

Weights in statistics

Thomas Lumley^{1,2*}

Abstract

There are roughly three and half distinct uses of the term weights in statistical methodology, and it's a problem for software documentation and software development. Here, I want to distinguish the different uses and clarify when the differences are a problem. I also want to talk about the settings where we know how to use these sorts of weights, and the ones where we don't. In the interests of doing one thing at a time, I'm going to assume the weights are the right weights and you do actually want to use them; we can have the other discussion some other time (Gelman, 2007a,b).

Keywords

keyword1, keyword2, etc

¹Faculty of Science, Statistics / University of Auckland

SCIENCE CENTRE - MATHPHYSIC - Bldg 303, 38 PRINCES ST, AUCKLAND, 1010, New Zealand.

²Department of Biostatistics / University of Washington

1410 NE Campus Parkway Seattle, WA , 98195, USA.

*Corresponding author: t.lumley@auckland.ac.nz, <https://notstatschat.rbind.io/2020/08/04/weights-in-statistics/>

Contents

1	Introduction	1
2	Means	1
3	Everything else (nearly)	2
4	Hints for guessing the type of weights	2
5	Importance weights	3
6	Everything else else	3

1. Introduction

Weights are a genuine cause of concern; people really do get this wrong, and also they spend time emailing me about it. The only software environment that currently does a competent job of systematically documenting these distinctions is Stata, and even there I don't like all of their terminology.

The three main types of weights are

- the ones that show up in the classical theory of weighted least squares. These describe the precision (1/variance) of observations. An observation with a weight of 10 was classically an average of 10 replicate observations, so it has 10 times the precision of an observation with a weight of 1. The idea was extended to handle other reasons for variance/precision to differ between observations, but the basic rule remains that ratios of weights describe ratios of precision. I call these precision weights; Stata calls them analytic weights.
- the ones that show up in categorical data analysis. These describe cell sizes in a data set, so a weight of 10 means that there are 10 identical observations in the dataset, which have been compressed to a covariate pattern plus a count. For example, `data(esoph)` in R

has three covariates and a binary outcome on about 1200 people summarised as 88 covariate patterns with a count of 0 and 1 outcomes for each pattern. Stata calls these frequency weights, and so do I.

- the ones that show up in classical survey sampling theory. These describe how the sample can be scaled up to the population. Classically, they were the reciprocals of sampling probabilities, so an observation with a weight of 10 was sampled with probability 1/10, and represents 10 people in the population. In real life, these are typically more complicated than just sampling probabilities, but they play the same role of trying to rescale the sample distribution to match the population distribution. I call these sampling weights, Stata calls them probability weights, other people call them design weights or grossing-up weights (I quite like this last one, too).

2. Means

Now, suppose you want to estimate the mean μ_Y of a variable Y from n observations.

- With precision weights, the Gauss-Markov theorem says you should take

$$\hat{\mu}_Y = \sum_{i=1}^n w_i^* Y_i = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}$$

where w_i^* are the weights rescaled to have unit sum.

- With frequency weights you need to uncompress the data and take the sample mean. Write $N = \sum_i w_i$ for the implied full data size, and we have

$$\hat{\mu}_Y = \frac{\sum_{i=1}^n w_i Y_i}{N} = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}$$

- With sampling weights you need to gross up to the population, estimate the population total, and then divide by the estimated population size. The estimated population size is the estimated population total of a variable that's identically 1, so

$$\hat{N} = \sum_{i=1}^N w_i$$

and

$$\hat{\mu}_Y = \frac{\sum_{i=1}^n w_i Y_i}{\hat{N}} = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}$$

The resulting formula is the same for all three; but it is constructed differently each time, out of components that mean different things.

What's different is **variance estimation**.

- With precision weights, we take advantage of the fact that precisions add. We have $\text{var}^{-1}[Y_i] = w_i/\sigma^2$ for some σ^2 , and

$$\text{var}^{-1}[\hat{\mu}] = \frac{\sum_i w_i}{\sigma^2}$$

To get an estimator, we use

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n w_i (Y_i - \hat{\mu}_Y)^2}{n - 1}$$

giving

$$\widehat{\text{var}}[\hat{\mu}] = \frac{\hat{\sigma}^2}{\sum_i w_i} = \frac{\sum_{i=1}^n w_i (Y_i - \hat{\mu}_Y)^2}{(\sum_{i=1}^n w_i)(n - 1)}$$

This is invariant under rescaling of the weights; only ratios of precision weights matter.

- With frequency weights, we just think of the uncompressed original sample

$$\widehat{\text{var}}[\hat{\mu}_Y] = \frac{\sum_i w_i (Y_i - \hat{\mu}_Y)^2}{N - 1} = \frac{\sum_i w_i (Y_i - \hat{\mu}_Y)^2}{(\sum_i w_i) - 1}$$

This is not invariant under rescaling; having ten copies of each observation is different from having one copy of each observation.

- With sampling weights, the randomness is in the sampling, not in the Y s. Write R_i for the indicator that person i is sampled. The estimated mean is really

$$\hat{\mu}_Y = \frac{\sum_{i=1}^N R_i w_i Y_i}{\sum_{i=1}^N w_i R_i}$$

and that's a ratio of two random quantities, so it gets messy. In the special case where the denominator $\hat{N} = \sum_i w_i R_i$ is non-random and equal to N , and the R_i are independent, the variance simplifies to

$$\text{var}[\hat{\mu}_Y] = \frac{1}{N^2} \sum_{i=1}^N \text{var}[R_i] w_i^2 Y_i^2$$

and if $w_i = 1/\pi_i$, the reciprocal of the sampling probability we have $\text{var}[R_i] = \pi_i(1 - \pi_i) = w_i^{-1}(1 - w_i^{-1})$ so

$$\text{var}[\hat{\mu}_Y] = \frac{1}{N^2} \sum_{i=1}^N (w_i - 1) Y_i^2$$

estimated by

$$\widehat{\text{var}}[\hat{\mu}_Y] = \frac{1}{N^2} \sum_{i=1}^N R_i w_i (w_i - 1) Y_i^2$$

In this special case, the variance is invariant under rescaling the weights (because $N = \sum_i w_i$), but the formula is different from the precision-weights one because it involves w_i^2 . The mean is more precisely estimated when the weights are all similar and less precisely estimated when the weights vary a lot.

3. Everything else (nearly)

Most univariate frequentist analyses – quantiles, tables, linear, generalised linear, or non-linear models, survival models, etc – can handle frequency weights and sampling weights. Analyses where the response variable has a free dispersion parameter can also handle precision weights.

If your estimator maximises or minimises $\sum_i m_i(\theta)$, or solves $\sum_i z_i(\theta) = 0$, where each m_i or z_i depends on only one observation's outcome Y_i , it's easy to put in frequency weights or sampling weights. You just maximise or minimise $\sum_i w_i m_i(\theta)$, or solve $\sum_i w_i z_i(\theta) = 0$. For slightly more complicated functions like the Cox partial likelihood, you need a bit more work, but it's still possible. Precision weights don't always make sense, but if there's a free scale parameter on the outcome variable they can be estimated the same way.

Standard error estimation is different for the three sets of weights. It's different in the same ways as for the mean, so implementors could handle any of the types of weights, but it would take extra work each time. This is where it's most important to document which type(s) of weight your code is expecting. If a user supplies the wrong sort of weights, they will get the wrong standard errors.

4. Hints for guessing the type of weights

As I've said, the type of weight should be documented. If it isn't, here are some hints

- If sampling weights are supported, the documentation will usually say so
- Methods that assume the Normal distribution for the outcome variable will typically be using precision weights
- Generalised linear models will often take precision weights (with Y given as the success proportion for binomial data). There will often also be an option for frequency weights (binomial counts) in binomial data.

- SPSS supports frequency weights for basically everything.
- the standard error estimates will change under rescaling of the weights if you have frequency weights, but not if you have precision weights.

5. Importance weights

Because sampling weights and frequency weights (and precision weights if they make sense) plug in to the analysis in the same way and give the same point estimators in common models, you might sometimes want a concept that means just that and doesn't commit you on standard error estimation. Stata uses the term importance weights, which is not bad. I'd prefer analytic weights, but Stata was already using that as a term for precision weights.

6. Everything else else

The big exception to the principle that point estimates are the same for all types of weights is mixed models. There are two issues

mixed models are about partitioning variance, and precision weights encode assumptions about variance, so precision weights matter for estimating residual variance vs variance components. the loglikelihood is not just a sum of terms involving one observation, so there's nowhere to stick sampling weights There are also major complications for Bayesian inference with sampling weights, and we'll just not go there for now.

References

- Gelman, A. (2007a, May). Rejoinder: Struggles with Survey Weighting and Regression Modeling. *Statistical Science* 22(2).
- Gelman, A. (2007b, May). Struggles with Survey Weighting and Regression Modeling. *Statistical Science* 22(2).