August 31, 2018

Master thesis proposal

Sensory shelf-life estimation for food products using survival analysis methods

Note: This proposal is for co-supervision including an international collaboration with a researcher from Argentina. Consequently, the master thesis needs to be in English, as will most of the communication be. Reasonable working knowledge of English is hence essential. Most if not all of the supervision will be remote (phone, email).

Sensory evaluation is the key factor for determining the shelf life of many food products. Survival analysis has been used to estimate sensory shelf life (SSL) of foods based on consumers' acceptance or rejection of samples with different storage times. (Hough et al., 2003, Hough et al., 2006). The number of consumers used for these estimations has varied between 50 and 80. Although SSL confidence intervals have been reasonable, there has being no systematic criteria in the choice of the number of consumers. The interval censored nature of the data does not allow regular methods to estimate the number of necessary consumers (*N*) for these studies. Hough et al. (2007) use simulations to determine sample size requirements for studies where the same consumers evaluate products at various storage times. A difficulty in this setup is that consumers might be inconsistent, e.g. a consumer accepted a yogurt with 4 h storage, rejected it with 12 h storage and accepted it with 24 h storage. Hough et al. (2007) aggregate the data by consumer into one interval in which the products changes from accept to reject, while due to inconsistencies, these intervals might be quite large. For 6 storage times, the simulations suggest using about 120 consumers, i.e. 720 observations in total.

In a subsequent paper, Libertino et al. (2011) consider the case where every consumer only evaluates a single sample. Apparently, inconsistencies cannot occur in this scenario (though the root causes for these inconsistencies will of course persist, we just cannot detect them anymore). They conclude that about 300 consumers across 6 storages times (i.e. 50 per storage time) will provide similar results.

Comparing the recommendations from the two studies, it appears that with exposing the same consumer to multiple storage times, more than twice as many observations are needed in total relative to exposing every consumer to just one single sample. This seems counterintuitive, as it relates to the ideas of independent observations vs. blocked data, where typically the latter will require smaller sample sizes in case there is a true block effect. With any consumer data, we can reasonable assume that a block effect is indeed at least moderately active.

The suspicion is that this is due to the repeated observations being aggregated by panelist, resulting in a single interval (possibly wide due to the inconsistencies) per assessor. With that, all other information from that consumer is ignored, thereby decreasing power.

The purpose of the master thesis is to consider alternative models for repeated measures by the same consumer at different storage times that will not require aggregation, but to include a blocking factor for the consumer instead. Using all individual observations w/o pre-processing (= more data) and taking the assessor effect into account (= lower residual error) should boost

efficiency. The first suggestions would be a generalized linear model similar to the one used by Libertino et al (2011) but adding a blocking factor, or alternatively a generalized mixed model, treating subject as a random effect. Based on these models, simulations should be run to determine adequate sample sizes.

The practical implications of this thesis imply the possibility of reducing sample sizes for SLL estimations. This would specially benefit small and medium sized food companies who don't usually have the resources to perform large sized consumer studies. Indirectly, consumers would also benefit from more reliable SSL estimations.

Co-Supervision: **PD Dr. Michael Meyners**, Procter & Gamble, Kronberg (Taunus)

Dr. Guillermo Hough, Comisión de Investigaciones Científicas, Buenos Aires, Argentina

Contact: PD Dr. Michael Meyners

Email: meyners.m [at domain] pg.com

Phone: 06173-302123

Starting date: anytime

References

- G. Hough, K. Langohr, G. Gómez, A. Curia. **Survival analysis applied to sensory shelf life of foods.** Journal of Food Science, 68 (2003), pp. 359-362
- G. Hough, L. Garitta, G. Gómez. **Sensory shelf-life predictions by survival analysis accelerated storage models.** Food Quality and Preference, 17 (2006), pp. 468-473
- G. Hough, M.L. Calle, C. Serrat, A. Curia. **Number of consumers necessary for shelf-life estimations based on survival analysis statistics.** Food Quality and Preference, 18 (2007), pp. 771-775
- L.M. Libertino, M.M. López Osorniom, G. Hough. **Number of consumers necessary for survival analysis estimations based on each consumer evaluating a single sample.** Food Quality and Preference, 22 (2011), pp. 24-30