date. This database, called ASHRAE Global Thermal Comfort Database II (downloaded from the University of California's DASH repository [49]), is an open-source database that includes approximately 107,500 sets of paired subjective comfort votes and objective instrumental measurements of the thermal environment. These observations were derived from field studies conducted worldwide between 1995 and 2016. A quality assurance check was performed on each dataset before its inclusion in the final database (see Ref [50] for more details).

To achieve the aim of this study, the dependent variable needs to be measured on the ordinal scale. Unfortunately, the ASHRAE Global Thermal Comfort Database II does not distinguish between scales, and ordinal and continuous measurements are mixed. Additionally, even if all datasets composing the database went through a rigorous quality assurance process to harmonise their contents, it is reasonable to assume that each dataset has some unique peculiarities – different measurement protocols, questionnaires, or instruments. This aspect of the database would require that analysis of the entire database be carried out with an 'appropriate' method that considers these peculiarities (e.g., multilevel modelling) because, otherwise, the results may be unpredictably affected. For the purpose of this study, in order to reduce the uncertainty due to the unique peculiarities of different datasets, the following analysis was carried out on the data deriving from a single study.

Among the subjective thermal comfort votes available in ASHRAE Global Thermal Comfort Database II, the highest number of observations are thermal sensation votes (TSV). For this reason, TSV was selected as the dependent variable. However, the same analysis could be applied to the other rating scales if measured on the ordinal scale. For simplicity, only the two variables (one categorical and one continuous) presented in Table 2 were utilised as covariates during the analysis. Indeed, thermal sensation depends on other variables, such as clothing, metabolic rate, air movement, radiant temperature, and relative humidity, and perhaps on several variables not yet clearly identified. Also, it is likely that not accounting for possible confounders affects the estimation of the models' coefficients. However, given that this study is an illustrative example, which aims to highlight the issue of analysing ordinal data as they were continuous, the issue of including/excluding variables (regardless of their importance) from the model can be overlooked.

To lessen the uncertainty caused by the distinctive characteristics of the various datasets and thus improve the completeness and homogeneity of the data to be analysed, the following steps were followed to select the dataset:

1. Among all datasets included in the ASHRAE Global Thermal Comfort Database II, only those having the TSV measured as an ordinal variable were selected. This evaluation was performed graphically by plotting the TSV distribution and led to the selection of 43 datasets.
2. All rows containing missing values for thermal sensation votes, gender, or air temperature were deleted.
3. The datasets were subsequently filtered by selecting only those with at least one observation per gender (i.e., one observation for males and one for females) in each category of the TSV. Only 16 datasets met this criterion.
4. To increase the reliability of the results, the dataset with the largest sample size was selected from the remaining ones.

Table 2
List of covariates used in the model.

Variable	Code	Type	Unit
Gender	*Gender*	Categorical	female (reference) / male
Air temperature	*Tair*	Continuous	°C

This procedure led to the selection of the dataset of Indraganti et al. [37]. The dataset comprised 6048 observations (~27 % female) collected during 14 months from 2787 individuals (all Indian nationals within the age group of 18–48 years) with TSV distribution shown in Fig. 4. This dataset also had no missing values for thermal sensation votes, gender and air temperature, and data were collected under a wide range of indoor air temperatures (min = 20.80 °C; 1st quartile = 25.80 °C; median = 26.80 °C; mean = 27.06 °C; 3rd quartile = 28.30 °C; max = 36.50 °C) with no detectable outliers. More details regarding the field survey can be found in Indraganti et al. [37].

All statistical analyses were performed using R [51] with the RStudio integrated development environment [52]. Regression analyses, using both the cumulative probit and classical linear regression, were performed with the *brms* package [53], and the respective graphs were created with the *ggplot2* package [54] via the *tidybayes* package [55].

2.3. Model parametrisation

Before proceeding with the analysis, it is essential to briefly explain how *brms* parameterises the cumulative probit model because this has repercussions on its interpretation. The cumulative distribution function (CDF) of an ordinal model based on cumulative probabilities with probit link (i.e., cumulative probit model) can generally be stated as:

$$\begin{aligned}&\Pr(Y_i \le k|\{\tau_k\},\eta_i,\sigma_i) = \Phi\left(\frac{\tau_k-\eta_i}{\sigma_i}\right)\\&\eta_i = \beta_0 + \sum\nolimits_1^l \beta_l x_{l,i}\\&\log(\sigma_i) = \delta_0 + \sum\nolimits_1^m \delta_m x_{m,i}\end{aligned} \tag{4}$$

where Φ indicates the cumulative normal distribution function, τ_k are the thresholds parameters, η_i is the linear predictor term and σ_i is the standard deviation. The subscript i is to stress the dependency on the i^{th} observation. With $K+1$ ordinal values, a model has $(K+1)+1$ parameters ($\tau_1,\cdots,\tau_k$, η_i and σ_i) and is undetermined. Therefore, two parameters need to be fixed. *Brms* parametrises the model by fixing $\beta_0 = 0$ and $\delta_0 = 0$ and freely estimating all the thresholds, $\tau_1,\cdots,\tau_k$. When there are no predictors for η_i and σ_i in the model (i.e., unconditional model), $\eta_i = \beta_0 = 0$ and $\sigma_i = \exp(\delta_0) = 1$. Therefore, instead of estimating η_i and σ_i from a normal cumulative distribution function, *brms* uses the standard normal cumulative distribution function $\Phi(z)$. As a consequence, the parameters are expressed on the latent variable scale, that is, in units of ordered probit. Furthermore, since *brms* parametrise the model as:

$$\text{Probit}(\Pr(Y_i \le k|\{\tau_k\},\eta_i,\sigma_i)) = \frac{\tau_k-\eta_i}{\sigma_i} = \frac{\tau_k-(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{\sigma_i} \tag{5}$$

a positive coefficient for β indicates that an increase of 1-unit of the associated variable x_i increases the thermal sensation vote. Stated analogously, voting in higher categories is more likely. The interpretation would have been the opposite if the model was parametrised differently (i.e., with a '+' instead of a '−'). A positive coefficient for β would have indicated that an increase of 1-unit of the associated variable x_i would decrease the thermal sensation vote.

For comparison, the CDF for the ordinary linear regression model can generally be stated as:

$$\begin{aligned}&\Pr(Y_i \le y|\eta_i,\sigma_i) = \Phi\left(\frac{y-\eta_i}{\sigma_i}\right)\\&\eta_i = \beta_0 + \sum\nolimits_1^l \beta_l x_{l,i}\\&\log(\sigma_i) = \delta_0 + \sum\nolimits_1^m \delta_m x_{m,i}\end{aligned} \tag{6}$$

where Φ indicates the cumulative normal, η_i is the linear predictor term and σ_i is the standard deviation. The subscript i is to stress the dependency on the i^{th} observation. Here the β_0 and δ_0 are not fixed and therefore freely estimated by the model.

The following modelling steps were carried out for the cumulative probit model and compared with an ordinary linear regression, referred to as Gaussian (ordinal-as-metric) model.

2.4. Modelling steps

The analysis of this illustrative example was carried out following the subsequent steps:

1. *Fitting an unconditional model*: The goal of a modelling strategy is to try to reproduce or predict an observable phenomenon via the lens of a model. Before incorporating a predictor, the unconditional model can be used to test the 'goodness' of the modelling technique. For example, if a model makes implausible predictions that are unobservable in reality, perhaps a different technique should be adopted. In this model, the linear predictor term and standard deviation (SD) are given by:

$$\begin{aligned}&\eta = \beta_0\\&\log(\sigma) = \delta_0\end{aligned} \tag{7}$$

where for the cumulative probit model $\beta_0 = 0$ and $\delta_0 = 0$, whereas for the Gaussian (ordinal-as-metric) model β_0 and δ_0 are freely estimated. The subscript i is absent from both η and σ because, in this model, they do not depend on the i^{th} observation.

2. *Fitting a categorical variable*: In this step, the categorical variable *Gender* was added to the previous model. In this model, the linear predictor term and SD are given by:

$$\begin{aligned}&\eta_i = \beta_0 + \beta_1(\text{gender}_i)\\&\log(\sigma) = \delta_0\end{aligned} \tag{8}$$

where for the cumulative probit model $\beta_0 = 0$ and $\delta_0 = 0$, whereas for the Gaussian (ordinal-as-metric) model β_0 and δ_0 are freely estimated. *Gender* is a dummy variable, coded as 0 for females (i.e., the reference category) and 1 for males. The subscript i is absent from σ because, in this model, the standard deviation does not depend on the i^{th} observation (i.e., it is assumed to be a constant).

Unequal standard deviations can be included in the model by specifying an additional regression formula for the standard deviation component. In the context of this example, allowing for unequal standard deviations implies inquiring whether the standard deviations for TSV differ across the two categories of *Gender*. Consequently, the linear predictor term and SD are given by:

$$\begin{aligned}&\eta_i = \beta_0 + \beta_1(\text{gender}_i)\\&\log(\sigma_i) = \delta_0 + \delta_1(\text{gender}_i)\end{aligned} \tag{9}$$

where for the cumulative probit model $\beta_0 = 0$ and $\delta_0 = 0$, whereas for the Gaussian (ordinal-as-metric) model β_0 and δ_0 are freely estimated. *Gender* is a dummy variable, coded as 0 for females (i.e., the reference category) and 1 for males. The subscript i is to stress the dependency on the i^{th} observation.

3. *Fitting a linear predictor*: In this step, the continuous variable *Tair* was added to the previous model. However, *Tair* was standardised before entering the model (i.e., *Tair_s)*. Standardisation (i.e., subtracting the mean and dividing by its standard deviation) is done to improve the efficiency of MCMC sampling, that is, to

reduce autocorrelation in the chains. In principle, it is unnecessary to standardise, but it would take more time for the chains to produce a reasonable, effective sample size. Furthermore, standardising does not change the parameter estimates.

In this model, the linear predictor term and SD are given by:

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1(\text{gender}_i) + \beta_2 \text{Tair\_s}_i \\ \log(\sigma_i) &= \delta_0 + \delta_1(\text{gender}_i) \end{aligned} \tag{10}$$

where for the cumulative probit model $\beta_0 = 0$ and $\delta_0 = 0$, whereas for the Gaussian (ordinal-as-metric) model β_0 and δ_0 are freely estimated. *Gender* is a dummy variable, coded as 0 for females (i.e., the reference category) and 1 for males, and *Tair_s* is the standardised predictor of air temperature. The subscript i is to stress the dependency on the i^{th} observation.

4. *Structured thresholds*: In all the previous cumulative probit models, the thresholds $\{\tau_k\}$ were defined as 'flexible' providing the standard unstructured thresholds. However, restrictions such as equidistance can be imposed on the thresholds, which restricts the distance between consecutive thresholds to be of the same size (i.e., equally spaced). This allows assessing the assumptions that the subjects used the response scale (i.e., TSV) in such a way that the distance between adjacent response categories is the same, that is, $\tau_k - \tau_{k-1} =$ constant for $k \in \{1, \cdots, K\}$. This check was performed graphically by comparing the spacing of the equidistant thresholds with the average distance between consecutive unstructured thresholds and, formally, by comparing the computed models' relative fit to the data. The method used to assess the relative fit was approximate leave-one-out cross-validation (LOOCV) [56], where smaller values indicate a better fit. In particular, the LOO information criterion (LOOIC) for the two models and their differences was calculated. In the context of model selection, a LOOIC difference higher than four times its associated standard error suggests that the model with the lower LOOIC value fits the data significantly better.

3. Results of the statistical analysis of subjective thermal comfort data

3.1. Unconditional model

The unconditional model for the cumulative probit and gaussian (ordinal-as-metric) models are the thresholds-only and intercept-only models, respectively. The unconditional model results are shown in Table 3, while Fig. 5 shows its posterior prediction. Here the data generated from the thresholds-only and intercept-only models are compared with the empirical data.

The posterior predictive distribution for the cumulative probit model (Fig. 5.a) visually describes the distribution of the outcomes. Conversely, the posterior predictions for the Gaussian (ordinal-as-metric) model (Fig. 5.b) are not a good fit, and they also have impossible predictive outcomes (i.e., value below the category '1′ that is, 'cold' and above the category '7′, that is, 'hot'). Fig. 6 shows the standard normal distribution underlying the ordinal data and the position of the estimated thresholds $\{\tau_k\}$ (see Table 3). The area under the curve in each bin represents the probability of the corresponding observed ordinal response (see Fig. 4).

A 'pseudo' CDF is plotted in Fig. 7 for illustrative purposes only[4] to inspect further and compare the two models. This direct contrast shows that the cumulative probit model better describes the data than the Gaussian (ordinal-as-metric) model.

3.2. Fitting a categorical variable

In this section, the categorical variable *Gender* is added to the unconditional model. As described previously, *brms* parametrises the cumulative probit model by fixing $\eta_i = \beta_0 = 0$ and $\sigma_i = \exp(\delta_0) = 1$. Therefore, the underlying Gaussian for the reference category of *Gender* (i.e., female) will be Normal$(0, 1)$. Thus, the parameter value for the other category of *Gender* (i.e., male) is the difference in means expressed on the latent variable scale for the reference category. The results of this model are shown in Table 4.

The model above presumes that the standard deviation of the latent variable is the same throughout the model (see Fig. 8.a). However, this assumption can be relaxed by allowing unequal standard deviations for *Gender*.

Table 5 shows the results of the fitted cumulative probit model with group-specific η_i and σ_i values for the underlying normal distributions of the ordinal variable, Y (the results for the gaussian (ordinal-as-metric) model are added for comparison). There is a difference in the approach that *brms* uses to model unequal standard deviation for the cumulative probit and the conventional Gaussian model. The SD of both is modelled on the log scale to constrain its value to be 0 or larger. However, the parameter related to the latent standard deviations (i.e., for the cumulative probit model) is called *disc* (a contraction of 'discrimination'), following the conventions in item response theory. This parameter is not related to the standard deviation itself, but to the inverse of the SD, that is, $\sigma = 1/disc$. Consequently, the estimated SD for male is $\sigma_{male} = 1/\exp(0.14) = 0.87$ and $\sigma_{male} = \exp(0.28) = 1.32$ for the cumulative probit and gaussian (ordinal-as-metric) model, respectively (values from Table 5).

Fig. 8 shows the density plot of the two underlying latent distributions for *TSV* given *Gender*, expressed in terms of the posterior means of each parameter. The underlying distribution for the reference category (i.e., female) is the standard normal, while the mean and SD for the other category (i.e., male) are estimated from the model. In Fig. 8.b, the parameter value for *Male* is still the difference in means expressed on the latent variable scale for the reference group, but this time in terms of the SD of the reference group's latent variable (i.e., female). The SD for the two categories of *Gender* is not assumed to be the same, but it is allowed to vary. Also, the thresholds, $\{\tau_k\}$, are on the scale of the reference category's latent variable and are assumed to be the same for the two categories of *Gender*.

Table 5 shows that the coefficient for *Disc.Male* is positive without zero overlapping the 95 %-CI. This indicates that the SD for male is smaller than the female (i.e., $\sigma_{male} = 1/\exp(0.14) = 0.87 < \sigma_{female} = 1/\exp(0) = 1$) and the evidence based on the data and the applied model is sufficient in terms of 'standard decision rules'. As such, in this sample, the standard deviations for TSV differ across the two categories of *Gender*.

Fig. 9 shows the marginal posterior distribution of the parameters (i.e., the means and standard deviations) and the effect sizes for the cumulative probit (green) and gaussian (ordinal-as-metric) (orange) models, respectively. The cumulative probit model does not have a distribution for female because this is the reference category and its mean and standard deviation are fixed.

Here, the effect size is computed by dividing the difference of the means of the two groups by the pooled standard deviation given in Eq. (11):

$$\sigma_p = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \tag{11}$$

which is defined for two groups with unequal sample sizes (where n_1 and n_2 are the group-based sample sizes). In Fig. 9, the black line and dot at the bottom of each distribution represent the highest density interval (HDI) and the mode, respec-

[4] Strictly speaking, a cumulative distribution function is defined as a continuous function only for continuous variables. For discrete variables, it should be a step function.

Table 3
Regression coefficients for the unconditional model.

	Estimate	Est. Error	L-95 % CI*	U-95 % CI*
Cumulative probit model				
Threshold 1,τ_1	−2.54	0.06	−2.66	−2.42
Threshold 2,τ_2	−1.25	0.02	−1.29	−1.21
Threshold 3,τ_3	−0.35	0.02	−0.38	−0.32
Threshold 4,τ_4	0.48	0.02	0.45	0.52
Threshold 5,τ_5	0.96	0.02	0.92	1.00
Threshold 6,τ_6	1.47	0.02	1.42	1.52
Gaussian (ordinal-as-metric) model				
Intercept	4.08	0.02	4.04	4.11
Sigma	1.37	0.01	1.35	1.39

*CI stands for credible interval (based on quantiles).

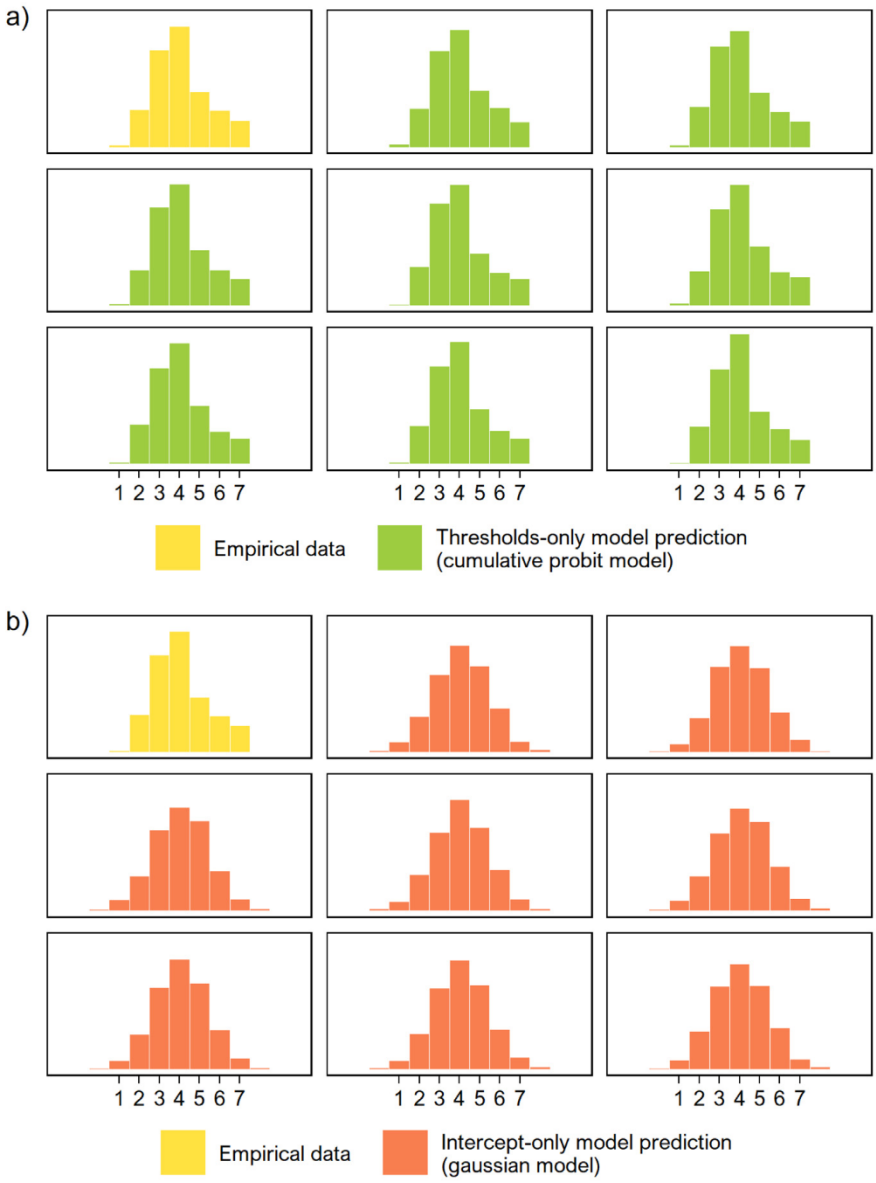

Fig. 5. Posterior prediction for (a) the thresholds-only and (b) intercept-only model. Note. The green and red histograms are obtained from 8 draws from the posterior predictive distribution of the thresholds-only and intercept-only models, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tively. The HDI is a way to summarise the distribution by defining an interval that spans over the distribution so that every point inside the interval has higher credibility than any point outside it. These intervals (i.e., the black lines) are defined here to span over 95 % of the distribution; therefore, they represent the 95 % HDIs.

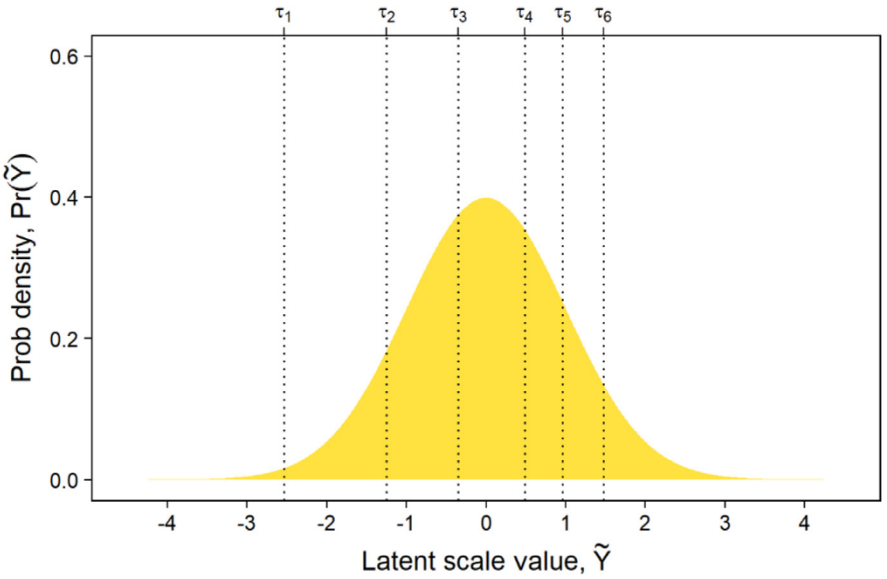

Fig. 6. Standard normal distribution underlying the ordinal data. Note. The linear predictor term η and the standard deviation σ are fixed at 0 and 1, respectively.

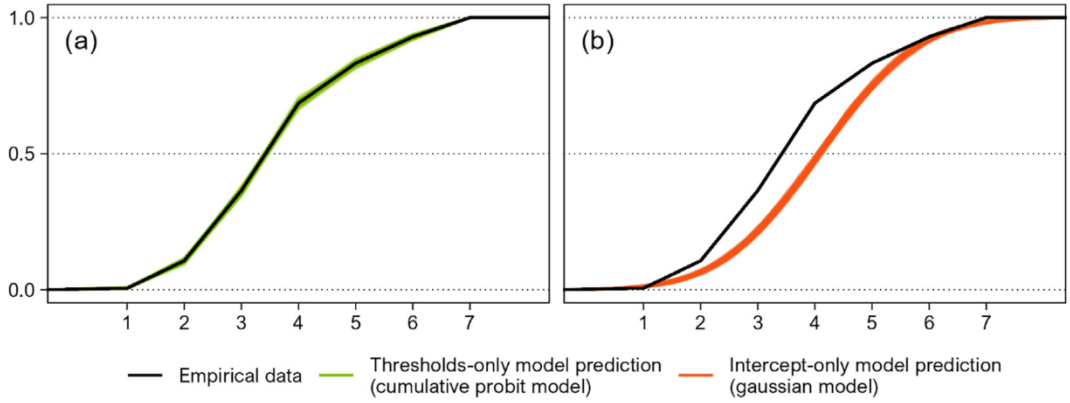

Fig. 7. Superimposition of the CDF for the (a) cumulative probit and (b) Gaussian (ordinal-as-metric) model. Note. The green and red lines are obtained from 100 draws from the posterior predictive distribution of the thresholds-only and intercept-only models, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Regression coefficients for the model with only a categorical variable (assuming constant standard deviation).

		Estimate	Est. Error	L-95 % CI*	U-95 % CI*
Cumulative probit model					
Threshold 1,τ_1		−2.49	0.06	−2.62	−2.37
Threshold 2,τ_2		−1.20	0.03	−1.26	−1.14
Threshold 3,τ_3		−0.30	0.03	−0.35	−0.24
Threshold 4,τ_4		0.53	0.03	0.48	0.59
Threshold 5,τ_5		1.01	0.03	0.96	1.07
Threshold 6,τ_6		1.52	0.03	1.46	1.58
Gender	female	reference			
	male	0.07	0.03	0.01	0.13
Gaussian (ordinal-as-metric) model					
Intercept		4.03	0.03	3.96	4.10
Gender	female	reference			
	male	0.06	0.04	−0.02	0.14
Sigma		1.37	0.01	1.35	1.39

*CI stands for credible interval (based on quantiles).

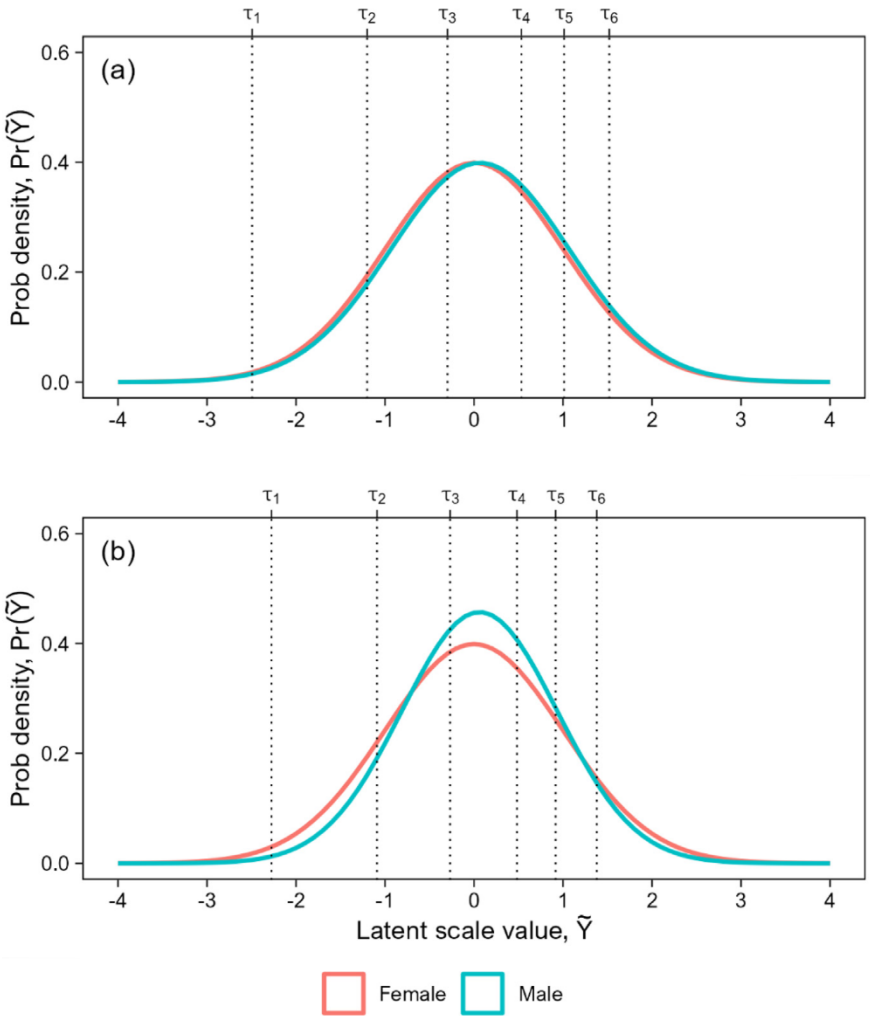

Fig. 8. Density plot of the two underlying latent distributions for *TSV* with constant (a) and unconstrained (b) standard deviation for *Gender*. Note. For both graphs, for the reference category of *Gender* (i.e., female), the linear predictor term η and the standard deviation σ are fixed at 0 and 1, respectively.

Table 5
Regression coefficients for the model with only a categorical variable (allowing the standard deviation to vary by group).

		Estimate	Est. Error	L-95 % CI*	U-95 % CI*
Cumulative probit model					
Threshold 1,τ_1		−2.28	0.07	−2.41	−2.14
Threshold 2,τ_2		−1.09	0.03	−1.16	−1.02
Threshold 3,τ_3		−0.27	0.03	−0.32	−0.21
Threshold 4,τ_4		0.48	0.03	0.43	0.54
Threshold 5,τ_5		0.92	0.03	0.85	0.98
Threshold 6,τ_6		1.38	0.04	1.30	1.46
Gender	female	reference			
	male	0.06	0.03	0.00	0.12
Disc.Male		0.14**	0.02	0.09	0.18
Gaussian (ordinal-as-metric) model					
Intercept		4.04	0.04	3.96	4.11
Gender	female	reference			
	male	0.06	0.04	−0.02	0.14
Sigma.Female		0.39**	0.02	0.36	0.42
Sigma.Male		0.28**	0.01	0.26	0.31

*CI stands for credible interval (based on quantiles).
**Values expressed on the logarithmic scale.

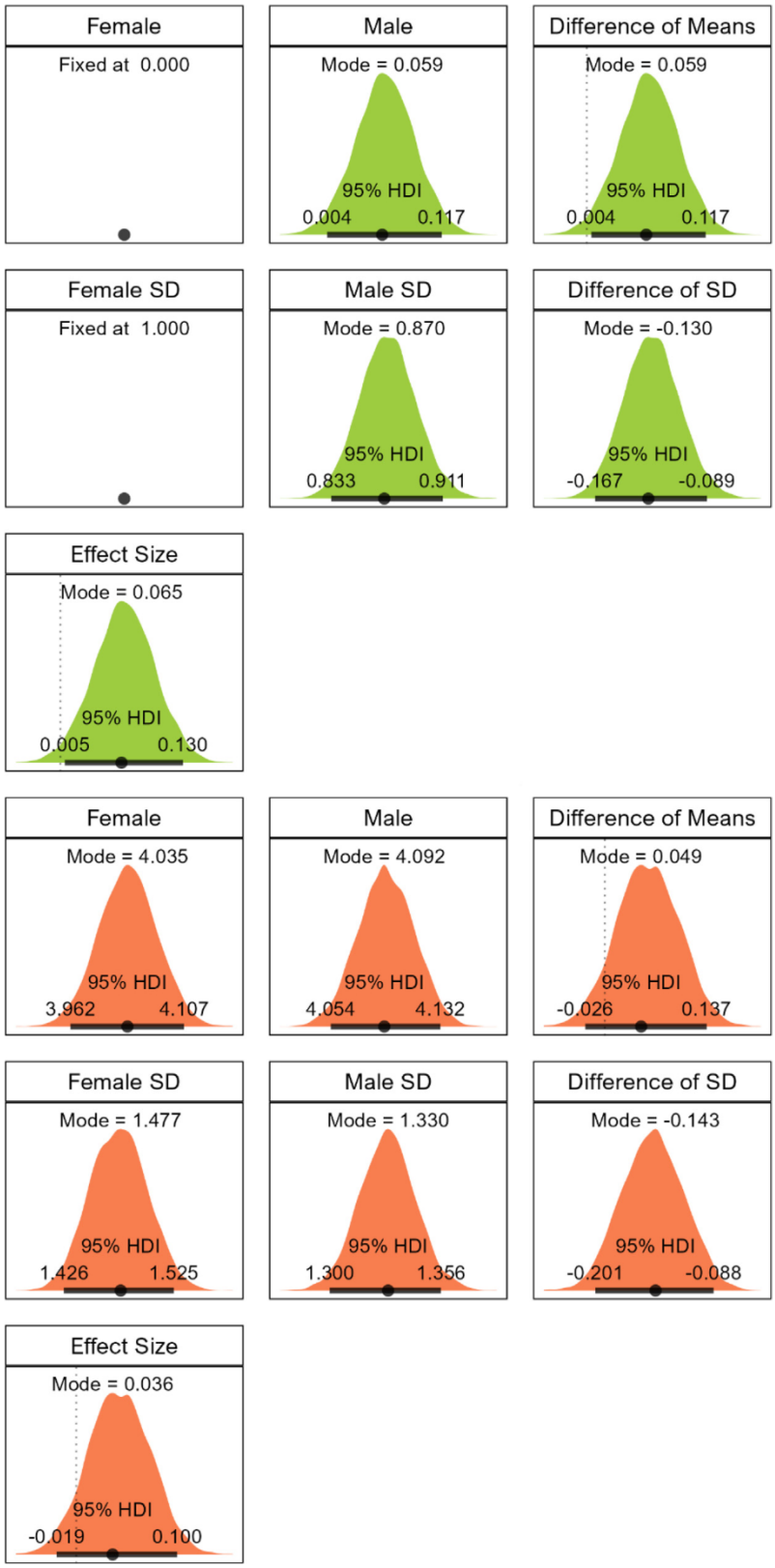

Fig. 9. Posterior distributions for the model that include the variable *Gender*: cumulative probit (green) and Gaussian (ordinal-as-metric) model (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Focusing on effect sizes and differences in means and standard deviations, two different results can be observed from Fig. 9. For the cumulative probit model, it can be seen that zero is outside the 95 % HDI for the effect size and the difference in means and SD. However, in the gaussian (ordinal-as-metric) model, zero is included in the 95 % HDIs for the effect size and the difference in

SD while it is outside the 95 % HDI for the difference in means. As a result, at this stage of the modelling phase, the same data analysed with two models convey different conclusions in terms of 'standard decision rules'. While the cumulative probit model conveys a difference in effects size and difference in means for *Gender*, the gaussian (ordinal-as-metric) model does not.

3.3. Fitting a linear predictor

In this section, the continuous variable *Tair_s* was added to the previous model, that is, the model with the variable *Gender* and unconstrained standard deviation (i.e., where the standard deviation is allowed to vary by *Gender*). The results of fitting this model are presented in Table 6. Here can be seen that after adding *Tair_s* as a predictor, the upper and lower 95 % CI (i.e., L-95 % CI and U-95 % CI) for the male coefficient of the gaussian (ordinal-as-metric) model does not include zero. Consequently, at this stage of the modelling phase, the same data analysed with the two models now convey the same conclusions regarding *Gender*; both the Gaussian (ordinal-as-metric) and cumulative probit models show a difference between males and females.

The marginal distribution of the standardised regression coefficient for *Tair_s* is shown in Fig. 10. As explained in Section 2.4, this is a standardised regression coefficient and represents a sort of effect size for air temperature. The two models give a different distribution for the coefficient, with a distinct mode and 95 % HDIs. The coefficient of the cumulative probit model is expressed on the underlying latent scale, while the Gaussian (ordinal-as-metric) coefficient refers to the ordinal scale. As a consequence, the Gaussian (ordinal-as-metric) coefficient for air temperature is overestimated.

3.4. Structured thresholds

Fig. 11.a shows the spacing of the equidistant threshold (i.e., structured thresholds), while Fig. 11.b shows the average distance between consecutive unstructured thresholds (i.e., $\frac{1}{k}\sum_{1}^{k}(\tau_k - \tau_{k-1})$). It can be seen that zero is outside the 95 % HDI for the difference between the spacing for structured and unstructured thresholds (Fig. 11.c), suggesting that, in terms of 'standard decision rules', the thresholds should not be approximated as equidistant.

Furthermore, whether the restriction on the thresholds is warranted by the data can be assessed formally by comparing the relative fit of the computed models to the data. Table 7 shows the estimated LOO information criterion (LOOIC) for the two models and their differences. It can be seen that the cumulative probit model with unstructured thresholds has a significantly better fit (smaller LOOIC value) than the structured thresholds one since the difference in LOOIC (i.e., LOOIC.diff) is very large (more than 12 times the corresponding standard error, SE.diff). This provides evidence that, in this sample, the thresholds should not be assumed to be equally spaced.

4. Discussion

This study aimed to highlight the ordinal-as-metric issue during the subjective thermal comfort data analysis. Here, the method used to assess the reliability of the two models (i.e., the cumulative probit and gaussian (ordinal-as-metric) approach) is the so-called posterior predictive checks, a commonly used technique in Bayesian analysis. In essence, after computing the posterior distribution of the parameters, many simulated data are generated and compared with the observed ones. Therefore, the posterior predictive check is used to look for 'systematic discrepancies that would be meaningful to address' [47]. This approach has the evident drawback of evaluating a model against the same data used to estimate its parameters. Unsurprisingly, the model predicts the data used to fit the parameters, but even this simple test fails when the model's assumptions are severely violated. These systematic discrepancies are clearly shown in Section 3.1 when fitting the unconditional model for both the cumulative probit and gaussian (ordinal-as-metric) approach.

The influence of the statistical analysis on the conclusions has also been shown by Schweiker et al. [57]. In this study, the same thermal sensation votes were analysed with both linear and ordinal regression. The authors showed that the two statistical methods led to differences in the thermal conditions perceived as 'optimal' as well as between gender (i.e., female and male). However, compared with our study, important distinctions need to be made. To begin with, in Schweiker et al. [57], the analysis was carried out using mixed-effect models (linear and ordinal mixed-effect regression, specifically). This modelling strategy (also known as multilevel modelling) was applied to account for repeated measures (i.e., multiple observations for each subject). Moreover, the analysis was carried out within a frequentist framework. In our study, results are obtained by utilising a cumulative probit model in a Bayesian framework instead of the 'classic' frequentist approach. However, we emphasise that we are neither advocating a Bayesian approach as better than the frequentist approach nor that the cumulative probit model is the correct model to analyse ordinal data. Ordinal models in a frequentist framework provide another valid solution for analysing ordinal data (see *ordinal* package [58]). Also, other link functions besides probit are possible (e.g., logit or cloglog) and can be used. In addition, in Schweiker et al. [57], the linear mixed-effect regression model applied to ordinal data suggested a difference in means between genders. In contrast, the ordinal mixed-effect regression model led to the opposite conclusion. In our study, we obtained the opposite result: under given conditions (see Sections 3.2), the gaussian (ordinal-as-metric) approach inferred non-differences in gender, whereas the cumulative probit model showed a difference. Together, these results demonstrate the issue highlighted in this study: linear regression should not be used in place of ordinal regression to analyse ordinal data. It is essential to point out that we are not claiming that a difference between gender exists. In the literature (e.g., Refs [59;60]), many factors other than gender might lead to individual differences in thermal comfort - for example, age, circadian rhythm, physical disabilities, and fitness [59]. Furthermore, thermal sensation depends on other variables (e.g., clothing, metabolic rate, etc.), and we are aware that not accounting for these factors may have likely confounded the estimation of the models' coefficients. However, given the aim of the study, this last issue can be overlooked. Here, our claim is that the same data (measured on an ordinal scale) analysed with two different methods (i.e., linear and ordinal regression) lead to different results, regardless of the specific outcome. As shown in great detail by Liddell and Kruschke [25] (see also Section 1.3), analysing ordinal data with linear regression may, generally, lead to serious errors in inference. As such, it is not a problem concerning some specific variables (e.g., gender and air temperature in our illustrative example) but a more general issue.

One of the objectives of this work was to highlight that analysing ordinal data as they were continuous may lead to serious errors in inference (i.e., testing theoretical hypotheses). However, regression-type models can address different substantive goals and are therefore well suited to handle distinct purposes. For instance, Shmueli [61] separates a model's aim into descriptive, predictive, and causal explanations. Each of these distinct aims significantly impacts each step of the statistical modelling process and its consequences [61]. For instance, if the purpose is predictive modelling, the exact form of the data-generation process is not of

Table 6
Regression coefficients for the model with a categorical and continuous variable (allowing the standard deviation to vary by group).

		Estimate	Est. Error	L-95 % CI*	U-95 % CI*
Cumulative probit model					
Threshold 1,τ_1		−2.39	0.07	−2.53	−2.25
Threshold 2,τ_2		−1.15	0.04	−1.23	−1.08
Threshold 3,τ_3		−0.28	0.03	−0.34	−0.23
Threshold 4,τ_4		0.53	0.03	0.47	0.59
Threshold 5,τ_5		1.01	0.03	0.94	1.08
Threshold 6,τ_6		1.52	0.04	1.44	1.60
Gender	female	reference			
	male	0.09	0.03	0.03	0.14
Tair_s		0.34	0.01	0.31	0.37
Disc.Male		0.12**	0.02	0.07	0.16
Gaussian (ordinal-as-metric) model					
Intercept		4.01	0.03	3.95	4.08
Gender	female	reference			
	male	0.09	0.04	0.01	0.16
Tair_s		0.47	0.02	0.44	0.51
Sigma.Female		0.32**	0.02	0.28	0.35
Sigma.Male		0.22**	0.01	0.20	0.25

*CI stands for credible interval (based on quantiles).
**Values expressed on the logarithmic scale.

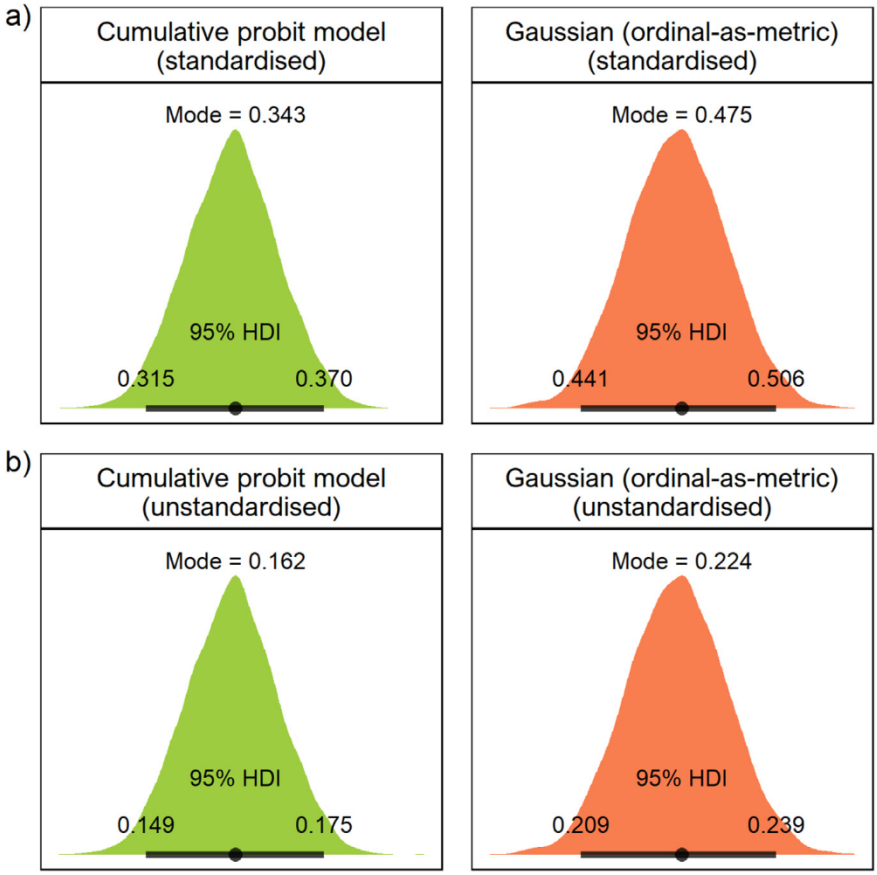

Fig. 10. (a) Standardised and (b) 'original' regression coefficient for air temperature for the cumulative probit (green) and Gaussian (ordinal-as-metric) (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interest, provided that it yields accurate predictions for the dependent variable. If the aim is inference (e.g., explanatory modelling), the estimate of the data-generation process is of interest, while making predictions of the dependent variable is not. In this study, it is not possible to draw specific conclusions regarding the accuracy of the prediction of TSVs. In this regard, Lai and Chen [62] analysed the predictive capability of linear regression compared with ordinal and multinomial regression. Using two separate datasets, the authors demonstrated that ordinal and multinomial regression predicted around half of the individual TSVs, whereas the accuracy of the linear regression model was only around 20 to 40 %. Furthermore, chi-square statistics demonstrate that the ordinal and multinomial regression model outperformed the linear regression model in predicting TSV distributions.

In Section 3.4, the assumption of equidistance between the categories of the ASHRAE 7-point thermal sensation scale was checked. Fig. 11 shows that in terms of 'standard decision rules', the estimated thresholds $\{\tau_k\}$ should not be approximated as

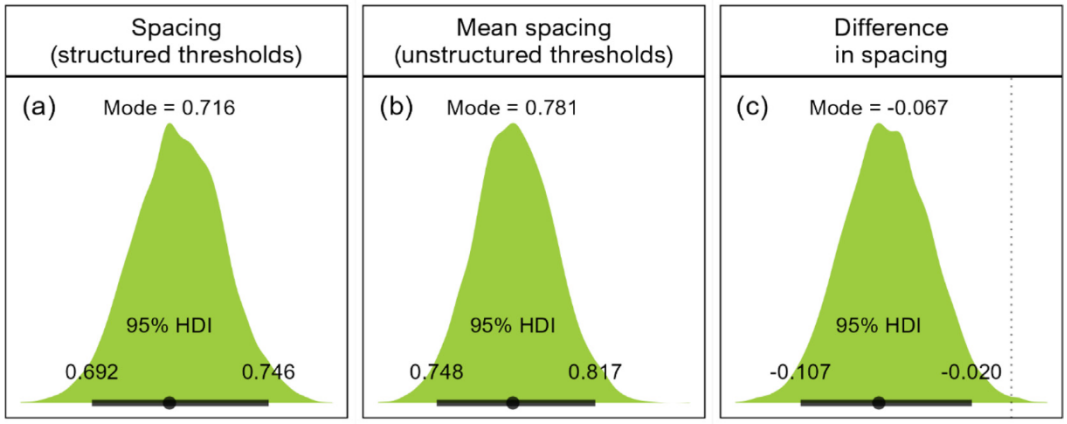

Fig. 11. Spacing for (a) structured and (b) unstructured thresholds and (c) their difference.

Table 7
Values of the Leave-One-Out Information Criterion (LOOIC) and their difference for the cumulative probit model with structured and unstructured thresholds.

Model	LOOIC	SE	LOOIC.diff*	SE.diff**
Cumulative probit model (unstructured thresholds)	19,449.2	100.0	0.0	0.0
Cumulative probit model (structured thresholds)	20,014.0	97.2	564.81	44.39

*LOOIC.diff is the difference between the two LOOIC scores.
**SE.diff is the standard error of the LOOIC.diff.

equidistant, suggesting that, in this sample, the TSV is not interval-scaled. This result was corroborated by the formal analysis presented in Table 7. Here the cumulative probit model with flexible (i.e., unstructured) thresholds fitted the data significantly better than the one with equidistant (i.e., structured) thresholds. It has to be noted that the distances between the thresholds are affected by the form of the latent distribution, which is defined by the link function used. For instance, if the thresholds were found to be equidistant under a latent symmetric distribution (e.g., probit or logit link), under a latent skew distribution (e.g., clog-log link), they will generally not be equidistant. However, since an underlying normal distribution (i.e., probit link) was used in our example, this issue did not affect the results. The inappropriateness of the assumption of equidistance between the categories of the ASHRAE 7-point scale was also found in Schweiker et al. [7]. From a large international collaborative questionnaire study (8225 questionnaires), the authors concluded that significant differences appeared between groups of participants in relation to the distances of the anchors of the thermal sensation scale (and other scales commonly used in thermal comfort studies). Nonetheless, we cannot claim that treating ordinal data as continuous always yields a different result or conclusion than treating them as ordinal. However, knowing in advance that a difference exists is impossible; a different result can be detected only if an ordinal analysis is also performed. Therefore, it is recommended to perform an ordinal analysis directly. Furthermore, since the arguments used by McIntyre [9], which are included in ISO 10551:2019 [2], to legitimise treating ordinal data from the ASHRAE 7-point scale as a continuous variable are disputable (see Section 1 for more detail), **we strongly discourage the use of linear regression for analysing thermal comfort data measured on an ordinal scale. To improve the reliability of the results, we encourage researchers to use ordinal models.**

Moreover, ordinal models offer additional modelling possibilities that this paper has not discussed. For instance, the proportional odds assumptions can be relaxed, and the threshold parameters can depend on some regression variables. In the context of thermal comfort studies, this can be translated to having, for example, different threshold parameters for gender or season.

4.1. Limitations

A fundamental aspect that is usually overlooked is the assumption of independence: residuals, and thus observations, are assumed to be independent. Non-independence can arise, for example, from temporal and spatial autocorrelation. When underlying spatial or temporal processes have the potential to impact a response, the data are autocorrelated – the closer the observations are in space or time, the more highly correlated they are. These sources of non-independence can be apparent or far less so. The response of one sampling unit influencing the response of other sampling units is an example of evident non-independence. The non-independence caused by non-measured confounding influences that vary spatially or temporally is less obvious to detect. Dealing with temporal (or spatial) autocorrelation or analysing temporal (or spatial) trends is different. The former endeavours to deal with the lack of independence associated with temporal (or spatial) data, while the latter tries to model the effect of temporal (or spatial) patterns. During the data analysis stage, it was impossible to identify either spatial or temporal autocorrelation to test the assumption of independence because there was no temporal (e.g., subject ID and timestamp) or spatial (e.g., building ID) information available. As a consequence, this assumption was not checked. Given that the analysis was carried out for illustrative purposes only, this issue can be overlooked**. However, in a real-world analysis, the assumption of independence needs to be verified.** Furthermore, other issues, such as functional form misspecification, collinearity and omitted variable, were not considered during the analysis because they were outside the scope of this article. Nevertheless, when developing a model, depending on the aim of the study, these issues can play an important role and need to be considered. In addition, as stated in Section 2.2, the ASHRAE Global Thermal Comfort Database II does not distinguish between scales, and ordinal and continuous measurements are mixed. Consequently, there is a lack of homogeneity throughout the database that affects its integrity. Furthermore, there are conspicuous missing values in the ASHRAE Global Thermal Comfort Database II. This issue does not derive from the database itself but originates from the lack of explicit agreement on measuring

the 'essential' variables in thermal comfort studies. If this lack of agreement continues, it could affect the future usefulness of the database because the information being added would continue to be non-homogeneous, thus limiting its usability and the new knowledge that could be extracted from it.

5. Conclusions and future perspectives

One of the aims of thermal comfort research is to establish the relationship between the thermal environment and the human sensation of warmth. Typically, this is accomplished by evaluating a subject's subjective thermal reaction to various temperature settings. Diverse rating scales are generally used to measure different aspects of thermal comfort, such as thermal sensation, thermal comfort, thermal preference, and thermal acceptability. While the problem of comparison of different scales (i.e., semantic equivalence) is an issue the thermal comfort research community is aware of, the use of reliable statistical methods to analyse the latter appears to be less discussed. In the thermal comfort domain, it is common practice to analyse subjective human thermal responses independently of how they have been measured. That is, the statistical analysis is unrelated to the modalities of the data that have been acquired. For example, even if measured on an ordinal scale, thermal sensation vote is generally treated as continuous and analysed with linear regression or other statistical tests that assume (conditional) normality. This approximation might be a concurrent factor to explain different results found in previous studies where, for example, gender was found to be or not an influential factor in explaining human responses to the thermal environment.

In this study, we first discussed why the arguments used in ISO 10551:2019 [2] to legitimise treating ordinal data from the ASHRAE 7-point thermal sensation scale as a continuous variable are disputable (see Section 1 for more detail). Secondly, to highlight the ordinal-as-metric issue during the subjective thermal comfort data analysis, the results obtained by utilising a cumulative probit and linear regression model were compared. Based on the analysis carried out on the dataset, the following conclusions can be drawn:

- Compared to the cumulative probit model, the linear regression model inferences non-differences in gender under given conditions.
- Compared to the cumulative probit model, the linear regression model distorts the estimate for the regression coefficient for the air temperature.
- The cumulative probit model shows that subjects used the ASHRAE 7-point thermal sensation in such a way that the distance between adjacent response categories is not the same; that is, they are not equidistant. Consequently, the cumulative probit model with flexible thresholds fitted the data significantly better than the one with equidistant thresholds.

As far as we know, in the field of thermal comfort research, the statistical issues highlighted in this paper are not usually mentioned because the modelling steps are rarely presented, and only the final model is described. However, this is a limitation because researchers can neither assess the reliability of the model nor completely understand the limits of its applicability. Furthermore, while not a primary output of this article, it emerged that there is a lack of homogeneity in the collection of common variables within the ASHRAE Global Thermal Comfort Database II. We recommend that guidelines be developed for defining specific variables to measure. Although there is generally no one-size-fits-all method (e.g., questionnaire) valid for all purposes, agreeing on a 'minimum set' of variables to be consistently measured, possibly with a standardised protocol, would undoubtedly benefit the thermal comfort research community. This would also include agreeing on a specific standardised rating scale (e.g., categorical or visual analogue scale) to be used consistently for all thermal comfort studies. Different rating scales would require diverse statistical modelling approaches, which, in turn, may affect the results.

Data availability

The dataset used can be freely downloaded from the ASHRAE Global Thermal Comfort Database II (https://datadryad.org/stash/dataset/doi:10.6078/D1F671, file 'ashrae_db2.01.csv').

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Matteo Favero would like to thank the Research Centre on Zero Emission Neighbourhoods in Smart Cities (FME ZEN, Grant no. 257660) and the Research Council of Norway (Norges Forskingsrådet) for their support. Salvatore Carlucci and Antonio Luparelli would like to thank the European Commission for funding the Horizon 2020 project Collective Intelligence for Energy Flexibility (COLLECTiEF, Grant no. 101033683). The views and opinions expressed are those of the authors and do not reflect the views of the European Union.

References

[1] ANSI/ASHRAE Standard 55, Thermal environmental conditions for human occupancy, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2020.
[2] ISO 10551, Ergonomics of the physical environment — Subjective judgement scales for assessing physical environments, International Organization for Standardization, Geneva, Switzerland, 2019.
[3] R.J. de Dear, G.S. Brager, D. Cooper, Developing an adaptive model of thermal comfort and preference - Final Report on ASHRAE RP-884, 1997.
[4] P.O. Fanger, Thermal comfort: analysis and applications in environmental engineering, Danish Technical Press, Copenhagen, Denmark, 1970.
[5] L.L. Thurstone, Attitudes Can Be Measured, Am. J. Sociol. 33 (4) (1928) 529–554, https://doi.org/10.1086/214483.
[6] R. Likert, A technique for the measurement of attitudes, Arch. Psychol. 22 (140) (1932) 55.
[7] M. Schweiker, M. André, F. Al-Atrash, H. Al-Khatri, R.R. Alprianti, H. Alsaad, R. Amin, E. Ampatzi, A.Y. Arsano, M. Azadeh, E. Azar, B. Bahareh, A. Batagarawa, S. Becker, C. Buonocore, B. Cao, J.-H. Choi, C. Chun, H. Daanen, S.A. Damiati, L. Daniel, R.D. Vecchi, S. Dhaka, S. Domínguez-Amarillo, E. Dudkiewicz, L.P. Edappilly, J. Fernández-Agüera, M. Folkerts, A. Frijns, G. Gaona, V. Garg, S. Gauthier, S.G. Jabbari, D. Harimi, R.T. Hellwig, G.M. Huebner, Q. Jin, M. Jowkar, J. Kim, N. King, B. Kingma, M.D. Koerniawan, J. Kolarik, S. Kumar, A. Kwok, R. Lamberts, M. Laska, M.C.J. Lee, Y. Lee, V. Lindermayr, M. Mahaki, U. Marcel-Okafor, L. Marín-Restrepo, A. Marquardsen, F. Martellotta, J. Mathur, I. Mino-Rodriguez, D. Mou, B. Moujalled, M. Nakajima, E. Ng, M. Okafor, M. Olweny, W. Ouyang, A.L.P.D. Abreu, A. Pérez-Fargallo, I. Rajapaksha, G. Ramos, S. Rashid, C. F. Reinhart, M.I. Rivera, M. Salmanzadeh, K. Schakib-Ekbatan, S. Schiavon, S. Shooshtarian, M. Shukuya, V. Soebarto, S. Suhendri, M. Tahsildoost, F. Tartarini, D. Teli, P. Tewari, S. Thapa, M. Trebilcock, J. Trojan, R.B. Tukur, C. Voelker, Y. Yam, L. Yang, G. Zapata-Lancaster, Y. Zhai, Y. Zhu, Z. Zomorodian, Evaluating assumptions of scales for subjective assessment of thermal environments – do laypersons perceive them the way, we researchers believe?, Energy Build 211 (2020), https://doi.org/10.1016/j.enbuild.2020.109761.
[8] M. Schweiker, X. Fuchs, S. Becker, M. Shukuya, M. Dovjak, M. Hawighorst, J. Kolarik, Challenging the assumptions for thermal sensation scales, Build. Res. Inf. (2016) 1–18, https://doi.org/10.1080/09613218.2016.1183185.
[9] D.A. McIntyre, Seven point scales of warmth, Build. Serv. Eng. Res. Technol. 45 (1978) 215–226.
[10] A. Agresti, Categorical data analysis, third ed., Wiley, Hoboken, 2013.
[11] J.-Y. Lee, E.A. Stone, H. Wakabayashi, Y. Tochihara, Issues in combining the categorical and visual analog scale for the assessment of perceived thermal sensation: Methodological and conceptual considerations, (2010) 282–290. doi:10.1016/j.apergo.2009.07.007.

[12] G.A. Miller, The magical number seven, plus or minus two: some limits on our capacity for processing information, Psychol. Rev. 63 (2) (1956) 81–97, https://doi.org/10.1037/h0043158.
[13] Level of Measurement, in: W. Kirch (Ed.), Encyclopedia of Public Health, Springer Netherlands, Dordrecht, 2008, pp. 851–852.
[14] S.S. Stevens, On the Theory of Scales of Measurement, Science 103 (2684) (1946) 677–680, https://doi.org/10.1126/science.103.2684.677 PMID - 17750512.
[15] S.S. Stevens, On the Averaging of Data, Science 121 (3135) (1955) 113–116, https://doi.org/10.1126/science.121.3135.113 PMID - 13225751.
[16] P.F. Velleman, L. Wilkinson, Nominal, Ordinal, Interval, and Ratio Typologies are Misleading, Am. Statistician 47 (1) (1993) 65–72, https://doi.org/10.1080/00031305.1993.10475938.
[17] J.W. Tukey, Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments, The collected works of John W. Tukey1986, pp. 391–484.
[18] F.M. Lord, On the Statistical Treatment of Football Numbers, Am. Psychol. 8 (12) (1953) 750–751, https://doi.org/10.1037/h0063675.
[19] F. Mosteller, J.W. Tukey, Data analysis and regression: a second course in statistics, Addison-Wesley, London, 1977.
[20] N.R. Chrisman, Rethinking Levels of Measurement for Cartography, Cartogr. Geogr. Inf. Sc. 25 (4) (1998) 231–242, https://doi.org/10.1559/152304098782383043.
[21] P. Lavrakas, Encyclopedia of Survey Research Methods, (2008). doi:10.4135/9781412963947.n461.
[22] H.B. Rijal, M.A. Humphreys, J.F. Nicol, Towards an adaptive model for thermal comfort in Japanese offices, Build. Res. Inf. (2017) 1–13, https://doi.org/10.1080/09613218.2017.1288450.
[23] R. McElreath, Statistical Rethinking: A Bayesian Course with Examples in R and Stan, second ed.2020.
[24] P.-C. Bürkner, M. Vuorre, Ordinal Regression Models in Psychology: A Tutorial, Adv. Methods Pract. Psychol. Sci. 2 (1) (2018) 77–101, https://doi.org/10.1177/2515245918823199.
[25] T.M. Liddell, J.K. Kruschke, Analyzing ordinal data with metric models: What could possibly go wrong?, J Exp. Soc. Psychol. 79 (2018) 328–348, https://doi.org/10.1016/j.jesp.2018.08.009.
[26] F. Zhang, R.J. de Dear, Impacts of demographic, contextual and interaction effects on thermal sensation—Evidence from a global database, Build. Environ. 162 (2019), https://doi.org/10.1016/j.buildenv.2019.106286.
[27] S. Heidari, S. Sharples, A comparative analysis of short-term and long-term thermal comfort surveys in Iran, Energ. Build. 34 (6) (2002) 607–614, https://doi.org/10.1016/S0378-7788(02)00011-7.
[28] C. Bouden, N. Ghrab, An adaptive thermal comfort model for the Tunisian context: a field study results, Energ. Build. 37 (9) (2005) 952–963, https://doi.org/10.1016/j.enbuild.2004.12.003.
[29] Z. Wang, A field study of the thermal comfort in residential buildings in Harbin, Build. Environ. 41 (8) (2006) 1034–1039, https://doi.org/10.1016/j.buildenv.2005.04.020.
[30] B. Cao, Y. Zhu, Q. Ouyang, X. Zhou, L. Huang, Field study of human thermal comfort and thermal adaptability during the summer and winter in Beijing, Energ. Build. 43 (5) (2011) 1051–1056, https://doi.org/10.1016/j.enbuild.2010.09.025.
[31] J. Han, W. Yang, J. Zhou, G. Zhang, Q. Zhang, D.J. Moschandreas, A comparative analysis of urban and rural residential thermal comfort under natural ventilation environment, Energ. Build. 41 (2) (2009) 139–145, https://doi.org/10.1016/j.enbuild.2008.08.005.
[32] M.K. Singh, S. Mahapatra, S.K. Atreya, Thermal performance study and evaluation of comfort temperatures in vernacular buildings of North-East India, Build. Environ. 45 (2) (2010) 320–329, https://doi.org/10.1016/j.buildenv.2009.06.009.
[33] Z. Wang, L. Zhang, J. Zhao, Y. He, A. Li, Thermal responses to different residential environments in Harbin, Build. Environ. 46 (11) (2011) 2170–2178, https://doi.org/10.1016/j.buildenv.2011.04.029.
[34] D. Teli, M.F. Jentsch, P.A.B. James, Naturally ventilated classrooms: An assessment of existing comfort models for predicting the thermal sensation and preference of primary school children, Energ. Build. 53 (2012) 166–182, https://doi.org/10.1016/j.enbuild.2012.06.022.
[35] H. Djamila, C.-M. Chu, S. Kumaresan, Field study of thermal comfort in residential buildings in the equatorial hot-humid climate of Malaysia, Build. Environ. 62 (2013) 133–142.
[36] Z. Wang, A. Li, J. Ren, Y. He, Thermal adaptation and thermal environment in university classrooms and offices in Harbin, Energ. Build. 77 (2014) 192–196, https://doi.org/10.1016/j.enbuild.2014.03.054.
[37] M. Indraganti, R. Ooka, H.B. Rijal, G.S. Brager, Adaptive model of thermal comfort for offices in hot and humid climates of India, Build. Environ. 74 (2014) 39–53, https://doi.org/10.1016/j.buildenv.2014.01.002.
[38] M. Hawighorst, M. Schweiker, A. Wagner, Thermo-specific self-efficacy (specSE) in relation to perceived comfort and control, Build. Environ. 102 (2016) 193–206, https://doi.org/10.1016/j.buildenv.2016.03.014.
[39] C. Bae, H. Lee, C. Chun, Predicting indoor thermal sensation for the elderly in welfare centres in Korea using local skin temperatures, Indoor Built Environ. 26 (8) (2017) 1155–1167, https://doi.org/10.1177/1420326x16664563.
[40] CEN EN 15251, Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics, European Committee for Standardization, Brussels, Belgium, 2007.
[41] CEN EN 16798-1, Energy performance of buildings - Ventilation for buildings - Part 1: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics - Module MI-6, European Committee for Standardization, Brussels, Belgium, 2019.
[42] S. Crosby, A. Rysanek, Predicting thermal satisfaction as a function of indoor CO2 levels: Bayesian modelling of new field data, Build. Environ. 209 (2022), https://doi.org/10.1016/j.buildenv.2021.108569.
[43] J. Langevin, J. Wen, P.L. Gurian, Modeling thermal comfort holistically: Bayesian estimation of thermal sensation, acceptability, and preference distributions for office building occupants, Build. Environ. 69 (2013) 206–226, https://doi.org/10.1016/j.buildenv.2013.07.017.
[44] L.T. Wong, K.W. Mui, C.T. Cheung, Bayesian thermal comfort model, Build. Environ. 82 (2014) 171–179, https://doi.org/10.1016/j.buildenv.2014.08.018.
[45] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, third ed., 2013.
[46] L. Wasserman, All of Statistics: a Concise Course in Statistical Inference, Springer, New York, NY, 2004.
[47] J.K. Kruschke, Doing Bayesian Data Analysis, second ed.2015.
[48] J.K. Kruschke, T.M. Liddell, Bayesian data analysis for newcomers, Psychon. B Rev. 25 (1) (2018) 155–177, https://doi.org/10.3758/s13423-017-1272-1.
[49] V. Földváry Ličina, T. Cheung, H. Zhang, R.J. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G.S. Brager, P. Li, S. Kaam, ASHRAE Global Thermal Comfort Database II, (2018). doi:10.6078/D1F671.
[50] V. Földváry Ličina, T. Cheung, H. Zhang, R.J. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G.S. Brager, P. Li, S. Kaam, M.A. Adebamowo, M.M. Andamon, F. Babich, C. Bouden, H. Bukovianska, C. Candido, B. Cao, S. Carlucci, D.K.W. Cheong, J.-H. Choi, M. Cook, P. Cropper, M. Deuble, S. Heidari, M. Indraganti, Q. Jin, H. Kim, J. Kim, K. Konis, M.K. Singh, A. Kwok, R. Lamberts, D. Loveday, J. Langevin, S. Manu, C. Moosmann, F. Nicol, R. Ooka, N.A. Oseland, L. Pagliano, D. Petráš, R. Rawal, R. Romero, H.B. Rijal, C. Sekhar, M. Schweiker, F. Tartarini, S.-i. Tanabe, K.W. Tham, D. Teli, J. Toftum, L. Toledo, K. Tsuzuki, R.D. Vecchi, A. Wagner, Z. Wang, H. Wallbaum, L. Webb, L. Yang, Y. Zhu, Y. Zhai, Y. Zhang, X. Zhou, Development of the ASHRAE Global Thermal Comfort Database II, Building and Environment 142 (2018) 502–512. doi:10.1016/j.buildenv.2018.06.022.
[51] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2022. https://www.R-project.org/.
[52] RStudio Team, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA, 2022. http://www.rstudio.com/.
[53] P.-C. Bürkner, brms: An R Package for Bayesian Multilevel Models Using Stan, J. Stat. Softw. 80 (2017) 1–28, https://doi.org/10.18637/jss.v080.i01.
[54] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.
[55] Matthew Kay, tidybayes: Tidy Data and Geoms for Bayesian Models, 2021. http://mjskay.github.io/tidybayes/.
[56] A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, Stat. Comput. 27 (5) (2017) 1413–1432, https://doi.org/10.1007/s11222-016-9696-4.
[57] M. Schweiker, K. Schakib-Ekbatan, X. Fuchs, S. Becker, A seasonal approach to alliesthesia. Is there a conflict with thermal adaptation?, Energ Build. 212 (2020), https://doi.org/10.1016/j.enbuild.2019.109745.
[58] R.H.B. Christensen, ordinal—Regression Models for Ordinal Data, 2019. https://CRAN.R-project.org/package=ordinal.
[59] Z. Wang, R.J. de Dear, M. Luo, B. Lin, Y. He, A. Ghahramani, Y. Zhu, Individual difference in thermal comfort: A literature review, Build. Environ. 138 (2018) 181–193, https://doi.org/10.1016/j.buildenv.2018.04.040.
[60] S. Karjalainen, Thermal comfort and gender: a literature review, Indoor Air 22 (2) (2012) 96–109, https://doi.org/10.1111/j.1600-0668.2011.00747.x.
[61] G. Shmueli, To Explain or to Predict?, Stat Sci. 25 (3) (2010) 289–310, https://doi.org/10.1214/10-STS330.
[62] D. Lai, C. Chen, Comparison of the linear regression, multinomial logit, and ordered probability models for predicting the distribution of thermal sensation, Energy Build. 188–189 (2019) 269–277, https://doi.org/10.1016/j.enbuild.2019.02.027.