0.1 Statistical Models for Analysing Route-Finding Tasks

When designing the format of the empirical experiment, we had to consider the method by which we were to analyse the data afterwards. Three options presented themselves as to how the data could be collected and analysed:

1. Group the subjects into three groups and only present questions involving one variant of each map to each group. For example, group A would only see the geographical maps, group B would see the official published maps and group C would see the automatically-drawn maps. This has the advantage that a simple between-subjects analysis using analysis of variance (ANOVA) could be used.

2. Do not group any subjects but present all three map variants for each question to each subject. This would have allowed a within-subjects analysis of the data using either ANOVA or a t-test. However, this method would have required asking the same question three times to each subject (once for each map variant). The possible effect of subjects remembering the maps, questions and answers might have significantly skewed the results and could not be ignored.

3. Group the subjects into three groups and, for each question, present a different variant of each map to each group. This is effectively the same as a combined between- and within-subjects design and was the method that we used for our study. This method has the advantage that any learning of the maps and questions is minimised and should produce more meaningful results. However, the analysis of the data is more complex since each subject responded more than one task. Hence, in this case it is required the use of models capable to distinguish across strata and clusters, in a broad sense.

We decided that the most appropriate method for analysis was the third one and we are able to analyze the results using time-to-event (commonly named duration or survival) models. The variable of relevance would be the time for a person to complete the task of finding a correct metro route using one of the three different maps. In particular, we study the duration time to complete the task using proportional hazard models to distinguish stratification effects for using one of the three drew maps considering the effects of metro locations. We also used frailty models to measure random effect across individuals, since our data consists of more than one sample for each individual. Hence, the effect on the time duration to define a correct metro routes using different drawing system would be measured via its correspondent factor effect. Since the variable of relevance is the duration to complete the task correctly, incorrect answers are also considered in the analysis by treating them as right censored data.

0.2 Frailty Models

We used frailty models (introduced by Clayton and Cuzick **???**) which are a generalisation of Cox's proportional hazards model **???**. Survival models **?** are more often used for analysing the survival of biological organisms or the failure of mechanical systems, however their use extends to the study of duration data as well. The idea of survival analysis is to study time-to-event data (for example the survival time or the time taken to complete a task). For our experimental data we are modelling the duration time taken by an individual to find a correct metro station reading a map, T, that is taken to answer a question.

Typically, survival models are used to analyse the results of subjects undergoing some form of time-to-event test. Each subject in the study contributes to learn about the survival function of a certain underlying population which characterizes the time-to-event to complete a certain task or to get a specific test result. As we mentioned before, proportional hazard models are used to differentiate the effect of different stratification factors in the underlying survival function. Those effects multiplicatively affect the underlying law driven the duration. In our study, those factors are defined via dummy variables. The frailty component extends the proportional hazard model by the inclusion of an additional multiplicative random effect to study clustering patterns across observations. In our study, observation are clustered by individuals.

During the study, there were some individuals who did not respond correctly the task of finding the correct answer or responded an incorrect answer. We assumed that all these cases would eventually be answered correctly. Therefore, those cases were considered as truncated (right censored) duration times to get the right answer. In that way the information contained in all the observation of the sample is incorporated in the analysis.

The mathematical description of survival models with proportional hazards and frailties is as follows. Let T be the duration time to complete a task correctly. Let assume that T is a random variable having a cumulative distribution function

$$P(t) = Pr(T \leq t) \tag{1}$$

and probability density function

$$p(t) = \frac{d}{dt} P(T). \tag{2}$$

The survival function is defined as complement of P, i.e.,

$$\begin{aligned} S(t) &= 1 - P(t) \\ &= Pr(T > t). \end{aligned} \tag{3}$$

In other words, the survival function $S(t)$ gives the probability that someone would have to expend more time than t to complete the task given that at time t he/she did not complete it. The time that a question is answered can also be represented by the hazard function, $h(t)$, which gives the instantaneous rate of failure at time t on condition that something survives until at least t. The hazard function is defined by

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr\left(T \in [t, t + \Delta t) \mid T \geq t\right)}{\Delta t} \\ &= \frac{p(t)}{S(t)}. \end{aligned} \tag{4}$$

Equivalently, the density function is completely defined in term of its corresponding hazard and survival function via $p(t) = h(t)S(t)$.

As we mentioned before, proportional hazard models with frailties incorporate multiplicative effects given for stratified and clustered observations. Its analytic expression is defined by

$$h\left(t_{ij} \mid \boldsymbol{x}_{ij}, v_i\right) = h_0\left(t_{ij}\right) \exp\left(\boldsymbol{x}_{ij}' \boldsymbol{\beta}\right) v_i \tag{5}$$

where i indexes the group of individuals, j indexes individuals, $\boldsymbol{x}_{ij}$ is a p-dimensional covariate vector for stratification, $\boldsymbol{\beta}$ is a p-dimensional unknown parameter vector, h_0 is the baseline hazard function, common to all individuals in the study, and v_i is a multiplicative random effect for each cluster, i.e. the v's are independent random variables with common distribution functions. In our model, $\boldsymbol{x}_{ij}$ is defined as a vector containing seven dichotomous variables. Two ment to evaluate the effect of using the published and automatic metro map drawn and the rest to measure the effect across metro locations. The multiplicative variable is usually reparameterised as $\theta_i = \log v_i$. Whence θ_i can then be artificially incorporated in the linear proportional hazard, as follows

$$h\left(t_{ij} \mid \boldsymbol{x}_{ij}, v_i\right) = h_0\left(t_{ij}\right) \exp\left(\boldsymbol{x}_{ij}' \boldsymbol{\beta} + \theta_i\right), \tag{6}$$

with

$$\boldsymbol{x}_{ij}' \boldsymbol{\beta} = \beta_P \mathbf{1}_P + \beta_A \mathbf{1}_A + \sum_{k=2}^{6} \beta_k \mathbf{1}_{\text{Metro Sys } k}, \tag{7}$$

where $\mathbf{1}_Q$ denotes an indicator variable for the characteristic Q. A key feature of the model used is that h_0 has no analytical form, i.e. it is defined nonparametrically. Hence, it is determinated only by the observed data.

Considering a set of data, one learn from the model (6) via its corresponding likelihood functions. Is in that function where the censoring process of the data comes into the analysis. Lets define a latent variable δ_{ij} defined to be equal to 1 if the ijth duration time correspond to a correct answer and 0 otherwise. The likelihood function, L, is defined by

$$L = \prod_{\delta_{ij}=1} h(t_{ij}) \prod_{\delta_{ij}=0} S(t_{ij}) \tag{8}$$

$$= \prod h(t_{ij})^{\delta_{ij}} S(t_{ij}). \tag{9}$$

Here, the parameters to be estimated are h_0, β and v's. We fitted the model using the Bayesian framework, considering that h_0 is approximated by a B-spline and that all the v's have a Gamma distribution function. In particular we used the package `survBayes` in `R` (include references), working with the standard prior specification considered therein.

0.2.1 Contrast of map type effects

The idea of including frailties and other stratification variables in the model aims at differentiate effects attributable to the map type only without being perturbed by those factors. Working with the Map Type effect, we have considered to keep the geographic map as reference including only two dichotonomous variables for the published and automatic metro map, respectively. It is important to notice that these effects are invariant to any similar specification.

That effect is measured in terms of the ratio of hazard functions. For instance, the effect of using the published map vs the geographic map, given all the rest components, is given by

$$\frac{h(t|\mathrm{P}, v, ...)}{h(t|\mathrm{G}, v, ...)} = \exp(\beta_P). \tag{10}$$

Similarly, the effect of using the automatic drawn map vs the geographic one is given by

$$\frac{h(t|\mathrm{A}, v, ...)}{h(t|\mathrm{G}, v, ...)} = \exp(\beta_A). \tag{11}$$

Working in terms of our hypothesis, we can define that the Map Type A is more efficient than Map Type P if it is more probable to expend less time to find a given route using map A than using map P. In probabilistic terms, our hypothesis requires

$$Pr(T \leq t|A) \geq P(T \leq t|P), \tag{12}$$

for all t, i.e. $P_A(t) \geq P_P(t)$. That is equivalent to

$$S_A(t) \leq S_P(t), \tag{13}$$

for all t. Considering that all the remaining components are given, the above condition reduces to

$$S_0(t)^{\exp(\beta_A)} \leq S_0(t)^{\exp(\beta_P)}, \tag{14}$$

which is equivalent to

$$\beta_P \leq \beta_A, \tag{15}$$

since $S_0(t)$ is bounded at $(0, 1)$ for all t.

To conclude, in a Bayesian framework, we can compute the probability of our hypothesis given the observed data. We would say that there is enough evidence in favor of our hypothesis if that probability is greater or equal than 0.5. Recall that the parameters β_P and β_A are assumed random. Therefore, an equivalent rule to validate our hypothesis might be given by

$$\Pr(\beta_P \leq \beta_A|\text{data}) \geq 0.5. \tag{16}$$

Main references

Clayton, D. G. and Cruzick, J. (1985) “Multivariate generalization of the proportional hazards model (with discussion).” *Journal of the Royal Statistical Society, Series A*, **148**: 82-117.

Cox, D. R. (1972) “Regression models and life-tables (with discussion).” *Journal of the Royal Statistical Society, Series B*, **34**: 187-220.

Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data.* London: Chapman & Hall.