Modeling Population Distributions of Subjective Ratings

Matthew P. Reed and Carol A. C. Flannagan
University of Michigan Transportation Research Institute

GLOBAL MOBILITY DATABASE

All SAE papers, standards, and selected books are abstracted and indexed in the Global Mobility Database

Modeling Population Distributions of Subjective Ratings

Matthew P. Reed and Carol A. C. Flannagan
University of Michigan Transportation Research Institute

ABSTRACT

Most human figure models used in ergonomic analyses present postural comfort ratings based on joint angles, and present a single comfort score for the whole body or on a joint-by-joint basis. The source data for these ratings is generally derived from laboratory studies that link posture to ratings. Lacking in many of these models is a thorough treatment of the distribution of ratings for the population of users. Information about ratings distributions is necessary to make cost-effective tradeoffs when design changes affect subjective responses. This paper presents experimental and analytic methods used to develop distribution models for incorporating subjective rating data in ergonomic assessments.

INTRODUCTION

Human figure models provide a flexible and powerful means of assessing product and workstation designs with respect to human requirements. Figure models facilitate analyses of:

- spatial fit and clearance,
- strength and load tolerance, and
- vision requirements.
-

Most figure models also produce comfort scores. The comfort scores are usually derived from studies in which posture data and comfort ratings are obtained simultaneously for a range of workstation geometries (1-4). Local postural comfort is calculated from joint angles (e.g., elbow flexion or shoulder flexion), and the overall score is calculated as a weighted sum of body region scores. Changing the posture of the figure model usually changes the scores, because the scores are driven by the joint angles of the figure. Changing the figure anthropometry while maintaining the same task constraints (hand positions, for example) will often change the joint angles and the comfort scores.

The usefulness of comfort scores is substantially limited because the scores apply only to a single figure (set of anthropometric data) in a particular posture (set of joint angles). Comfort ratings will vary even among people who match the figure anthropometry and who have the specified posture. More generally, a product or workstation is designed for a population of users, each of whom will have a unique set of anthropometric variables and a unique set of postures, along with their own personal comfort preferences. Comfort evaluations with a single figure, or even a family of figure models, will not capture the range of comfort ratings for the user population. As a consequence, it is not possible to optimize designs for comfort using the existing models. For example, a particular layout for a driver's workstation in a car will produce a different posture for each member of a family of figure models selected to span a wide range of anthropometry. The comfort ratings for each model will consequently be different, yet the distribution of ratings for the family of models does not provide a useful quantitative indication of the distribution of ratings for the actual user population, because

1. the figure models do not encompass the complete distribution of anthropometric variables,
2. the figure models do not capture the range of posture variance that is unrelated to anthropometry, and
3. the models do not represent the variance in subjective preference that is unrelated to anthropometry or posture.

These limitations apply even if the anthropometric scaling is correct and the posture prediction is accurate.

This paper presents a general approach to the modeling of population distributions of subjective ratings that can be applied to the evaluation of any workstation, posture, or environment. Some examples from the auto industry are subjective evaluations of steering wheel position, headroom, legroom, lumbar support prominence, and lower extremity posture. The scales used to measure

these variables are different, but they share the general characteristic of being amenable to rating using an ordinal scale. For example, steering wheel position could be rated using an integer from -5 to +5, with anchors at the ends indicating "too close" and "too far." Lower extremity posture could be rated on a discomfort scale, with zero representing "no discomfort" and 10 representing "severe discomfort." The effort required to operate a control could be rated as "very unacceptable," "somewhat unacceptable," "somewhat acceptable," or "very acceptable." Data from an experiment concerning subjective perception of vehicle headroom are used to illustrate the modeling method, followed by a more general statement of the steps in creating a useful model of subjective rating distributions.

ILLUSTRATIVE ANALYSIS

DATA SOURCE – In this paper, methods for measuring and modeling subjective evaluations are presented using the example of automobile passenger headroom. In a laboratory study, 90 men and women rated the headroom provided by a range of roof configurations on a five-point sufficiency scale, shown in Figure 1. Although the study involved a variety of different roof shapes and positions, this analysis will focus on the effects of roof height on headroom perception, with the objective of developing a model to predict the distribution of occupant headroom ratings.

very insufficient	insufficient	barely sufficient	sufficient	more than sufficient
1	2	3	4	5

Figure 1. Headroom sufficiency scale.

The headroom rating is a useful dependent measure for this example because, while the rating is likely to be affected by anthropometry, it is related to posture only in extreme conditions. That is, a lower roof position may affect the subjective perception of headroom without affecting the occupant's posture. This is an example of a subjective measure that is best predicted independent of posture.

EXAMPLE DATA – Figure 2 shows distributions of sufficiency ratings from 90 men and women for three roof positions (high, medium, and low). The mean rating drops from 4.8 for the highest roof position to 4.3 and 3.0 for the middle and lowest roof positions, respectively. A typical subjective rating model would produce these mean values for roof positions matching the test conditions. The mean ratings do not provide much information about the cost, in terms of subjective ratings, of changing the roof height. Suppose the highest and middle roof positions were two candidate designs. The difference in mean sufficiency rating is 0.5. Is that an important shift? Suppose the design population is different from the original test population. What effect will that have on the predicted ratings?

The example data in Figure 2 are typical of subjective ratings data in several ways. The data are not normally or even symmetrically distributed, and the shape of the distribution is affected by the independent variable (roof height, in this case). Because the distribution shape changes, changes in the mean value of the distribution do not provide useful information about changes in, for example, the median or 90th percentile of the distribution.
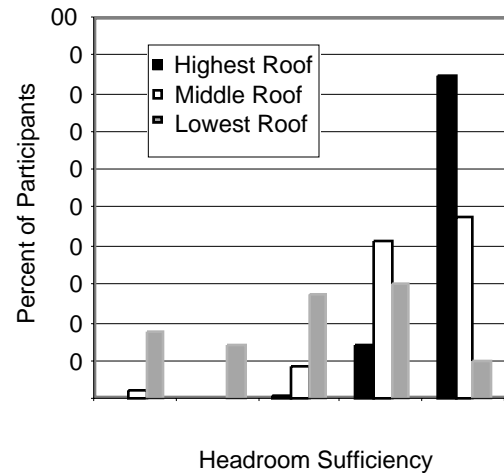

Figure 2. Ratings of headroom sufficiency for three roof heights. Mean ratings are 4.8, 4.3, and 3.0 for the highest, middle, and lowest roof positions, respectively.

MODELING EXPERIMENTAL DATA – An ordinal scale like the one in Figure 2 can be transformed to a set of binary variables by choosing cutoff points. For example, analyses can be conducted on a variable that takes on a value of one if the rating is equal to 4 or higher and zero otherwise. With a binary dependent measure, logistic regression can be used to examine the influence of subject variables (e.g., stature or sitting height) and test variables (roof height, in this example) on the distribution of ratings.

A logistic regression fits the experimental data with a function of the form

$$P[x] = \frac{e^f}{1 + e^f} = \frac{1}{1 + e^{-f}} \qquad [1]$$

where P[x] is the percentage of subjects rating at the selected criterion level and f is a polynomial function of the independent variables. Figure 3 shows logistic regression functions predicting the percentage of subjects rating the headroom as sufficient or better as a function of subject stature (erect standing height) for five levels of roof height. Similar functions could be generated for other criterion levels (e.g., sufficiency ≥ 3). The figure shows a nonlinear effect of roof height on headroom ratings and an interaction with stature. That is, the effect of stature on the roof ratings varies with roof height. At the highest roof height, 100% of people of all

statures rate the roof height as sufficient. At the lowest roof height, only about 10 percent of people 1800 mm tall would rate the roof height as sufficient compared with over 60 percent of people 1500 mm tall.
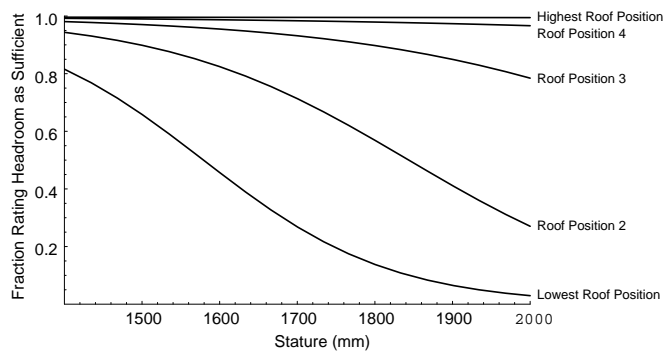

Figure 3. Logistic regression models showing the percentage of people rating the headroom as sufficient as a function of stature for five different roof heights.

The two independent variables in Figure 3, stature and roof height, can be combined to create a single logistic regression function,

$$f = -16984 + 20\,\text{Stature} - 0.078\,(\text{Stature})\,(Z) + 0.1723\,(Z^2) \qquad [2]$$

where stature is in mm and Z (roof height) is in mm relative to an arbitrary reference. Note the second-order Z term that accounts for a substantial nonlinearity in the Z effect (see Figure 3). Nonlinearities are typical in subjective data on roominess, clearances, and control positions.

For a particular roof position (the second-most restrictive condition from Figure 3), the fraction of people predicted to rate the headroom as sufficient is given by

$$P[\text{stature}] = 1 - \frac{1}{1 + e^{16894 - 20\,Stature}} \qquad [3]$$

MODELING POPULATION RATING DISTRIBUTIONS – For purposes of this analysis, the occupant population can be described by the stature distributions of males and females plus the gender mix. The distribution of stature within gender is reasonably modeled by a normal distribution. Figure 4 shows adult U.S. male and female stature distributions together with equation 3 predicting the fraction of people of each stature who will rate the headroom as acceptable. For this population and roof condition, what percentage of the total occupant population will rate the headroom as acceptable?
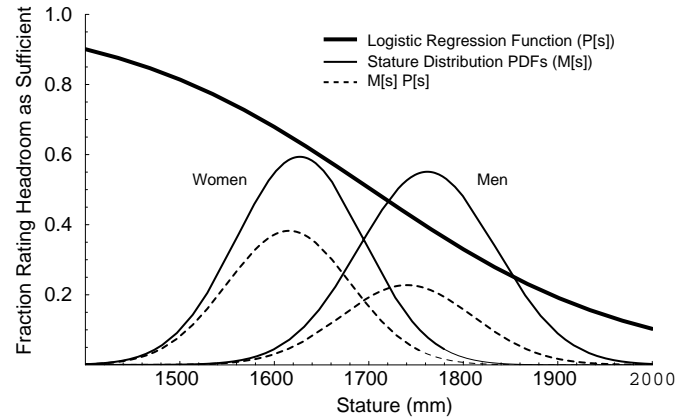

Figure 4. Logistic regression predicting the fraction of people rating the headroom as acceptable as a function of stature, along with U.S. adult male and female stature distributions. Stature probability density functions are multiplied by 1000 for clarity. Distributions shown with dashed lines are obtained by multiplying the stature distributions by the logistic regression function.

The normal distribution probability density function (PDF) gives the probability of sampling a person of a particular stature from the population. (Or, more accurately, the integral of the PDF from Stature1 to Stature2 is the probability of sampling someone whose stature lies between Stature1 and Stature2.) If M[S] is the probability of someone in the male occupant population having stature S, and P[S] is the probability of someone of that stature rating the headroom as sufficient, the M[S] P[S] is the combined probability of sampling someone of stature S *and* having them rate the headroom as sufficient. Summing over all statures is accomplished by integrating M[S] P[S]. If the probability of a "sufficient" rating is unity for all statures, then M[S] P[S] = M[S], i.e., the normal PDF, and the integral is unity. If P[S] = 0.5 (half of people rate the headroom as sufficient independent of stature), then the integral is equal to 0.5.

Figure 4 shows the functions resulting from multiplying each stature distribution by the logistic regression function. Note that the area under the male curve is smaller, indicating that a smaller percentage of men are predicted to rate the headroom as sufficient for this roof height. The male and female fractions (integrals of the male and female curves) are combined using the ratio of males to females in the population. Defining m = fraction of males in the occupant population, the fraction of the total population rating the headroom as sufficient is given by

$$H = m\,F_m + (1-m)\,F_f \qquad [4]$$

where F_m is the fraction of males rating at the criterion level and F_f is the fraction of females. In this example, if 50 percent of the occupant population is male, 71 percent of the population are predicted to rate the headroom as sufficient.

GENERAL OUTLINE OF APPROACH

The headroom-rating example above illustrates the general approach to creating population rating models. The approach to collecting and applying subjective rating data is summarized as follows:

1. Select a subjective rating scale that is appropriate for the current application.

This step is critically important, yet is often given insufficient attention. The subjective rating models predict the distribution of ratings that would be expected if an identical questionnaire or rating scale were used with the design being assessed. A comfort model, for example, does not predict comfort, but rather comfort ratings. Hence, a good rating scale is one that can be used repeatedly in many studies. It should be readily and consistently understood by any potential study participants, and should also produce values that designers and managers find useful for establishing design criteria. Often this means using anchor words that embody value judgments (such as "sufficient" in the headroom scale).

2. Identify design population descriptors.

A product or workstation is designed for a particular population of users, who are described using a number of variables. Some variables are categorical (gender, ethnic group), while others are continuous (anthropometric variables, age). Identification of variables that will be used to describe the target population is important because the study population must be selected to include sufficient independent variance in these variables. For example, if stature and age are both potential population descriptors, care must be taken to select old and young study participants who span a wide range of stature within each age group. The selected population descriptors will be input variables for the population models. At a minimum, a population will usually be described by gender mix (fraction of males and females in the population) and the stature distributions for each gender.

3. Identify variables used to describe test conditions.

Along with population descriptors, the other inputs to the population ratings models are variables that describe the physical environment and task constraints that comprise the situation to be rated. Two general approaches have been used in comfort modeling. One is to use the postural response to the test conditions as the descriptor. In this case, the subjective response is tied to the subject's posture. An alternative is to use environment and task descriptors directly. In the automobile situation, variables such as seat height, steering wheel position, and headroom are used routinely. As with the population descriptors and the rating scale selection, it is critical that the test conditions

be described using variables that have long-term meaning. For example, many studies are conducted comparing subjective ratings among several vehicles. However, absent any quantitative descriptions of the characteristics of the vehicles, the data (or models derived from the data) have no generality.

4. Conduct a study to gather subjective rating data.

Effective experiment design is a complex area, but there are certain issues that are particularly important for the development of subjective rating models. First, representative sampling from the target population is almost never the best approach. For most ergonomic analyses, anthropometry is (or is believed to be) the most important subject variable that will influence the ratings. Stratification on anthropometry, with oversampling on the tails of the distribution, provides a means of increasing the statistical power for estimating anthropometric effects. Stratification and oversampling should also be applied to other variables that are hypothesized to affect ratings. For example, a study of vehicle ingress-egress might oversample older drivers, relative to their representation in the population, to provide more power to estimate age effects. If the hypothesized anthropometric effects are not observed in the study data, the stratification does not diminish the value of the study. A population ratings model can be constructed without anthropometric factors, using either the pooled data from the stratified sample, or, more conservatively, a resampled population configured to match the target population.

Test condition variables should also be manipulated independently, but a full-factorial approach is appropriate only a few variables are involved. Experienced practitioners can usually eliminate from consideration many higher-order interactions, enabling a smaller, more efficient experiment to examine the main effects and interaction of interest. As noted above, many subjective phenomena (such as headroom ratings) are nonlinear in the test condition descriptors, so three or more levels of such variables may be required.

The collection of experimental data used for creating and validating population ratings models should be viewed as an ongoing process. If the rating scale, population descriptors, and test conditions are selected and specified with an appropriate level of generality, it will be possible to build up a database over a series of studies that expand the range of applicability of the resulting models.

5. Construct population models.

The statistical analysis proceeds as in the example above, with two basic steps:

a. Select subjective cutoffs to create binary variables from the rating scale data (e.g., "sufficient" or better).

b. Conduct logistic regression analyses using population and test condition descriptors.

Separate models are generated from each potential binary cutoff on the rating scale. Each predicts the fraction of the population who will rate at or above the specified criterion.

6. Make predictions for candidate designs.

The models are applied by combining the population descriptors with values for design parameters (such as roof position in the headroom example). Often, the same design population (e.g., U.S. adults) will be used for many analyses. In that case, the ratings models are a function only of the design geometry (or other test condition descriptors).

DISCUSSION

PREDICTING POPULATION RATINGS DISTRIBUTIONS – This paper presents a general approach to predicting population distributions of subjective ratings based on logistic regression. The conventional approach to modeling subjective data gives a single value for a set of input conditions, which is usually intended to be representative of the average rating. Changing the input variable values changes the response, but it is difficult to interpret the importance of the change. As noted above, the change in mean could be driven by large changes in ratings by a few people or small changes by a large number of people.

Design optimization is more readily performed when the distribution of ratings is predicted. The method proposed above gives two scales on which design criteria can be set. First, a subjective rating criterion can be selected (e.g., "sufficient" or better on the headroom scale). Second, a population accommodation level can be chosen (e.g., 95 percent of the target population rating at the subjective rating criterion). One method for applying these two interrelated scales is to set a subjective criterion level according to the type of product or workstation, and then to make design tradeoffs using the accommodation level. For example, the rear seat of a sports car might be assessed using a "somewhat sufficient" criterion level for headroom, while the front seat might be assessed using "sufficient." Alternative designs can then be evaluated based on the percentage of the population who are predicted to rate at the specified criterion level. Suppose an increase in roof padding is required to meet safety regulations, but that change will reduce headroom by 20 mm. What is the cost of the change with respect to headroom perception? The question can be answered quantitatively, in terms of the reduction in the percentage of people rating the headroom as sufficient.

CHOICE OF INPUT VARIABLES – The subjective assessment models in most human figure models report postural comfort based on joint angles. This approach has a substantial advantage in that the input variables are always directly available to the software algorithm. The most important disadvantage is that population ratings cannot usually be generated from the posture of a single figure. Given population anthropometry, it would be possible to model the distribution of ratings expected for a population of people in the specified posture, but that result would have little meaning. People of different sizes in the same task environment will have different postures. In general, comfort predictions based on joint angles are useful primarily as an indicator of the extent to which the specified posture approaches the limits of joint ranges of motion. Such predictions should be interpreted in the context of the population variability in posture and in joint range of motion (even within individuals of similar body size).

An alternative approach is to specify the task constraints and environment (i.e., the inputs to an individual's rating process) using variables and dimensions external to the figure model. In the automobile accommodation case, such variables might include seat height, fore-aft steering wheel position, and the height of the roof above the seat. The primary advantages of this approach are (1) input parameters can be expressed in terms of engineering design variables rather than with respect to postural responses to those variables, (2) population ratings distributions can be calculated, since the environment variables apply to all members of the target population, and (3) the effects of variables that have negligible effects on posture, but important effects on subjective ratings, can be quantified. Headroom is an example of a variable that affects overall comfort ratings substantially, even among individuals whose posture is not affected.

Basing the subjective models on external variables requires that the model user provide additional input — the figure model posture is no longer sufficient. There is also a perceived reduction in generality, in that the posture-based ratings models will provide a rating for virtually any posture. However, external variables are frequently required for posture prediction, and hence are already needed for figure model applications. Moreover, mean comfort ratings for postures that are obtained in task conditions other than those used to generate the original subjective data are unlikely to be valid.

IMPLEMENTATION IN HUMAN MODEL SOFTWARE – The population-based modeling approach described in this paper can be implemented outside of human figure modeling software, but the models can also be integrated with modeling software in a manner that facilitates ergonomic assessments. A user interface to the models can be constructed that prompts the user for the population description and for the workspace variables that are inputs to the models. For each figure selected, the distribution of ratings expected for people

of the selected size can be calculated and displayed along with the ratings expected for the entire target population.

CONCLUSIONS

Current methods of calculating and presenting subjective ratings data in human figure models are inadequate because they do not capture the distribution of ratings expected for a user population. Further, using only the figure posture as input unnecessarily limits the subjective models by precluding the use of design variables that do not substantially affect user postures. The modeling approach described in this paper is applicable to any ordinal subjective rating scale, and predicts population ratings distributions as a function of the task environment and the population description. These models are valuable tools for assessing factor tradeoffs during design optimization.

REFERENCES

1. Grandjean, E. (1980). Sitting posture of car drivers from the point of view of ergonomics. In Human Factors in Transport Research. Volume 2 - User Factors: Comfort, the Environment and Behaviour, ed. D.J. Osborne and J.A. Levis, 205-213. New York: Academic Press.
2. Krist, R. (1994). Modellierung des Sitzkomforts: Eine experimentelle Studie. Doctoral Dissertation. .
3. Porter, M.J. and Gyi, D.E. (1998). Exploring the optimum posture for driver comfort. Int. J. Vehicle Design, Vol. 19, No. 3.
4. Rebiffe, R. (1980). General reflections on the postural comfort of the driver and passengers: Consequences on seat design. In Human Factors in Transport Research. Volume 2 - User Factors: Comfort, the Environment and Behaviour, ed. D.J. Oborne and J.A. Levis, 240-248. New York: Academic Press.

CONTACT

Matthew P. Reed, Ph.D.
University of Michigan Transportation Research Institute
2901 Baxter Rd.
Ann Arbor, Michigan 48109
U.S.A
mreed@umich.edu