Marginal Estimates

Terry Therneau

19 Sept 2023

> "Comparative experiments are mandatory in order to not view coincidences as cause-effect relationships. ... The comparative experiment requires, to be of some value, to be run in the same time and on as similar as possible patients, else the physician walks at random and becomes the sport of illusions." C.Bernard, Introduction à L'Etude de la Médicine Expérimantale, 1866

Statistics is the art of clever averaging.

To avoid the messiness of multiple covariate-specific curves and to provide an illustration of the difference between groups after adjustment for confounders, it would be useful to create a single overall curve for each FLC group. These curves need to be both adjusted for the other covariates and properly calibrated (i.e., the overall average value is correct). The key idea is to impose balance. To illustrate the idea we start with a simple example.

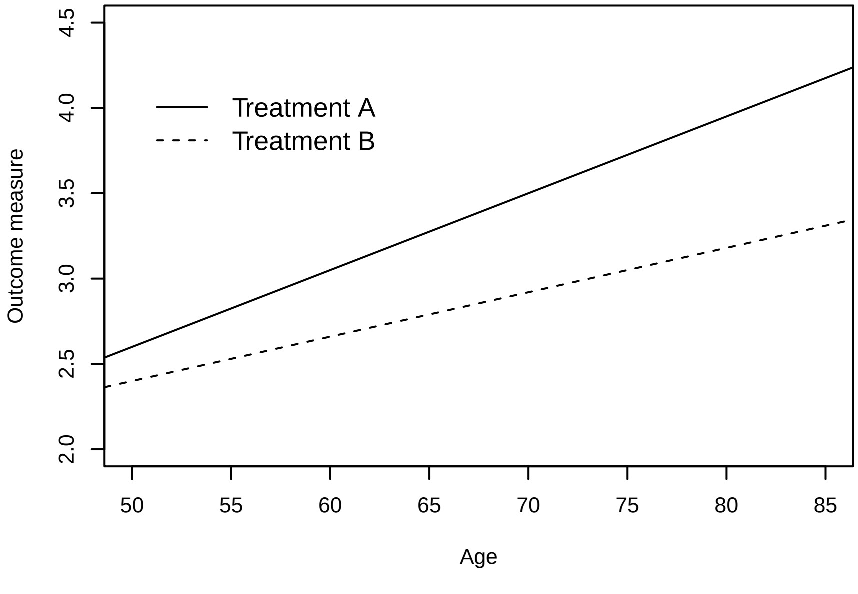

Figure 1: Hypothetical clinical trial comparing 2 treatment arms (A and B)

Consider the hypothetical data shown in Figure 1 comparing two treatment arms, A and B, with age as a confounder. What is a succinct but useful summary of the treatment effect for arms A and B and of the difference between them? One approach is to select a fixed *population* for the age distribution, and then compute the mean effect over that population.

More formally, assume we have a fitted model. We want to compute the conditional expectations

$$m_A = E_F\left(\hat\theta | \mathrm{trt} = A\right) \tag{1}$$

$$m_B = E_F\left(\hat\theta | \mathrm{trt} = B\right) \tag{2}$$

where F is some chosen population for the covariates other than treatment.

Key questions are

1. What statistic $\hat\theta$ to average.
2. What population to use for the adjusting variables
3. Statistical properties of m_A, m_B, $m_A - m_B$
4. How to compute all this
5. What to call it

With repect to 5 there have been dozens of names: population marginal means, g-estimates, marginal effect, Yates' sum of squares, standardized incidednce ratio, least squares means, ... This basic idea has been discovered dozens of times.

The most imortant qustion is what to average. One possible guiding princple is what I call a "poor man's definition" of a causal estimate.

- The prediction can, at least in theory, be assessed in an individal. If I say that Terry's hazard ratio for death is 1.4, and then follow him for 30 years, there is nothing I can measure to evaluate the statement. On the other hand, a statement that Terry's 5 year P(survival) is .5, the observed survival provides data.
- The average over a group is informative for the individual, $\hat\theta_G = (1/n)\sum \hat\theta_i$

Under this definition, predicted survival curves, E(sojourn times), etc are causal.

The second important question is what choice to give for the population F, and this depends critically on what question we want to answer. For instance, in the simple example of Figure 1, if we were considering deployment of these two treatments in nursing home patients, then it would make sense to use an average that gives larger weights to older ages, e.g., a known age distribution for nursing homes. Three common choices for F are:

1. Empirical: the dataset itself or a specific subset.
    - For the simple example above, this would be the distribution of all n ages in the dataset, irrespective of treatment.
    - For a case-control study, it is common to use the distribution of the cases.

2. External: An external reference population, such as:
    - A fixed external reference, e.g., the age/sex distribution of the 2000 US Census. This approach is common in epidemiologic studies.
    - Data from a prior study. This can be useful for comparison of one study to another.
3. Factorial or Yates: the dataset for a balanced factorial experiment. This is only applicable if the adjusting variables are all categorical; the population consists of all unique combinations of the adjusters.
4. Theoretical, e.g., using the Gaussian for a random effect term.
5. SAS type III. Factorial for all categorical covariates, emprical for all continuous ones.

The modern computation is simple brute force. Create a copy of the data set (n obs)

- set x=1 for all obs, compute the n predicted values, average
- set x=2 for all obs, compute the n predicted values, average
- ...

The problem is variance. Bootstrap, IJ, ...?

For $\hat\theta = \eta = X\hat\beta$

$$\begin{aligned}
m_1 &= 1'(X^{\dagger 1}\hat\beta)/n \\
&= (1/n)(1'X^{\dagger 1})\hat\beta \\
&= c_1\hat\beta \\
\text{var}(m_1) &= c_1' V c_1
\end{aligned}$$

The math is nice, but for a Cox (or logistic) model the population average linear predictor is not a useful quantity.

Yates Assume that the X matrix is in standard order: intercept, then main effects, 2 way interactions, 3 way interactions, etc; and that all the variables are factors. If there are any empty cells the Yates' estimate is not defined, so assume none. Let Z be a balanced subset of X, i.e., all combinations of the factors appear equally (one row per combination will suffice). Let C be a matrix such that $C(Z'Z)^{-1}C' = I$; e.g., the Cholesky decompostion of $Z'Z$, and C_g be the rows of C corresponding to one of the main effects. Then $C_g\beta$ is the constrast corresponding to the Yates' SS for that margin. Proof: assign to a postdoc.

This is the heart of the SAS type 3. But the documentation is tailored to the form of X and the computational "leftovers" from the original SAS GLM procedure. If there are missing cells the details of the compuation, and the final results are opaque. The phglm procedure uses an incorrect version of the algorithm, and the type 3 tests are complete garbage.