American Journal of Epidemiology, 2025, 00, 1–3
https://doi.org/10.1093/aje/kwaf050
Advance access publication date March 7, 2025
The AJE Classroom

Inverse probability weighting for categorical exposures

Ashley I. Naimi*,[1] and Brian Whitcomb[2]

[1]Department of Epidemiology, Emory University, Atlanta, GA 30322, United States
[2]Department of Epidemiology, University of Massachusetts at Amherst, Amherst, MA 01003, United States
*Corresponding author: Ashley I. Naimi, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road, CNR 3023, Atlanta, GA 30322 (ashley.naimi@emory.edu)

Inverse probability weights (IPW) are invaluably useful tools for addressing issues including censoring, missing data, selection bias, or confounding.[1] In complex scenarios, such as with time-varying confounding, IPW can be used to estimate causal effects and overcome shortcomings of conventional regression-based adjustment. When used to address confounding in an observational study evaluating binary treatment, IPW functions similar to standardization, and the weights are the inverse of the observed probability of receiving treatment. In this case, the IPW weights can be determined straightforwardly using a logistic regression model of treatment as a function of measured confounders. From this model, one can obtain an estimate of the conditional probability of being treated for the treated individuals, and one minus the conditional probability of being treated for untreated individuals. These probabilities can then be used to construct weights.[1]

In studies considering a categorical exposure, IPW can still be used to address confounding. However, the procedure for determining IPW in the setting of an exposure with more than two levels requires a different regression modeling strategy. In this Classroom article, we demonstrate how to construct stabilized IPW for a categorical exposure using multinomial logistic regression. We begin with the publicly available National Health and Nutrition Examination Survey (NHANES) Epidemiologic Follow up Study data (NHEFS) and estimate the adjusted mean difference between exercise measured in three categories in 1972 (the treatment), and weight change between 1981 and 1972 (the outcome) using IPWs for control of confounding. We explain the data setting, explain how to construct IPWs for the three-level treatment variable and describe an approach for evaluation of potential violations of the positivity assumption. We provide code in a linked GitHub repository demonstrating how to deploy the procedures in the R programming language.

Data setting: NHEFS

Data for the example were obtained from Hernán and Robins,[2] linked in the GitHub repository. Suppose we are interested in using IPW to address confounding when estimating the effect of exercise on weight change in these data. The exercise variable in the NHEFS data are classified as:

exercise = 0: "much exercise."
exercise = 1: "moderate exercise."
exercise = 2: "little or no exercise."

We can estimate the effect of exercise on weight change, as follows:

$$\psi_1 = E\left(Y^{x=0} - Y^{x=2}\right)$$

$$\psi_2 = E\left(Y^{x=1} - Y^{x=2}\right)$$

$$\psi_3 = E\left(Y^{x=0} - Y^{x=1}\right)$$

where Y denotes the weight change variable measured in kilograms, X denotes the exercise variable, and Y^x denotes the potential outcome that would be observed if X were set to x. These counterfactual contrasts can be interpreted as the difference in mean outcomes that would be observed if everyone engaged in: "much exercise" versus "little or no exercise" (ψ_1); "moderate exercise" versus "little or no exercise" (ψ_2); and "much exercise" versus "moderate exercise" (ψ_3). These contrasts are akin to what might be estimated in a three-arm randomized trial. However, because of potential confounding by other influences of weight change that vary by exercise category, these contrasts are not directly estimable. Yet we can use the NHEFS data to estimate these quantities using IPW under the relevant identification assumptions.[3]

Multinomial models for weight construction

To construct weights for the categorical exercise exposure, we can fit a multinomial logistic regression model that regresses the three-category exercise variable against confounders, here denoted C:

$$\log\left[\frac{P\left(X = k \mid C\right)}{P\left(X = k_0 \mid C\right)}\right] = \beta_{0\,k} + \beta_k C$$

This model compares the log of the probability of being in exposure category k relative to some baseline exposure category k_0 as a function of the confounders included in the model. Other models (eg, ordinal logistic regression) could also be deployed. The key is to obtain an estimate of the probability of being exposed to each exposure category for each person. This is the propensity score model, in which the intercept and coefficients in this model are specific to the exposure category k (all relative to exposure category k_0).

Received: July 6, 2024. **Accepted:** October 27, 2024

Table 1. First five observations from the NHEFS data showing the outcome (weight change between 1971 and 1982), the exposure (exercise status), and select confounders (sex, age, race, income, school), as well as three columns representing the propensity score for being in each exposure category.

ID	pEx0	pEx1	pEx2	wt82_71	Exercise	Sex	Age	Race	Income	School
1	0.18	0.31	0.51	−10.09	2	0	42	1	1	7
2	0.33	0.39	0.28	2.60	0	0	36	0	1	9
3	0.04	0.27	0.69	9.41	2	1	56	1	0	11
4	0.05	0.20	0.75	4.99	2	0	68	1	0	5
5	0.30	0.43	0.27	4.99	1	0	40	0	1	11

Abbreviation: NHEFS, NHANES Epidemiologic Follow up Study data.
Propensity scores that are highlighted gray correspond to the score for the observed exposure category.

Depending on the software package used, different sets of probabilities can be obtained from the fit of the propensity score model, including the conditional probabilities of a given observation being in each exposure category. These are shown in Table 1 (as pEx0, pEx1, and pEx2), along with NHEFS data we used to fit the multinomial logistic model. This table highlights the predicted probability that corresponds to the observed exercise category for that individual. For example, the exercise category for ID = 1 is exercise = 2: "little to no exercise." The estimated conditional probabilities of each category of exercise (ie, 0, 1, and 2) are 0.18, 0.31, and 0.51, respectively. Because this individual's observed exposure category is exercise = 2, the relevant conditional probability for constructing the IPW is 0.51, the estimated probability that exercise = 2 conditional on covariates.

The inverse of the conditional probability of being in the observed exposure category is the IPW. Thus, for the five observations in Table 1, IPWs are 1/0.51, 1/0.33, 1/0.69, 1/0.75, and 1/0.43.

The potential benefits of incorporating covariates to yield stabilized IPW have been described.[1] Stabilized weights can be obtained by multiplying the unstabilized weights by the unconditional probability of being in the observed exposure category. These probabilities correspond to the simple proportion in each exposure category, which can be obtained from an intercept-only multinomial logistic regression model. In the NHEFS data, these unconditional probabilities are 0.19, 0.43, and 0.38 for exercise categories 0, 1, and 2, respectively. This leads to stabilized weights (sIPW) for the five observations in Table 1 of 0.38/0.51 = 0.74, 0.19/0.33 = 0.59, 0.38/0.69 = 0.55, 0.38/0.74 = 0.51, and 0.43/0.43 = 0.99, respectively (Table 2).

Propensity score overlap for evaluating positivity

When using sIPW for control of confounding to obtain causal estimates, violations of positivity can be evaluated because the weights are determined directly from the data, even if the contrasts of interest are not. One can obtain propensity score overlap plots that provide insight into the positivity assumption for the estimands of interest. One way to evaluate positivity is to explore the overlap in the probability of being exposed to each exercise category among those in different observed exposure categories. This propensity score overlap plot is demonstrated in Figure 1 and shows good overlap in the distribution of the propensity scores. Another technique known as ternary plots, which works for three-category exposures, is illustrated in the GitHub repository.

Estimating the contrasts of interest

We can now use the stabilized inverse probability weights to estimate ψ_1 and ψ_2, the contrasts of interest. Here, we use a simple weighted least squares estimator with a model that regresses weight change against indicator variables for the three-category exercise variable, setting the reference category as exercise = 2: "little to no exercise." To estimate ψ_3, we set exercise = 1: "moderate exercise" as the referent.

$$E(Y \mid X) = \alpha_0 + \psi_1 I(X = 0) + \psi_2 I(X = 1)$$
$$E(Y \mid X) = \beta_0 + \psi_3 I(X = 0) + \beta_1 I(X = 2)$$

Implementation of weights results in a pseudo-population, for which standard errors for the point estimates such as ψ_1, ψ_2, and ψ_3 are typically obtained using a robust variance estimator.[4] However, this does not account for the variability in the propensity score model. One can alternatively use the bootstrap to account for this variability and obtain correct standard errors and/or confidence intervals. Table 2 presents the mean differences, and 95% normal interval confidence intervals obtained via the robust variance estimator (HC3 estimator with no small sample correction), and a bootstrapped standard error with 2000 replicates (Table 3).

Table 2. First five observations from the NHEFS data showing the exposure (exercise status), three columns representing the propensity score for being in each exposure category, and the unstabilized and stabilized inverse probability weights computed from each relevant propensity score entry.

ID	Exercise	pEx0	pEx1	pEx2	IPW	sIPW
1	2	0.18	0.31	0.51	1/0.51	0.38/0.51
2	0	0.33	0.39	0.28	1/0.33	0.19/0.33
3	2	0.04	0.27	0.69	1/0.69	0.38/0.69
4	2	0.05	0.20	0.75	1/0.75	0.38/0.74
5	1	0.30	0.43	0.27	1/0.43	0.43/0.43

Abbreviation: NHEFS, NHANES Epidemiologic Follow up Study data.
Propensity scores that are highlighted gray correspond to the score for the observed exposure category.

Table 3. Estimated average treatment effects (contrast) obtained from the NHEFS data for each exercise category, along with lower and upper 95% confidence intervals obtained using the robust variance estimator and the bootstrap.

Contrast	Estimate	Variance estimator	Lower 95% CI	Upper 95% CI
$\psi_1 = E\left(Y^{x=0} - Y^{x=2}\right)$	−0.22	HC3	−1.36	0.92
		Bootstrap	−1.30	0.88
$\psi_2 = E\left(Y^{x=1} - Y^{x=2}\right)$	−0.01	HC3	−0.93	0.93
		Bootstrap	−0.91	0.91
$\psi_3 = E\left(Y^{x=0} - Y^{x=1}\right)$	−0.22	HC3	−1.30	0.85
		Bootstrap	−1.24	0.83

Abbreviation: NHEFS, NHANES Epidemiologic Follow up Study data.

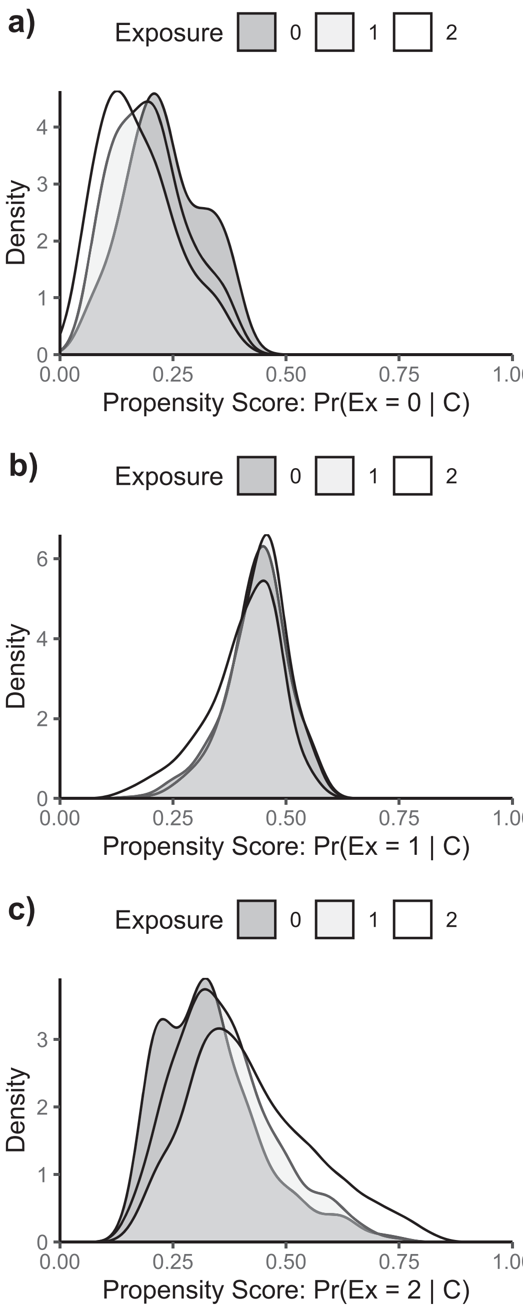

Figure 1. Propensity score overlap plots for the three-category exercise variable in the NHEFS data.

Conclusion

Inverse probability weights are a useful tool to address confounding and other biases in epidemiologic research. To address potential complication for use with categorical exposures, we have illustrated the use of multinomial logistic regression to construct stabilized inverse probability weights for a categorical exposure in a simple data setting. We also demonstrated how the positivity assumption can be explored using propensity score overlap plots with a categorical exposure.

Importantly, this simple example can be generalized in a number of ways to exposure with many more categories, and beyond simple continuous outcome data. A variation of this procedure referred to as the quantile binning method has been proposed to construct inverse probability weights for continuous exposures as well.[5] We hope this simple illustration serves as a useful starting point for researchers interested in more complicated settings.

Acknowledgments

We thank Andreas Stang for prompting us to write this short Classroom article.

Funding

None declared.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6):656-664. https://doi.org/10.1093/aje/kwn164
2. Hernan MA, Robins JM. *Causal Inference: What If?* Boca Raton: Chapman & Hall / CRC; 2020.
3. Naimi AI, Whitcomb BW. Defining and identifying average treatment effects. *Am J Epidemiol.* 2023;192(5):685-687. https://doi.org/10.1093/aje/kwad012
4. Mansournia MA, Nazemipour M, Naimi AI, et al. Reflection on modern methods: demystifying robust standard errors for epidemiologists. *Int J Epidemiol.* 2021;50(1):346-351. https://doi.org/10.1093/ije/dyaa260
5. Naimi AI, Moodie EEM, Auger N, et al. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology.* 2014;25(2):292-299. https://doi.org/10.1097/EDE.0000000000000053